frontiers | Frontiers in Aging Neuroscience

TYPE Original Research
PUBLISHED 21 September 2022
DOI 10.3389/fnagi.2022.945024

OPEN ACCESS

EDITED BY
Redha Taiar,
Université de Reims
Champagne-Ardenne, France

REVIEWED BY
Monisha Dey,
Port City International University,
Bangladesh
Yuanluo An,
Beijing Jiaotong University, China

*CORRESPONDENCE
Noriaki Kuwahara
nkuwahar@kit.ac.jp

SPECIALTY SECTION
This article was submitted to Alzheimer's Disease and Related Dementias,
a section of the journal
Frontiers in Aging Neuroscience

RECEIVED 16 May 2022
ACCEPTED 29 August 2022
PUBLISHED 21 September 2022

CITATION
Jiang L, Siriaraya P, Choi D, Zeng F and Kuwahara N (2022) Electroencephalogram signals emotion recognition based on convolutional neural network-recurrent neural network framework with channel-temporal attention mechanism for older adults. *Front. Aging Neurosci.* 14:945024. doi: 10.3389/fnagi.2022.945024

Electroencephalogram signals emotion recognition based on convolutional neural network-recurrent neural network framework with channel-temporal attention mechanism for older adults

Lei Jiang[1], Panote Siriaraya[1], Dongeun Choi[2], Fangmeng Zeng[3] and Noriaki Kuwahara[1]*

[1]Graduate School of Science and Technology, Kyoto Institute of Technology, Kyoto, Japan, [2]Faculty of Informatics, The University of Fukuchiyama, Kyoto, Japan, [3]College of Textile Science and Engineering, Zhejiang Sci-Tech University, Hangzhou, China

Reminiscence and conversation between older adults and younger volunteers using past photographs are very effective in improving the emotional state of older adults and alleviating depression. However, we need to evaluate the emotional state of the older adult while conversing on the past photographs. While electroencephalogram (EEG) has a significantly stronger association with emotion than other physiological signals, the challenge is to eliminate muscle artifacts in the EEG during speech as well as to reduce the number of dry electrodes to improve user comfort while maintaining high emotion recognition accuracy. Therefore, we proposed the CTA-CNN-Bi-LSTM emotion recognition framework. EEG signals of eight channels (P3, P4, F3, F4, F7, F8, T7, and T8) were first implemented in the MEMD-CCA method on three brain regions separately (Frontal, Temporal, Parietal) to remove the muscle artifacts then were fed into the Channel-Temporal attention module to get the weights of channels and temporal points most relevant to the positive, negative and neutral emotions to recode the EEG data. A Convolutional Neural Networks (CNNs) module then extracted the spatial information in the new EEG data to obtain the spatial feature maps which were then sequentially inputted into a Bi-LSTM module to learn the bi-directional temporal information for emotion recognition. Finally, we designed four group experiments to demonstrate that the proposed CTA-CNN-Bi-LSTM framework outperforms the previous works. And the highest average recognition accuracy of the positive, negative, and neutral emotions achieved 98.75%.

KEYWORDS

electroencephalogram (EEG), emotion recognition, channel-temporal attention, CNN-RNN, older adults

Introduction

Background

Japanese family norms based on the traditional culture of filial piety form a social support network centered on kinship ties, which differs sharply from the individual-centric social networks of Western countries (Sugisawa et al., 2002; Knight and Sayegh, 2010). As a result, Japanese older adults are more likely to feel socially isolated at a rate of 15.3% compared to 5.3% in the UK (Noguchi et al., 2021). Poor interaction and lack of social participation are among the contributing factors to social isolation which are closely associated with depression, one of the major risk factors for the development of Alzheimer's dementia (Santini et al., 2015). Many studies (Westermann et al., 1996; Thierry and Roberts, 2007; Sitaram et al., 2011; Iwamoto et al., 2015; Leahy et al., 2018) have shown that reminiscence and communication about past photographs between older adults and younger volunteers, healthcare workers, or families encourage positive interaction and social engagement. And therefore, they are highly effective in improving the emotional state and alleviating depression in older adults. However, it is necessary to evaluate the emotional state of the older person when talking about the photographs to (a) ensure that the communication is positive, as long-term negative emotions may cause changes in feelings and state of mind (**Figure 1A**) leading to various mental illnesses (Van Dis et al., 2020). For example, mania is easily caused by a prolonged state of ecstasy (high positive) and euphoria (high arousal) as shown in **Figure 1B** and (b) estimate whether the photographs in the conversation are effective in improving the emotion of the older person, and replace them with other photographs if they are not effective. While numerous studies focus on the evaluation of emotions in older adults, earlier studies generally used self-assessment in the form of verbal or questionnaires and were found to be intermittent and influenced by social expectations or demand characteristics (the idea that participants or stimulators will develop similar or specific emotions in response to perceived expectations) (Orne, 1962). Later developments use smart wearable devices (physiological signals) (Kouris et al., 2020), facial expression (Caroppo et al., 2020), and speech recognition (Boateng and Kowatsch, 2020) to monitor and recognize emotions. However, variances and continuities such as facial aging in older adults and differing accents among various groups of people (e.g., different dialects spoken throughout Japan) make it difficult to distinguish and unify such features and expressions. These inevitably result in unreliable emotion recognition results for older adults.

Thus, while physiological signals to monitor emotions seem to be a more suitable approach for older adults, not all physiological signals are suitable for distinguishing between different emotional experiences. For example, although excitement and panic are different emotions generated in response to different stimuli of award and threat, both exhibit the same physiological changes (i.e., increased heart rate, increased blood pressure, body shaking, etc.). Moreover, time is also necessary for the autonomic and sympathetic nervous systems to switch on and off, resulting in outward physiological changes that are slow-acting and insufficient in keeping up with the emotional changes (Liu and Cai, 2010). By conducting the EEG signals through the electrodes on the scalp we can collect EEG signals with a high temporal resolution that reflect different emotional states and variances between these moments (Alarcao and Fonseca, 2017) (Note: all commercially available acquisition devices have a sampling rate of at least 160 Hz/s). As we know from the widely accepted cognitive-evaluation theory of emotions (two-factor theory) (Cornelius, 1991), when we are stimulated by the external environment, we immediately generate physical reactions and simultaneously evaluate them with past knowledge and experience (cognitive process) and finally integrate them into the cerebral cortex obtaining the emotional state (the whole process shown in **Figure 1C**). Therefore, we can say that EEG signals have a significantly stronger association with emotions than other physiological signals. They are also objective, non-invasive, and safe.

The general process and principles of EEG signals for the emotion recognition system (shown in **Figure 1D**) are (1) stimulus materials elicit emotions in the subject while collecting EEG signals, (2) the computer sequentially preprocesses and extracts features from the received EEG signals, and (3) an EEG-based emotion recognition classifier is trained using task-relevant EEG features. The emotion label of EEG features in training emotion classifiers is based primarily on the SAM scale using the valance-arousal emotion model proposed by Posner et al. (2005). The subject is exposed to stimuli and their emotional state is evaluated by oneself using the SAM scale (Valence: positive to negative emotional state; Arousal: difference in the level of physiological activity and mental alertness), which is mapped to the valence-arousal emotion model (**Figure 1B**) to obtain a corresponding "emotion label." In this way, the subjective experience of different emotions (emotion labels) and subjects' objective physiological responses (EEG signals) are matched one-to-one. Nowadays, many inexpensive solutions for portable EEG acquisition devices are available on the market (Stytsenko et al., 2011; Surangsrirat and Intarapanich, 2015; Athavipach et al., 2019), and thus EEG signal-based emotion recognition has a promising application and research value. For this study in the conversation scenario using EEG signals for emotion recognition is extremely challenging. Especially, as the facial muscle activity during the conversation will evoke high-energy artifacts that may distort the intrinsic EEG signal. Such artifacts will hide the rhythm of the real EEG signal and cause perturbation in an EEG system that makes EEG signal processing difficult in all respects (Kamel and Malik, 2014). Therefore, in EEG-based emotion

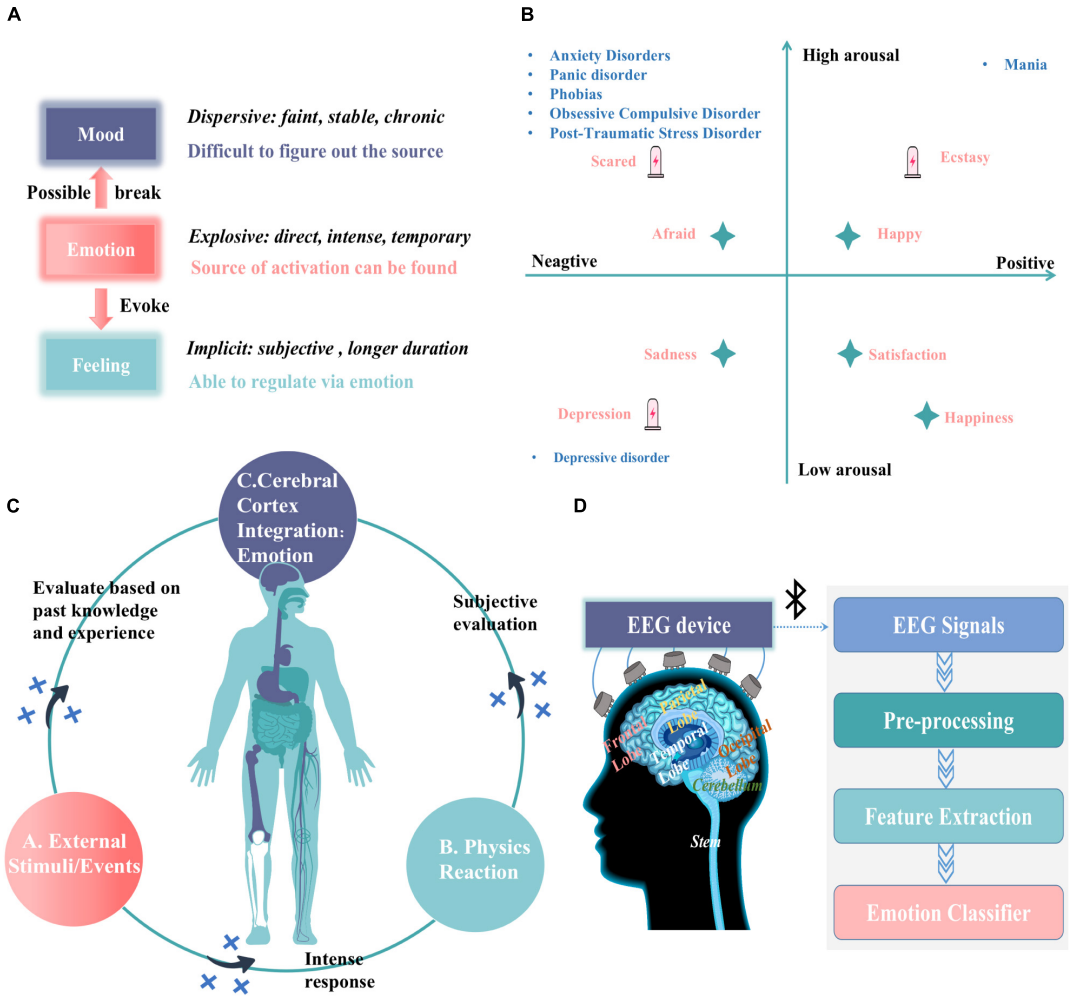

FIGURE 1
Overview of the causes and methods of monitoring individual (subject) emotions. **(A)** Characteristics of mood, emotion, and feeling. **(B)** Expression of emotions space model and mental disorders. **(C)** The formation of emotion. **(D)** EEG emotion recognition system.

recognition, appropriate signal pre-processing methods must be first adopted to remove artifacts and make the EEG data as clean as possible simply reflects the brain's activity. Meanwhile, the challenge to reduce the number of dry electrodes to improve user comfort while ensuring a high emotion recognition rate remains.

In this paper, we propose a CNN-RNN framework combined with a channel-temporal attention mechanism (CTA-CNN-Bi-LSTM) for EEG emotion recognition inspired by the channel-spatial attention module (CBAM) proposed in the field of computer vision research (Woo et al., 2018). The primary contributions of this study are summarized as follows.

(1) In the EEG signal pre-processing stage, due to the EMG and EOG artifacts contribute differently to different brain regions and attenuate as the distance from the scalp gets more remote. We divided the 8-channel EEG signals into Frontal, Temporal, and Parietal groups according to brain regions. And then remove multiple biological artifacts from raw EEG signals in each group separately based on the MEMD-CCA method (Xu et al., 2017; Chen et al., 2018).

(2) In the phase of assigning emotional labels to EEG signals, the emotion labels (positive, neutral, and negative) of EEG signals were automatically obtained by the K-means method based on the ratings of the emotion scale [Valence (-4,4), Arousal(-4,4) and Stress (1,7)] of each participant. The advantage of using this method is not to use the same rating classification criteria for all participants, but to use each participant's rating to classify their own emotions.

(3) For data-driven EEG-based emotion recognition without feature engineering, we developed a CTA-CNN-Bi-LSTM framework. This framework integrates the channel-temporal attention mechanism (CTA) into the CNN-Bi-LSTM module to explore using spatial-temporal information of different important channels (channel attention) and time points

(temporal attention) of EEG signals to achieve EEG-based emotion recognition. And the proposed framework achieved average emotion recognition accuracy of 98%, 98%, and 99% in the negative, neutral, and positive emotions.

(4) We conducted four group experiments on the OCER dataset to explore the contribution of each module to EEG-based emotion recognition. The experimental results indicate that the CNN module provided the largest contribution to the accuracy improvement (21.29%) of the proposed framework, the Bi-LSTM module after the CNN module provided little enhancement (8%) of the framework and the addition of the Channel-Temporal attention module before the CNN-RNN module led to a further significant improvement (11%).

Related works

In this part, we first describe the artifacts that typically emerge during EEG acquisition and existing effective methods to remove them. Then we introduce EEG emotion recognition systems which have evolved from traditional hand-crafted feature extraction to end-to-end deep learning frameworks with channel selection mechanisms.

Electroencephalogram artifacts and removal methods

Due to the potential technical and biological artifacts (**Figure 2**) in the EEG acquisition process will cause the oscillating discharge larger than the neuronal discharge (Kamel and Malik, 2014). Before proceeding with electroencephalography (EEG) data analysis, it is important to make sure that the EEG data is as clean as possible, meaning that the data collected simply reflects the brain's activity.

Technical artifacts mainly include three types: impedance fluctuation (Rodriguez-Bermudez and Garcia-Laencina, 2015), line interference (Huhta and Webster, 1973), and wire movement (Urigüen and Garcia-Zapirain, 2015). Technical artifacts can be avoided by paying attention during the acquisition of the EEG signals. Biological artifacts mainly include two types: muscular artifacts [Electromyogram (EMG), Electrocardiogram (ECG)], ocular artifacts [Electrooculogram (EOG)] including eye movement and eye blinking. Such biological artifacts are inevitable contaminations due to the conductivity of the scalp (Kamel and Malik, 2014), and the closer the artifact's sources are to the electrodes, the more significant is their effect on the EEG data. In particular, the activity of the facial muscles (forehead, cheeks, mouth), neck muscles and jaw musculature (EMG) have a serious effect on the EEG, with a broadband frequency distribution of 0–200 Hz (Halliday et al., 1998; Van Boxtel, 2001). In addition, the heart also is muscular (ECG) and continuously active, which also affects the quality of the EEG data. The artifact has a broadband frequency distribution of 0–75 Hz (Lee and Lee, 2013), but has less effect on the EEG because of the large distance between the scalp and the heart. The eyes have a powerful electromagnetic field, which is formed by millions of neurons in the retina, thus eye movement (horizontal, vertical, and rotation) and eyeblink will affect the electric field received by the electrodes resulting in electrooculogram (EOG) artifacts. Similar to eye movements, eye blinking can interfere with brain signals to a large extent, one, due to the proximity of the eye to the brain, two, as individuals would blink 20 times per minute to keep the ocular moisture of their eyes (Karson, 1983), and these artifacts are unavoidable for prolonged tasks.

Therefore, for our task, the removal of EMG and EOG artifacts from raw EEG signals can be considered the top issue to address. There are already many algorithms (Narasimhan and Dutt, 1996; Jung et al., 2000; Schlögl et al., 2007; Ferdousy et al., 2010; Vos et al., 2010; Safieddine et al., 2012; Sweeney et al., 2012; Teng et al., 2014; Zhao et al., 2014; Chen et al., 2017; Paradeshi et al., 2017; Yang et al., 2017) for removing these two artifacts (summarized in **Table 1**), the BSS-based techniques are widely proposed because they do not require a priori knowledge and reference electrodes for EMG/EOG signals acquisition and they could separate related artifacts from EEG signal by statistical inference. Among them, CCA-based methods which more effective than ICA-based methods and other filters, taking advantage of the fact that the autocorrelation coefficient of EEG is larger than that of EMG, so it is possible to separate task-related EEG and EMG artifacts. Moreover, relevant studies (Vos et al., 2010; Urigüen and Garcia-Zapirain, 2015) have demonstrated the effectiveness of the CCA method in removing muscle artifacts during speech. EEMD-CCA (Sweeney et al., 2012) is one of the best methods the for removal of EMG and EOG artifacts for single-signals EEG signals. Although for non-single channel EEG signals, EEMD-CCA can be applied channel-by-channel, the inter-channel correlation is not captured. The later proposed MEMD-CCA (Chen et al., 2017) addressed the challenge by decomposing all channels together and then aligning the same frequency components of each channel to form multivariate IMFs before applying CCA (by setting the autocorrelation coefficient threshold, generally less than 0.9 components are set to 0) to remove the artifacts to reconstruct the EEG signals. However, it does not take into account the different degrees of influence on the EEG signals due to the distance of the artifact source from the location of the scalp electrodes (shown in **Figure 2**). Therefore, it is necessary to group the EEG channels based on brain areas and then use MEMD-CCA on each group separately.

Electroencephalogram emotion recognition systems

Electroencephalogram emotion recognition systems, mainly differ in their approach to feature extraction and choice of classifiers: a step-by-step machine learning framework (hand-crafted feature extraction, feature fusion, modeling

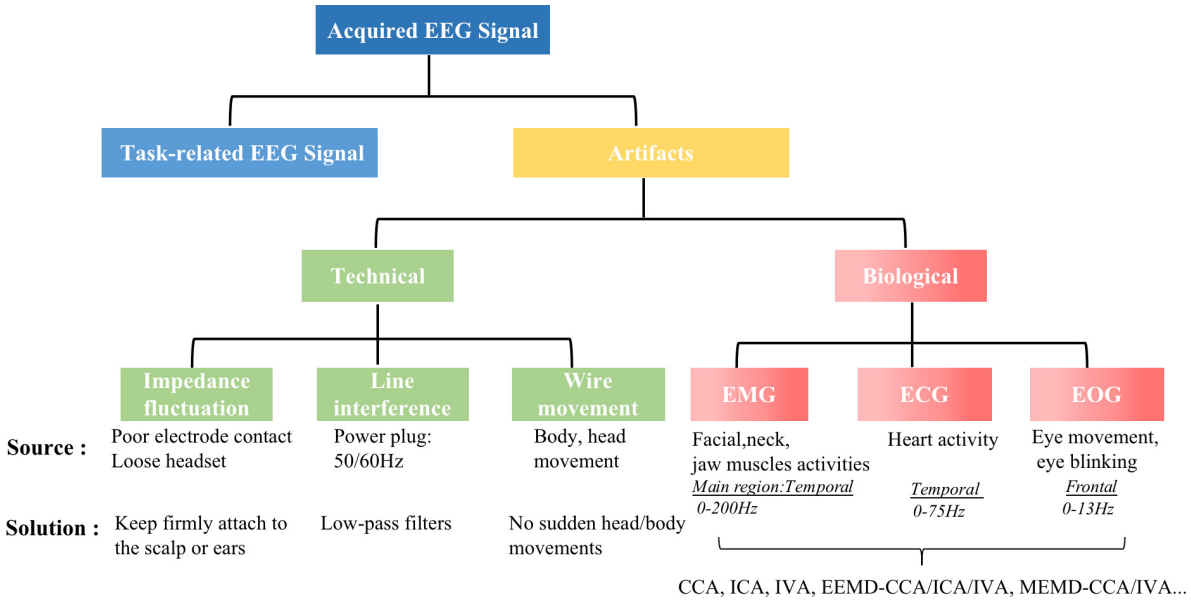

FIGURE 2
The sources of artifacts in EEG signal and the removal methods.

TABLE 1 Comparison of EMG and EOG artifacts removal techniques.

Methods		Ref. E	Channel	Comparison results
PK	NPK (BSS-based)			(Better than)
Adaptive filtering		√	All	EMG: Low-pass filter EOG: WPT, ICA, DWT, ANC (Narasimhan and Dutt, 1996; Zhao et al., 2014)
Linear regression		√	All	EOG: Visual identification (Schlögl et al., 2007)
	ICA	×	Multi	PCA, LR, Wavelet (Jung et al., 2000; Paradeshi et al., 2017)
	CCA	×	Multi	EMG: low-pass filter + Robust ICA; EOG: equivalent to ICA (Ferdousy et al., 2010; Vos et al., 2010)
	EMD	×	Single	ICA, CCA, WT (Safieddine et al., 2012)
	EEMD-CCA	×	Single	EMD, EMD-ICA, EMD-CCA, EEMD, EEMD-ICA (Sweeney et al., 2012)
	MEMD	×	Few	ICA (Teng et al., 2014)
	MEMD-CCA	×	Few	EMG: ICA, EEMD-ICA, MEMD-ICA CCA, EEMD-CCA (Chen et al., 2017)
	CCA-MEMD	×	Few	EOG:ICA, CCA (Yang et al., 2017)

PK, prior knowledge; NPK, no prior knowledge; BSS, blind source separation; Ref. E, reference electrode; ICA, independent component analysis; CCA, canonical correlation analysis; EMD, empirical mode decomposition; EEMD, ensemble empirical mode decomposition; MEMD, multivariate empirical mode decomposition.

classification) and an end-to-end deep learning framework (automatic feature extraction, feature fusion, modeling classification).

Step-by-step machine learning framework

The performance of machine learning frameworks largely depends on the quality of hand-crafted extracted features (Hosseini et al., 2020). Generally, researchers extract the EEG features from parts of the brain regions considered to contribute the most to emotions based on *a priori* knowledge of the combinatorial design. Of the most used in emotion recognition are the following two theories based on asymmetric behavior: (1) the right hemisphere dominance theory which posits right hemispheric dominance over the expression and perception, and (2) the valence theory which asserts that the right hemisphere predominantly processes negative emotions and left hemisphere predominantly processes positive emotions (Coan and Allen, 2003; Demaree et al., 2005). For example, in the study (Wang et al., 2014), the authors subtracted the power spectrum (PSD) of obtained brain waves collected from 27 pairs of symmetrical electrodes in the left and right brain regions to obtain 27 asymmetrical PSD features input to SVM classifiers. The negative and positive emotion recognition accuracy average rate was 82.38%. Later studies

(Duan et al., 2013; Zheng et al., 2014, 2017) demonstrated that the following six features: PSD, differential entropy (DE), DASM (DE(L_{eft})-DE(R_{ight})), RASM (DE(L_{eft})/DE(R_{ight})), ASM ([DASM, RASM]), DCAU(DE(F_{rontal})-DE($P_{osterior}$)) were robust and effective for EEG emotion recognition. However, the DE features achieved the highest recognition accuracy of 91.07% which is higher than the other four asymmetric features. This demonstrates ambiguity as to what degree the stimuli (pictures, music, videos, etc.) elicit neuronal processes similar to those occurring in real-life emotional experiences; making it difficult to cover all the implied features by hand-extracted features.

End-to-end deep learning framework

Recently studies began to focus on end-to-end deep learning frameworks (Craik et al., 2019). In a study (Alhagry et al., 2017), the authors proposed the use of LSTM models to automatically learn features of emotions from the context of EEG signals. They achieved average recognition accuracy of 85.65% in the valence dimension. Later in a study (Zhang et al., 2020) the authors further considered that the spatial information in the EEG signal could be used to improve the accuracy of emotion recognition and thus proposed a CNN-LSTM model. EEG raw data was first input into a CNN module (1-dimensional convolutional layer, maximum pooling layer) to extract the local spatially features which were then input into a two-layer LSTM to learn the temporal information in the spatial features. The result was an average recognition accuracy of 94.17% with a four-emotion classification. In addition, the authors input EEG raw data separately into CNN (four-layer of two-dimensional convolution; spatial features) and LSTM (four-layer; temporal features) to achieve accuracies of 90.12% and 67.47%, respectively. A later study (Sheykhivand et al., 2020) also proposed the use of CNN-LSTM for EEG raw data with the main structure of 10-1D convolutional layers plus 3 LSTM layers, achieving a recognition average accuracy of 97.42% with a two emotion classification.

From the results of the above-related studies, it was found that (a) the model automatically learns emotional features from EEG raw data better than hand-crafted extracted features, and (b) the model emotion classification recognition performance using EEG spatial-temporal features demonstrates improvements across a wide range. In addition, there are also studies that combine feature extraction and deep learning models, such as a DECNN model (Liu et al., 2020) was proposed that focuses on subject-independent emotion recognition and used extracted DDE (dynamic differential entropy) features fed into the CNNs for emotion classification. Finally, the average accuracy achieved 97.56% in EEG subject-independent emotion recognition on the SEED public dataset.

Channel selection mechanism

The number of dry electrodes used in the EEG emotion recognition systems studied above is, in general, excessive and not conducive to prolonged wear from a comfort perspective. Moreover, the EEG signals obtained with multichannel EEG devices often contain redundant, irrelevant, or interfering information (noise, overlapping/interference of signals from different electrodes) for affective analysis (Alotaiby et al., 2015). Thus, selecting the most relevant channel for emotion analysis is essential for enhancing comfort and emotion recognition accuracy.

A study (Tong et al., 2018), utilized the Relief algorithm to calculate the weight values of each channel according to the time-domain features of the EEG signal. At the cost of losing 1.6% accuracy, 13 channels with the highest contribution to emotion classification under time-domain features were selected from the initial 32 channels. Later, a study (Dura et al., 2021) used the reverse correlation algorithm applied to the band-time-domain features of 32 channels to construct a subset of electrodes with the smallest band correlation for each subject. The number of occurrences of each subset in each subject was then calculated to obtain the most common subset of channels. The smallest subset contained only four electrodes and accuracy was not affected. However, the accuracy of such channel selection methods would depend entirely on the quality of hand-extracted features. In response, the latest has research proposed to apply an attention mechanism to channel selection to prompt the network to automatically learn the most important information and improve the performance of important features. In a study (Tao et al., 2020), the authors added the channel attention module before the CNN-LSTM model to automatically learn the importance of each channel to the EEG emotion signals and then assigned weights to each channel. It was found that the FC5, P3, C4, and P8 channels contributed the most to emotion classification on the DEAP dataset (32 channels) and had an average accuracy improvement of 28.57% compared to the CNN-LSTM model without the channel attention. Later, a 3DCANN (Liu et al., 2021) framework was proposed, in which five consecutive 1s-62-channel EEG signals were fed as 3D data inputted to a CNNs module with two convolutional layers to extract spatial features, which were later output to two attention modules in the channel dimension to enhance or weaken the effect of different electrodes on emotion recognition. The model achieved an average accuracy of 96.37% for positive, negative, and neutral emotions. It is demonstrated that the attention mechanism enhances the information of the important channels and suppresses the information of the irrelevant channels for emotion analysis. However, the shortcoming is that, to get the global perspective of the temporal dimension (Time × Sample point) of the EEG signals (Time × Sample point × Channel), the channel attention module pools the EEG signals globally into 1 × 1 × Channel to get the weight matrix of the channel. This directly ignores the specific temporal information of the EEG signals, if a channel contains more noise/artifacts, it may get larger weight values instead of being conducive to the later model learning.

As our purpose is to perform emotion recognition during the conversation, even if the removal of artifacts is implemented in EEG signals, some artifacts may be still present. Therefore, attention mechanisms need to be applied simultaneously in the temporal dimension. The temporal attention mechanism will play an important role in determining "where" the need to focus attention exists. It can improve the expressiveness of the time points of changing emotional states in the EEG signal while suppressing noise/artifacts information.

Materials and methods

In this section, first, we describe the EEG dataset, the method of division of the EEG dataset, and the preprocessing of EEG signals. Then, we describe in detail the structure of each module of the proposed CTA-CNN-Bi-LSTM.

The division and preprocessing of electroencephalogram dataset

Our experiments were conducted on the dataset from the previous study (Jiang et al., 2022). Eleven older adults (six males and five females) and seven younger adults (five males and two females) were randomly pair-matched into 11 groups, and each group engaged in 36 photo conversations. The young person guided the older adult in a 1-min conversation around each photo during which time the EEG signals from the older adult were collected. After each photo conversation, the older adult also filled out an emotion evaluation form (rating of valence, arousal from −4 to 4, and the level of stress from 1 to 7). A detailed description of the EEG dataset (here named OCER) is presented in **Table 2**.

As individual differences in gender, age, economic, educational, and life circumstances would result in differences in the benchmarks for evaluating emotions, we did not classify samples by uniformly setting thresholds for the rating values on each dimension. Instead, in our experiments, the ratings of valence, arousal, and stress were first standardized using the standard deviation standardization method (*Z*-score). And the K-means method (Likas et al., 2003) was then applied to the three standardized scores for each individual, and the 36 samples were divided into three groups of positive, neutral, and negative emotions samples. Finally, the pre-processing was applied to the EEG signals in the dataset:

TABLE 2 Summary of experiment dataset (OCER).

Conversation experiment	
Trails	36 trails × 60 s
Subject	Older: 11 (M = 71.25 ± 4.66) Young: 7 (M = 22.4 ± 1.51)
Rating	Valence (−4,4), Arousal (−4,4), Stress (1,7)
EEG dataset	
Device	OpenBCI Cyton board (250 Hz/s)
Channel	F3, F4, F7, F8, T7, T8, P3, P4 (10–20 system)
Array	396(Samples) × 60 (s) × 250(Hz/s) × 8 (Channels)

Removal of technical artifacts

Electroencephalogram signals used a 1–45 Hz bandpass filter (removal of the line interference) and a Chebyshev I high-pass filter to remove baseline drift (Jiang et al., 2022).

Removal of biological artifacts

Electroencephalogram signals were divided into three groups: The frontal group (F7, F8, F3, F4), the temporal group (T7, T8), and the parietal group (P3, P4). And then each group used MEMD-CCA (Xu et al., 2017; Chen et al., 2018) methods to remove multiple artifacts (detailed in **Figure 3A**).

Segmented data

Used a 3s-non-repetitive window for segmentation of each sample (60 s trail) in the dataset. The reason is that normally the duration of an adult's emotional state does not exceed 12 s and in studies (Li et al., 2017; Tao et al., 2020) a 3s-non-repetitive sliding window applied to the EEG signals achieved excellent results for the emotion recognition task.

Preprocessing of proposed framework

First of all, before inputting the raw clean EEG dataset into the first module of the proposed framework, we normalized each raw EEG sample along the channel direction with zero-mean normalization to eliminate subject and channel differences in EEG signals and reduce computational complexity. Thus, the mean value of the processed raw EEG signals sample for each channel is 0 and the standard deviation is 1. The normalization formula for each channel is as follows:

$$X_{i,j}^{k*} = \frac{X_{i,j}^{k} - \overline{X^{k}}}{\sigma^{C}}. \quad (1)$$

Where $X_{i,j}^{k}\left(i = 1, 2, 3\ Time\left(second\right); j = 1, 2, \ldots, 250\ Sample\ Point\ \left(250Hz/s\right); k = 1, 2, \ldots, 8\ EEG\ Channel\right) \in R^{T\times P}$ represents the data of the k-*th* channel of a 3s-EEG sample. T and P are the time length and the sample point of the 3s-EEG sample respectively. $\overline{X^{k}}$ and σ^{C} are the mean and standard deviation of the k-th channel respectively.

Modules of proposed CTA-CNN-Bi-LSTM framework

The proposed framework consists of the following three modules: channel-temporal attention module, spatial feature

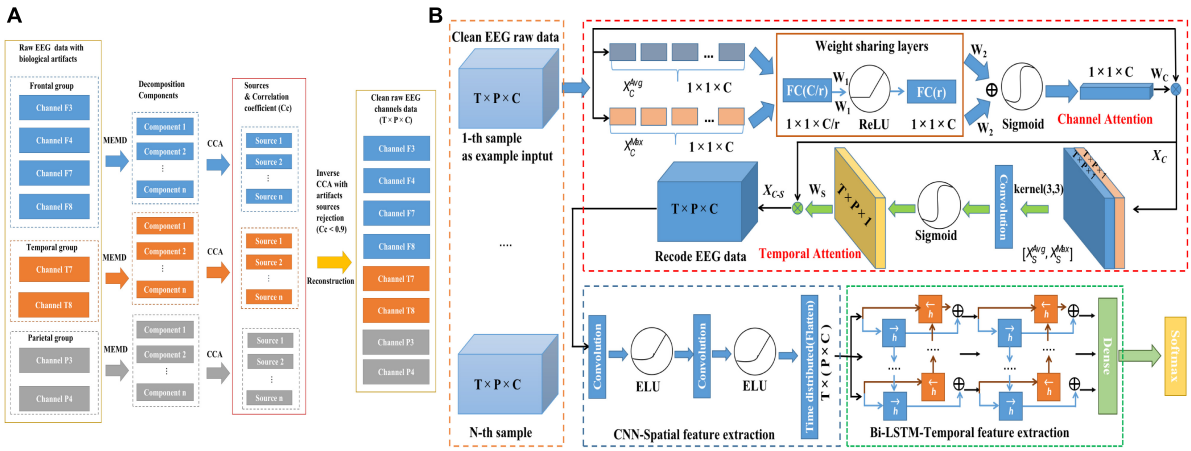

FIGURE 3
Illustration of the proposed framework based on raw EEG data for emotion recognition. **(A)** Removal of biological artifacts by MEMD-CCA. **(B)** CTA-CNN-Bi-LSTM framework.

extraction module (CNNs) and bi-directional temporal feature extraction module (Bi-LSTM). The structure of the proposed CTA-CNN-Bi-LSTM framework is shown in Figure 3B. The specific calculation process and description are as follows.

Channel-temporal attention module

An EEG sample is defined as $X \in R^{T\times P\times C}$ whereby T denotes the time duration of one EEG sample, P is the sampling points per second and C denotes the number of EEG channels. The output after the channel-temporal attention module is $X_{c\_s} \in R^{T\times P\times C}$, and the specific calculation process and descriptions are as follows. Here for our EEG dataset, the T is 3 s, P is 250 Hz/s, and C is 8 channels.

Channel attention

The global average pooling and maximum pooling are performed separately in the temporal dimension on the channel direction of X to obtain two channel statistical descriptions $X_C^{\text{Avg}}, X_C^{\text{Max}} \in R^{1\times 1\times C}$. They are then fed into a two-layer weight-sharing multi-layer perceptron (MLP): the first layer is the compression layer (the number of neurons is set to C/r to get the weight $W_1 \in R^{1\times 1\times C/r}$ and ReLU is used as the activation function. r represents the reduction ratio and here r is set to 2); The second layer is the excitation layer (the number of neurons is set to C to get the weight $W_2 \in R^{1\times 1\times C}$). Finally, these two combined features are mapped using a sigmoid activation function to generate the channel attention mapping matrix $W_c \in R^{1\times 1\times C}$ as follows:

$$W_c(\mathrm{X}) = \mathrm{Sigmoid}(W_2(\mathrm{ReLu}(W_1 \cdot X_c^{Avg}) + W_2(\mathrm{ReLu}(W_1 \cdot X_c^{Max}))). \quad (2)$$

And the output of channel attention $X_c \in R^{T\times P\times C}$ is as follows:

$$X_c = \mathrm{W}_c(X) \otimes X. \quad (3)$$

Temporal attention

Average pooling and maximum pooling are used along the channel dimension on the temporal direction to obtain $X_s^{\text{Avg}}, X_s^{\text{Max}} \in T \times P \times 1$ to stitch them together, and convolutional layers (a convolutional kernel of size 3 × 3, $K^{3\times 3}$). The sigmoid activation functions are used to generate the temporal ($T \times P$, Time × Sample Point) attention mapping matrix $W_s \in R^{T\times P\times 1}$ as follows:

$$W_s(X) = Sigmoid(K^{3\times 3}([X_s^{Avg}, X_s^{Max}])). \quad (4)$$

Thus, our final output $X_{c\_s} \in R^{T\times P\times C}$ is as follows:

$$X_{c-s} = W_s(\mathrm{X_c}) \otimes X_c. \quad (5)$$

In this way, the output shape of $X_{c-s} \in R^{T\times P\times C}$ remain unchanged and has learned what channels are important and at which time points in the channel and temporal dimension.

Convolution neural networks module

The convolution neural networks (CNNs) and their essential characteristics (spatially dependencies/local connection and weight sharing) have been widely used in various fields, especially for image tasks (object segmentation, image classification, style conversion, etc.) (Hijazi et al., 2015; Rawat and Wang, 2017). All of these applications were built based on the feature maps after the CNNs performed feature extraction for the task. Thus, essentially the role of CNNs models is to extract local spatial features of EEG signals. The specific steps of our CNNs module are as follows.

Step 1: Convolution layer

The input recorded EEG signals $X_{c-s} \in R^{T\times P\times C}$ and the convolution kernel of CNNs is defined as $filter^{k}_{(i,j)}$. k represents the number of filters, which is the same as the number of EEG channels. (i, j) is the size of the convolutional sliding window in the temporal-spatial ($T \times P$, Time × Sample Point) dimension of multi-channel EEG signals. More specifically, the k-th filter is convolved with the corresponding region in $T \times P$ dimension of the k-th channel of X_{c-s} with a window size of $i \times j$ sliding in step 1 (direction from left to right and top to bottom). The output value is obtained by adding the sum of the k channels. Finally, the feature map X^{f}_{C-S} is as follows:

$$X^{f}_{C-S} = \mathrm{f}\Big(\sum X_{C-S} \otimes K + b\Big). \tag{6}$$

The bias term is represented by b. A convolution kernel produces a feature map, and the closer the value in X^{f}_{C-S} is to 1, the more it is associated with the feature, and the closer it is to −1, the less it is associated.

In our dataset for EEG emotion recognition, the k was set 8 corresponding to the number of EEG channels. And the size of the sliding window (i, j) was set "1×10" and sliding in step 1, where the "i" was set 1 in order not to destroy the temporal features of the EEG signals at different seconds and the convolutional window to constantly move at the same second as the sampling points when sliding. Therefore, later the generated spatial feature maps have the following characteristics: (a) different spatial locations on the same channel were sharing convolutional kernel parameters (spatial independence), and (b) different convolutional kernels were used on different channels (channel specificity). This allowed each feature map of the output CNNs to learn different spatial emotion features with temporal information preserved.

Step 2: Exponential linear units layer

The exponential linear units (ELU) was selected as the activation function after the convolution layer because it is continuous and differentiable at all points and its gradient is non-zero for all negative values, meaning it does not encounter the problem of exploding or disappearing gradients during deep network learning. It achieves higher accuracy compared with other activation functions such as ReLU, Sigmoid, and tanh (Clevert et al., 2015). The ELU activation function can be written as:

$$f(x) = \begin{cases} e^{x} - 1, & x < 0 \\ x, & x \geq 0 \end{cases}. \tag{7}$$

As can be deferred from (7), the ELU function retains the values greater than or equal to 0 in the feature map X^{f}_{C-S} and assigns $e^{x} - 1$ to all the remaining values less than 0. This further suppresses the uncorrelated data in the feature map using a non-linear activation function.

To ensure that the temporal information contained in the extracted spatial feature maps is not reduced during the input temporal feature extraction module (Bi-LSTM), we did not use the pooling layer often used in CNN structures. Thus, our spatial feature extraction module used two convolution-ELU layers.

Bi-directional long short-term memory module

LSTM networks (Hochreiter and Schmidhuber, 1997) have been widely used in time series related tasks, such as disease prediction (Chimmula and Zhang, 2020) and air quality prediction (Yan et al., 2021). This is because, unlike previous feedforward neural networks (one-way propagation, where the input and output are independent of each other), LSTM networks have internally inclusive memory units (the state of the current time step is jointly determined by the input of that time step and the output of the previous time step). LSTM is, effectively, a gating algorithm added to the memory unit of a traditional RNN which solves the problem of long sequences in which the gradient disappears and explodes during the training process of the RNN model (Bengio et al., 1994). The memory unit of LSTM is as follows:

$$Z_{forget} = sigmoid(W_f \left[h_{t-1}, X_t\right] + b_f), \tag{8}$$

$$Z_{input} = sigmoid(W_i \left[h_{t-1}, X_t\right] + b_i), \tag{9}$$

$$Z_{output} = sigmoid(W_o \left[h_{t-1}, X_t\right] + b_o), \tag{10}$$

$$Z = tanh(W \left[h_{t-1}, X_t\right] + b_c), \tag{11}$$

$$C_t = Z_{forget} C_{t-1} + Z_{input} Z, \tag{12}$$

$$h_t = Z_{output} tanh(C_t), \tag{13}$$

$$y_t = \sigma\left(W_t h_t\right). \tag{14}$$

where Z_{forget}, Z_{input}, Z_{output} are vectors of data input from the current state and input data received from the previous node multiplied by the weights and then converted to values from 0 to 1 by a sigmoid activation function which acts as a gating function (0 means complete discard of information, 1 means complete retention of information). Thus Z_{forget} determines which information in C_{t-1} needs to be forgotten; Z_{input} determines which new information in X_t needs to be recorded; Z_{output} determines which information is the output of the current state. Z is converted to a value between −1 and 1 by the tanh activation function as the input data for C_t. In addition, C_t (cell state) and h_t (hidden state) represent the two transmission states of the memory cell to the next cell of the LSTM. y_t is obtained from h_t by σ transformation and represents the output of the memory cell.

However, for the EEG emotion recognition task, the current emotional state is correlated with both previous and subsequent information due to the latency of the device during signal acquisition. Bi-LSTM (Schuster and Paliwal, 1997) is an extension of LSTM consisting of a forward LSTM layer (fed the sequence, left to right) and a backward LSTM layer (reversed fed the sequence, from right to left) which can solve this problem. The out layer of the memory unit of Bi-LSTM is as follows:

$$y_t = \sigma(W_t(\underset{h_t}{\rightarrow} + \underset{h_t}{\leftarrow}). \quad (15)$$

TABLE 3 Division of OCER into three motions by K-MEANS.

Subject	Rating	Clustering center (Z-score)		
ID	scale	1	2	3
1	Valence	−1.58	0.26	1.27
	Arousal	−0.98	0.65	1.85
	Stress	−0.90	−0.90	−0.90
2	Valence	−1.58	0.55	1.16
	Arousal	−0.98	−1.12	1.08
	Stress	−0.90	−0.90	−0.90
3	Valence	0.63	0.61	0.63
	Arousal	−2.40	−0.96	0.44
	Stress	0.23	−0.80	−0.90
4	Valence	−0.20	0.63	1.02
	Arousal	−0.41	−0.56	0.62
	Stress	0.23	0.46	0.38
5	Valence	−1.58	−1.53	−0.11
	Arousal	−3.11	−1.07	−0.98
	Stress	0.23	0.78	1.36
6	Valence	−0.35	0.22	0.63
	Arousal	−0.63	0.44	0.60
	Stress	2.30	0.73	2.01
7	Valence	−3.05	−1.58	−0.48
	Arousal	−0.98	−0.98	−0.98
	Stress	−0.90	−0.83	−0.90
8	Valence	0.14	−0.11	0.35
	Arousal	0.91	−0.27	0.44
	Stress	1.36	−0.90	0.23
9	Valence	−1.39	−0.45	0.51
	Arousal	−0.81	−0.60	−0.20
	Stress	2.78	1.10	0.41
10	Valence	−0.48	0.49	1.36
	Arousal	−0.27	0.91	1.69
	Stress	−0.90	−0.85	−0.90
11	Valence	−0.11	0.63	1.36
	Arousal	0.40	0.76	1.14
	Stress	0.23	0.23	0.23

All results are retained to 2 decimal places. The larger the score of Valence indicates the more positive; the larger the score of Arousal indicates the greater emotional intensity (no positive or negative directionality); the larger the score of Stress indicates the greater stress (negative directionality).

In addition, the LSTM contains overly numerous parameters, and the Bi-LSTM is twice as large as the LSTM, so it is easy to overlearn to produce the overfitting problem. The most common solution in deep learning is the utilization of dropout regularization (Hinton et al., 2012) which temporarily disconnects the input-hidden layer-output layer with a certain probability. However, temporarily dropping layer-to-layer connections in recurrent neural networks may cause direct loss of some of the previous memory. Therefore, we use the recurrent dropout method (Semeniuta et al., 2016) to act on the memory units; temporarily dropping a part of the links in h_t (hidden state) with probability p at each time step. This ensures that the output y_t does not lose the earlier important information while simultaneously solving the overfitting problem. Therefore, the features of past and future emotion information through this structure were combined in the out layer. Here, our temporal feature extraction module consists of two layers of internal memory cell units (32 and 16 respectively) with a 0.2 recurrent dropout rate of the bidirectional LSTM.

In summary, our proposed framework can automatically extract meaningful features for emotion classification from raw clean EEG data. Firstly, a channel-temporal attention mechanism is used to infer attention weights for raw EEG signals X successively along the channel and temporal dimensions and got re-coded EEG signals X_{c-s} , which improves the points of time representation of significant channel and emotional state changes. Next, CNNs (including two convolution-ELU layers) are used to extract spatial features of X_{c-s} to get feature maps X^f_{C-S}. Finally, all spatial feature maps X^f_{C-S} were packaged in time series input into a two-layer Bi-LSTM with a recurrent dropout function to learn temporal information from the spatial features maps for EEG emotion recognition.

Results and analysis

Firstly, we describe the division and preprocessing of the EEG dataset. Secondly, we displayed the result of the channel attention weights in the channel-temporal attention module. Finally, we introduce designed four groups of deep learning methods for demonstrating the validity of each module of our proposed method.

The division and preprocessing of electroencephalogram dataset

The standardization of the scores of the rating scale [Valence (–4,4), Arousal (–4,4) and Stress (1,7)] and the classification of emotions using K-means for 36 samples of each participant was completed in IBM SPSS Statistics (version 26). The related results were displayed in **Table 3**. Each participant's 36 trials were divided into three categories respectively: Clustering "1"

TABLE 4 The emotion classification of OCER and the data arrays.

Emotion classification	
Negative	72 samples (60 s)
Neutral	180 samples (60 s)
Positive	138 samples (60 s)
3s-dataset arrays	
Dataset	7800(seg) × 3(s) × 250(Hz/s) × 8(channels)
Label	7800 × 3(Negative, Neutral, Positive)

represented the "negative emotion"; Clustering "2" represented the "neutral emotion"; and Clustering "3" represented the "positive emotion". The advantage of using this method is that instead of using the equal criteria for all participants, each participant's criteria was used to classify the emotions. Therefore, there are 73 negative samples, 182 neutral samples and 141 positive samples in the EEG dataset (OCER). Then, after removing the technical and biological artifacts in the EEG dataset by using the method mentioned in section the division and preprocessing of electroencephalogram dataset, we found that artifacts of the 36-th trail from subject 4, the 11-th trail from subject 6, the 31-th and 32-th trails from subject 8, the 23-th trail from subject 9 and the 35-th trail from subject 10 could not be removed cleanly (the amplitude of EEG signals more than 200 μV) and they were excluded. Finally, we cut each clean trail using a 3-s non-repeating window. Therefore, the array of the EEG dataset became 7800 (390 × 20 segments) × 3 (seconds) × 250 (sample points/s) × 8 (channels). The details were shown in **Table 4**.

Electroencephalogram channel attention weights

To illustrate the different degrees of importance of each EEG signal channel for emotion recognition, the mean of ten times weight calculations of the channel attention in the channel-temporal attention module for OCER are shown in **Figure 4**. As shown, the weights of the channels for different emotions were significantly different. The EEG signals of the channels corresponding to the right brain regions (except F4 is less than 0.5) contributed more to positive emotions. The EEG signals the channels corresponding to the left brain regions (except P3 less than 0.5) contributed more to negative emotions. And the weights of channel F4 and channel F3 achieved significant advantages in neutral emotion. Further to demonstrate the contribution of different channels to the emotions, the one-way ANOVA was implemented on the 8-channel weights of the three emotions respectively. The weights of channel F8 and F7 had a significant ($F = 3.55$, $p < 0.01$) in negative emotion, channel P4 and T7 had a significant ($F = 3.39$, $p < 0.01$) in positive emotion, a non-significant for 8-channel in neutral emotion. This suggests that there are variances in the contribution of channels to different emotions and utilizing channel attention mechanisms could enhance the ability to discriminate between different emotions.

Parameters of proposed and baseline methods

The proposed framework mainly was implemented with the Keras module based on the TensorFlow framework and trained on NVIDIA GeForce GTX GPU. At first, each batch size (here denoted by None) of samples defined as (None, 3,250,8) was input into the CTA module and the output shape was the same as (None, 3,250,8). The samples were then fed into the CNN module which used the AdaBelief optimizer (Zhuang et al., 2020) with a learning rate of 1e-3 and the epsilon of 1e-7 to minimize the cross-entropy loss function. In order not to destroy the temporal information in the EEG signals, the size of the convolution kernel was set to 1×10×8 (height, width, depth) and the number of kernels was 8 the same as the number of EEG signal channels. This makes the number of channels of the feature maps of the output CNNs consistent with the number of channels of the EEG raw data. This causes each feature map in the input Bi-LSTM with a shape of (None, 3,250,8), which means that the time step is 3 and the features map was split up into three feeds that can be expressed as (None, 3,250 × 8). Therefore, the Bi-LSTM module further extracts the temporal features from the spatial feature maps containing temporal information. In the Bi-LSTM module, the dimension of the hidden states of the LSTM in each direction (forward/backward) of the two layers were 32 and 16, respectively. In addition, in every LSTM the recurrent dropout rate was set as 0.2. Initially, the input batch size is 10 and the epoch is set at 200. And the early stopping technique (Prechelt, 1998) is used during the training process: the training is stopped when the loss value of the test set no longer decreases in two epochs.

To demonstrate the validity of the three modules in the proposed framework, four groups of experiments were implemented: **Group A:** LSTM, Channel Attention-LSTM (CA-LSTM) and Channel-temporal Attention (CTA)-LSTM;

TABLE 5 Baseline and proposed method for EEG dataset emotion recognition.

	Channel attention	**Temporal attention**	**CNN**	**LSTM/Bi-LSTM**
RNN	×	×	×	√
C-RNN	√	×	×	√
CTA-RNN	√	√	×	√
CNN-RNN	×	×	√	√
C-CNN-RNN	√	×	√	√
Proposed method	√	√	√	√

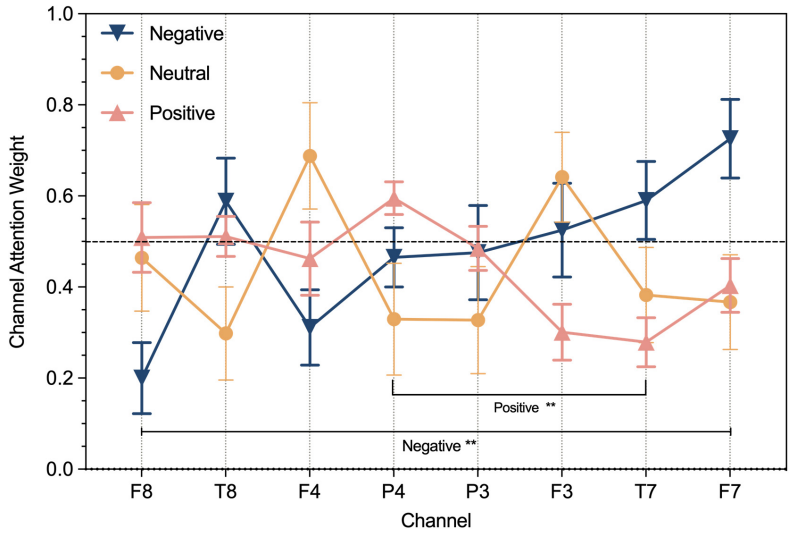

FIGURE 4
The result of channel weight on OCER dataset respectively for negative, neutral, and positive emotions. $^{**}P < 0.01$.

Group B: Bi-LSTM, CA-Bi-LSTM, and CTA-Bi-LSTM; **Group C:** CNN-LSTM, CA-CNN-LSTM, CTA-CNN-LSTM; **Group D:** CNN-Bi-LSTM, CA-CNN-Bi-LSTM and CTA-CNN-Bi-LSTM. Here the LSTM and Bi-LSTM are referred to generically as RNN. Except for the different number of RNN layers, the model including the CNN-RNN module used two layers of LSTM/Bi-LSTM with hidden states of dimensions 32 and 16. The model including the RNN module used 3 layers of LSTM/Bi-LSTM with hidden states of dimensions 64, 32, and 16. Because each layer of Bi-LSTM is combined with two directions of LSTM, the training parameters are twice as large as the LSTM. All models use the same parameter settings as the proposed method. Specific structures and parameters were shown in Tables 5, 6.

Results of experiments

Our work aimed to evaluate individuals' emotions during the periodic implementation of reminiscence therapy, in which our focus was on the individual's emotion recognition accuracy, rather than an emotional recognition model to accommodate all older adults. Therefore, the subject-dependent method was utilized for EEG emotion recognition. All samples of the OCER dataset (Table 4) were divided into training sets and test set based on the 10-fold cross-validation method. This method randomly divided the dataset into ten equal parts (nine parts as the training set and one part as the test set) and this process repeated 10 times. Finally, the number of samples in the training set was 7020 and the number of samples in the test set was 780.

The average accuracy of the results of the 10 test sets was used as the evaluation metric for the performance of the proposed framework and baseline methods. Further, a one-way ANOVA was performed on the results of average emotion recognition accuracy rate for four groups of 12 models to explore whether there was a significant in EEG emotion recognition across models. The detailed results were shown in Figure 5. As seen from the figure, for the average accuracy of emotion recognition on the OCER dataset: (1) Significance among groups A, B, C, D ($F = 372.8, p < 0.0001$); (2) Significance among the models within groups A, B, C, D ($p < 0.01$ or $p < 0.0001$), except between the CNN-LSTM model and the CA-CNN-LSTM model in group C ($p = 0.7756$, non-significant). (3) The proposed framework CTA-CNNN-Bi-LSTM in group D achieved the best emotion recognition accuracy with 98.75% on three emotions.

To demonstrate the contribution and performance of each module of the proposed framework on the recognition of negative, neutral, and positive emotions, we implemented confusion matrices on all models (see Figure 6). As can be seen, vertically, from the base models (LSTM, Bi-LSTM, CNN-LSTM, CNN-Bi-LSTM) to the front of the models adding the channel attention mechanism (CA) and adding the channel-temporal module (CTA), the recognition accuracy improves for almost in the three emotions. Specifically, the recognition accuracy of negative, neutral and positive emotions improved by at least 18%, 2%, and 9%, respectively. It proved the effectiveness of the CTA module in improving the model's performance in distinguishing between different emotions. Horizontally, from left to right, from the LSTM series models to the CNN-Bi-LSTM series models (except for the CA-CNN-LSTM model), the recognition accuracy improves for almost the three emotions. Specifically, the recognition accuracy of negative, neutral and positive emotions

10.3389/fnagi.2022.945024

TABLE 6 Array and total parameters of 3s-dataset (OCER) fed into different models.

Model	Input array	Main layers	Total params
RNN	None × 3 × 2000	3 Unit (64,32,16)	544,243/1108,963
C/CTA-RNN	(None × 3 × 250 × 8)reshape (None × 3 × 2000)	3 Unit (64,32,16)	544,243/1108,963
-CNN-RNN	(None × 3 × 250 × 8)reshape (None × 3 × 2000)	2 Conv (K = 8(1,10))2 Unit (32,16)	648 × 2 + 263411/530915

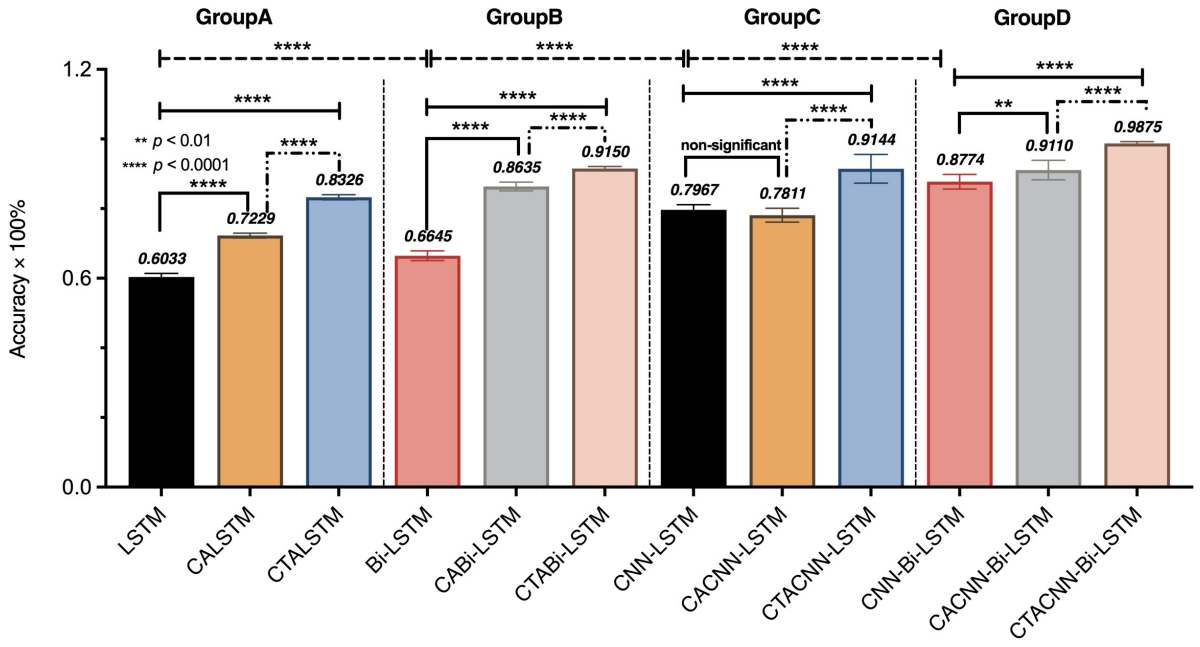

FIGURE 5
The results of mean accuracy (%) and one-way ANOVA in baseline and proposed methods on 3s-data set. The symbol ** means that $P < 0.01$ and is a statistically significant difference. The symbol **** means that $P < 0.0001$ is an extremely significant statistical difference.

improved by at least 21%, 9%, and 10%, respectively. It sufficiently demonstrated the superiority of CNN-Bi-LSTM in integrating the bi-directional temporal features (past and future information features) on the spatial features information in the EEG signals information to determine the current emotional state. The CTA-CNN-Bi-LSTM model achieved the best accuracy of emotion recognition for negative, neutral, and positive emotions.

Furthermore, to indicate the performance of the proposed framework on the emotion recognition of each individual, we conducted experiments on each individual. As Figures 7–9 show, the proposed framework CTA-CNN-Bi-LSTM almost achieved the best accuracy of emotion recognition for negative, neutral and positive emotions on each subject. In addition, for the base models, the RNN models did not perform well below 60% for each individual on negative emotions, but after adding the CTA module before the RNN models the individual's negative emotion recognition rate with an accuracy of more than 80%. And the CNN-RNN series models perform better than the RNN series in terms of positive emotion for each subject. There was no such significant tendency in negative emotion and neutral emotion for each subject. For the negative emotion, the CA-Bi-LSTM model performed better than the CA-CNN-Bi-LSTM on subjects 5, 6, and 11. For the neutral emotion, the CTA-Bi-LSTM model performed better than the CTA-CNN-LSTM model on subjects 1, 5, 6, 7, and 8. However, the proposed framework performed best on each individual for the three emotions.

Discussion

Principal results and limitations

The experimental results reveal that the CTA-CNN-Bi-LSTM framework performs better in EEG emotion recognition as the proposed framework combined consideration of the spatial features and two directions' temporal features which were extracted from the channels and temporal dimension of EEG signals most relevant to emotions.

In the first module of our proposed framework, the channel-temporal attention module applied to the clean EEG

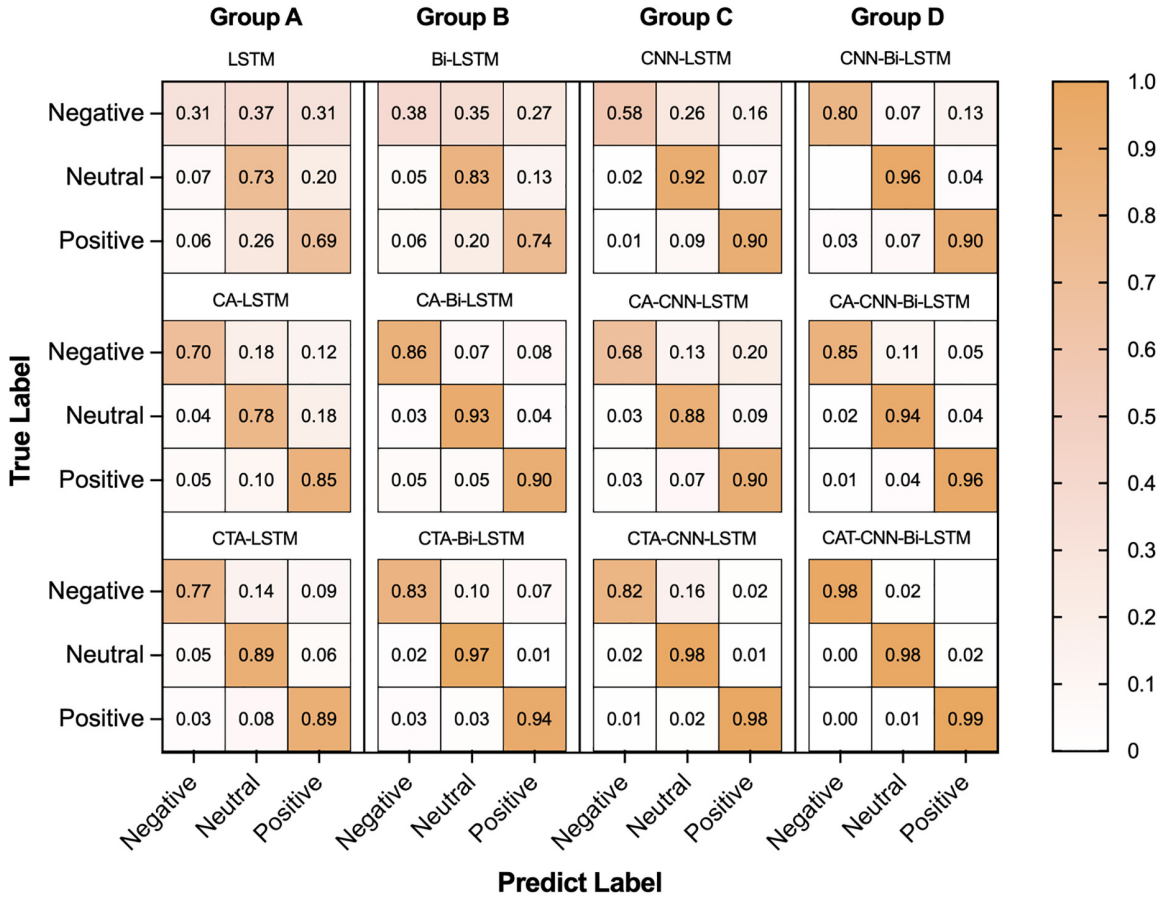

FIGURE 6
The result of confusion matrixes of negative, neutral, and positive emotions in baseline and proposed method.

raw data emphasized meaningful feature information and suppresses irrelevant information in both channel and temporal dimensions. Firstly, the channel weights were calculated under the global average pooling and global maximum pooling in the temporal dimension to obtain two channel statistical descriptions (two different angles of the global field of view). In contrast, the channel attention module in the previous study (Tao et al., 2020) only conducted global average pooling on the temporal dimension (T8 and F8 dominated among 14 electrodes in the DREAMER dataset; FC5, P3, C4, and P8 dominated among 40 electrodes in DEAP), which may have resulted in the inability to distinguish the contribution of channels to different emotions. The channel weights in this study were calculated so that the weights of F3 and F4 achieved significant advantages in neutral emotion. For negative emotions, channel weights greater than 0.5 are F3, T7, F7, and T8, meanwhile the weights of channels F8 (0.2) and F7 (0.73) had a significant ($F = 3.55$, $p < 0.01$) in negative emotion, which both suggested that they played a major role in the channels corresponding to the left-brain area. For positive emotions, the channel weights greater than 0.5 were F8, T8, and P4, and channel P4 (0.6) and T7 (0.28) had a significant ($F = 3.39$, $p < 0.01$) in positive emotion, which indicate that the right-brain area corresponding to the channel was dominant. These findings are consistent with previous studies: (1) The valence theory stated that left-brain areas predominantly process negative emotions and right-brain areas process positive emotions (Demaree et al., 2005); (2) EEG signals in the frontal lobe, lateral temporal lobe, and parietal lobe brain regions of the brain were the most informative on different emotions (Lin et al., 2010; Zheng et al., 2017; Özerdem and Polat, 2017; Tong et al., 2018). If it is necessary to reduce electrodes while ensuring a high recognition rate of emotions, the intersection of all emotion-dominated channels or channels with significant differences can be selected. This means that F3, F4, F7, and F8 can be chosen for the task of our study. Other tasks can recalculate channel weights according to this method.

When the recoded EEG data obtained directly using the channel attention is used for subsequent model learning, as shown in **Figure 5**, the average accuracy of the CA-RNN/CA-CNN-RNN model improved only slightly compared to RNN/CNN-RNN, except the CA-CNN-LSTM model was slightly lower than the CNN-LSTM model. However, from

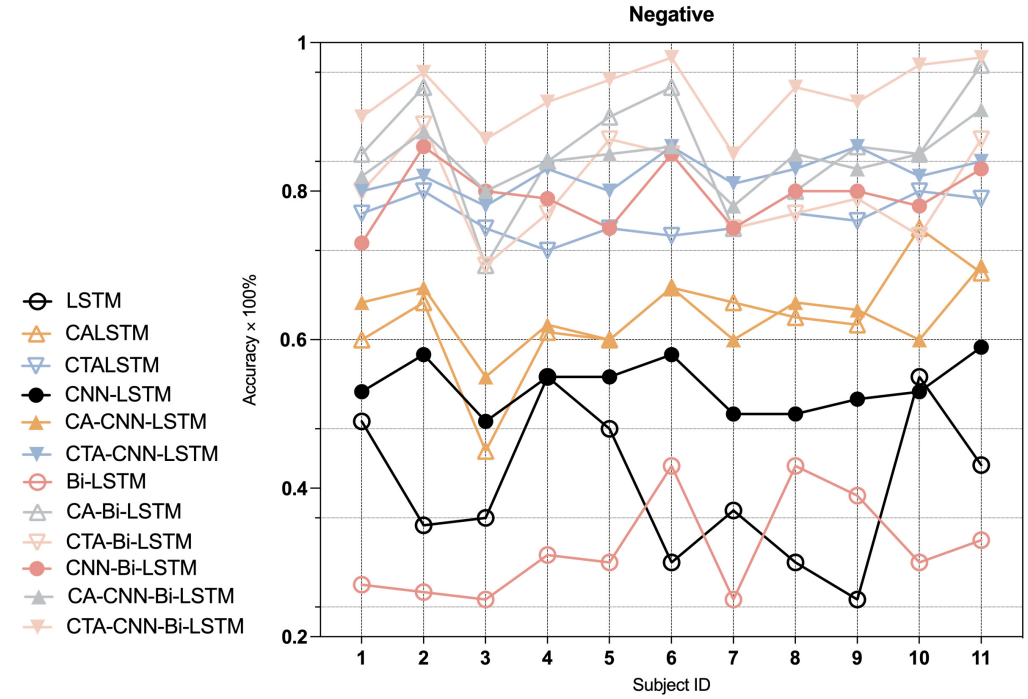

FIGURE 7
Average accuracy (%) of baseline and proposed method on the recognition of negative emotion in each individual.

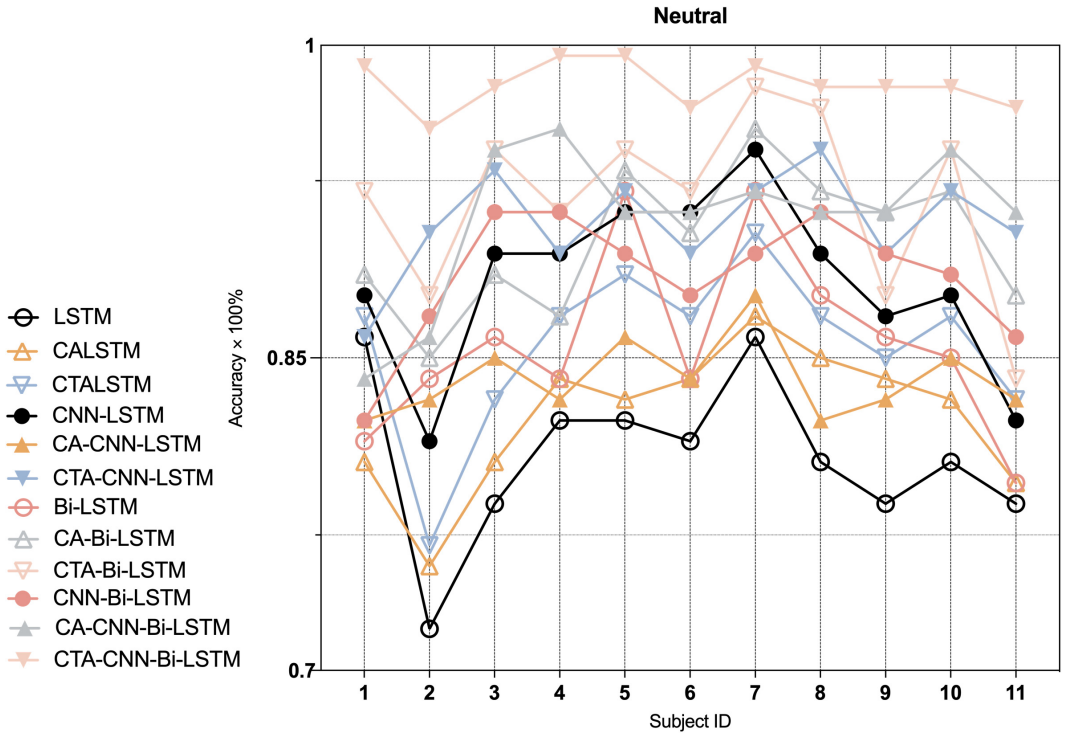

FIGURE 8
Average accuracy (%) of baseline and proposed method on the recognition of neutral emotion in each individual.

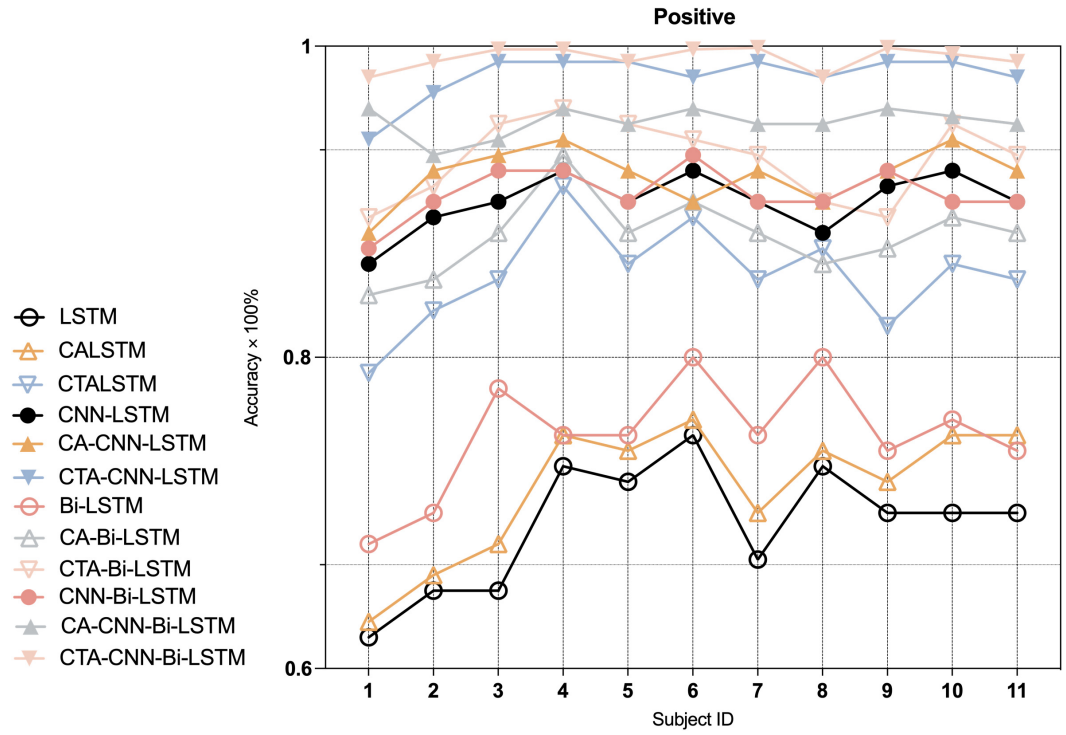

FIGURE 9
Average accuracy (%) of baseline and proposed method on the recognition of positive emotion in each individual.

Figure 6, the CA-CNN-LSTM improved the recognition accuracy of negative emotion by 10% over the CNN-LSTM model. The average accuracy of CTA-RNN/CTA-CNN-RNN not only increased but also achieved the minimum variance, demonstrating that the temporal attention mechanism did improve the representation of emotional state change time points in EEG signals while further suppressing the noise/artifact information. And the results were higher than the accuracy results of the previously mentioned related studies using channel selection (Alotaiby et al., 2015; Tong et al., 2018; Tao et al., 2020; Dura et al., 2021). Therefore, the EEG raw data was processed by the channel-temporal attention module to emphasize meaningful feature information and suppress irrelevant information in both channel and temporal dimensions.

In the second and third modules of our proposed framework, the recoded EEG signals (containing information on the most relevant channel and temporal dimensions to the task) from the channel-temporal attention module were fed into the CNNs and RNN to extract spatial and temporal features for emotion recognition. As **Table 6** and **Figure 5** shown, the training parameters of CNN-RNN without the channel-temporal attention mechanism (264,707/532,211) were much smaller than those of RNN (544,243/1108,963), while the average accuracy was substantially higher than that of RNN (19.34% and 21.29% improvement). This, as has been shown in previous studies (Sheykhivand et al., 2020; Zhang et al., 2020; Ramzan and Dawn, 2021), demonstrates that it is necessary to consider both spatial and temporal information of EEG signals for emotion recognition. And the CA-CNN-RNN models achieved an average accuracy of 78.11% (1.56% lower than CNN-LSTM model and 10% improvement on negative emotion) and 91.11% (3.37% improvement over CNN-Bi-LSTM model), respectively. It was further demonstrated that channel attention suppresses the information of irrelevant channels and enhances emotional information. Finally, the results of the CTA-CNN-Bi-LSTM model proposed in this study achieved the highest average accuracy of 98.75%. It further demonstrated that channel-temporal attention suppresses both the information of irrelevant channels and the irrelevant information of temporal dimensions. The CTA-CNN-Bi-LSTM model with an improvement of 7.25% over the CTA-CNN-LSTM model. The reason is that Bi-LSTM model learned the temporal information on the spatial feature map from both forward and reverse directions while LSTM model learned from only one direction in the forward direction. This is consistent with the conclusions in the study (Siami-Namini et al., 2019): Bi-LSTM model outperforms the LSTM model on temporal series forecasting tasks. As our experiments also employed 10-fold cross-validation, the average accuracy standard deviation values can more objectively demonstrate that the proposed

framework has a high emotion recognition performance. From the result of confusion matrixes of negative emotion (Figure 6), it was found that the recognition rate of the basic RNN and CNN-RNN models on negative emotion was far lower than the other two emotions, firstly, the number of samples of negative emotion was lower than the other two emotions, and secondly, negative emotion seemed to be easily misclassified as neutral emotion. However, through the channel attention mechanism (CA) and the channel-temporal attention (CTA), the recognition of negative emotions with small samples is enhanced and the accuracy rate is further improved. Finally, we conducted experiments on each individual, and the proposed framework CTA-CNN-Bi-LSTM almost achieved the best accuracy of emotion recognition for negative, neutral, and positive emotions on each subject.

In summary, EEG raw 3s-dataset achieved the highest accuracy of 98.75% by the proposed method CTA-CNN-Bi-LSTM. It included channel-temporal attention module (CTA), spatial feature extraction (CNNs) and Bi-LSTM. The proposed method improved the average accuracy by 38.42% compared to the LSTM model. Of which, the channel-temporal attention module (CTA-CNN-Bi-LSTM) led to an average accuracy improvement of 11.01% for CNN-Bi-LSTM. The convolutional module (CNN-Bi-LSTM) resulted in an average accuracy improvement of 21.29% for Bi-LSTM. And the bi-directional LSTM module (CNN-Bi-LSTM) led to an improvement of 8.07% in CNN-LSTM. It indicates that the convolution module (spatial information of the EEG signal) provides the largest contribution (21.29%) to the accuracy improvement of the framework. The bi-directional LSTM module after the CNN module provides little enhancement (8.07%) to the framework. However, the addition of the channel-temporal attention module before the convolution module (by suppressing the irrelevant channel information and temporal dimensional noise) led to a further significant improvement (11.01%) in the accuracy of the model while reducing the std. dev. to a minimum. Thus, our proposed framework was demonstrated to be effective in extracting spatial and temporal information from recoded EEG signals (including most relevant channels and temporal dimensional information to emotion) for emotion recognition. However, our framework used the dataset divided using the subject-dependent method as the usage scenario of our task, and it has not been demonstrated whether the same high performance of emotion recognition can be achieved on the dataset divided by the subject-independent method.

Conclusion and future work

The proposed framework in this paper used clean raw EEG signals (removal of muscle artifacts by MEMD-CCA) as input to an end-to-end deep learning method (without feature engineering) for emotion recognition. The proposed CTA-CNN-Bi-LSTM framework considered both spatial features and bidirectional temporal features in the channel dimension and temporal dimensions that were most relevant to emotions in the raw EEG signals. At first, the channel-temporal attention module suppresses the channel information in both the EEG signal that is not related to emotion and the noise in the spatial dimension in each channel. Later, the CNN-RNN module first extracts the spatial features in the recoded EEG signals and then feeds them into the Bi-LSTM network in order. Therefore, the Bi-LSTM learned the temporal information simultaneously from two directions (forward LSTM for previous information and reverse LSTM for future information) on the spatial feature maps. Finally, the results of four group experiments have demonstrated that CTA-CNN-Bi-LSTM improved EEG emotion recognition compared to other methods and achieved the highest average accuracy of 98.75% for negative, positive, and neutral emotion recognition. Therefore, the proposed framework reduces the emotion-independent information and noise in the channel and temporal dimensions, CTA-CNN-Bi-LSTM significantly improved the accuracy of emotion recognition in the dataset compared with existing methods.

However, this work may not achieve high emotion recognition accuracy for new individuals and requires retraining the model/fine-tuning the model to achieve it, which is not conducive to later applications of real-time emotion monitoring. In future work, after collecting EEG signals from more individuals, perhaps self-supervised learning models such as a contrastive learning model which learns knowledge on its own from unlabeled data, could be used to potentially realize a plug-and-play real-time EEG emotion recognition system. It focuses on learning the common features between similar examples and distinguishing the differences between non-similar examples to construct an encoder. This encoder has the ability to encode similar data of the same category and make the encoding results of different categories of data as different as possible.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LJ and NK designed the experiments. LJ carried out the experiments. LJ, PS, and FZ analyzed the results. LJ and DC

prepared the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by JSPS KAKENHI grant number 19H04154.

Acknowledgments

We would like to express our gratitude to the older people and the young people who assisted in completing the experiment.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alarcao, S. M., and Fonseca, M. J. (2017). Emotions recognition using EEG signals: A survey. *IEEE Trans. Affect. Comput.* 10, 374–393. doi: 10.1109/TAFFC.2017.2714671

Alhagry, S., Fahmy, A. A., and El-Khoribi, R. A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. *Emotion* 8, 355–358. doi: 10.14569/IJACSA.2017.081046

Alotaiby, T., Abd El-Samie, F. E., Alshebeili, S. A., and Ahmad, I. (2015). A review of channel selection algorithms for EEG signal processing. *EURASIP J. Adv. Signal Process.* 2015:66. doi: 10.1186/s13634-015-0251-9

Athavipach, C., Pan-Ngum, S., and Israsena, P. (2019). A wearable in-ear EEG device for emotion monitoring. *Sensors* 19:4014. doi: 10.3390/s19184014

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* 5, 157–166. doi: 10.1109/72.279181

Boateng, G., and Kowatsch, T. (2020). "Speech emotion recognition among elderly individuals using transfer learning," in *Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI'20 Companion)*, (Switzerland: University of St. Gallen), 17–21. doi: 10.1145/3395035.3425255

Caroppo, A., Leone, A., and Siciliano, P. (2020). Comparison between deep learning models and traditional machine learning approaches for facial expression recognition in ageing adults. *J. Comput. Sci. Technol.* 35, 1127–1146. doi: 10.1007/s11390-020-9665-4

Chen, X., Chen, Q., Zhang, Y., and Wang, Z. J. (2018). A novel EEMD-CCA approach to removing muscle artifacts for pervasive EEG. *IEEE Sens. J.* 19, 8420–8431. doi: 10.1109/JSEN.2018.2872623

Chen, X., Xu, X., Liu, A., McKeown, M. J., and Wang, Z. J. (2017). The use of multivariate EMD and CCA for denoising muscle artifacts from few-channel EEG recordings. *IEEE Trans. Instrument. Meas.* 67, 359–370. doi: 10.1109/TIM.2017.2759398

Chimmula, V. K. R., and Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* 135:109864. doi: 10.1016/j.chaos.2020.109864

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv*. [Preprint].

Coan, J. A., and Allen, J. J. (2003). "The state and trait nature of frontal EEG asymmetry in emotion. The state and trait nature of frontal EEG asymmetry in emotion," in *The Asymmetrical Brain*, eds K. Hugdahl and R. J. Davidson (Cambridge, MA: MIT Press), 565–615.

Cornelius, R. R. (1991). Gregorio Marafion's Two-Factor Theory of Emotion. *Pers. Soc. Psychol. Bull.* 17, 65–69. doi: 10.1177/0146167291171010

Craik, A., He, Y., and Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* 16:031001. doi: 10.1088/1741-2552/ab0ab5

Demaree, H. A., Everhart, D. E., Youngstrom, E. A., and Harrison, D. W. (2005). Brain lateralization of emotional processing: Historical roots and a future incorporating "dominance. *Behav. Cogn. Neurosci. Rev.* 4, 3–20. doi: 10.1177/1534582305276837

Duan, R. N., Zhu, J. Y., and Lu, B. L. (2013). "Differential entropy feature for EEG-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, (New York, NY: IEEE), 81–84. doi: 10.1109/NER.2013.6695876

Dura, A., Wosiak, A., Stasiak, B., Wojciechowski, A., and Rogowski, J. (2021). "Reversed Correlation-Based Pairwised EEG Channel Selection in Emotional State Recognition"," in *International Conference on Computational Science*, (Germany: Springer), 528–541. doi: 10.1007/978-3-030-77967-2_44

Ferdousy, R., Choudhory, A. I, Islam, M. S., Rab, M. A., and Chowdhory, M. E. H. (2010). "Electrooculographic and electromyographic artifacts removal from EEG," in *2010 2nd International Conference on Chemical, Biological and Environmental Engineering*, (New York, NY: IEEE), 163–167. doi: 10.1109/ICBEE.2010.5651351

Halliday, D. M., Conway, B. A., Farmer, S. F., and Rosenberg, J. R. (1998). Using electroencephalography to study functional coupling between cortical activity and electromyograms during voluntary contractions in humans. *Neurosci. Lett.* 241, 5–8. doi: 10.1016/S0304-3940(97)00964-6

Hijazi, S., Kumar, R., and Rowen, C. (2015). *Using Convolutional Neural Networks for Image Recognition*. San Jose, CA: Cadence Design Systems Inc

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv*. [Preprint].

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hosseini, M.-P., Hosseini, A., and Ahi, K. (2020). A review on machine learning for EEG signal processing in bioengineering. *IEEE Rev. Biomed. Eng.* 14, 204–218. doi: 10.1109/RBME.2020.2969915

Huhta, J. C., and Webster, J. G. (1973). "60-Hz interference in electrocardiography," in *IEEE Transactions on Biomedical Engineering*, (New York, NY: IEEE:), 91–101. doi: 10.1109/TBME.1973.324169

Iwamoto, M., Kuwahara, N., and Morimoto, K. (2015). Comparison of burden on youth in communicating with elderly using images versus photographs. *Int. J. Advan. Comput. Sci. Appl.* 6, 168–172. doi: 10.14569/IJACSA.2015.061023

Jiang, L., Siriaraya, P., Choi, D., and Kuwahara, N. (2022). Emotion Recognition Using Electroencephalography Signals of Older People for Reminiscence Therapy. *Front. Physiol.* 12:823013. doi: 10.3389/fphys.2021.823013

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178. doi: 10.1111/1469-8986.3720163

Kamel, N., and Malik, A. (2014). "The fundamentals of EEG signal processing," in *EEG/ERP Analysis: Methods and Applications*, eds N. Kamel and A. Malik (Boca Raton, FL: CRC Press), 21–71. doi: 10.1201/b17605-3

Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain* 106, 643–653. doi: 10.1093/brain/106.3.643

Knight, B. G., and Sayegh, P. (2010). Cultural values and caregiving: The updated sociocultural stress and coping model. *J. Gerontol. B* 65, 5–13. doi: 10.1093/geronb/gbp096

Kouris, I., Vellidou, E., and Koutsouris, D. (2020). "SMART BEAR: A large scale pilot supporting the independent living of the seniors in a smart environment," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, (New York, NY: IEEE), 5777–5780. doi: 10.1109/EMBC44109.2020.9176248

Leahy, F., Ridout, N., Mushtaq, F., and Holland, C. (2018). Improving specific autobiographical memory in older adults: Impacts on mood, social problem solving, and functional limitations. *Aging Neuropsychol. Cogn.* 25, 695–723. doi: 10.1080/13825585.2017.1365815

Lee, K. J., and Lee, B. (2013). "Removing ECG artifacts from the EMG: A comparison between combining empirical-mode decomposition and independent component analysis and other filtering methods," in *2013 13th International Conference on Control, Automation and Systems (ICCAS 2013)*, (New York, NY: IEEE), 181–184. doi: 10.1109/ICCAS.2013.6703888

Li, Y., Huang, J., Zhou, H., and Zhong, N. (2017). Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. *Appl. Sci.* 7:1060. doi: 10.3390/app7101060

Likas, A., Vlassis, N., and Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern Recognize.* 36, 451–461. doi: 10.1016/S0031-3203(02)00060-2

Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., et al. (2010). EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* 57, 1798–1806. doi: 10.1109/TBME.2010.2048568

Liu, F., and Cai, H.-D. (2010). Integration Models of Peripheral and Central Nervous System in Research on Physiological Mechanisms of Emotions. *Adv. Psychol. Sci.* 18:616.

Liu, S., Wang, X., Zhao, L., Li, B., Hu, W., Yu, J., et al. (2021). "3DCANN: A spatio-temporal convolution attention neural network for EEG emotion recognition," in *IEEE Journal of Biomedical and Health Informatics*, (New York, NY: IEEE), 1–1. doi: 10.1109/JBHI.2021.3083525

Liu, S., Wang, X., Zhao, L., Zhao, J., Xin, Q., and Wang, S.-H. (2020). Subject-independent emotion recognition of EEG signals based on dynamic empirical convolutional neural network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1710–1721. doi: 10.1109/TCBB.2020.3018137

Narasimhan, S., and Dutt, D. N. (1996). Application of LMS adaptive predictive filtering for muscle artifact (noise) cancellation from EEG signals. *Comput. Electr. Eng.* 22, 13–30. doi: 10.1016/0045-7906(95)00030-5

Noguchi, T., Saito, M., Aida, J., Cable, N., Tsuji, T., Koyama, S., et al. (2021). Association between social isolation and depression onset among older adults: A cross-national longitudinal study in England and Japan. *BMJ Open* 11:e045834. doi: 10.1136/bmjopen-2020-045834

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Am. Psychol.* 17:776. doi: 10.1037/h0043424

Özerdem, M. S., and Polat, H. (2017). Emotion recognition based on EEG features in movie clips with channel selection. *Brain Inform.* 4, 241–252. doi: 10.1007/s40708-017-0069-3

Paradeshi, K., Scholar, R., and Kolekar, U. (2017). "Removal of ocular artifacts from multichannel EEG signal using wavelet enhanced ICA," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, (New York, NY: IEEE), 383–387. doi: 10.1109/ICECDS.2017.8390150

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* 17, 715–734. doi: 10.1017/S0954579405050340

Prechelt, L. (1998). "Early stopping-but when," in *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, eds G. Montavon, G. B. Orr, and K. R. Müller (Berlin: Springer). doi: 10.1007/3-540-49430-8_3

Ramzan, M., and Dawn, S. (2021). Fused CNN-LSTM deep learning emotion recognition model using electroencephalography signals. *Int. J. Neurosci.* 27, 1–11. doi: 10.1080/00207454.2021.1941947

Rawat, W., and Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* 29, 2352–2449. doi: 10.1162/neco_a_00990

Rodriguez-Bermudez, G., and Garcia-Laencina, P. J. (2015). Analysis of EEG signals using nonlinear dynamics and chaos: A review. *Appl. Math. Inf. Sci.* 9:2309.

Safieddine, D., Kachenoura, A., Albera, L., Birot, G., Karfoul, A., Pasnicu, A., et al. (2012). Removal of muscle artifact from EEG data: Comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches. *EURASIP J. Adv. Signal Process.* 2012, 1–15. doi: 10.1186/1687-6180-2012-127

Santini, Z. I., Koyanagi, A., Tyrovolas, S., Mason, C., and Haro, J. M. (2015). The association between social relationships and depression: A systematic review. *J. Affect. Disord.* 175, 53–65. doi: 10.1016/j.jad.2014.12.049

Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., and Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clin. Neurophysiol.* 118, 98–104. doi: 10.1016/j.clinph.2006.09.003

Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45, 2673–2681. doi: 10.1109/78.650093

Semeniuta, S., Severyn, A., and Barth, E. (2016). "Recurrent dropout without memory loss," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, (Osak: The COLING 2016 Organizing Committee), 1757–1766.

Sheykhivand, S., Mousavi, Z., Rezaii, T. Y., and Farzamnia, A. (2020). Recognizing emotions evoked by music using CNN-LSTM networks on EEG signals. *IEEE Access* 8, 139332–139345. doi: 10.1109/ACCESS.2020.3011882

Siami-Namini, S., Tavakoli, N., and Namin, A. S. (2019). "The performance of LSTM and BiLSTM in forecasting time series," in *2019 IEEE International Conference on Big Data (Big Data)*, (New York NY: IEEE), 3285–3292. doi: 10.1109/BigData47090.2019.9005997

Sitaram, R., Lee, S., Ruiz, S., Rana, M., Veit, R., and Birbaumer, N. (2011). Real-time support vector classification and feedback of multiple emotional brain states. *Neuroimage* 56, 753–765. doi: 10.1016/j.neuroimage.2010.08.007

Stytsenko, K., Jablonskis, E., and Prahm, C. (2011). "Evaluation of consumer EEG device Emotiv EPOC," In *Proceedings of the MEI CogSci conference*, Ljubljana, 99.

Sugisawa, H., Shibata, H., Hougham, G. W., Sugihara, Y., and Liang, J. (2002). The impact of social ties on depressive symptoms in US and Japanese elderly. *J. Soc. Issues* 58, 785–804. doi: 10.1111/1540-4560.00290

Surangsrirat, D., and Intarapanich, A. (2015). "Analysis of the meditation brainwave from consumer EEG device," in *SoutheastCon 2015*, (New York, NY: IEEE), 1–6. doi: 10.1109/SECON.2015.7133005

Sweeney, K. T., McLoone, S. F., and Ward, T. E. (2012). The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique. *IEEE Trans. Biomed. Eng.* 60, 97–105. doi: 10.1109/TBME.2012.2225427

Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., et al. (2020). "EEG-based emotion recognition *via* channel-wise attention and self attention," in *IEEE Transactions on Affective Computing*, (New York, NY: IEEE). doi: 10.1109/TAFFC.2020.3025777

Teng, C., Zhang, Y., and Wang, G. (2014). "The removal of EMG artifact from EEG signals by the multivariate empirical mode decomposition," in *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, (New York, NY: IEEE), 873–876. doi: 10.1109/ICSPCC.2014.6986322

Thierry, G., and Roberts, M. V. (2007). Event-related potential study of attention capture by affective sounds. *Neuroreport* 18, 245–248. doi: 10.1097/WNR.0b013e328011dc95

Tong, L., Zhao, J., and Fu, W. (2018). "Emotion recognition and channel selection based on EEG Signal"," in *2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, (New York, NY: IEEE), 101–105. doi: 10.1109/ICICTA.2018.00031

Urigüen, J. A., and Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *J. Neural Eng.* 12:031001. doi: 10.1088/1741-2560/12/3/031001

Van Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. *Psychophysiology* 38, 22–34. doi: 10.1111/1469-8986.3810022

Van Dis, E. A., Van Veen, S. C., Hagenaars, M. A., Batelaan, N. M., Bockting, C. L., Van Den Heuvel, R. M., et al. (2020). Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders: A systematic review and meta-analysis. *JAMA Psychiatry* 77, 265–273. doi: 10.1001/jamapsychiatry.2019.3986

Vos, D. M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.-X., Huffel, V. S., et al. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics* 8, 135–150. doi: 10.1007/s12021-010-9071-0

Wang, X.-W., Nie, D., and Lu, B.-L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing* 129, 94–106.

Westermann, R., Spies, K., Stahl, G., and Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *Eur. J. Soc. Psychol.* 26, 557–580. doi: 10.1002/(SICI)1099-0992(199607)26:4<557::AID-EJSP769>3.0.CO;2-4

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, eds V. Ferrari., M. Hebert., C. Sminchisescu., Y. Weiss, (Cham: Springer), 3–19. doi: 10.1007/978-3-030-01234-2_1

Xu, X., Liu, A., and Chen, X. (2017). "A novel few-channel strategy for removing muscle artifacts from multichannel EEG data," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, (New York, NY: IEEE), 976–980. doi: 10.1109/GlobalSIP.2017.8309106

Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., and Li, F. (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Syst. Appl.* 169:114513. doi: 10.1016/j.eswa.2020.114513

Yang, B., Zhang, T., Zhang, Y., Liu, W., Wang, J., and Duan, K. (2017). Removal of electrooculogram artifacts from electroencephalogram using canonical correlation analysis with ensemble empirical mode decomposition. *Cogn. Comput.* 9, 626–633. doi: 10.1007/s12559-017-9478-0

Zhang, Y., Chen, J., Tan, J. H., Chen, Y., Chen, Y., Li, D., et al. (2020). An investigation of deep learning models for EEG-based emotion recognition. *Front. Neurosci.* 14:622759. doi: 10.3389/fnins.2020.622759

Zhao, Q., Hu, B., Shi, Y., Li, Y., Moore, P., Sun, M., et al. (2014). Automatic identification and removal of ocular artifacts in EEG—improved adaptive predictor filtering for portable applications. *IEEE Trans. Nanobioscience* 13, 109–117. doi: 10.1109/TNB.2014.2316811

Zheng, W.-L., Zhu, J.-Y., and Lu, B.-L. (2017). Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 10, 417–429. doi: 10.1109/TAFFC.2017.2712143

Zheng, W. L., Zhu, J. Y., Peng, Y., and Lu, B. L. (2014). "EEG-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, (New York, NY: IEEE), 1–6. doi: 10.1109/ICME.2014.6890166

Zhuang, J., Tang, T., Ding, Y., Tatikonda, S. C., Dvornek, N., Papademetris, X., et al. (2020). Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. *Adv. Neural Inform. Process. Syst.* 33, 18795–18806.