EEG-Based Emotion Classification Using a Deep Neural Network and Sparse Autoencoder

Junxiu Liu [1,2], Guopei Wu [1,2], Yuling Luo [1,2], Senhui Qiu [1,2,3], Su Yang [4], Wei Li [5,6] and Yifei Bi [7,8]*

[1] School of Electronic Engineering, Guangxi Normal University, Guilin, China, [2] Guangxi Key Lab of Multi-Source Information Mining & Security, Guangxi Normal University, Guilin, China, [3] Guangxi Key Laboratory of Wireless Wideband Communication and Signal Processing, Guilin, China, [4] Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China, [5] Academy for Engineering & Technology, Fudan University, Shanghai, China, [6] Department of Electronic Engineering, The University of York, York, United Kingdom, [7] College of Foreign Languages, University of Shanghai for Science and Technology, Shanghai, China, [8] Department of Psychology, The University of York, York, United Kingdom

Emotion classification based on brain–computer interface (BCI) systems is an appealing research topic. Recently, deep learning has been employed for the emotion classifications of BCI systems and compared to traditional classification methods improved results have been obtained. In this paper, a novel deep neural network is proposed for emotion classification using EEG systems, which combines the Convolutional Neural Network (CNN), Sparse Autoencoder (SAE), and Deep Neural Network (DNN) together. In the proposed network, the features extracted by the CNN are first sent to SAE for encoding and decoding. Then the data with reduced redundancy are used as the input features of a DNN for classification task. The public datasets of DEAP and SEED are used for testing. Experimental results show that the proposed network is more effective than conventional CNN methods on the emotion recognitions. For the DEAP dataset, the highest recognition accuracies of 89.49% and 92.86% are achieved for valence and arousal, respectively. For the SEED dataset, however, the best recognition accuracy reaches 96.77%. By combining the CNN, SAE, and DNN and training them separately, the proposed network is shown as an efficient method with a faster convergence than the conventional CNN.

Keywords: EEG, emotion recognition, convolutional neural network, sparse autoencoder, deep neural network

1. INTRODUCTION

The Brain–Computer Interface (BCI) directly connects human (or animal) brain activity with artificial effectors (Kübler et al., 2009), which provides an interactive pathway between the human brain and external devices for various applications. The process of such an interaction starts by recording the brain activity through the signal processing and analysis to detect the users' intent (Tabar and Halici, 2016). BCI systems and their various implementations have been subjects of ongoing study for decades, and one of the most appealing research directions is emotion recognition due to its potential applications in numerous scenarios. Both non-physiological and physiological signals could be employed for emotion detections. Non-physiological signals include facial expression images (Lane et al., 1997), voice signals (Scherer, 1995), and body gesture (Cheng and Liu, 2008). Compared to the non-physiological signals, physiological signals can be detected by some wearable devices, such as an electroencephalogram (EEG) (Zheng, 2017),

electromyogram (Hiraiwa et al., 1989), electrocardiogram (Agrafioti et al., 2012), the galvanic skin response, blood volume pressure, and a photoplethysmogram. Among these physiological signals, EEG signals have been widely used for research into emotion recognition (Chi et al., 2012; Huang et al., 2015; Li et al., 2016; Liu et al., 2018c). Captured from the scalp by a number of EEG electrodes, emotion could be reflected immediately by an EEG signal once a subject receives the stimulations.

There are two conventional rules to follow when categorizing human emotions, namely, the discrete basic emotion description and the dimension approaches. According to the discrete basic emotion description approach, emotions can be classified into six basic emotions: sadness, joy, surprise, anger, disgust, and fear (van den Broek, 2013). For the dimension approach, the emotions can be classified into two (valence and arousal) or three dimensions (valence, arousal, and dominance) (Zheng and Lu, 2015). Among these dimensions, valence describes the level of positivity or negativity of one person, and arousal describes the level of excitement or apathy of emotion. The scale of dominance ranges from submissive (without control) to dominance (empowered). The emotion recognition is usually based on the dimension approach because of its simplicity compared to the discrete basic emotion description (Zheng and Lu, 2015).

Early works on emotion recognition through analysing EEG signal could be traced back to more than 50 years ago (Fink, 1969). Many new methods on feature extraction and classification have recently been proposed for emotion detection (Petrantonakis and Hadjileontiadis, 2010). For the feature extraction, two types of feature are commonly used to analyze EEG signals: time-domain and frequency-domain features. Time-domain features capture the temporal information of signals, such as the fractal dimension (Hjorth, 1970), Hjorth, and higher-order crossing features (Petrantonakis and Hadjileontiadis, 2010). The frequency-domain features can extract the useful information from the frequency perspective under different frequency bands. For instance, the EEG signal could be decomposed into δ (1–3 Hz), θ (4–7 Hz), α (8–13 Hz), β (14–30 Hz), and γ bands (31–50 Hz) (Hjorth, 1970; Li and Lu, 2009; Petrantonakis and Hadjileontiadis, 2010; Nie et al., 2011), where the features can be extracted from each of them. In addition, other features, such as Deep Forest (Zhou and Feng, 2017), Statistical Characteristics (SC), Differential Entropy (DE) feature (Zheng et al., 2014), Pearson Correlation Coefficient (PCC) feature (Lewis et al., 2007), and Principal Component Analysis (PCA) (Subasi and Gursoy, 2010), are also used in emotion recognitions.

In the meantime, various classification methods have been used for emotion recognition, such as k-Nearest Neighbor (Bahari and Janghorbani, 2013), Multi-Layer Perceptron (Orhan et al., 2011). A Support Vector Machine (SVM) and Linear Regression (LR) were used in Wang et al. (2019), but recognition accuracy can be improved. In recent years, deep neural networks (DNN) (Tripathi et al., 2017) has been developed into one of the most effective and popular methods in many research fields (Fu et al., 2017; Liu et al., 2018a,b, 2019; Luo et al., 2018). Convolutional Neural Networks (CNN) are widely used in

computer vision, image classifications, visual tracking (Danelljan et al., 2016), segmentation, and object detections (Girshick et al., 2014). EEG emotion classification using the CNN method was also explored in the approaches of Tripathi et al. (2017). Cascade and parallel convolutional recurrent neural networks have been used for EEG human-intended movements classification tasks (Zhang et al., 2018). Additionally, before applying the CNN, EEG data could be converted to image representation after feature extraction (Tabar and Halici, 2016). However, the accuracy of emotion recognition by using only CNN is not high. In the work of Zhang et al. (2017), a deep learning framework consisting of the sparse autoencoder (SAE) and logistic regression was used to classify EEG emotion status. The sparse autoencoder was employed for feature extraction, and logistic regression was used to predict affective states. The SAE is an unsupervised machine learning algorithm. By calculating the error between the output of the SAE and original input, data could be reconstructed and useful features could be extracted for classification task. However, accuracy of that work is not high and there are no experiments for comparing to verify the work of the SAE.

In this work, a novel network model combining the CNN, SAE, and DNN to convert EEG time series into 2D images for a good emotion classification performance is proposed. The EEG signal is decomposed into several different bands. Based on frequency, time, and location information, the 2D features are extracted from EEG data. Then convolutional layers of the CNN are trained and used for further extracting features. The SAE is used for reconstructing data obtained from convolutional layers, and the DNN is used for classification. Compared to other approaches, the proposed neural network model, which leverages the benefits of convolutional layers of the CNN and sparsity of the SAE, demonstrates a good classification accuracy and fast convergence. The procedure of the proposed method is summarized in **Figure 1**. Original EEG data are pre-processed, and features are extracted for deep learning model. After training and testing on the model, final classification results are obtained.

The rest of this paper is organized as follows: the proposed neural network model is presented in section 2. Datasets and experimental results are provided in section 3. Section 4 summaries the work and discusses the future work.

2. DEEP LEARNING FRAMEWORK

In this section, fundamental principles and essential network modules are presented. The novel model is also introduced in detail.

2.1. Convolution Neural Network (CNN)
The features extracted from original EEG data are sent to the CNN first. The CNN model includes several convolution-pooling layer pairs and one output layer. Before sending to the CNN, features are concatenated into image form which is then convolved with several one-dimensional filters in convolution layers. After the pooling layer, the data are further subsampled to images with smaller size. Network weights and filters in the convolution layers are learned through back-propagation algorithm.

FIGURE 1 | Emotion classification procedure in this work.
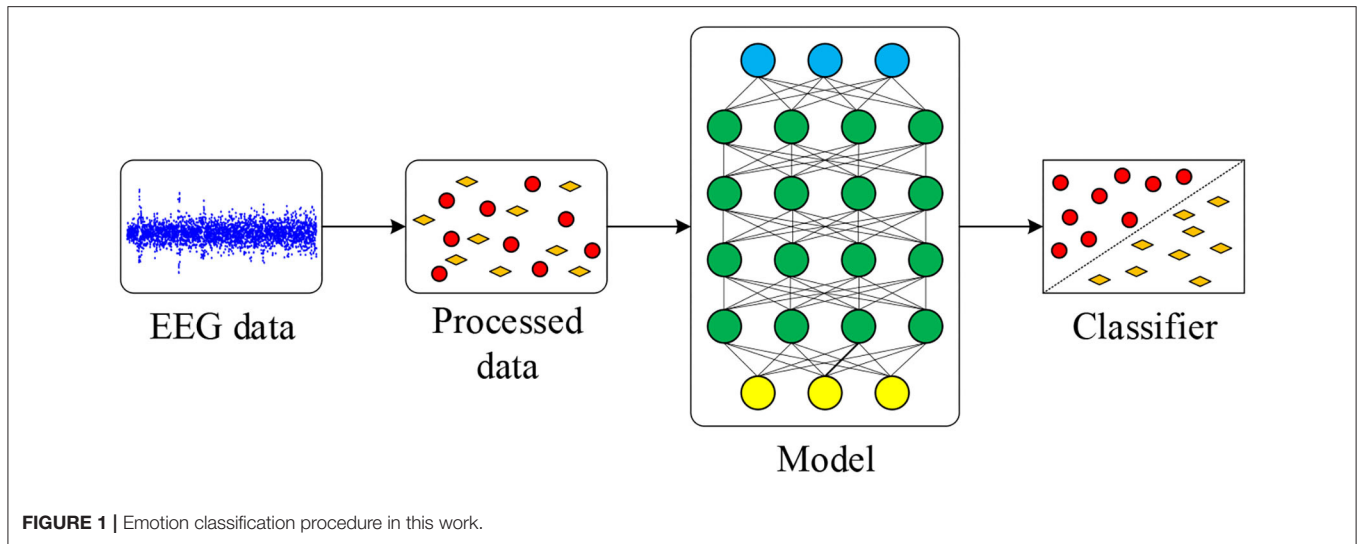
In our experiments, data extracted from EEG signal are from four typical frequency bands, which include α (1–7 Hz), β (8–13 Hz), θ (14–30 Hz), and γ bands (30–45 Hz), using a Butterworth band-pass filter. After that, data are reformed into two-dimensional features, such as PCC, which are the input for CNN. Detailed methods of this procedure is presented in sections 3.3 and 3.4. It is worth noting that the two-dimensional features contain not only the frequency but also spatial location information of each electrode (Tabar and Halici, 2016). To preserve this information, one-dimensional filtering is applied in this work instead of two-dimensional filtering.

The CNN structure is relatively straightforward. Input vector is two-dimensional feature, which can be given by

$$x = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}, \tag{1}$$

where $m \times n$ is the shape of input vector x. The input two-dimensional feature is convolved with filters W_k at the convolution layer, which is given by

$$W_k = \begin{pmatrix} W_{11} \\ W_{21} \\ \dots \\ W_{i1} \end{pmatrix}, \tag{2}$$

where i is length of W_k and $i < m$ in Equation (1). After the image convolution, output map is formed and the feature map at the given layer is obtained by

$$f(\alpha) = f(W_k \times x + b_k), \tag{3}$$

where $W_k \in R^{i \times 1}$ is the weight matrix and b_k is the bias value, k denotes the filter, for $k = 1, 2, \dots, n$ and n denotes the total number of filtering in convolutional layer. The activation function is f, which is a rectified linear unit (ReLU) function

in this work. Compared with the traditional neural network activation functions, such as *sigmoid* and *tanh*, *ReLU* is more efficient in avoiding gradient disappearance. *ReLU* function is defined by

$$f(\alpha) = ReLU(\alpha) = \ln(1 + e^\alpha), \tag{4}$$

where α is defined in Equation (3). At the max-pooling layer, the feature map is down sampled through the max-pooling function. Max-pooling is used because it is found that the maximum value from the selected values of a given feature map could be effectively extracted using this function.

After the last pooling layer, a fully connected layer follows in which output data from pooling layer is flattened. After that, fully connected layers named DNN are followed. In DNN, the activation function of each layer is also *ReLU*. For the output layer, because there are two classification tasks, including binary classification and multi-class classification, *sigmoid* and *softmax* are used, respectively. For the binary-classification task, *Adadelta* is used as an optimizer, and loss is calculated by binary crossentropy, which is given by

$$loss = -\sum_{n=1}^{N} \hat{y}_i log y_i + (1 - \hat{y}_i) log(1 - \hat{y}_i), \tag{5}$$

where N is number of samples, y_i is the value, which is a form of one-hot code, and $\hat{y}_i$ is the output from the output layer where *sigmoid* is used. For multi-class classification, such as three-class classification, *Adam* is used as an optimizer, and loss is calculated by categorical crossentropy, which is given by

$$loss = -\sum_{n=1}^{N} \hat{y}_{i1} log y_{i1} + \hat{y}_{i2} log y_{i2} + \hat{y}_{i3} log y_{i3}, \tag{6}$$

where N is number of samples, y_{i1}, y_{i2}, y_{i3} are values of the label, which is also a form of one-hot code, and $\hat{y}_{i1}, \hat{y}_{i2}$, and $\hat{y}_{i3}$ are three outputs from the output layer where *softmax* is used. Parameters

FIGURE 2 | The autoencoder includes one input, one hidden, and one output layer.
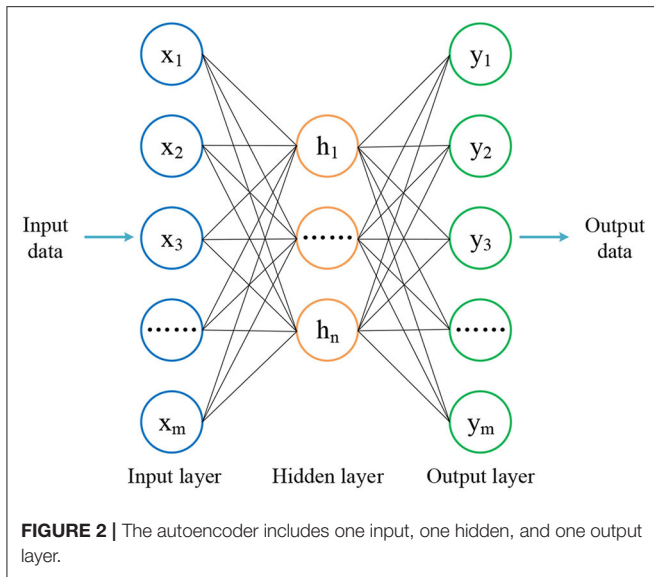
in the model are updated by using back-propagation algorithm. The error between the desired output and the actual output is computed and the gradient descent method is applied to update parameters in order to minimize the error. Functions to update the weight and bias are shown by

$$W_k = W_k - \eta \frac{\partial E}{\partial W_k}, \qquad (7)$$

$$b_k = b_k - \eta \frac{\partial E}{\partial b_k}, \qquad (8)$$

where W_k is the weight matrix, b_k is the bias, and η represents the learning rate, E is the error. E is equal to *loss* in Equations (5) and (6). The results obtained from this CNN will be used as a benchmark for the performance comparison in section 3.

2.2. Sparse Autoencoder (SAE)

An autoencoder is a network including one input, one hidden, and one output layer, which is used to preserve the essence of the input data as much as possible and remove the potential noise in an unsupervised manner. The output data are therefore simplified, and important information from the input data are retained, which is beneficial for classification.

The structure of autoencoder is shown in **Figure 2**. The whole data processing is divided into encoding and decoding phases. In the encoding phase, the dimension of input data are reduced in one layer. When the decoded data arrives at the hidden layer, the dimension of input data reaches the same as the number of neurons predefined for this layer. The encoding function of the hidden layer, h, is defined by

$$h = encoder(x) = f(W_k \times x + b_k), \qquad (9)$$

where $W_k \in R^{m \times n}$ is the weight matrix between input layer and the next layer. As defined previously in CNN, b_k is also the bias

vector, and f represents the output function. The output function used in this part is *ReLU*, which is similar to the activation of the CNN. Differently from the encoding phase, in the decoding phase, the same number of neurons in output layers should be set as that of layers in encoding phase, in order to guarantee the output data has the same dimension as the input data. The decoding function is shown by

$$y = decoder(x) = g(W_k \times x + b_k), \qquad (10)$$

where $W_k \in R^{n \times m}$. After encoding and decoding phases, the model is trained, and the parameters could be obtained by minimizing the cost function, which is defined by

$$min \sum \left| E(x^i, y^i) \right|, \qquad (11)$$

where y^i is output data and x^i is original input data. When the network is trained, output values are reconstructed, and the shape of which is equal to that of input data. Parameters of the model could be updated according to

$$W_k = W_k - \eta \frac{\partial E(x^i, y^i)}{\partial W_k}, \qquad (12)$$

$$b_k = b_k - \eta \frac{\partial E(x^i, y^i)}{\partial b_k}, \qquad (13)$$

where η denotes the learning rate of the network. E is an error in the SAE. For details of optimizer and E, they are the same as that in section 2.1 in binary-classification task.

In order to increase the generalization of the network and improve the training efficiency of the proposed network, a sparse constraint on the activity of the hidden representations is added in this work. Sparse constraint helps suppress activation of neurons in the hidden layer, and useful features can be extracted by autoencoder. Thus, the cost function in sparse autoencoder is described by

$$J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{m} KL(\rho || \hat{\rho}_j), \qquad (14)$$

where $\hat{\rho}_j$ is the average activation of hidden unit j, ρ is the sparsity level, and β is the weight of the sparsity penalty term. *KL* is the Kullback–Leibler divergence, which ensures the sparsity of neurons in hidden layer. *KL* is defined by

$$KL(\rho || \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \frac{1 - \rho}{1 - \hat{\rho}_j}, \qquad (15)$$

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} f_j(x^i), \qquad (16)$$

where m denotes the number of samples at unite j in the hidden layer, and f_j denotes the activation of hidden neuron j.
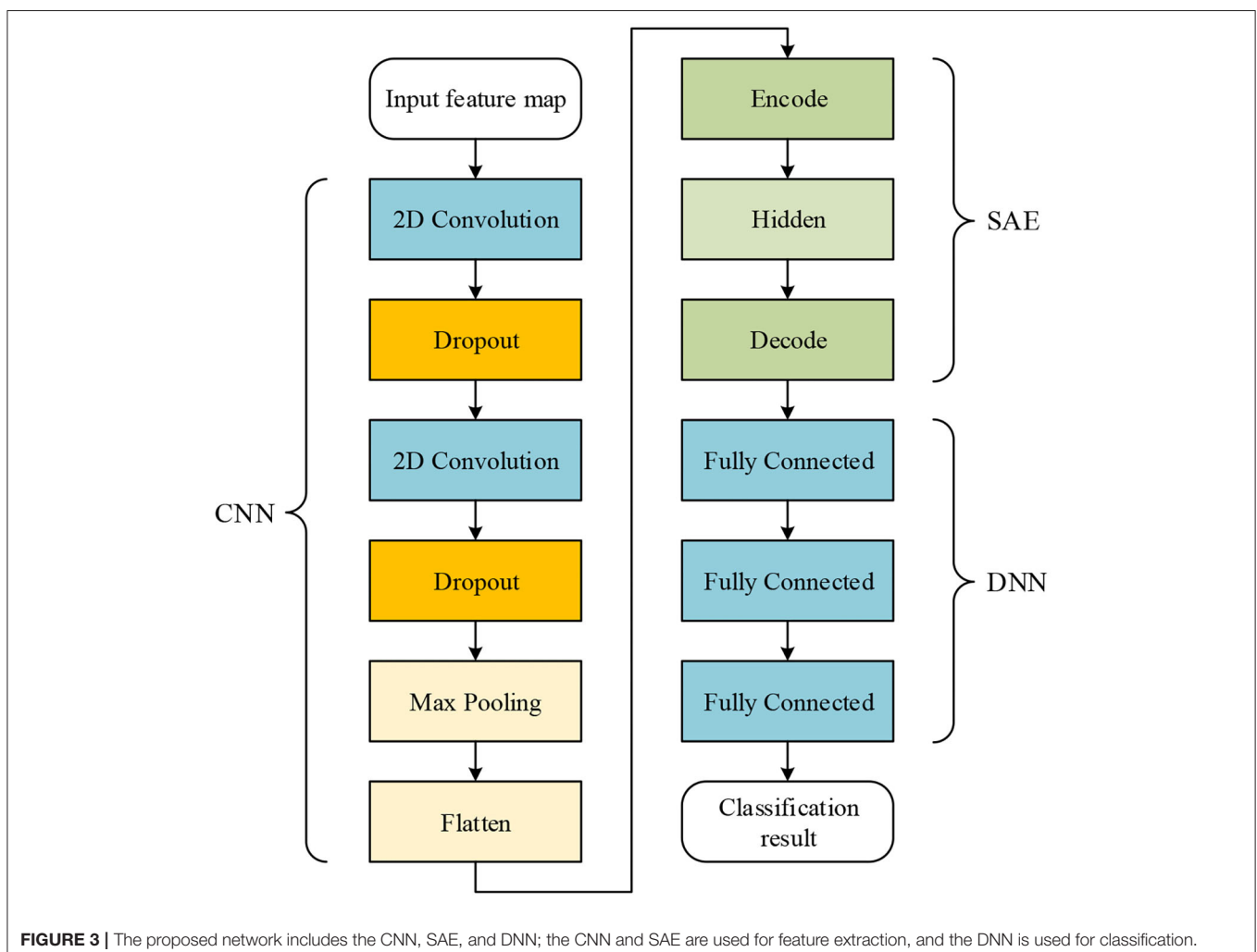
2.3. Combined CNN-SAE-DNN

EEG signal is quite sensitive to a variety of factors during acquisition, such as environmental interference and the emotional fluctuations of humans. Therefore, EEG signals may be mixed with a variety of noise, which would undoubtedly influence the required brain patterns and the experimental results. In addition, in some experiments, subjects were unable to perform the emotion collection task successfully and the experimental results were deviated greatly. In order to overcome these problems, a deep learning network structure is proposed in this work. The structure of the proposed network is shown in **Figure 3**.

As shown by **Figure 3**, in the proposed network, the CNN structure consists of two convolutional layers and one max-pooling layer. Dropout connects to each convolutional layer. The SAE consists of one encode, one hidden, and one decode layer. In the DNN, there are three fully connected layers used for classification. Given features, such as PCC for input of the proposed network, the output of max-pooling layer is used as the input for the SAE. Finally, the output of the SAE is used as the input of the DNN for classification.

The training procedure is that the CNN with one fully-connected output layer are trained for some epochs using all samples and all features, and the output layer is abandoned after training. Then, by sending features to input the trained CNN, the output of the max-pooling layer can be obtained. The output is flattened to one-dimension data, and it is set as the input of SAE. After unsupervised learning of the SAE, data are reconstructed. The reconstructed data are divided for training and testing in the DNN, i.e., the CNN and SAE are trained separately. Thus, before data are classified in the DNN, training in the CNN and SAE can be seen as a part of feature extraction. It should be noticed that the DNN used for finally classification is not the fully-connected output layer abandoned from the CNN in the first step. The DNN is never trained before output of the SAE is obtained as input data for the DNN.

Another CNN with the same parameters and structure as the whole proposed network is set as comparison in order to test the performance of the proposed network fairly. When adding more layers into this CNN, accuracy does not improve and leads to an overfitting problem. For experiments on this CNN, features are split directly into 80% for training and the rest for testing.

FIGURE 3 | The proposed network includes the CNN, SAE, and DNN; the CNN and SAE are used for feature extraction, and the DNN is used for classification.

3. DATASETS AND EXPERIMENTS

In this section, two datasets of DEAP (Koelstra et al., 2012) and SEED (Zheng and Lu, 2015) are used to evaluate the proposed network model. Data processing methods and experiment results are presented.

3.1. Emotional EEG Datasets

The DEAP dataset was collected from 32 subjects when they were watching 40 sets of 1-min music and video clips. The age of the subjects ranges between 19 and 37 years old, and half of them were males. During the 40 trials for each subject, various signals were recorded as 40-channel data, including EEG, electromyograms, breathing zone, plethysmographs, temperature, and so on (Koelstra et al., 2012). The EEG signal was recorded at 512 Hz. The data was segmented into trials of 60 s, and a bandpass frequency filter was applied after that. After each trial, the participants were asked to do a self-assessment about their emotional levels, including four different scales, such as valence, arousal, dominance, and liking.

The EEG signal is downsampled into 128 Hz for the experiments in this work, where the frequency of EEG data are from 4.0 to 45.0 Hz. Valence and arousal are the two scales chosen for this work. Each of them ranges from one (low) to nine (high), and scales are divided into two parts to construct our binary-classification tasks. Similarly to the work in Koelstra et al. (2012), valence is divided into high (ranging from five to nine) and low valence (range from one to five) according to the valence scale, and according to the arousal scale, arousal is divided into high (ranging from five to nine) and low arousal (ranging from one to five).

The SEED dataset was collected from 15 subjects (seven males) when they were asked to watch 15 film clips. The duration of each film clip was about 4 min, and each film as easily understood in order to elicit emotion of 15 subjects participating in the experiments effectively. There were 15 trials for each subject and each trial lasted for 305 s consisting of a hint of start for 5 s, a movie clip for 4 min, a self-assessment for 45 s, and a rest for 15 s. EEG data in SEED dataset was collected from 62 electrodes, which includes more information than the DEAP dataset. After collection, EEG data was downsampled to 200 Hz and applied with a bandpass filter from 0 to 75 Hz.

Similar to the DEAP dataset, in this dataset, the data are applied with a frequency filter from 4.0 to 45.0 Hz in order to equitably evaluate the proposed network. Negative, positive, and neutral are emotion labels in this dataset that represent the subjects' emotion states during each experiment. Label value of negative, positive and neutral is -1, 1, and 0, respectively. Thus, labels in the SEED dataset include three categories.

3.2. Experiment Setting

In order to test the efficacy of the proposed network, the CNN model and the proposed network are trained by using data obtained from two time windows of different lengths; in total, four groups of experiments were conducted. For experiments in the CNN used for comparison, after feature extraction of EEG data, 80% samples are used as training data and the rest

samples are used as test data among all of the data. Average accuracy is calculated from accuracies of the last 10 epochs in each experiment. For the proposed network, before training data and testing data were divided, the CNN and SAE in the proposed network were trained using features. After that, features are sent to the input of the CNN, and the output data of SAE is obtained. The output data after feature extraction were divided into 80% for training and 20% for testing in the DNN. In this work, Keras and Tensorflow (Abadi et al., 2016) ere used for the proposed network implementation. For detailed free parameters in the proposed network, they are described in sections 3.3 and 3.4, respectively.

3.3. Experiments on the DEAP Dataset

Length of data in the DEAP dataset is 63 s, and the first 3 s are removed in the experiments. Then band pass filtering is then applied. Among 40 channels, EEG data are contained in 32 channels, which are chosen for experiments. After that, EEG signals are decomposed into α (1–7 Hz), β (8–13 Hz), θ (14–30 Hz), and γ bands (30–45 Hz). After band pass filtering, signal windowing on four frequency bands is applied. EEG signals are divided into short time frames in order to facilitate signal processing, thus time windows with different overlaps are applied to EEG data in order to increase samples for training. Two window sizes, 8 and 12 s, are used for evaluating the proposed network. From the start of each recorded EEG signal, data are segmented by a sliding time window with an overlap for each frequency band. For each trial of 60 s, 14 segments are obtained using an 8-s time window moving every 4 s, and seven segments are obtained using a 12-s time window moving every 8 s. Finally, from a total of 32 participants, 17,920 (14 segments $\times$ 40 trials $\times$ 32 participants) and 8,960 (seven segments $\times$ 40 trials $\times$ 32 participants) samples are obtained using time windows of 8 and 12 s, respectively. Segment labels are the same as the label of the original sample.

After that, three different features, namely PCC, PCA, and SC, are extracted to evaluate the proposed network. For PCC-based features, PCC of data in every two channels are calculated, and a 32 $\times$ 32 PCC matrix is constructed for one sample. For PCA-based features, dimension of data from each channel is reduced into 32, and features with the shape of 32 $\times$ 32 are obtained. For SC-based features, four different characteristics are extracted, including variance, mean, kurtosis, and skewness. These statistical characteristics of data are calculated together, and a 32 $\times$ 4 matrix is finally obtained. In the proposed work, the features are separately extracted in each of the frequency bands (α, β, θ, and γ bands). According to the work in Wang et al. (2018) and other similar researches, data of four frequency bands are used together in order to get the best results. After data are processed, for the data obtained using a time window of 8 s, the shapes of the above three different feature matrixes are 17,920 $\times$ 4 $\times$ 32 $\times$ 32, 17,920 $\times$ 4 $\times$ 32 $\times$ 32, and 17,920 $\times$ 4 $\times$ 32 $\times$ 4, respectively. For data obtained using a time window of 12 s, they are 8,960 $\times$ 4 $\times$ 32 $\times$ 32, 8,960 $\times$ 4 $\times$ 32 $\times$ 32, and 8,960 $\times$ 4 $\times$ 32 $\times$ 4, respectively. Detailed configuration of the proposed network for DEAP dataset is shown by **Figure 4**. For SC, input

FIGURE 4 | Configuration of the proposed network for the DEAP dataset.
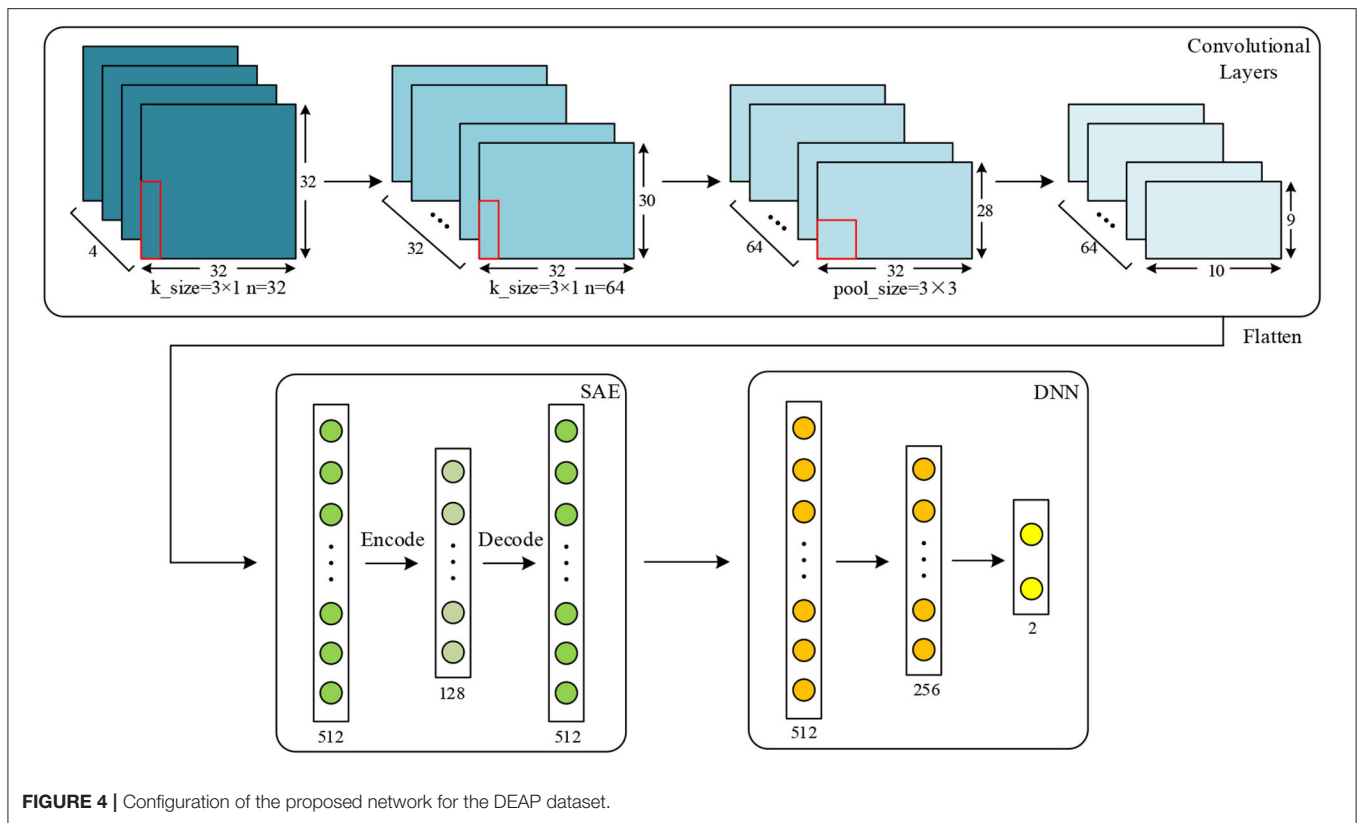
shape is 32 × 4. These features are two-dimensional, which are suitable inputs for the CNN and the proposed network.

As shown in **Figure 4**, for the DEAP dataset, two convolutional layers and one max-pooling layer are applied for the proposed network. Kernel size is set to 3 × 1, and pooling size is set to 3 × 3. The input data shape is 32 × 32. The numbers of kernels in convolutional layer are set to 32 and 64, respectively. In the SAE, the numbers of neurons in encode, hidden, and decode layers are set to 512, 128, and 512, respectively. In the DNN, the numbers of three fully connected layers are set to 512, 256, and 2, respectively. In the proposed network, the training epochs, batch size, and learning rate in the CNN are set to 50, 128, and 0.01. Epoch, batch size, and learning rate in the SAE are set to 100, 64, and 0.01, respectively. For those of the DNN, they are set to 100, 128, and 0.01, respectively.

In the proposed network, the training epochs are carried out in convolutional layers, and the SAE for features extraction, training, and testing epochs are carried out in the DNN for classification. Another CNN with the same parameters and structure as the proposed network served as a baseline method to evaluate the performance of the proposed network. The epoch, batch size, and learning rate of this CNN were set to 100, 128, and 0.01. Parameters in this CNN were the same as that of the proposed network. The data results of the experiments using a time window of 8 s are shown by **Table 1**.

From **Table 1**, among all features extracted from EEG data, we can see the PCC feature was demonstrated to be better
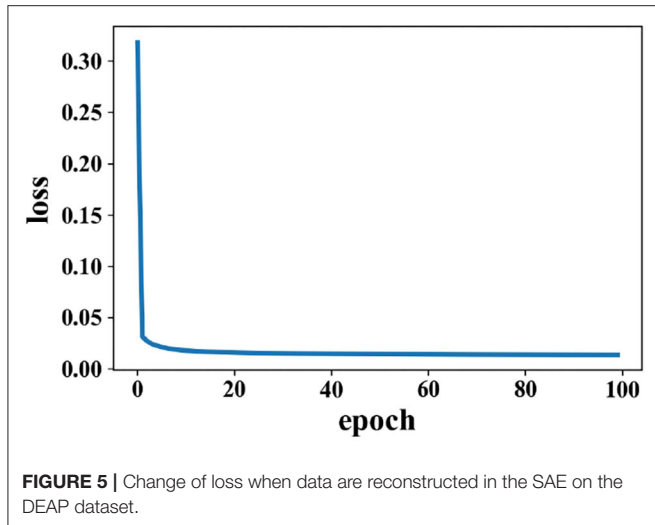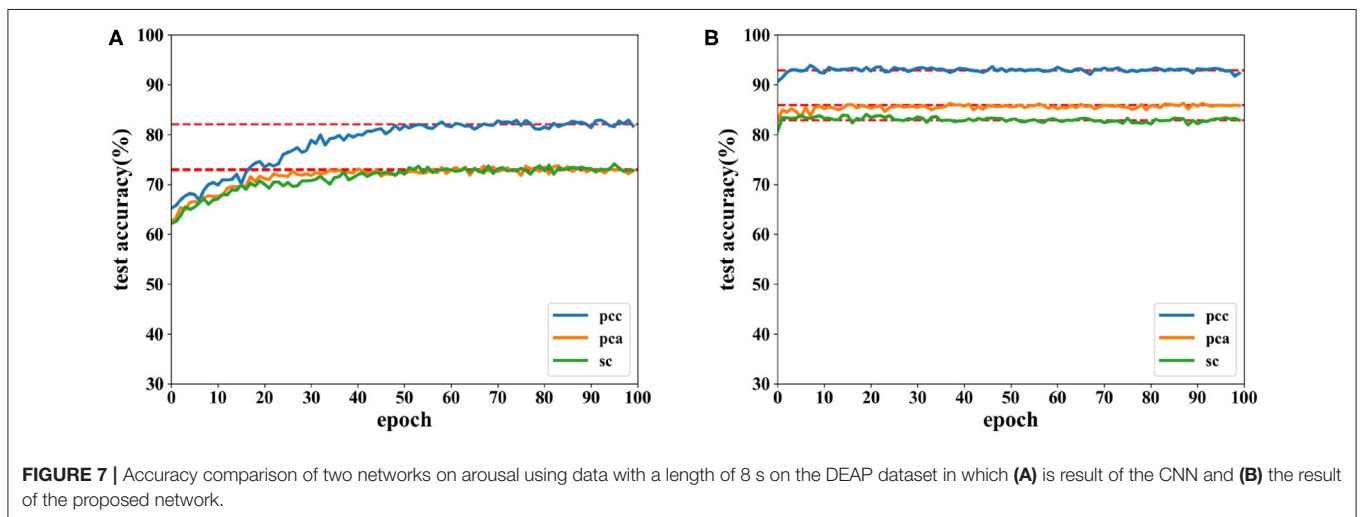
TABLE 1 | Average accuracies comparisons of the DEAP dataset using different features extracted from the data with a length of 8 s between two networks.

Network	Labels	PCC (%)	PCA (%)	SC (%)
CNN	Valence	78.80	73.32	71.10
	Arousal	82.25	72.76	73.04
Proposed network	Valence	89.49	75.59	81.93
	Arousal	92.86	85.87	82.94

than most of the other features on both the CNN and the proposed network. The proposed network can reach a recognition accuracy of 92.86% on arousal by using PCC. Moreover, recognition accuracies of most experiments on the proposed network are better than the CNN (3.27–13.11% improvement). As described previously, this is due to the inclusion of SAE, which can not only reconstruct data from convolutional layers and pooling layer but can also extract features further and make the data easier to be recognized than the CNN.

Training for loss of SAE is shown in **Figure 5**; data reconstruction is achieved when the loss does not change sharply, and data reconstruction is fast during the training process of SAE. For other extracted features (except PCC), the recognition accuracy of each method is better than the work in Zhang et al. (2016a) (81.21% for valence and 81.26% for arousal).

Figures 6, **7** show the accuracies of the CNN and the proposed network. Red lines in figures denote the average

accuracy of the last 10 epochs. It can be seen that the accuracy of the CNN gradually converges. For the proposed

network, the accuracy converges rapidly at the beginning of the epoch after fewer than 10 epochs. This is because features are easy to recognize using output data obtained from the SAE before they are classified by the DNN. For features extracted by the PCC and other methods, the accuracy of a proposed network has a faster convergence than CNN.

Similarly, results using data obtained from a time window of 12 s are shown in **Table 2**. From **Table 2**, we can see that the accuracy obtained using data with a length of 12 s is lower than
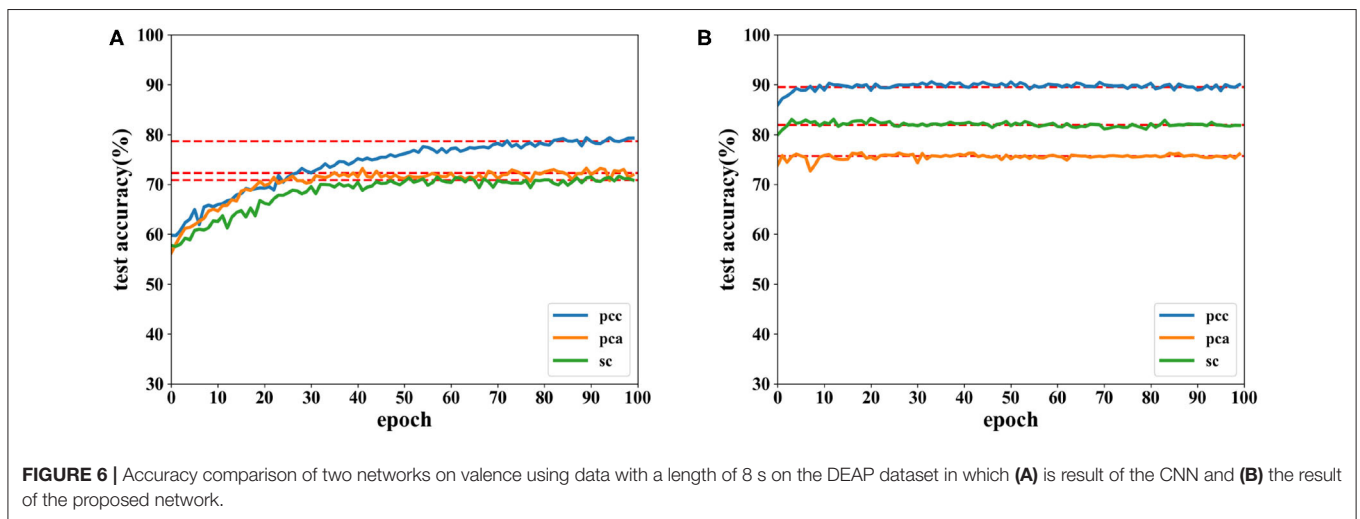

FIGURE 5 | Change of loss when data are reconstructed in the SAE on the DEAP dataset.

TABLE 2 | Average accuracy comparisons on the DEAP dataset using different features extracted from data with a length of 12 s between two networks.

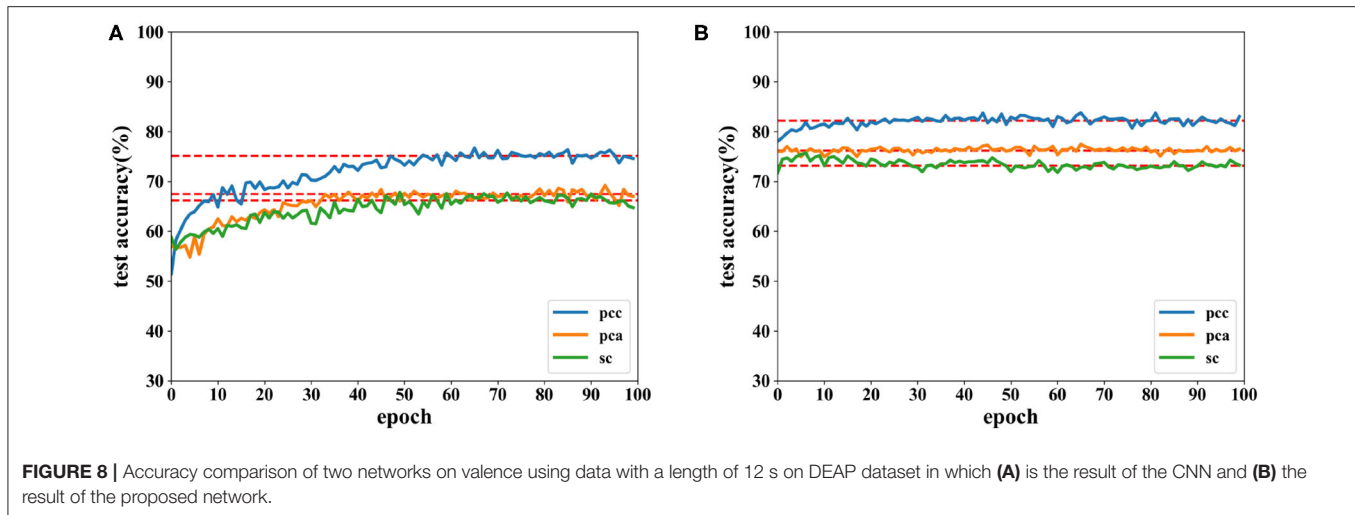| Network	Labels	PCC (%)	PCA (%)	SC (%)
CNN	Valence	75.13	67.23	66.09
	Arousal	76.12	69.20	69.48
Proposed network	Valence	82.16	76.34	73.41
	Arousal	85.47	79.11	75.44

FIGURE 6 | Accuracy comparison of two networks on valence using data with a length of 8 s on the DEAP dataset in which **(A)** is result of the CNN and **(B)** the result of the proposed network.
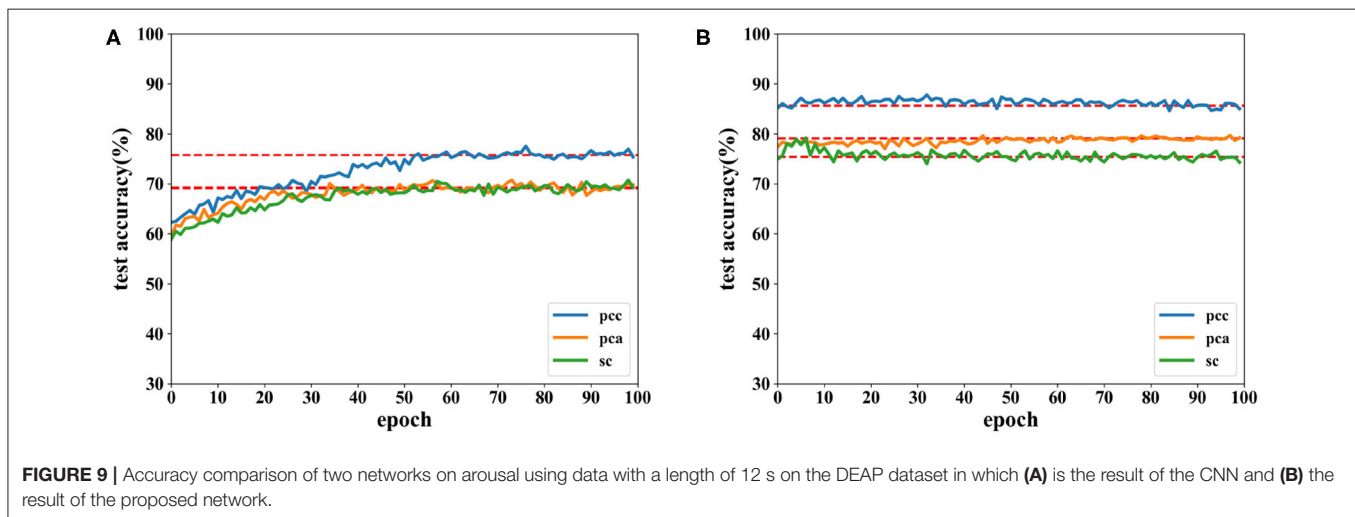

FIGURE 7 | Accuracy comparison of two networks on arousal using data with a length of 8 s on the DEAP dataset in which **(A)** is result of the CNN and **(B)** the result of the proposed network.

FIGURE 8 | Accuracy comparison of two networks on valence using data with a length of 12 s on DEAP dataset in which **(A)** is the result of the CNN and **(B)** the result of the proposed network.

FIGURE 9 | Accuracy comparison of two networks on arousal using data with a length of 12 s on the DEAP dataset in which **(A)** is the result of the CNN and **(B)** the result of the proposed network.

that of 8 s. It is more difficult to collect emotion information when the stimulation time is increasing. Most studies related to the classification of EEG data was focused on a short length of time. In this experiment, higher classification accuracy is achieved on data of 12 s than that of shorter length in other studies; this is like the work in Zhang et al. (2017), which exhibits the effectiveness of the proposed network.

Accuracies for classification on data of 12 s on both CNN and the proposed network are shown by **Figures 8**, **9**. It can be found that higher recognition accuracy is obtained by the proposed network. Moreover, the classification accuracy of the proposed network has a faster convergence in each experiment.

The results in this subsection demonstrate that accuracies can reach 92.86% for data of 8 s and 85.47% for data of 12 s. When the same feature is used for comparison, the proposed network is more powerful in classifying the EEG emotion data than the CNN. Finally, the proposed network has a quicker convergence speed.

3.4. Experiments on SEED Dataset

There are a total of 675 trials in the SEED dataset. According to the work in Zheng and Lu (2015), the first sample of each subject was chosen, and a total of 225 samples were then obtained. Due to the different data length of each channel, the 80 s data segment was chosen to reduce the influence of unstable signals at the beginning and end of the whole signal; finally, data with the shape of $16,000 \times 225 \times 62$ were obtained. Moreover, the data were processed as the same way as in the DEAP dataset: each sample was divided into different frames with different time windows. Two time windows, 8 and 12 s, were also used in the SEED dataset. A total of 19 and nine segments were obtained separately from data using a time window of 8 s moving every 4 s and a time window of 12 s moving every 8 s for each sample, respectively. Thus, in a total of 225 trials, 4,275 (19 segments $\times$ 15 trials $\times$ 15 participants) and 2,025 (9 segments $\times$ 15 trials $\times$ 15 participants) samples were obtained, respectively.

The detailed configuration of the proposed network for the SEED dataset is shown in **Figure 10**. The amount of data

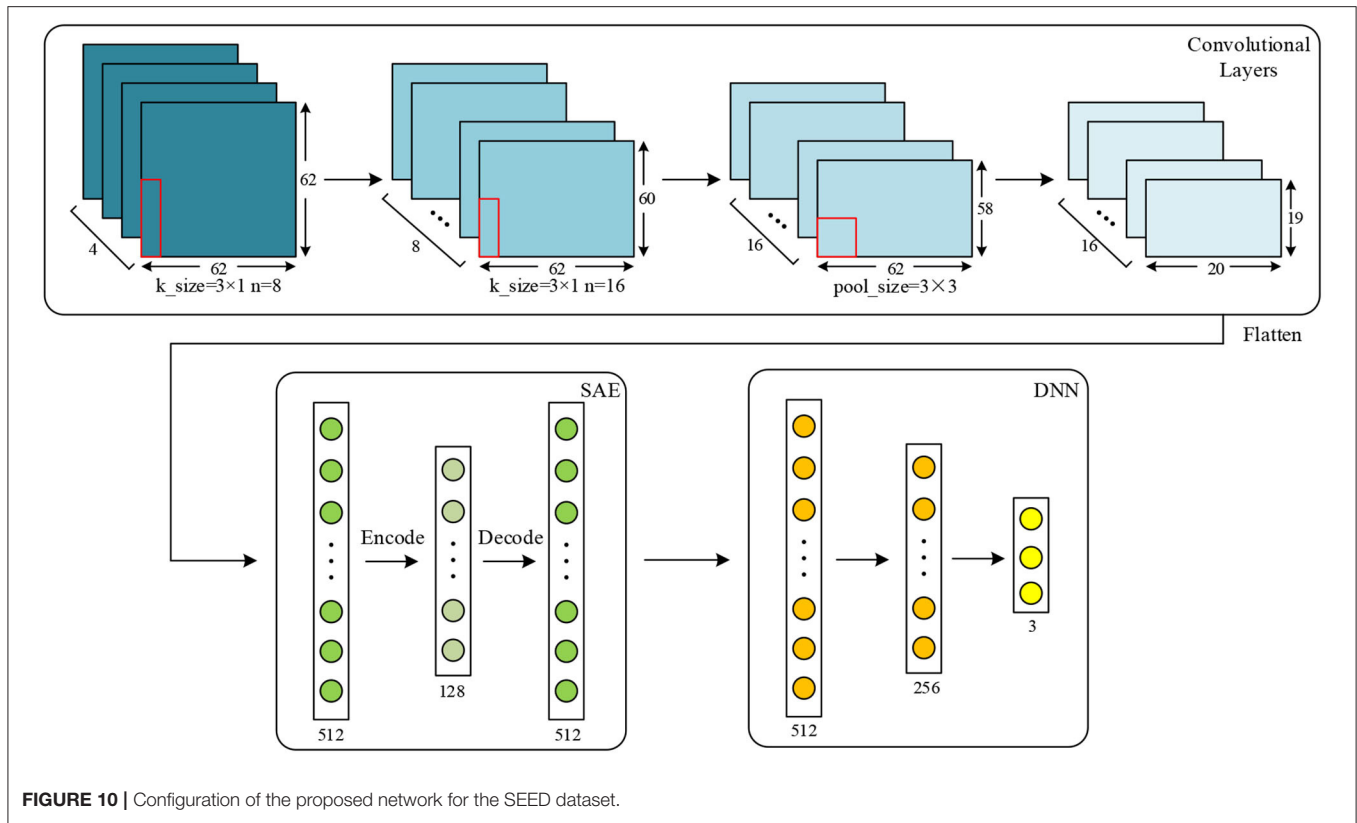**FIGURE 10 |** Configuration of the proposed network for the SEED dataset.

TABLE 3 | Average accuracy comparisons on SEED dataset using different features extracted from data with a length of 8 s between two networks.

Network	Data length (s)	PCC (%)	PCA (%)	SC (%)
CNN	8	93.71	77.66	83.32
	12	91.53	70.59	75.48
Proposed network	8	96.77	88.90	87.73
	12	94.62	70.62	79.09

extracted from this dataset is much less than from the DEAP dataset, and the two classifiers used on this dataset are thus a little different. For PCC, input data are 62×62, and the numbers of kernels are separately set to eight and 16 in two convolutional layers. In the SAE and DNN, the number of each layer is set the same as that on the DEAP dataset except that the number of the output layer is set to three because this is a three-classification task on a SEED dataset. After training in the CNN and SAE for feature extraction, the DNN is used for the final classification. Similarly to the DEAP dataset, the same features are extracted for the SEED dataset. For PCA and SC, the input shape is 62×62 and 62×4, respectively. For the CNN used for comparison, parameters are also set as the same as the proposed network.

The experiment results of data obtained from time windows of 8 and 12 s are shown in **Table 3**. The accuracy under

FIGURE 11 | Change of loss when data are reconstructed by the SAE on the SEED dataset.
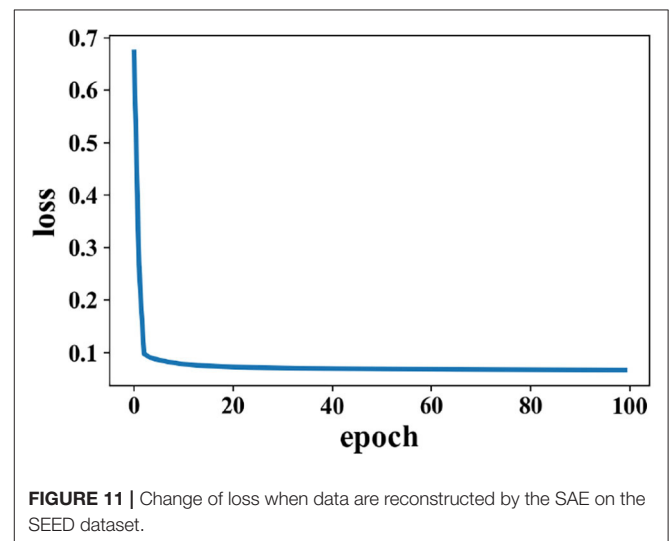
this dataset is higher than the DEAP dataset. The highest average accuracy could reach 96.77%, which is better than the work in Wang et al. (2018), 90.2%. For the data obtained from time window of 12 s, the best accuracy could reach 94.62%, which shows that PCC-based features exhibit a better performance than others. The reconstruction of data by the SAE due to the change in loss on the SEED dataset is shown by **Figure 11**. Loss drops immediately following several epochs, i.e.,

the data reconstruction can be achieved quickly when the SAE is being trained.

Accuracies under different features extracted from data of 8 and 12 s are depicted in **Figures 12**, **13**. It is shown that recognition accuracies of the proposed network are better than the CNN for almost all features, especially the PCC-based features. The proposed network can achieve faster convergence on classification accuracy than the CNN on the SEED dataset. Experiments on these two datasets shows that the proposed network performs better than original the CNN in emotion recognition.

Moreover, EEG data divided by a fixed time window with different overlaps on the SEED dataset are tested. Besides a time window of 8 s with an overlap of 4 s, overlaps of 6 and 8 s are also tested. Due to the highest accuracy, PCC-based features are used in these experiments, and classification results are displayed in **Figure 14**.

As seen in **Figure 14**, recognition accuracy could reach the highest value while the overlap is 4 s. The shorter the overlap is, the more similar the neighboring data segments are, i.e., features could be learned better when similar information is included in each trial. However, when the overlap is too short, the number of data segments increases, which requires longer time for training. In this experiment, data with a length of 8 s and overlap of 4 s could achieve the best result.

In a short summary, the best recognition could reach 96.77% on the three-class classification. The proposed network is demonstrates to be more powerful in classifying EEG emotion data than the CNN on the SEED dataset. For the data with the same length, length of overlap has an impact on recognition accuracy where 4-s overlap obtained the best performance. In addition to this, the proposed network is also compared with other research works using the DEAP and SEED datasets, and the results can be seen in **Table 4**. For complexity analysis, the number of parameters are 7.55×10^5 and 7.50×10^5 for the networks used for DEAP and SEED, respectively.
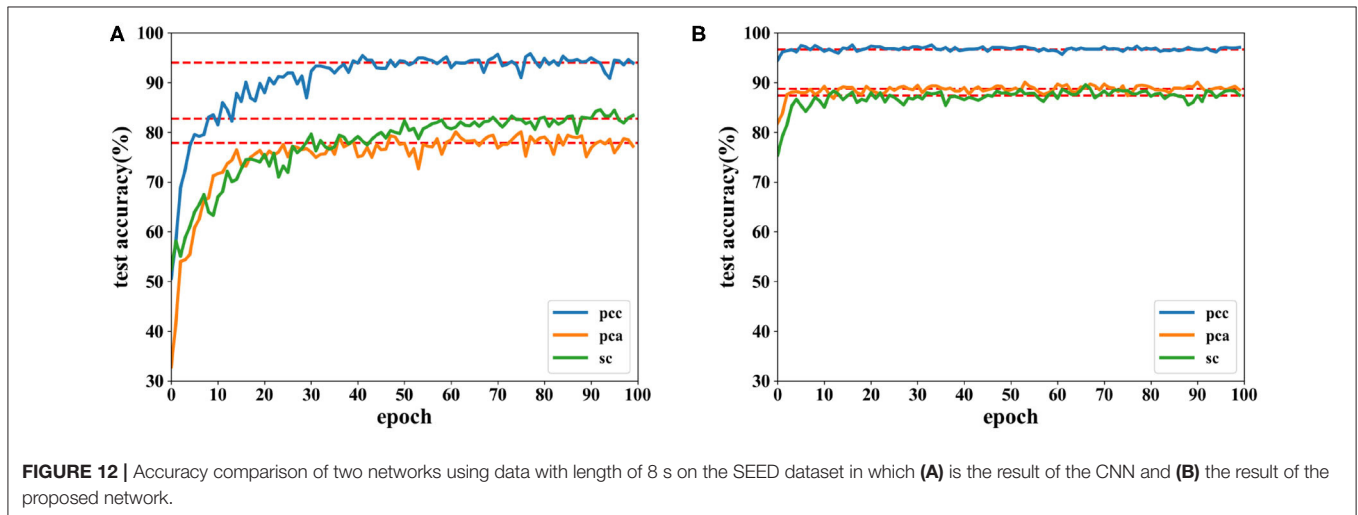

FIGURE 12 | Accuracy comparison of two networks using data with length of 8 s on the SEED dataset in which **(A)** is the result of the CNN and **(B)** the result of the proposed network.
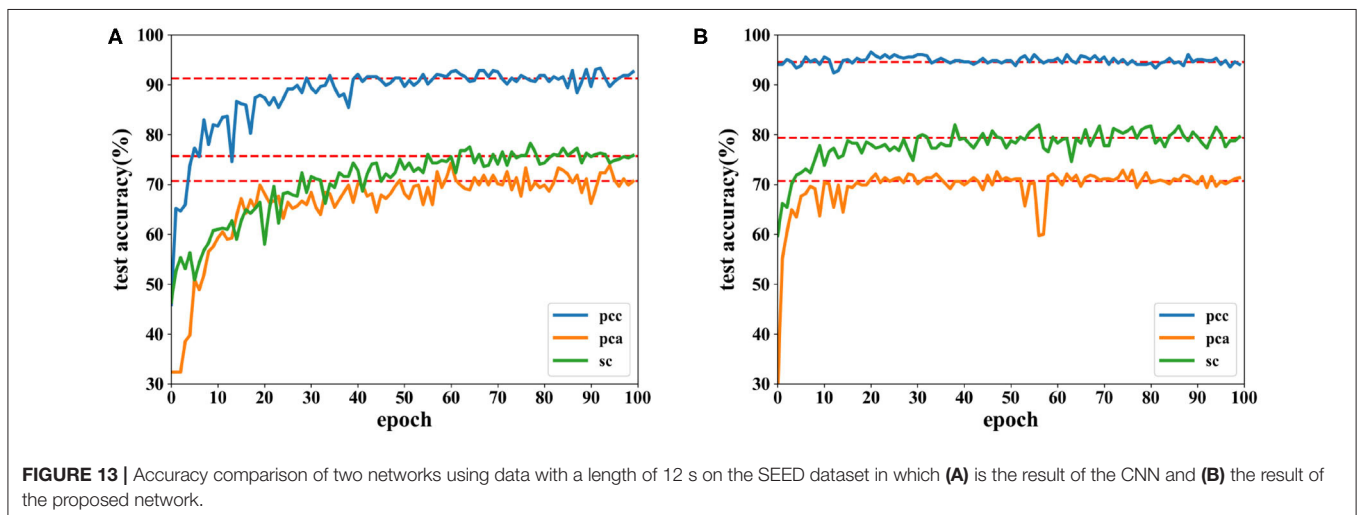

FIGURE 13 | Accuracy comparison of two networks using data with a length of 12 s on the SEED dataset in which **(A)** is the result of the CNN and **(B)** the result of the proposed network.

FIGURE 14 | Accuracy comparison of two networks using features extracted from different lengths of overlap on the SEED dataset in which **(A)** is the result of the CNN and **(B)** the result of the proposed network.
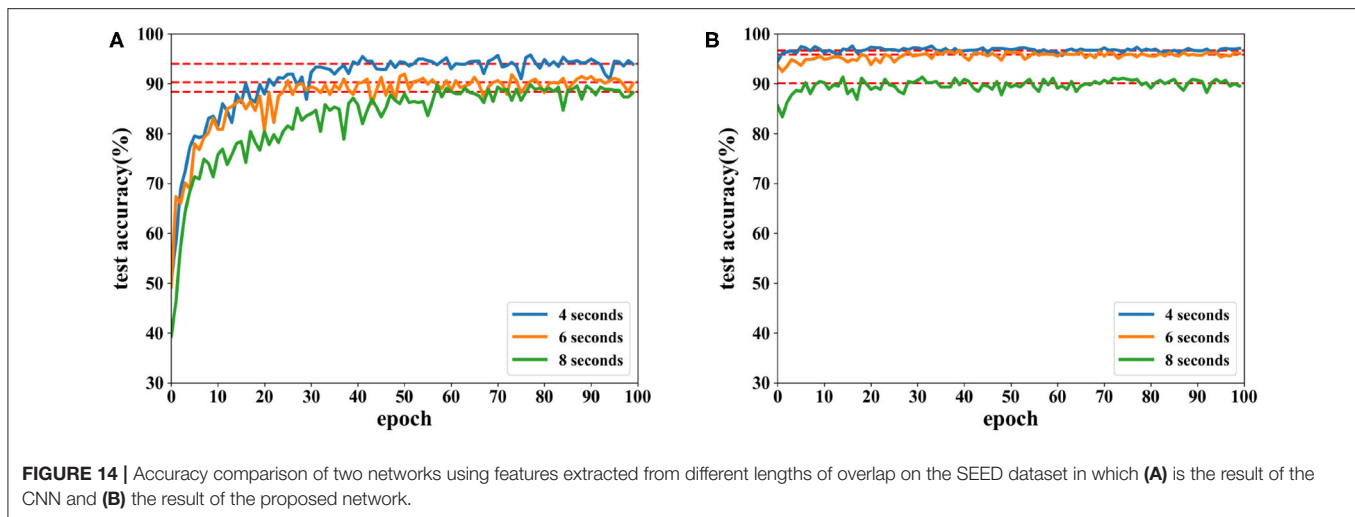
Table 4 shows the results of García et al. (2016) achieved 88.3% on valence and 90.6% on arousal. However, data used for experiments are limited, and the classification model is a better fit for classifying small amounts of data with high dimensions. The approach of Koelstra et al. (2012) used a Gaussian Bayes classifier, and experiment results proved that EEG signals are effective in emotion recognition of the DEAP dataset. The recent study (Tripathi et al., 2017) used extracted data for classification, and better accuracy results were obtained by using the CNN, where the classification accuracy of valence and arousal is 81.4 and 73.4%, respectively. The approaches of García et al. (2016) and Wang et al. (2018) used DE-based features and dynamical graph convolutional neural networks, and the accuracy achieved 93.7%. In the approach of Wang et al. (2019), BLSTM and other machine learning classifiers, such as SVM and LR were used for emotion recognition. BLSTM achieved the best accuracy of 94.96% on the SEED dataset, which is better than SVM and LR. In the approach of Soroush et al. (2019), phase space dynamics were introduced to classify emotions, achieving 87.42% on arousal and 84.59% on valence, respectively. A sparse discriminative ensemble was used for feature extraction in Ullah et al. (2019) and achieved 82.81% on valence and 74.53% on arousal, respectively. In this work, both the DEAP and SEED datasets are used for experiments, where accuracies achieve 89.49% and 92.86% in valence and arousal on the DEAP dataset, respectively, and 96.77% on the SEED dataset. Results demonstrate that the proposed network is more powerful than the CNN and other approaches.

4. DISCUSSION AND CONCLUSION

There are some points worth discussing. First, the proposed model can be trained using an end-to-end method, which is different from this work. The end-to-end training method was tested, and it obtained a similar performance. However, the training model can be further investigated

TABLE 4 | Performance comparisons with other approaches.

Classification methods	DEAP dataset		SEED dataset (%)
	Valence (%)	Arousal (%)	
CNN + statistical methods (Tripathi et al., 2017)	81.4	73.4	/
Gaussian Bayes (Koelstra et al., 2012)	57.6	62.0	/
Deep SAE + RSP (Zhang et al., 2017)	73.1	80.8	/
BDGLS + DE (Wang et al., 2018)	/	/	93.7
DGCNN + DE (Zhang et al., 2016)	/	/	90.4
GP + LVM (García et al., 2016)	88.3	90.6	/
BLSTM + DE (Wang et al., 2019)	/	/	94.96
Physe Space Dynamics (Soroush et al., 2019)	84.6	87.4	/
SDEL + PCA (Ullah et al., 2019)	82.8	74.5	/
This work [PCC]	89.49	92.86	96.77

and optimized in a future work. Second, labels are used in feature extraction. It should be noted that many feature extraction algorithms use labels such Relief and ReliefF (Kira and Rendell, 1992), where feature weights are calculated according to samples in the same and different classes. Label information has been used in the feature extraction process (Bohgaki et al., 2014; Zhang et al., 2016b). Third, constructing an autoencoder-like structure is another method of emotion recognition, and this can be investigated in a future work.

In this work, a new deep network is proposed to classify EEG signals for emotion recognition. The CNN and the proposed network are applied for two different datasets, i.e., the DEAP and SEED datasets. In the proposed network, the CNN and SAE are trained for feature extraction in which, by combining supervised learning of the CNN and unsupervised learning of the SAE, more useful features are extracted. Experimental results show that the proposed network achieves a better performance than the CNN and other approaches. It also shows that when

embedding an SAE structure into a CNN, the accuracy is better compared to a CNN with the same parameters and structure as the proposed network. In the proposed network, three different features are extracted for classifications. Results showed that, by using PCC-based features, the average recognition accuracy of the proposed network can reach 89.49% on valence and 92.86% on arousal for DEAP and 96.77% for SEED, where the proposed network has a faster convergence speed. In addition, overlap length also affects the performance, and results under the SEED dataset showed that data of 8 s with an overlap of 4 s can achieve the best result. It is also found that the data processed by the SAE is easily classified in the proposed network, which indicates that the SAE is effective in extracting features from EEG data. Future works will consider using the SAE and other classifiers to further improve the classification performance.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

JL, GW, SQ, and YL developed, implemented, and evaluated the neural network algorithm. JL and GW wrote and revised the manuscript. SY, WL, and YB analyzed the performance of the proposed network and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "Tensorflow: a system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation* (Savannah, GA: USENIX Association), 265–283.

Agrafioti, F., Hatzinakos, D., and Anderson, A. K. (2012). ECG pattern analysis for emotion detection. *IEEE Trans. Affect. Comput.* 3, 102–115. doi: 10.1109/T-AFFC.2011.28

Bahari, F., and Janghorbani, A. (2013). "EEG-based emotion recognition using recurrence plot analysis and k nearest neighbor classifier," in *2013 20th Iranian Conference on Biomedical Engineering (ICBME)* (Tehran: IEEE), 228–233. doi: 10.1109/ICBME.2013.6782224

Bohgaki, T., Katagiri, Y., and Usami, M. (2014). Pain-relief effects of aroma touch therapy with citrus junos oil evaluated by quantitative EEG occipital alpha-2 rhythm powers. *J. Behav. Brain Sci.* 4, 11–22. doi: 10.4236/jbbs.2014.41002

Cheng, B., and Liu, G. (2008). "Emotion recognition from surface EMG signal using wavelet transform and neural network," in *2008 2nd International Conference on Bioinformatics and Biomedical Engineering* (Shanghai: IEEE), 1363–1366. doi: 10.1109/ICBBE.2008.670

Chi, Y. M., Wang, Y.-T., Wang, Y., Maier, C., Jung, T.-P., and Cauwenbe, G. (2012). Dry and noncontact EEG sensors for mobile brain–computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 228–235. doi: 10.1109/TNSRE.2011.2174652

Danelljan, M., Robinson, A., Khan, F. S., and Felsberg, M. (2016). "Beyond correlation filters: learning continuous convolution operators for visual tracking," in *European Conference on Computer Vision*, eds B. Leibe, J. Matas, N. Sebe, and M. Welling (Amsterdam: Springer), 472–488. doi: 10.1007/978-3-319-46454-1_29

Fink, M. (1969). EEG and human psychopharmacology. *IEEE Trans. Inform. Technol. Biomed.* 9, 241–258. doi: 10.1146/annurev.pa.09.040169.001325

Fu, Q., Luo, Y., Liu, J., Bi, J., Qiu, S., Cao, Y., et al. (2017). "Improving learning algorithm performance for spiking neural networks," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)* (Chengdu: IEEE), 1916–1919. doi: 10.1109/ICCT.2017.8359963

García, H. F., Álvarez, M. A., and Orozco, Á. A. (2016). "Gaussian process dynamical models for multimodal affect recognition," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Orlando, FL: IEEE), 850–853. doi: 10.1109/EMBC.2016.7590834

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition* (Washington: IEEE), 580–587. doi: 10.1109/CVPR.2014.81

Hiraiwa, A., Shimohara, K., and Tokunaga, Y. (1989). "EMG pattern analysis and classification by neural network," in *Conference Proceedings, IEEE International Conference on Systems, Man and Cybernetics* (Cambridge, MA: IEEE), 1113–1115. doi: 10.1109/ICSMC.1989.71472

Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* 29, 306–310. doi: 10.1016/0013-4694(70)90143-4

Huang, Y.-J., Wu, C.-Y., Wong, A. M.-K., and Lin, B.-S. (2015). Novel active comb-shaped dry electrode for EEG measurement in hairy site. *IEEE Trans. Biomed. Eng.* 62, 256–263. doi: 10.1109/TBME.2014.2347318

Kira, K., and Rendell, L. A. (1992). "A practical approach to feature selection," in *Machine Learning Proceedings* (San Francisco, CA), 249–256. doi: 10.1016/B978-1-55860-247-2.50037-1

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdan, A., Ebrahimi, T., et al. (2012). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Kübler, A., Furdea, A., Halder, S., Hammer, E. M., Nijboer, F., and Kotchoubey, B. (2009). A brain–computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Ann. N. Y. Acad. Sci.* 1157, 90–100. doi: 10.1111/j.1749-6632.2008.04122.x

Lane, R. D., Reiman, E. M., Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J., et al. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia* 35, 1437–1444. doi: 10.1016/S0028-3932(97)00070-5

Lewis, R. S., Weekes, N. Y., and Wang, T. H. (2007). The effect of a naturalistic stressor on frontal EEG asymmetry, stress, and health. *Biol. Psychol.* 75, 239–247. doi: 10.1016/j.biopsycho.2007.03.004

Li, M., and Lu, B.-L. (2009). "Emotion classification based on gamma-band EEG," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Minneapolis, MN: IEEE), 1223–1226.

Li, X., Hu, B., Sun, S., and Cai, H. (2016). EEG-based mild depressive detection using feature selection methods and classifiers. *Comput. Methods Prog. Biomed.* 136, 151–161. doi: 10.1016/j.cmpb.2016.08.010

Liu, J., Huang, Y., Luo, Y., Harkin, J., and McDaid, L. (2019). Bio-inspired fault detection circuits based on synapse and spiking neuron models. *Nerocomputing* 331, 473–482. doi: 10.1016/j.neucom.2018.11.078

Liu, J., Mcdaid, L. J., Harkin, J., Karim, S., Johnson, A. P., Millard, A. G., et al. (2018a). Exploring self-repair in a coupled spiking astrocyte neural network. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 865–875. doi: 10.1109/TNNLS.2018.2854291

Liu, J., Sun, T., Luo, Y., Fu, Q., Cao, Y., Zhai, J., et al. (2018b). "Financial data forecasting using optimized echo state network," in *25th International Conference on Neural Information Processing (ICONIP)*, eds L. Cheng, A. C. S. Leung, and S. Ozawa (Cham: Springer), 138–149. doi: 10.1007/978-3-030-04221-9_13

Liu, Y.-J., Yu, M., Zhao, G., Song, J., Ge, Y., and Shi, Y. (2018c). Real–time movie–induced discrete emotion recognition from EEG signals. *IEEE Trans. Affect. Comput.* 9, 550–562. doi: 10.1109/TAFFC.2017.2660485

Luo, Y., Lu, Q., Liu, J., Fu, Q., Harkin, J., McDaid, L., et al. (2018). "Forest fire detection using spiking neural networks," in *Proceedings of the 15th ACM International Conference on Computing Frontiers* (New York, NY: ACM), 371–375. doi: 10.1145/3203217.3203231

Nie, D., Wang, X.-W., Shi, L.-C., and Lu, B.-L. (2011). "EEG-based emotion recognition during watching movies," in *2011 5th International IEEE/EMBS Conference on Neural Engineering* (Cancun: IEEE), 667–670. doi: 10.1109/NER.2011.5910636

Orhan, U., Hekim, M., and Ozer, M. (2011). EEG signals classification using the k-means clustering and a multilayer perceptron neural network model. *Expert Syst. Appl.* 38, 13475–13481. doi: 10.1016/j.eswa.2011.04.149

Petrantonakis, P. C., and Hadjileontiadis, L. J. (2010). Emotion recognition from EEG using higher order crossings. *IEEE Trans. Inform. Technol. Biomed.* 14, 186–197. doi: 10.1109/TITB.2009.2034649

Scherer, K. R. (1995). Expression of emotion in voice and music. *J. Voice* 9, 235–248. doi: 10.1016/S0892-1997(05)80231-0

Soroush, M. Z., Maghooli, K., Setarehdan, S. K., and Nasrabadi, A. M. (2019). A novel EEG-based approach to classify emotions through phase space dynamics. *Signal Image Video Process.* 13, 1149–1156. doi: 10.1007/s11760-019-01455-y

Subasi, A., and Gursoy, M. I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Syst. Appl.* 37, 8659–8666. doi: 10.1016/j.eswa.2010.06.065

Tabar, Y. R., and Halici, U. (2016). A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* 14, 90–100. doi: 10.1088/1741-2560/14/1/016003

Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., and Bhattacharya, S. (2017). "Using deep and convolutional neural networks for accurate emotion classification on deap dataset," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (Hawaiian, HI), 4746–4752.

Ullah, H., Uzair, M., Mahmood, A., Ullah, M., Khan, S. D., and Cheikh, F. A. (2019). Internal emotion classification using EEG signal with sparse discriminative ensemble. *IEEE Access* 7, 40144–40153. doi: 10.1109/ACCESS.2019.2904400

van den Broek, E. L. (2013). Ubiquitous emotion-aware computing. *Pers. Ubiquit. Comput.* 17, 53–67. doi: 10.1007/s00779-011-0479-9

Wang, X., Zhang, T., Xu, X., Chen, L., Xing, X., and Chen, C. L. P. (2018). "EEG emotion recognition using dynamical graph convolutional neural networks and broad learning system," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Madrid: IEEE), 1240–1244. doi: 10.1109/BIBM.2018.8621147

Wang, Y., Qiu, S., Li, J., Ma, X., Liang, Z., Li, H., et al. (2019). "EEG-based emotion recognition with similarity learning network," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Berlin: IEEE), 1209–1212. doi: 10.1109/EMBC.2019.8857499

Zhang, D., Yao, L., Zhang, X., Wang, S., Chen, W., and Boots, R. (2018). "Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface," in *32nd AAAI Conference on Artificial Intelligence, AAAI 2018* (New Orleans, LA).

Zhang, J., Chen, M., Hu, S., Cao, Y., and Kozma, R. (2016). "PNN for EEG-based emotion recognition," in *2016 IEEE International Conference on Systems, Man, and Cybernetics* (Budapest), 2319–2323.

Zhang, J., Chen, M., Hu, S., Cao, Y., and Kozma, R. (2016a). "PNN for EEG-based emotion recognition," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Budapest: IEEE), 002319–002323.

Zhang, J., Chen, M., Zhao, S., Hu, S., Shi, Z., and Cao, Y. (2016b). Relieff-based EEG sensor selection methods for emotion recognition. *Sensors* 16, 40144–40153. doi: 10.3390/s16101558

Zhang, Q., Chen, X., Zhan, Q., Yang, T., and Xia, S. (2017). Respiration-based emotion recognition with deep learning. *Comput. Ind.* 92–93, 84–90. doi: 10.1016/j.compind.2017.04.005

Zheng, W. (2017). Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis. *IEEE Trans. Cogn. Dev. Syst.* 9, 281–290. doi: 10.1109/TCDS.2016.2587290

Zheng, W.-L., and Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Mental Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

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* (Chengdu: IEEE), 1–6. doi: 10.1109/ICME.2014.6890166

Zhou, Z. H., and Feng, J. (2017). "Deep forest: towards an alternative to deep neural networks," in *IJCAI International Joint Conference on Artificial Intelligence* (Melbourne, VIC), 3553–3559. doi: 10.24963/ijcai.2017/497