

Channel Semantic Enhancement-Based Emotional Recognition Method Using SCLE-2D-CNN

Dan Fu, Sichuan Vocational College of Finance and Economics, China

Weisi Yang, Sichuan Vocational College of Finance and Economics, China*

Li Pan, UCSI University, Malaysia

ABSTRACT

The existing EEG emotion classification methods have some problems, such as insufficient emotion representation and lack of targeted channel enhancement module due to feature redundancy. To this end, a novel EEG emotion recognition method (SCLE-2D-CNN) combining scaled convolutional layer (SCLs), enhanced channel module and two-dimensional convolutional neural network (2D-CNN) is proposed. Firstly, the time-frequency features of multi-channel EEG emotional signals were extracted by stacking scl layer by layer. Secondly, channel enhancement module is used to reassign different importance to all EEG physical channels. Finally, 2D-CNN was used to obtain deep local spatiotemporal features and complete emotion classification. The experimental results show that the accuracy of SEED data set and F1 are 98.09% and 97.00%, respectively, and the binary classification accuracy of DEAP data set is 98.06% and 96.83%, respectively, which are superior to other comparison methods. The proposed method has a certain application prospect in the recognition of human mental state.

KEYWORDS

Channel Semantic EEG Signals, Emotional Recognition, Enhancement, Psychology, Time-Frequency Features, Two-Dimensional Convolutional Neural Network

INTRODUCTION

Emotions can represent the current psychological and physiological states of humans, and they play a vital role in cognition, decision-making, and communication. Stable and optimistic emotions are important indicators of mental and physical health. For example, positive emotions such as joy, excitement, surprise, satisfaction, love, and friendship can help stimulate healthy psychology, enhancing happiness, satisfaction, and self-confidence (Li et al., 2022b). On the contrary, negative emotions, including anger, sadness, anxiety, fear, shame, and disgust can easily lead to negative psychology and unhealthy physiology. With the rapid development of artificial intelligence (Tan et al., 2022; Jiao et al., 2022), big data (Thirumalaisamy et al., 2022), cloud computing (Vijayakumar

DOI: 10.4018/IJSWIS.337286

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

et al., 2022; Dwivedi, 2022), fog computing (Thoumi & Haraty, 2022), the Internet of Things (Chamra & Harmanani, 2020), smart homes (Madhu et al., 2022; Guebli & Belkhir, 2021), intelligent communication (Samir et al., 2020) and other fields, the application scope of human emotion recognition is becoming increasingly widespread. Therefore, conveniently, effectively, and accurately recognizing human emotions is significantly important for promoting the development of new eras such as artificial intelligence, Web 3.0, and the metaverse (Kouti et al., 2022).

Usually, human facial expressions, speech signals, or posture and gait are used for emotion recognition. However, these signals are easily influenced by human subjective characteristics and may not reflect the true emotional state (Zhao et al., 2023). Emotion recognition using electroencephalography (EEG) signals can avoid human camouflage of emotions and can provide a more accurate analysis of emotions by detecting electrophysiological signals (Islam et al. 2021; Li et al., 2019). Single channel approach (Dan, 2021) and multi-channel approach (Jie et al., 2022) can be distinguished according to the number of channels present in EEG signals. The main advantage of a single channel is high efficiency, while the main advantage of multiple channels is multi-dimensional, comprehensive, and high recognition rate. From the perspective of feature extraction, it can be divided into manual feature extraction methods and deep learning-based automatic feature extraction methods (Ghosh et al., 2021; Appati et al., 2021). The manual feature extraction approach has rich prior knowledge and can fully utilize non-stationary non-linear EEG signals to achieve feature extraction. The deep learning-based approach can extract various types of information from EEG signals and usually has a higher recognition rate compared to the manual extraction approach (Tan et al., 2021a; Zali-Vargahan et al., 2023).

However, whether single-channel or multi-channel approaches are adopted, many existing approaches still have some potential problems. The multi-channel EEG data is large and contains a lot of redundant information, which is not highly relevant to the emotion recognition task of EEG, resulting in insufficient representation of emotion information. Since the importance of each channel in the original EEG signal is the same, it is difficult to effectively enhance the EEG signal of different channels to improve the recognition performance.

To resolve the above-mentioned issues, this paper proposes a method for emotion recognition from EEG signals (SCLE-2D-CNN) that incorporates SCL, enhanced channel modules, and 2D-CNN. The proposed method achieves accurate emotion recognition through multi-channel time-frequency feature extraction and enhancement of EEG signal channels. The main innovation points are:

- 1) The proposed method aims to improve the representation of emotional information in EEG signals by automatically extracting time-frequency features from multi-channel EEG signals. This is achieved by integrating the time and frequency dimensions of the original EEG signals into the stacked SCLs.
- 2) To enhance the effective emotional information in time-frequency features and suppress the negative impact of redundant information, the proposed method constructs an EEG enhancement module utilizing an attention mechanism. The enhancement module assigns different weights to each EEG channel in the network, this contributes to improved accuracy in recognizing emotions.
- 3) Through the direct connection of four convolutional layers, a two-dimensional CNN (2D-CNN) model is employed to accurately identify the emotions of EEG signals and investigate the deep local spatiotemporal features of EEG signals.

This method shows great potential for various practical applications. For example, the medical field (Mandle et al., 2022), smart home, education field, intelligent transportation, mental illness warning, and so on.

RELATED WORKS

Numerous studies have been conducted on feature extraction and emotion recognition of EEG signals, which have achieved certain research results.

Emotion Recognition Based on Machine Learning

Bhosale et al. (2022) proposed a novel approach to emotion classification by introducing modifications to the EEG signal-based emotion recognition process, which ensured the accuracy of emotion classification in different spaces while improving the classification efficiency. However, this method lacks high-performance classifiers. Salankar et al. (2021) proposed a high-precision classifier for recognizing emotions from EEG signals. However, this extraction method is relatively simple and difficult to apply to complex EEG signals. Subbasi et al. (2021) proposed a new lightweight automatic emotion recognition framework that combines discrete wavelet transform with a rotating forest set classifier to effectively achieve lightweight recognition and classification of emotion signals. Tan et al. (2021b) presented a multimodal approach for emotion recognition that combines facial expressions and electroencephalography (EEG). To address the issue of limited data, they employed the Monte Carlo method. However, this approach may not effectively capture deep local spatiotemporal features from EEG signals. In contrast, Liu and Fu (2021) proposed an emotion recognition method that integrates support vector machine (SVM) with multi-channel EEG and text features in the time domain. However, this method solely focuses on time domain features and overlooks frequency domain features, leading to potential limitations in recognition performance.

Emotion Recognition Based on Deep Learning

Shi et al. (2021) introduced a method for EEG-based emotion recognition that utilizes 2D-CNN and attention mechanisms. However, this method does not consider the influence of multi-channel factors, resulting in poor recognition effect. Olamat et al. (2022) developed an emotion detection method that involves multivariate empirical mode decomposition of EEG signals, leading to improved accuracy in emotion classification. However, the effectiveness of one-dimensional CNN in extracting deep local spatiotemporal features from EEG signals may be limited. In a different approach, Kumar and Suresh (2021) introduced a multi-objective classifier based on deep neural networks for the feature subset obtained from mixed feature extraction of EEG. This method enhances the accuracy and predictability of emotion classification in various conversation scenarios. Furthermore, Chen et al. (2021) proposed a multi-modal fusion method for emotion recognition using the Dempster-Shafer evidence theory. This approach effectively enhances the accuracy of emotion recognition by combining multiple modalities. However, this method cannot distinguish the importance of different channels, and it is difficult to effectively use important channels to improve the recognition performance.

Motivation

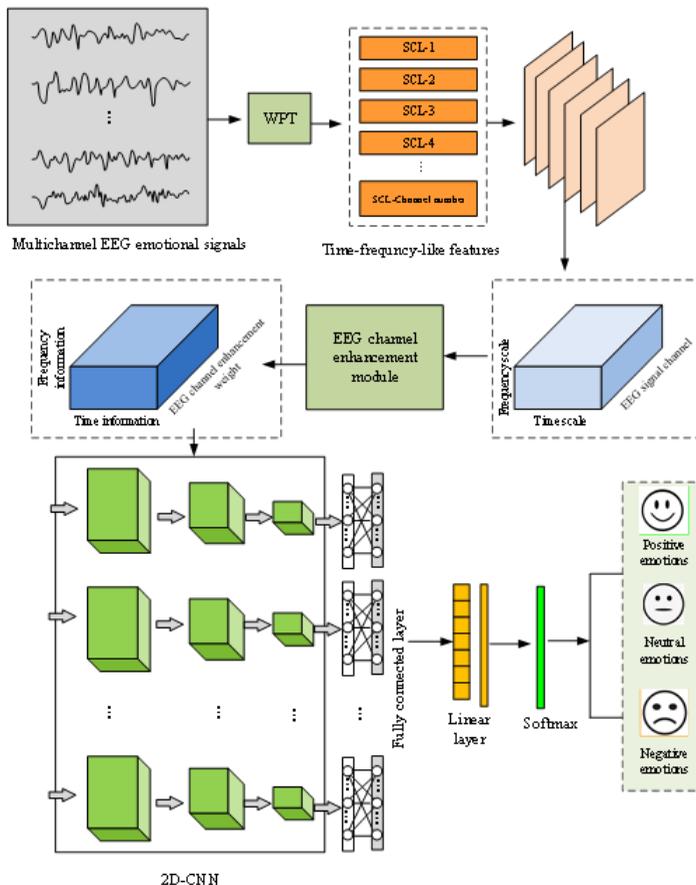
The above analysis highlights that the existing emotion classification methods for EEG signals often have several issues, such as insufficient emotional representation caused by feature redundancy, lack of targeted channel semantic enhancement modules, and insufficient enrichment of deep-level features extracted. To better address these issues, a fusion emotion recognition network structure based on SCLs, EEG channel semantic enhancement module, and 2D-CNN is proposed in this paper. The proposed network can integrate the time and frequency dimensions of the initial EEG signal to automatically extract features, strengthening the learning ability of the network for important channels through the EEG channel semantic enhancement module and reducing the influence of unimportant channels. Hence, the proposed framework achieves improved EEG emotion recognition performance.

PROPOSED EEG SIGNAL EMOTION RECOGNITION METHOD

Overall Structural Model of SCLE-2D-CNN

Figure 1 depicts the proposed EEG emotion recognition network, with multi-channel raw EEG emotion signals as inputs and positive, neutral, or negative emotions as outputs (Chen et al., 2021). In the proposed framework, the automatic time-frequency feature extraction module consists of SCLs that are independent of multiple channels and have the same number as the inputted EEG channels. The EEG channel semantic enhancement module consists of a global average pooling (GAP) layer, two full connection layers, ReLU activation function, and sigmoid activation function layers. The depth feature change module consists of three layers of 2D-CNN layers. Finally, the Softmax layer is used to classify EEG emotions. The SCL module is designed to capture features across various frequencies and time scales, enabling the model to gain a deeper understanding of the time-frequency characteristics present in EEG signals. The EEG channel enhancement module further processes and enhances the features obtained from the automatic time-frequency feature extraction module. Additionally, the global averaging pooling layer reduces the feature dimension and extracts the overall EEG feature. The fully connected layer, along with the activation function layer, is responsible for learning and extracting higher-level features that effectively represent the emotional information contained within the EEG signal.

Figure 1. Overall framework of the SCLE-2D-CNN



The SCLE-2D-CNN is an end-to-end emotion recognition network, where each EEG channel is independently assigned a SCL to obtain its time-frequency features (Jiang et al., 2021). The EEG physical channel is directly enhanced by utilizing the attention mechanism. The EEG channel semantic enhancement module produces a one-dimensional vector of importance, which is then weighted and multiplied with time-frequency features to obtain the enhanced recalibrated EEG channel features. Then, the deep information of the EEG signal is extracted through 2D-CNN, and finally, a fully connected neural network is employed for emotion recognition and classification.

EEG Data Preprocessing

Generally, signal enhancement techniques (Oscar et al., 2022) are used to pre-process the original data set. Commonly used signal enhancement techniques include wavelet transform (Subasi, et al., 2021) and wavelet packet transform (WPT) (Kumar et al., 2022). Wavelet transforms may not be efficient in decomposing and representing signals with abundant detailed information. On the other hand, the WPT offers a more refined decomposition of the high-frequency component. It can dynamically select the optimal wavelet basis function based on the signal's characteristics. As a result, the WPT outperforms the standard wavelet transform in terms of time-frequency localization analysis, particularly for signals that contain substantial middle and high-frequency information. EEG signals are non-stationary and contain a great deal of detailed information. Therefore, the proposed method uses WPT to enhance the signal and improve the time-frequency resolution of the signal.

WPT is a further optimization of the wavelet transform, which mainly achieves more accurate signal decomposition than the wavelet transform by halving the various scales and sub-bands of the decomposition filter. Given the orthogonal scaling function $\varphi(t)$ and wavelet function $\psi(t)$, the two scaling equations are:

$$\begin{cases} \phi(x) = \sqrt{2} \sum_t v_t \phi(2x - t) \\ \psi(x) = \sqrt{2} \sum_t o_t \phi(2x - t) \end{cases} \quad (1)$$

where, v_t, o_t are pulse response coefficients.

Note that d_j^{2n} and d_j^{2n+1} are wavelet packet coefficients on the j scale, and the recursive formula for wavelet packet coefficients is:

$$\begin{cases} d_j^{2n}[t] = \sum_{\tau \in Z} v_\tau - 2td_j^{2n} + 1[\tau] \\ d_j^{2n+1}[t] = \sum_{\tau \in Z} o_\tau - 2td_j^{2n} + 1[\tau] \end{cases} \quad (2)$$

The reconstruction formula based on wavelet packet coefficients is:

$$d_{j+1}^n[t] = \sum_{\tau \in Z} v_t - 2\tau d_j^{2n}[\tau] + \sum_{\tau \in Z} o_t - 2\tau d_j^{2n+1}[\tau] \quad (3)$$

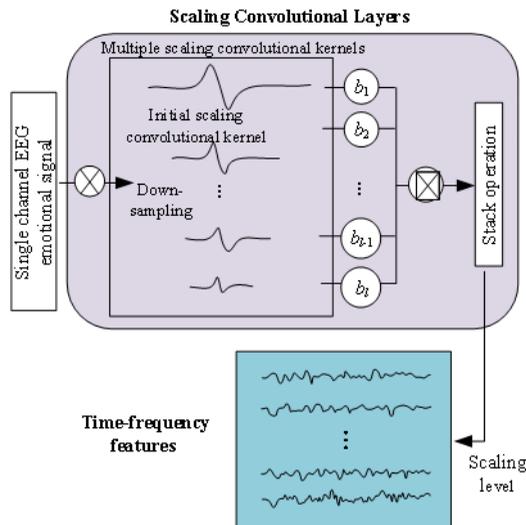
WPT uses aperiodic signals and scales the basis function coefficients of the translation transform to create a time-frequency scale graph. This graph can be used to analyze the position and frequency of the sudden changes in the signal. WPT can also be used to filter the signal by selectively removing a portion of the coefficients and then transforming them back into the time domain.

Time-Frequency Feature Extraction

The multi-channel EEG emotion signals after EEG preprocessing become time series signals and have the problem of insufficient emotional information representation. To combine the abundant information of time and frequency dimensions in EEG signals, emotion features can be derived automatically from temporal EEG signals. At the same time, deep learning technology can be used to automatically extract further high-dimensional feature tensors from EEG signals. After obtaining preprocessed multi-channel EEG signals, a feature extraction module utilizing SCLs is employed to automatically extract time-frequency features from multi-channel EEG emotion signals.

SCL is a neural network layer that extracts features of one-dimensional time series signals. The overall structure of the SCL model is illustrated in Figure 2. The input of SCL can be any length of EEG emotion signals (Hu et al., 2021). For each SCL, there will be a scaled convolutional kernel (SCK) for cross-correlation calculation with EEG emotion signals. Each EEG channel will be independently assigned an SCL to extract time-frequency features of emotion signals from different EEG channels.

Figure 2. Overall structure of the SCL model



For each SCL of the EEG channel, the initial SCK is used for cross-correlation calculation with the EEG emotion signal. To maintain the same output length after each cross-correlation calculation, the initial SCK is set as an odd number in order to ensure consistency. Then, the SCK of the EEG channel is down sampled by continuously scaling the size of the convolutional kernel until the set lower bound of the SCK is reached. Next, the time-frequency characteristic map corresponding to the EEG channel is obtained. The SCL is calculated as follows:

$$F_{out}(l) = \delta(b(l) + \zeta(\omega, l) \otimes \lambda F_{in}) \quad (4)$$

where F_{in} and F_{out} are the input one-dimensional EEG emotion signals and the output time-frequency feature matrices, respectively; b is the offset generated by scaling the convolutional kernel each time; λ is the weight of the input one-dimensional time series signal, ζ is the pooling operator that down samples the weight of an average convolution check with a window size of 2, and is executed once

until the set lower bound of the scaled convolution kernel is reached; l is the level of control scaling; $\delta(\cdot)$ is the activation function of scaling convolution layer; ω is the weight of the SCK; and \otimes represents a cross-correlation operator, which is expressed as follows:

$$(S \otimes F_{in})(n) \triangleq \sum_{s=0}^{K-1} \overline{f[s]} \lambda F_{in} [(s+n) \bmod K] \quad (5)$$

where S is $\zeta(\omega, l)$, which is the SCK after down sampling; F_{in} is the input one-dimensional time series signal in this network layer; K represents the total length of the time series signal data; n is an independent variable; and s is the sum variable. S slides on F_{in} , that is, the input one-dimensional EEG emotion signal, and the SCK continuously slides and performs cross-correlation calculations.

For each EEG channel, independent SCL is used to obtain the two-dimensional time-frequency information on that channel, and then the two-dimensional time-frequency features are stacked on the channel dimension, obtaining the three-dimensional time-frequency feature tensor of the entire EEG emotion signal (Rahman et al., 2021).

EEG Channel Semantic Enhancement

For enhancing the relationship between distinct EEG channels (Yu et al., 2022), an EEG channel semantic enhancement module was designed based on the attention mechanism. Different EEG channels were assigned different weights to enhance the channels related to emotion recognition tasks while suppressing the unrelated channels. For EEG emotion signals, $H = [h_1, h_2, \dots, h_N]$ represents the input information vector of the EEG channel semantic enhancement module. Given an information query vector q , a variable $L \in [1, N]$. In order to maintain consistent output length after each cross-correlation calculation, the initial SCK is chosen as an odd number. This ensures that the output length of the SCL remains unchanged in H . When given the input information H and query vector q , the probability distribution p_i of selecting the i -th EEG input information is defined as:

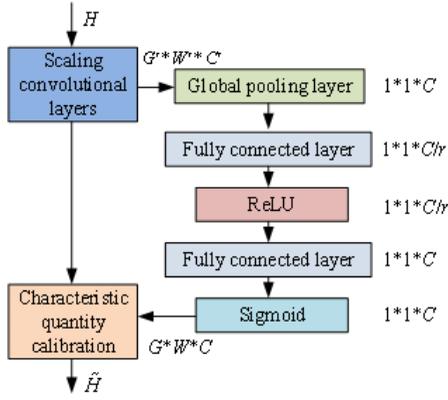
$$p_i = p(L = i | H, q) = \frac{\exp(e(h_i, q))}{\sum_{j=1}^N \exp(e(h_j, q))} \quad (6)$$

where $e(h_i, q)$ is the attention scoring function, and various methods such as additive models and dot product models can be used to achieve calculations.

In the EEG channel semantic enhancement module, the Squeeze and Excitation Network (SENet) attention mechanism is used for EEG physical channel enhancement. The SENet is more easily embedded into existing mainstream deep neural networks than other attention mechanisms (Naser & Saha, 2021). Figure 3 depicts the network structure of the EEG channel semantic enhancement module.

The EEG channel semantic enhancement module is established on top of the time-frequency feature extraction module, where the input and output of the time-frequency feature extraction module are $H_{in} \in R^{G' \times W' \times C'}$ and $H_{out} \in R^{G \times W \times C}$, respectively, where C is the physical channel dimension of the EEG. Since the extracted time-frequency feature is the local Receptive field of each EEG channel, information outside the Receptive field cannot be used. Therefore, the GAP operation is performed on the time-frequency feature maps obtained from each EEG channel. $z \in R^C$ represents the result

Figure 3. Network structure of EEG channel semantic enhancement module



of performing GAP on the $W \times G$ spatial dimension of the time-frequency feature, and each element z_c in z is represented as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{G \times W} \sum_{i=1}^G \sum_{j=1}^W \alpha u_c(i, j) \quad (7)$$

To simultaneously enhance the nonlinearity of features after GAP and understand the correlation between the EEG channels, it is necessary to ensure that the importance of multiple EEG channels is enhanced, rather than just a single EEG channel. This is done by adding two fully connected layers after the GAP layer. The FC output ϕ is represented as:

$$\phi = F_{ex}(z, W) = \sigma(W_2 \kappa(W_1 z)) \quad (8)$$

where σ and κ are the Sigmoid and ReLU activation functions, respectively, $W_1 \in R^{r \times C}$, $W_2 \in R^{C \times r}$, and r denotes the hyperparameter used to reduce the FC dimension.

Finally, the one-dimensional vector output from the Sigmoid activation function is weighted and multiplied with the three-dimensional time-frequency feature tensor. The final output result $\tilde{H} = [\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_C]$ of the EEG channel semantic enhancement module is obtained through Equation (9):

$$\tilde{h}_c = F_{scale}(u_c, \phi_c) = \phi_c u_c \quad (9)$$

where, F_{scale} is the product of each EEG channel and $u_c \in R^{H \times W}$.

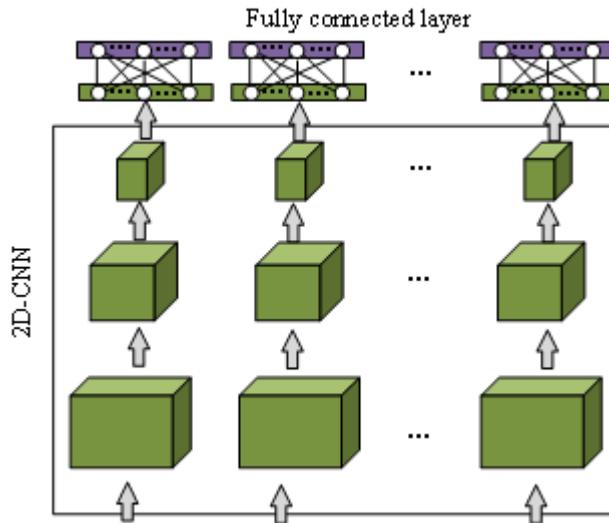
Setting the dimensionality reduction compression ratio r in the EEG channel semantic enhancement module to decrease the dimensions of the fully connected layer, dimensionality reduction techniques that are employed can decrease the computational complexity of the EEG channel semantic enhancement emotion recognition network, and enhancing the network's nonlinear capability is achieved to boost its nonlinear capacity. The GAP of time-frequency feature tensors further enhances the global receptive

field of the network. Additionally, the use of two fully connected layers in this module also enhances the nonlinear transformation ability of the network. Ultimately, the importance of EEG channels related to emotion recognition is enhanced, while suppressing irrelevant EEG channels.

Deep Local Spatiotemporal Feature Extraction Using 2D-CNN

The proposed method utilizes the 2D-CNN model to mine deeper local spatiotemporal features of EEG channels. The 2D-CNN network structure is depicted in Figure 4. The 2D-CNN model takes in a three-dimensional data structure that includes both spatial and temporal information. The CNN network captures spatial features from each two-dimensional mesh matrix of EEG data and subsequently organizes these extracted spatial features in a chronological sequence. The data is then inputted into a CNN network to continue extracting deep local spatiotemporal features of EEG data. Finally, the output of the CNN network is passed through a fully connected layer. The resulting feature vectors are then fed into a Softmax layer to predict the emotion category.

Figure 4. Overall structure of 2D-CNN model

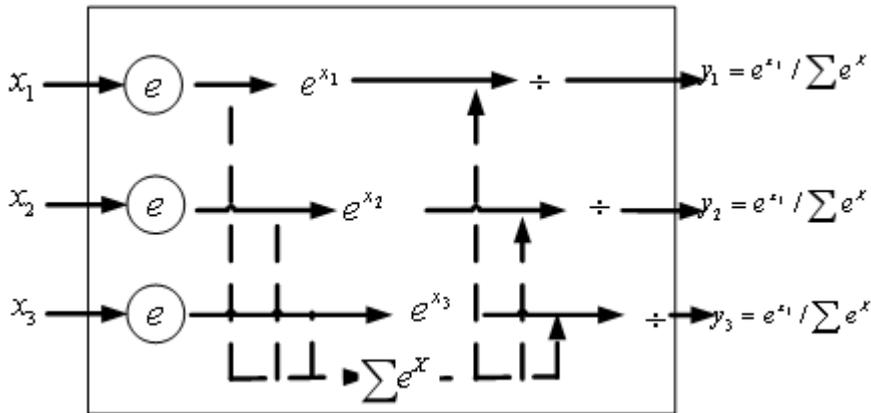


The proposed method utilizes a CNN network consisting of four convolutional layers and one fully connected layer to process a sequence of two-dimensional mesh matrices as input, converting it into a sequence with spatial feature vectors. Instead of using pooling operations for data dimensionality reduction, the proposed method directly connects four convolutional layers. The 2D-CNN model consists of four convolutional layers, each layer using 16, 32, 64, and 128 convolutional kernels of size 3*3 for non-filled convolution operations.

Design of Emotion Classifier

The SCLE-2D-CNN model employs a loss function to measure the disparity between the predicted outcomes and the ground truth values (Jana et al., 2022; Javidan et al., 2021). The Softmax function calculates the probability of a class of elements by dividing the exponent of a class of elements by the sum of the exponents of all elements, as shown in Figure 5. The input to the Softmax function is the vector x , and the calculated output is the vector y , where each element of y is the probability between 0 and 1. The sum of Softmax values for all elements is 1.

Figure 5. Calculation principle of softmax function



Given that emotions in EEG signals are categorized into three classes: positive, neutral, and negative, it is a multi-classification problem. For input data x , the probability size calculated for each classification category k is $P = (y = k | x)$, so the Softmax regression formula is expressed as:

$$h_{\theta}(x^{(i)}) = \begin{pmatrix} P(y^{(i)} = 1 | x^{(i)}; \theta) \\ P(y^{(i)} = 2 | x^{(i)}; \theta) \\ P(y^{(i)} = 3 | x^{(i)}; \theta) \end{pmatrix} = \frac{1}{\sum_{k=1}^3 e^{\theta_k^T x^{(i)}}} \begin{pmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ e^{\theta_3^T x^{(i)}} \end{pmatrix} \quad (10)$$

where, θ represents the parameter for model training.

EXPERIMENT AND ANALYSIS

Experimental Environment

Table 1 outlines the particular environment settings.

Table 1. Experimental Environment Setting

System environment	Settings
Processor	I7-13800H
Graphics card	NVIDIA GeForce GTX 1050Ti
Operating system	Windows10 64
Hard disk	SSD
Deep learning framework	Tensorflow-gpu2.0.0rc0,Opencv-python4.4.0.46,keras2.3.1

Datasets

The experiments were conducted using two datasets: the DEAP dataset and the SEED dataset.

The DEAP dataset, designed for emotion analysis using EEG, physiological, and video signals, is a publicly available dataset. It consists of 32 channels of EEG and eight channels of peripheral physiological signals from 32 participants. In the experiments, only the EEG signals from the 32 channels were utilized. The EEG signals were initially sampled at 512Hz, then down sampled to 128Hz. They were also subjected to bandpass filtering within the range of 4.0–45.0Hz, and any EOG artifacts were removed.

The SEED dataset is the processes involved in data collection, labeling, and other aspects adhering to rigorous standardization. Each participant watched 15 Chinese movie clips, each lasting about four minutes. The emotions in the movie clips were divided into three types: positive, neutral, and negative, with five movie clips for each emotion. Each participant participated in three separate experiments, each occurring approximately one week apart.

Evaluation Indicators

Performance measurement is an evaluation of the generalization ability of a classifier to generalize to new data, or its ability to make accurate predictions on data that it has not been trained on. According to the true and predicted categories of the sample, it is divided into True positive (TP), False positive (FP), True negative (TN), and False negative (FN).

- 1) Accuracy (A_{cc}) is the proportion of correctly predicted results to the overall sample size and is calculated as:

$$A_{cc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

- 2) Precision (P_{re}) represents the probability of the actual positive samples among the predicted positive samples. It is calculated as:

$$P_{re} = \frac{TP}{TP + FP} \quad (12)$$

- 3) Recall (R_e) is the probability of being predicted as a positive sample in the actual positive sample and is calculated as:

$$R_e = \frac{TP}{TP + FN} \quad (13)$$

- 4) F1 Score is the weighted harmonic average of Recall and Precision. It is calculated as:

$$F_1 = \frac{2 \times P_{re} \times R_e}{P_{re} + R_e} \quad (14)$$

The computational complexity of SCLE-2D-CNN mainly depends on three parts: the computational complexity of the wavelet packet transform is $O(n \log n)$, where n is the length of the original EEG signal. The computational complexity of scaling convolution layers depends mainly on the number of layers and the size of the convolution kernel, i.e. $O(n^2)$. The computational complexity of 2D-CNN depends mainly on the depth of the network and the number of parameters, which is $O(n^2)$. The total complexity is:

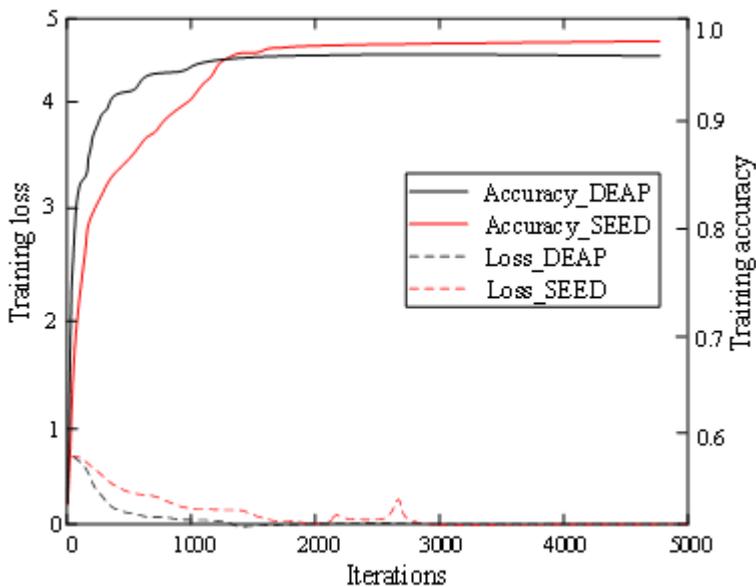
$$O(n \log n) + O(n^2) + O(n^2) = O(n \log n + 2n^2)$$

MODEL TRAINING

Training of Loss and Accuracy

Figure 6 illustrates the training progress curves of the SCLE-2D-CNN model on both the DEAP and SEED datasets.

Figure 6. Training curves of SCLE-2D-CNN model on two datasets



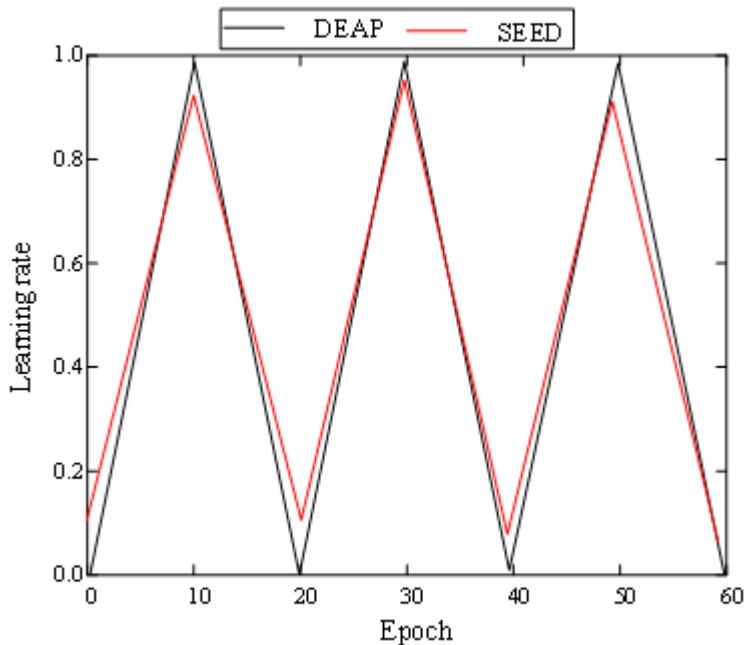
As shown in Figure 6, the average error loss on the SEED training set exhibited a notable surge followed by a swift decline, but overall, it decreased and continued to approach zero as the number of iterations increased. The loss value on the DEAP training set can quickly converge and approach zero. The training accuracy on the DEAP and SEED datasets ultimately converged to around 0.98 and 0.97, respectively. The training accuracy on the SEED dataset converges slowly and tends to stabilize when the number of iterations is about 2,000. The reason for the oscillation of the Loss_SEED curve may be that the hyperparameter rate for learning is set too high. At the beginning of the training, the performance of the model improves rapidly. It can be argued that the SCLE-2D-CNN model can achieve fast convergence on both datasets and has ideal training results.

Training of Learning Rate

To ensure a reasonable convergence process for the model, the SCLE-2D-CNN adopted the CyclicLR learning rate adjustment strategy, as shown in Figure 7.

Figure 7 shows that the learning rate gradually increases during the first few rounds with a small initial value. This prevents oscillation when the initial loss value is large. After the loss gradually decreases, the learning rate also gradually increases to the set maximum value, taking 0.01 as an example. At this point, the model may fall into a local optimum, and a higher learning rate can help it escape the local optimum and move toward the possible global optimum. By repeatedly changing the learning rate throughout the entire training process, the model is more likely to find the global optimal solution. In the final stage of training, a smaller learning rate can be used to approach the global optimum more precisely.

Figure 7. Change curve of CyclicLR learning rate



Performance Verification of Time-Frequency Feature Extraction

To demonstrate the efficacy of time-frequency feature extraction methods, taking the SEED dataset as an example, tenfold cross-validation was used for experimental comparison. The experimental samples were divided into training and validation set samples in a 5:1 ratio. Simultaneously, the proposed model compared SCL with differential entropy, power spectrum, and wavelet entropy. Table 2 presents the accuracy of emotion classification using different feature extraction methods.

Table 2 shows that all the models can achieve good recognition results. Among them, the proposed model which uses SCL to extract time-frequency features can obtain the best results, with an average accuracy of up to 98.09%.

Statistical analysis shows that the standard deviation of time-frequency features extracted using SCL is smaller than that of features extracted using other methods, indicating better stability. The EEG signal contains both time and frequency dimension information. When extracting time-frequency

Table 2. Accuracy of Emotion classification under different feature extraction methods

Subject	Wavelet entropy/%	Power spectrum/%	Differential entropy/%	SCL/%
1	90.45±2.5	91.38±1.5	92.09±1.2	94.12±1.1
2	90.48±2.9	92.91±1.7	93.37±2.9	96.77±1.5
3	96.51±1.8	95.72±2.9	97.85±0.8	98.05±1.0
4	89.61±3.9	91.84±3.9	93.45±1.9	96.95±1.3
5	96.98±1.7	96.75±0.1	98.52±2.8	98.14±1.2
6	90.69±1.4	92.98±3.6	93.25±1.4	96.92±1.2
7	94.07±3.5	94.46±3.3	95.87±2.5	98.20±0.5
8	98.93±0.7	97.13±1.9	98.02±0.3	98.23±1.2
9	95.81±2.4	96.52±2.9	97.71±1.4	97.64±0.9
10	95.96±2.5	97.06±0.1	98.38±1.5	98.43±0.8
11	96.05±2.1	96.59±1.4	97.65±1.9	98.19±0.6
12	92.52±4.8	94.74±0.4	95.70±0.3	97.17±1.2
13	93.89±3.0	94.28±2.6	95.93±2.0	97.49±1.3
14	94.56±3.2	95.57±2.3	96.08±1.2	97.12±1.0
15	95.20±2.9	95.92±1.5	96.31±1.6	97.52±1.3
Mean value	94.11±2.6	94.92±1.9	96.08±1.4	97.09±1.0

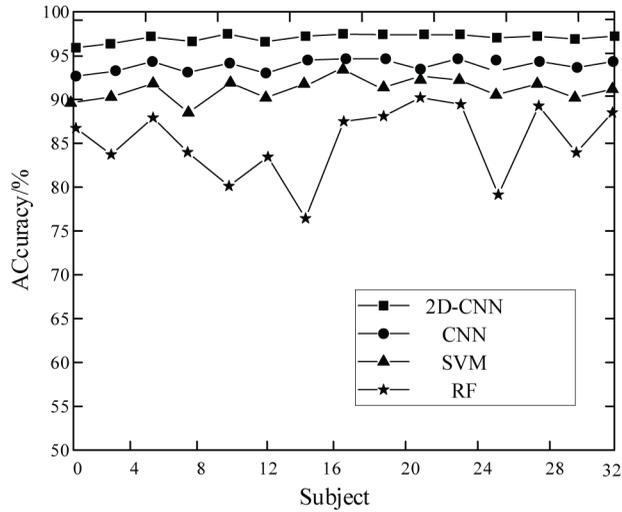
features from EEG signals using differential entropy, power spectrum, and wavelet entropy, it can be challenging to balance the information from both dimensions, resulting in inadequate emotional representation. To address this issue, the proposed SCLE-2D-CNN model effectively captures time and frequency information from multiple channels by employing stacked SCLs, resulting in richer emotional features and ensuring emotion classification accuracy.

Performance Verification of Deep Spatiotemporal Feature Extraction

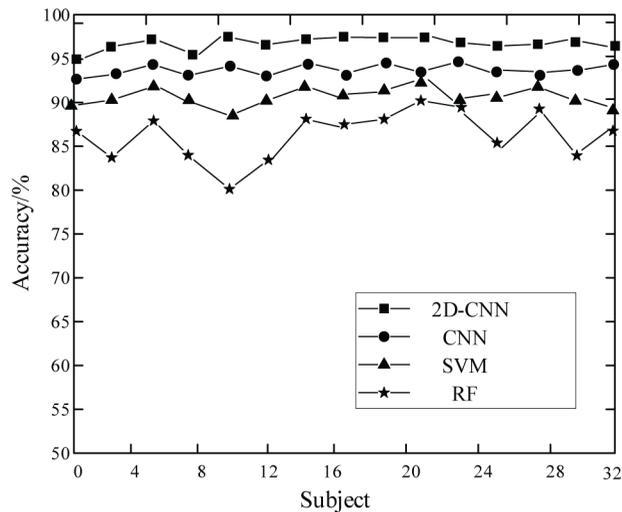
To demonstrate the classification performance of 2D-CNN, it was compared with the emotion recognition results of SVM, random forest algorithm (RF), and CNN classifier (Xing et al., 2019). Experiments were conducted on the DEAP dataset to perform binary classification in the arousal and valence dimensions. Figure 8 illustrates the recognition accuracy of subjects using various classifiers.

Figure 8 shows that the RF classifier has significantly lower accuracy than other models, with large fluctuations. Its average recognition accuracy is around 84%. This may be due to the excessive number of RF parameters, complex parameter tuning, and algorithm settings not applicable to all subjects. Compared with the RF algorithm, the other three models have achieved satisfactory results in emotion recognition on both datasets, with an accuracy rate of over 90%. Due to the capacity of the 2D-CNN model to understand deeper spatiotemporal features, the accuracy of emotion recognition for multi-channel raw EEG signals is higher, approaching 97%. However, the SVM model has certain shortcomings in recognizing multi-classification problems. Hence, the emotion recognition accuracy is about 90%. Since the learning depth of CNN is not deep enough, the emotion recognition accuracy is around 93%. The obtained results demonstrate that in multi-channel EEG emotion recognition tasks, 2D-CNN has better deep spatiotemporal feature extraction ability than other machine learning models.

Figure 8. Recognition accuracy of subjects under different classifiers



(a) Results of arousal on the DEAP dataset



(b) Results of valence on the DEAP dataset

Comparison With Other Advanced Models

Both subject-dependent and subject-independent experiments were conducted, designed to compare and analyze the emotion performance of four models including the 2DCNN-BiGRU model (Shi et al., 2021), the HF-DNN model (Kumar & Suresh, 2021), the SVM-LSTM model (Chen et al., 2021), and the proposed SCLE-2D-CNN. Experiments were designed on the DEAP dataset to perform binary classification and four-class classification under two dimensions, namely arousal and valence. The accuracy and F1 value of the three-class classification experiments were mainly given on the SEED dataset. The corresponding results on the two datasets are listed in Tables 3 and 4.

Table 3 shows that in the subject dependent experiment, the proposed model achieves high accuracy in binary classification, with 96.96% and 96.83% accuracy in the arousal and valence dimensions of

Table 3. Recognition results of different models on the DEAP datasets

Approach	Compared models	Binary classification (%)		Four-class classification (%)
		Arousal	Valence	
Subject-Dependent	2DCNN-BiGRU	96.01	95.65	93.58
	HF-DNN	91.88	91.27	88.57
	SVM-LSTM	95.32	95.06	93.69
	SCLE-2D-CNN (ours)	96.96	96.83	94.82
Subject-Independent	2DCNN-BiGRU	79.95	78.04	76.36
	HF-DNN	74.30	73.69	71.51
	SVM-LSTM	78.76	77.32	74.03
	SCLE-2D-CNN (ours)	80.29	80.58	78.75

Table 4. Recognition results of different models on the SEED dataset

Approach	Compared models	Accuracy/%	F1/%
Subject-Dependent	2DCNN-BiGRU	97.55	96.23
	HF-DNN	91.47	91.22
	SVM-LSTM	95.26	94.89
	SCLE-2D-CNN (ours)	98.09	97.05
Subject-Independent	2DCNN-BiGRU	84.25	84.03
	HF-DNN	79.78	78.82
	SVM-LSTM	82.10	81.86
	SCLE-2D-CNN (ours)	85.58	85.49

the DEAP dataset, respectively. Additionally, it achieves a quadrivariate classification accuracy of 94.82%. Table 4 demonstrates the model’s impressive performance in subject-related experiments on the SEED dataset, with accuracy and F1 values reaching 98.09% and 97.05%, respectively. On both datasets, most comparison models perform worse in subject-independent experiments than in subject-dependent experiments. This may be because these models are not specifically designed for scenarios that are independent of the subject.

However, mixed features contain redundant information that can lead to under expression of emotion. Taking DEAP data as an example, in the subject-dependent experiment, the binary results of high frequency dnn in the arousal dimension and titer dimension are only 91.88% and 91.27%, respectively. The SCLE-2D-CNN model proposed in this paper outperforms the 2DCNN-BiGRU model on the SEED dataset, and the F1 score of the subject correlation experiment reaches 96.96%. Although the 2DCNN-BiGRU model takes into account the spatio-temporal characteristics of the signal, the lack of attention to the multi-channel factors results in lower recognition results than the proposed model. In summary, the SCLE-2D-CNN model proposed in this paper has higher accuracy and better generalization ability.

Ablation Experiments

The effects of time-frequency features and EEG channel semantic enhancement module on emotion recognition performance were further discussed through ablation experiments and analysis. Firstly,

SCL and EEG channel semantic enhancement modules are removed from the model. Then, the time-frequency characteristics and the function of EEG channel semantic enhancement module are studied in turn. The results of the ablation experiment taking SEED data set as an example are shown in Table 5.

Table 5. Results of ablation experiment on SEED dataset

Model	Recognition accuracy/%				F1%
	Positive emotion	Neutral emotion	Negative emotion	Average	
baseline	95.18	94.45	95.37	95.00	95.23
baseline+SCL	97.56	96.81	97.29	97.22	96.86
baseline+SCL+SENet	98.71	97.38	98.18	98.09	97.05

Table 5 shows that adding an SCL module to the baseline can extract high-dimensional feature tensors from signals and improve recognition accuracy. After the introduction of SENet, SCL modules were stacked and SENet modules were configured for each channel of EEG signal. This enhanced the channels related to emotion recognition tasks and suppressed the channels unrelated to them, allowing EEG signals to be recognized more accurately. Consequently, the accuracy of positive, neutral, and negative emotions achieved by the suggested SCLE-2D-CNN model reached 98.71%, 97.38%, and 98.18%, respectively. These results indicate that the time-frequency features and EEG channel semantic enhancement module are effective in recognizing emotions on EEG signals.

CONCLUSION

This paper introduces a novel method for EEG signal emotion recognition called SCLE-2D-CNN. The proposed method effectively incorporates the time-frequency information of multi-channel EEG signals, enabling a comprehensive exploration of the relationship between emotional dimensions and EEG channels. The experimental results on the DEAP and SEED datasets demonstrate the method’s ability to achieve reliable recognition of EEG signal emotions.

- 1) By stacking SCL, the feature extraction effect can be improved; the proposed method ensures high accuracy in classifying emotions. The average accuracy of the proposed SCLE-2D-CNN in emotion classification on the SEED dataset reaches 98.09%.
- 2) The SCLE-2D-CNN model enhances the effective emotional information in time-frequency features through the EEG channel semantic enhancement module and introduces 2D-CNN for emotion classification. On the SEED dataset, the accuracies of positive, neutral, and negative emotions are 98.71%, 97.38%, and 98.18%, respectively, which are superior to other newer comparative models.

Human emotions are often expressed through multiple modalities such as EEG, images, speech, and text. The proposed model only considers the EEG signals, which cannot fully capture the complexity of human emotions. Therefore, in the future work, some other advanced deep learning networks, such as graph neural networks (GNN) (Li et al., 2022a), and BiLSTM (Rahman et al., 2022), new emotion recognition architectures will be introduced to enhance the classification performance of EEG emotion recognition tasks.

AUTHOR'S NOTE

The data used to support the findings of this study are included within the article.

The author declares no conflicts of interest.

This work was supported by the Soft Science Research Program of Henan Province (No. 222400410485).

Correspondence concerning this article should be addressed to Weisi Yang, Department of General and Liberal Arts Education, Sichuan Vocational College of Finance and Economics, Chengdu, 610000, China, E-mail: yangchuancai2023@163.com

REFERENCES

- Appati, J. K., Brown, G. A., Soli, M. A. T., & Denwar, I. W. (2021). A review of computational intelligence models for brain tumour classification and prediction. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, 13(4), 18–39. doi:10.4018/IJSSCI.2021100102
- Bhosale, S., Chakraborty, R., & Kopparapu, S. K. (2022). Calibration free meta learning based approach for subject-independent EEG emotion recognition. *Biomedical Signal Processing and Control*, 72(2), 103289–103295. doi:10.1016/j.bspc.2021.103289
- Chamra, A., & Harmanani, H. (2020). A smart green house control and management system using IoT. In *17th international conference on information technology–new generations (ITNG 2020)*. Springer International Publishing.
- Chen, T., Yin, H., Yuan, X., Gu, Y., Ren, F., & Sun, X. (2021). Emotion recognition based on fusion of long short-term memory networks and SVMs. *Digital Signal Processing*, 117(1), 103153–103162. doi:10.1016/j.dsp.2021.103153
- Dan, X. (2021). Emotion recognition of single electrode EEG based on multi feature combination in time frequency domain. *Journal of Physics: Conference Series*, 1827(1), 012031–012038. doi:10.1088/1742-6596/1827/1/012031
- Dwivedi, R. K. (2022). Density-based machine learning scheme for outlier detection in smart forest fire monitoring sensor cloud. [IJCAC]. *International Journal of Cloud Applications and Computing*, 12(1), 1–16. doi:10.4018/IJCAC.305218
- Ghosh, S.M., Bandyopadhyay, S., & Mitra, D. (2021). Nonlinear classification of emotion from EEG signal based on maximized mutual information. *Expert Systems with Application*, 185.
- Guebli, W., & Belkhir, A. (2021). Inconsistency detection-based LOD in smart homes. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, 17(4), 56–75. doi:10.4018/IJSWIS.2021100104
- Hu, J. Z., Wang, C., Jia, Q. M., Bu, Q. R., & Feng, J. (2021). ScalingNet: Extracting features from raw EEG data for emotion recognition. *Neurocomputing*, 463, 177–184. doi:10.1016/j.neucom.2021.08.018
- Islam, M. R., Moni, M. A., Islam, M. M., Rashed-Al-Mahfuz, M., Islam, M. S., Hasan, M. K., Hossain, M. S., Ahmad, M., Uddin, M. S., Azad, A., Alyami, S. A., Ahad, M., & Lió, P. (2021). Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques. *IEEE Access : Practical Innovations, Open Solutions*, 9(3), 94601–94624. doi:10.1109/ACCESS.2021.3091487
- Jana, G. C., Sabath, A., & Agrawal, A. (2022). Capsule neural networks on spatio-temporal EEG frames for cross-subject emotion recognition. *Biomedical Signal Processing and Control*, 72(7).
- Javidan, M., Yazdchi, M., Baharlouei, Z., & Mahnam, A. (2021). Feature and channel selection for designing a regression-based continuous-variable emotion recognition system with two EEG channels. *Biomedical Signal Processing and Control*, 70(7).
- Jiang, H., Jiao, R., Wang, Z., Zhang, T., & Wu, L. (2021). Construction and analysis of emotion computing model based on LSTM. *Complexity*, 4.
- Jiao, X., Yu, X., Peng, H., & Zhang, X. (2022). A smart learning assistant to promote learning outcomes in a programming course. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, 14(1), 1–23. doi:10.4018/IJSSCI.312557
- Jie, S., Wang, X., Zhao, K. N., Hao, S. Y., & Wang, T. Y. (2022). Multi-channel EEG emotion recognition based on parallel transformer and 3D-convolutional neural network. *Mathematics*, 10(3131).
- Kouti, M., Ansari-Asl, K., & Namjoo, E. (2022). Emotion discrimination using source connectivity analysis based on dynamic ROI identification. *Biomedical Signal Processing and Control*, 72(2), 103332.1-103332.12.
- Kumar, H., Ganapathy, N., Puthankattil, S. D., & Ramakrishnan, S. (2022). Classification of emotional states using EEG signals and wavelet packet transform features. *Studies in Health Technology and Informatics*, 294, 943–944. doi:10.3233/SHTI220632 PMID:35612250

- Kumar, M., & Suresh, K. (2021). Disabled person emotion recognition in EEG signal using deep neural network. *International Journal of Advanced Intelligence Paradigms*, 19(3/4), 295–313. doi:10.1504/IJAIP.2021.116358
- Li, C., Sheng, Y., Wang, H., Niu, M., Jing, P., Zhao, Z., & Schuller, B. W. (2022a). EEG emotion recognition based on self-attention dynamic graph neural networks. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. (pp. 292-296). IEEE.
- Li, D., Deng, L., Gupta, B. B., Wang, H., & Choi, C. (2019). A novel CNN based security guaranteed image watermarking generation scenario for smart city applications. *Information Sciences*, 479, 432–447. doi:10.1016/j.ins.2018.02.060
- Li, W., Tan, R., Xing, Y., Li, G., Li, S., Zeng, G., Wang, P., Zhang, B., Su, X., Pi, D., Guo, G., & Cao, D. (2022b). A multimodal psychological, physiological and behavioural dataset for human emotions in driving tasks. *Scientific Data*, 9(481), 481. doi:10.1038/s41597-022-01557-2 PMID:35933432
- Liu, Y., & Fu, G. (2021). Emotion recognition by deeply learned multi-channel textual and EEG features. *Future Generation Computer Systems*, 119(1), 1–6. doi:10.1016/j.future.2021.01.010
- Madhu, S., Padunnavalappil, S., Saajjal, P. P., Vasudevan, V. A., & Mathew, J. (2022). Powering up an IoT-enabled smart home: A solar powered smart inverter for sustainable development. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, 14(1), 1–21. doi:10.4018/IJSSCI.300362
- Mandle, A. K., Sahu, S. P., & Gupta, G. P. (2022). CNN-based deep learning technique for the brain tumor identification and classification in MRI images. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, 14(1), 1–20. doi:10.4018/IJSSCI.304438
- Naser, D. S., & Saha, G. (2021). Influence of music liking on EEG based emotion recognition. *Biomedical Signal Processing and Control*, 64(3), 102251.1-102251.15.
- Olamat, A., Ozel, P., & Atasever, S. (2022). Deep learning methods for multi-channel EEG-based emotion recognition. *International Journal of Neural Systems*, 32(5).
- Oscar, A. C., Dora, A. O., Jose, C. H., & Mario, I. M. (2022). Emotion recognition in EEG signals using the continuous wavelet transform and CNNs. *Neural Computing & Applications*, 35, 1409–1422.
- Rahman, A., Tubaishat, A., Al-Obeidat, F. N., Halim, Z., Tahir, M., & Qayum, F. (2022). Extended ICA and M-CSP with BiLSTM towards improved classification of EEG signals. *Soft Computing*, 26(20), 10687–10698. doi:10.1007/s00500-022-06847-w
- Rahman, M. M., Sarkar, A. K., Hossain, M. A., Hossain, M. S., Islam, M. R., Hossain, M. B., Quinn, J. M., & Moni, M. A. (2021). Recognition of human emotions using EEG signals: A review. *Computers in Biology and Medicine*, 136(2), 104696–104713. doi:10.1016/j.combiomed.2021.104696 PMID:34388471
- Salankar, N., Mishra, P., & Garg, L. (2021). Emotion recognition from EEG signals using empirical mode decomposition and second-order difference plot. *Biomedical Signal Processing and Control*, 65(8), 102389.1-102389.13.
- Samir, M., Ebrahimi, D., Assi, C., Sharafeddine, S., & Ghayeb, A. (2020). Leveraging UAVs for coverage in cell-free vehicular networks: A deep reinforcement learning approach. *IEEE Transactions on Mobile Computing*, 20(9), 2835–2847. doi:10.1109/TMC.2020.2991326
- Shi, H., Xie, H., & Wu, M. (2021). Emotion recognition based on multi-channel EEG signals. *Journal of Physics: Conference Series*, 2078(1), 012028–012037. doi:10.1088/1742-6596/2078/1/012028
- Subasi, A., Tuncer, T., Dogan, S., Tanko, D., & Sakoglu, Ü. (2021). EEG-based emotion recognition using tunable Q wavelet transform and rotation forest ensemble classified. *Biomedical Signal Processing and Control*, 68(7).
- Tan, C., Arlija, M., & Kasabov, N. (2021a). NeuroSense: Short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns. *Neurocomputing*, 434(4), 137–148. doi:10.1016/j.neucom.2020.12.098
- Tan, H., Gu, Z., Wang, L., Zhang, H., Gupta, B. B., & Tian, Z. (2022). Improving adversarial transferability by temporal and spatial momentum in urban speaker recognition systems. *Computers & Electrical Engineering*, 104, 108446. doi:10.1016/j.compeleceng.2022.108446

- Tan, Y., Sun, Z., Duan, F., Solé-Casals, J., & Caiafa, C.F. (2021b). A multimodal emotion recognition method based on facial expressions and electroencephalography. *Biomedical Signal Processing and Control*, 70(09), 103029.1-103029.11.
- Thirumalaisamy, M., Basheer, S., Selvarajan, S., Althubiti, S. A., Alenezi, F., Srivastava, G., & Lin, J. C. W. (2022). Interaction of secure cloud network and crowd computing for smart city data obfuscation. *Sensors (Basel)*, 22(19), 7169. doi:10.3390/s22197169 PMID:36236264
- Thoumi, S., & Haraty, R. A. (2022, November). Damage assessment and recovery in fog-based computing systems. In *2022 International Conference on Innovation and Intelligence for Informatics, IEEE. Computing, and Technologies (3ICT)*, (pp. 13-18). IEEE.
- Vijayakumar, P., Rajkumar, S. C., & Deborah, L. J. (2022). Passive-awake energy conscious power consumption in smart electric vehicles using cluster type cloud communication. [IJCAC]. *International Journal of Cloud Applications and Computing*, 12(1), 1–14. doi:10.4018/IJCAC.297108
- Xing, X., Li, Z., Xu, T., Shu, L., Hu, B., & Xu, X. (2019). SAE+ LSTM: A New framework for emotion recognition from multi-channel EEG. *Frontiers in Neurobotics*, 13, 37. doi:10.3389/fnbot.2019.00037 PMID:31244638
- Yu, M., Xiao, S., Hua, M., Wang, H., Chen, X., Tian, F., & Li, Y. (2022). EEG-based emotion recognition in an immersive virtual reality environment: From local activity to brain network features. *Biomedical Signal Processing and Control*, 72(3).
- Zali-Vargahan, B., Charmin, A., Kalbkhani, H., & Barghandan, S. (2023). Deep time-frequency features and semi-supervised dimension reduction for subject-independent emotion recognition from multi-channel EEG signals. *Biomedical Signal Processing and Control*, 85, 104806. doi:10.1016/j.bspc.2023.104806
- Zhao, Y., Guo, M., Sun, X. H., Chen, X. Y., & Zhao, F. (2023). Attention-based sensor fusion for emotion recognition from human motion by combining convolutional neural network and weighted kernel support vector machine and using inertial measurement unit signals. *IET Signal Processing*, 17(4), 17. doi:10.1049/sil2.12201

APPENDIX

Table 6. Equation reference

o_t	pulse response coefficients.
v_t	pulse response coefficients.
d_j^{2n}	wavelet packet coefficients
d_j^{2n+1}	wavelet packet coefficients
H_{in}	input signals
H_{in}	output time-frequency feature matrices
ζ	pooling operator
ω	weight of the SCK
K	total length of the time series signal data
s	sum variable
g	one-dimensional time series signal
q	query vector
p_i	probability distribution
z	result of performing GAP
σ	Sigmoid activation functions
κ	ReLU activation functions
θ	parameter for model training

Dan Fu was born in Sichuan, China, in 1985, Female, ethnic Han, associate professor, she received her master's degree in 2014. Her research interests are included psychological health education and positive psychological quality.

Weisi Yang was born in October 1987 in Pengzhou, Sichuan Province, China. Female, ethnic Han, lectureship. She received her master's degree from Sichuan Normal University in 2013. Since 2013, Weisi Yang has been working in Sichuan Vocational College of Finance and Economics. Her research interests: Development and Educational psychology, Psychological counseling.

Pan Li, graduated from Huazhong University of Science and Technology in December 2008, master's degree, associate professor. Research interests: Big data, cloud computing.