ORIGINAL ARTICLE

Enhancing the feasibility of cognitive load recognition in remote learning using physiological measures and an adaptive feature recalibration convolutional neural network

Chennan Wu[1] · Yang Liu[1] · Xiang Guo[1] · Tianshui Zhu[1] · Zongliang Bao[1]

Abstract

The precise assessment of cognitive load during a learning phase is an important pathway to improving students' learning efficiency and performance. Physiological measures make it possible to continuously monitor learners' cognitive load in remote learning during the COVID-19 outbreak. However, maintaining a good balance between performance and computational cost is still a major challenge in advancing cognitive load recognition technology to real-world applications. This paper introduced an adaptive feature recalibration (AFR) convolutional neural network to overcome this challenge by capturing the most discriminative physiological features (EEG and eye-tracking). The results revealed that the optimal average classification accuracy of the feature combination obtained by the AFR method reached 95.56% with only 60 feature dimensions. Additionally, compared with the best result of the conventional correlation-based feature selection (CFS) method, the introduced AFR algorithm achieved higher accuracy and cheaper computational cost, as well as a 2.06% improvement in accuracy and a 51.21% reduction in feature dimension, which is more in line with the requirements of low delay and real-time performance in practical BCI applications.

1 Introduction

According to cognitive load theory by Sweller et al. [1–3], people have limited cognitive resources for processing and holding information. Learners might fail to complete cognitive tasks that exceed human working memory capacity due to cognitive overload [4]. Therefore, controlling and optimizing learners' limited cognitive resources to achieve the best learning effect is important in engineering education [5].

The recent spread of COVID-19 has driven most schools to remote learning, leading to the emergence of a growing number of distance-learning applications. In a real-world classroom, experienced instructors can acutely sense whether the students understand the material and concepts being taught, which allows teachers to adjust teaching strategies immediately and reasonably. However, for remote teaching, it becomes difficult for teachers to judge the teaching effect through observation. Therefore, instructors desperately need new tools for monitoring and evaluating students' cognitive states to effectively conduct remote teaching.

Physiological signals have emerged as an alternative for overcoming this challenge. Traditionally, it is accepted that physiological signals are the most effective indicator of cognitive load because they provide efficient temporal resolution of long-term monitoring [6] and higher feasibility for estimating cognitive load compared to subjective rating methods [7]. In particular, electroencephalography (EEG) has been proven to be an effective, noninvasive method for detecting, estimating, or predicting human brain activities [8, 9]. The different rhythms generated by electrical brain activity, such as delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–50 Hz), are the most popular features in the context of cognitive load recognition [10], especially the theta and alpha range, which seem to involve higher brain function, reflecting task difficulty or cognitive load among diverse task demands [11–15]. Moreover, nonlinear EEG features (i.e., spectral entropy) also perform well in EEG cognitive analysis [16–18].

✉ Yang Liu
  liuyang@zust.edu.cn

[1] School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, China

In addition, eye-tracking is also a commonly used measure in analysing the effectiveness of learning materials [19]. For eye-tracking data, previous studies have confirmed that pupil diameter and fixation duration are sensitive indicators of variations in mental workload [20, 21]. Furthermore, combining EEG and eye-tracking signals can gather multimodal information, thereby leading to better accuracy and cognitive load measurement performance [22, 23].

However, improving classification performance and reducing computational costs is still a major challenge to advancing cognitive load recognition technology from laboratory to real-world applications. To address this, researchers have attempted to develop appropriate classifiers using different machine learning (ML) techniques. Nue et al. developed a lightweight CNN model to classify cognitive states for real-time computing environments, which reduces trainable parameters while maintaining high performance [24].

This paper focuses on improving the feature selection methods to select the most discriminative features to remedy the above challenge. In previous studies, the common algorithm used for feature selection was the CFS method, where the features are selected by ranking their correlation coefficient between the features and class labels [25, 26]. It has been shown to be effective in reducing feature dimensions [27].

Recently, the widespread use of deep learning in a variety of fields has shown its superiority over conventional machine learning, which motivates us to apply an adaptive feature recalibration (AFR) convolutional neural network to select the most discriminative features. The AFR algorithm is representative of the attention mechanism and was first proposed by Hu et al. in computer vision [28, 29]. In this research, we applied the AFR method to enhance feature learning in the EEG and eye-tracking feature selection field, which is realized by diverting attention to the most critical features and disregarding irrelevant features, thus determining what to focus. It is a novel perspective on enhancing cognitive load recognition performance since cognitive load recognition performance relies mainly on the quality of selected features that characterize cognitive activity [30].

To apply cognitive load recognition techniques to a remote learning environment, physiological signals are used as a medium, and the AFR method is introduced to maintain a good balance between performance and complexity. In this paper, we acquired a two-level cognitive load using EEG and eye-tracking from 42 healthy subjects. First, 123 kinds of handcrafted features were extracted, and statistical analysis was performed to verify whether physiological features can be used to discriminate different cognitive load levels in distance learning. Then, the AFR method was employed to select discriminate features for classification, which was accomplished by a radial basis function support vector machines (SVMs) classifier. Additionally, to demonstrate the advantages of the AFR method, the classification accuracy and feature dimensions used for classification are compared between the introduced AFR method and the traditional CFS method.

2 Material

2.1 Participants

This study was reviewed and approved by the Ethics Committee of Zhejiang University of Science and Technology. Forty-two undergraduate students (16 females, 26 males) with an average age of 20.81 ± 1.13 years provided written informed consent to participate in the study. In addition, participants' scores in a programming course were utilized to guarantee that they had similar prior knowledge ($F(1, 38) = .18, p = .67$). Participants with (1) a history of major craniocerebral injury or neurological disease or (2) low arousal levels prior to the trial phase were excluded from the experiments [31].

2.2 Task and stimuli

The participants were asked to complete a Python-themed online course learning task. According to the conclusion of Mayer et al., positive emotions in multimedia learning facilitate cognitive processes and learning, which can be produced by the design of various multimedia elements, such as the layout, colour, and sound [32, 33]. Therefore, we took colour as an independent variable to design two versions of the online course that would induce two different cognitive load levels. Screenshots of the two online courses are presented in Fig. 1. As shown in Fig. 1, the Neutral Emotional Design (NED) online course was achromatic, and the Positive Emotional Design (PED) adopted the "Palenight Theme" to highlight code format. All other variables were maintained at constant values (including course content and teacher teaching style). We hypothesized that an online course designed with positive emotion produces lower cognitive load levels.

2.3 Procedures

Subjects were randomly assigned to either the PED group or the NED group. Before the experiment began, we put electrode caps on them, injected conductive adhesive into the electrodes, and calibrated the eye-movement camera. We then asked participants to close their eyes for 3 min so that we could take a baseline measurement. Then, the formal experiment began: students were presented with the designated online course, which lasted an estimated 6 min. Upon viewing the assigned online course, the participants' perceptions of task difficulty were collected using a nine-point

Likert scale (i.e., how easy or difficult was the online course to understand? [34]), which has been shown to be effective in assessing learners' cognitive load during the learning process [35, 36]. Meanwhile, the participants completed the learning performance test to evaluate their ability to understand and retain the information presented in the courses. The entire process lasted an estimated 40 min.

3 Methods

The combined recording of EEG and eye-tracking data obtained in the experiment was analysed in a variety of ways, including signal preprocessing, feature extraction, feature selection, and cognitive load classification. An overview of the detailed analysis procedure is presented in Figs. 2 and 4.

3.1 Signal preprocessing

The EEG collection device used in this study was Open-BCI, and the eye-tracking equipment was Tobii T120. EEG was measured at 15 electrode positions with a sampling rate of 125 Hz (see Fig. 3), arranged according to the international 10/20 system [37]. The original EEG and eye-tracking data were preprocessed and analysed in MATLAB using customized MATLAB scripts. We visually examined the EEG and eye-tracking data and excluded four subjects' data due to poor quality signals, leaving the PED and NED groups with 19 valid subjects' data.

First, event markers at the start and the end of stimulus in both EEG and eye-tracking data served as synchronization events, and the eye-tracking data were integrated into the EEG data as additional channels with an upsampling rate operation (i.e., channels that contain the pupil diameter of each eye and the fixation duration); second, empirical mode decomposition (EMD) was used to decompose and filter the EEG signal; thereafter, the EEG signals were rereferenced by the average of all electrodes; finally, to correct artefacts such as blinking and muscle tension, the independent component correlation algorithm (ICA) was performed [38–40].

3.2 Feature extraction

The primary purpose of feature extraction is to derive salient features that can map physiological signals into consequent cognitive states. Additionally, extracting appropriate features plays an important role in classification tasks.

In this paper, we conducted two separate EEG data analyses (i.e., nonlinear dynamics and wavelet transform) and calculated cognitive-related eye-tracking measures (pupil

(a) Neutral Emotional Design (b) Positive Emotional Design

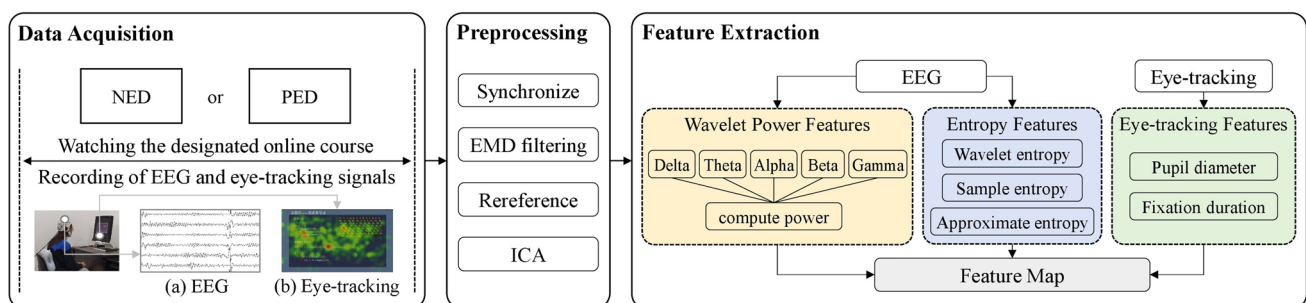**Fig. 1** Screenshots of two versions of the online courses

Fig. 2 The procedure of data processing (data acquisition, preprocessing, and feature extraction)
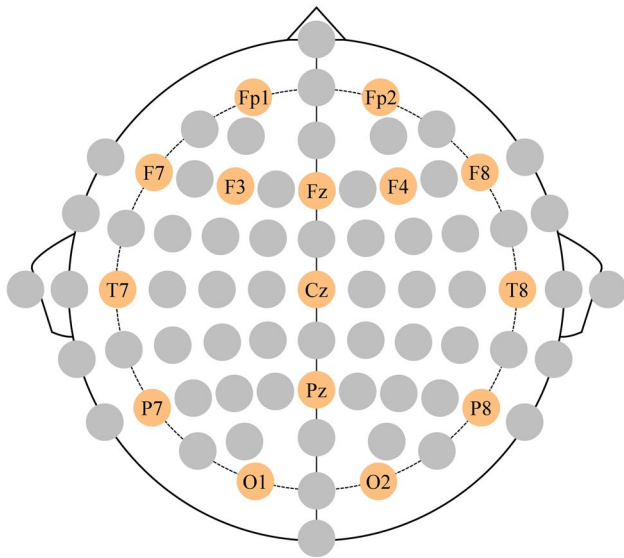
Fig. 3 Electrode placement configuration according to the 10–20 system

diameter and fixation duration) based on data epochs with a size of 4 s. Consequently, we extracted 123 kinds of cognitive load-related features, which consisted of 75 frequency band power features (5 rhythm × 15 channels), 45 entropy features (approximate, sample, and wavelet entropy × 15 channels), and 3 eye-tracking indicators (pupil diameter of each eye and fixation duration).

3.2.1 Nonlinear dynamics

EEG signals are highly complex, and nonlinear analysis is an especially important method to analyse physiological time series that contain complex dynamics. In this paper,

we studied EEG signals using two nonlinear dynamical methods: approximate entropy and sample entropy [41, 42].

Approximate entropy Approximate entropy (ApEn), proposed by Pincus in 1991, can be used to describe the irregularity of a time series [18]. The approximate entropy value reflects the degree of complexity corresponding to a time series, indicating the probability of generating a new pattern as the dimension increases. Intuitively, the more irregular and complex the time series is, the larger the corresponding ApEn value. ApEn can be reliably estimated from relatively short and noisy data. The details of the algorithm are shown below [43, 44].

A time series $\{x(N)\} = \{x(1), x(2), x(3), \ldots, x(N)\}$ consisting of N pieces of data is defined, and the corresponding approximate entropy is calculated according to the following steps:

Step 1: Each t successive data in a time series can form a t-dimensional vector, that is:

$$\{x_t(j)\} = \{x(j), x(j+1), \ldots, x(j+t-1)\} \quad (1)$$

where $1 \leq j \leq N - t + 1$.

Step 2: The maximum distance between time series $\{X_t(j)\}$ and $\{X_t(k)\}$ is denoted as $d[X_t(j), X_t(k)]$, which is the largest absolute value of the difference between each component:

$$d[X_t(j), X_t(k)] = \max\{|x(j+p) - x(k+p)|\} \quad (2)$$

where $0 \leq p \leq t - 1; 1 \leq j, k \leq N - m + 1, i \neq j$

Step 3: Given a positive threshold ω, if the maximum distance of two time series does not exceed the threshold ω, the two time series are considered similar. The number of time series that are similar to time series $\{X_t(j)\}$ can be
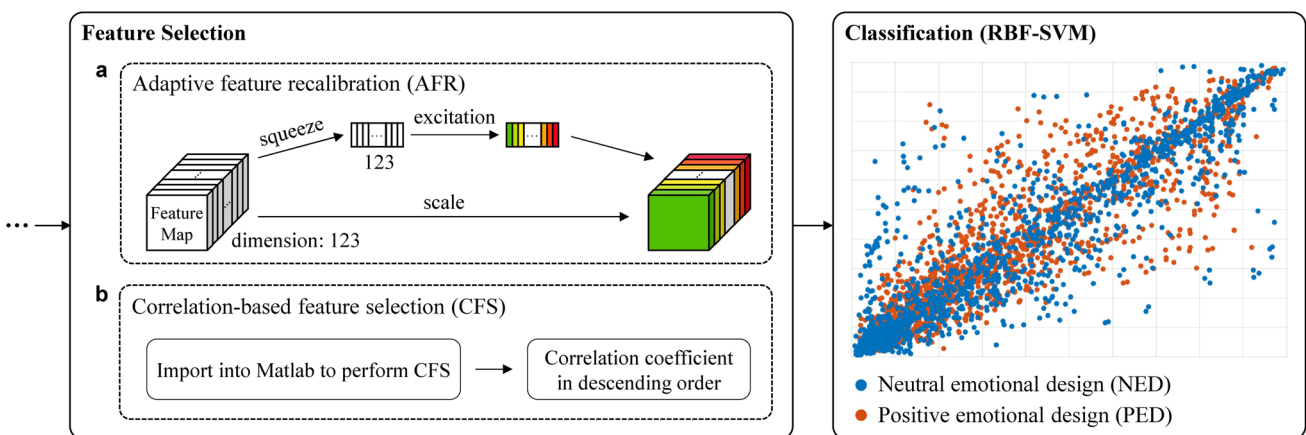

Fig. 4 The overall framework of the AFR and CFS models for cognitive load classification

counted, and its proportion to the total number of time series is given as:

$$\varnothing_j^t(w) = \frac{1}{N-t+1}\varnothing_j \tag{3}$$

where $\varnothing_j$ represents the number of time series that match $d\left[X_t(j), X_t(k)\right] < \omega$

Step 4: For each $\varnothing_j^t(w)$, calculate its logarithmic average:

$$\varnothing^t(w) = \frac{1}{N-t+1}\sum_{j=1}^{N-t+1}\ln\varnothing_j^t(w) \tag{4}$$

Step 5: Increase the dimension to $t+1$. Repeating steps 1–4, $\varnothing^{t+1}(w)$ can be calculated as:

$$\varnothing^{t+1}(w) = \frac{1}{N-t}\sum_{j=1}^{N-t}\ln\varnothing_j^{t+1}(w) \tag{5}$$

Step 6: The approximate entropy is defined as:

$$\text{ApEn}(t, w, N) = \varnothing^t(w) - \varnothing^{t+1}(w) \tag{6}$$

Sample entropy Sample entropy (SampEn), proposed by Richman et al. [45], is a time series complexity measure based on improved ApEn, in which it does not need to perform its own matching [46]. Intuitively, the value of SampEn reflects the degree of complexity in a time series. The sample entropy is defined as follows:

$$\phi^t(w) = \frac{1}{N-t}\sum_{i=1}^{N-t}\phi_i^t(w) \tag{7}$$

$$\text{SampEn}(t, w, N) = \ln\phi^t(w) - \ln\phi^{t+1}(w) \tag{8}$$

3.2.2 Wavelet transform

Wavelet transform is a typical and practicable time–frequency analysis method that has been widely used in analysing EEG signals [47–49]. By using the Daubechies wavelet as the basis function of the wavelet transform, the EEG signal was decomposed into five subbands [50]: delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–50 Hz); in this manner, the frequency components of five frequency bands in the EEG signal could be extracted. For EEG signals of 15 channels, three features were extracted from the wavelet coefficients of each subband, which were calculated as [51]:

(1) Wavelet energy

$$E(i) = \sum_{j=1}^{n_i} D_{i,j}^2 \tag{9}$$

(2) Wavelet energy ratio

$$R(i) = E(i)\Big/\sum_{j=1}^{n} E(j) \tag{10}$$

(3) Wavelet entropy

$$W_e = -\sum_{i=1}^{n} R(i)\ln R(i) \tag{11}$$

where $D_{i,j}$ represents wavelet coefficients of the corresponding decomposition levels.

3.3 Feature selection

Feature selection is a critical step in bringing cognitive load recognition techniques to practical use since selecting appropriate features can improve the model's learning efficiency and significantly benefit the performance of cognitive load recognition [52]. In this paper, two feature selection methods are compared. The first is the AFR method introduced in this paper, which utilizes a convolutional neural network to select discriminative features. The second is the conventional feature selection method represented by CFS, which selects prominent features according to statistical parameters (i.e., correlation coefficient). Both of them obtain a weight distribution, the former from squeeze-and-excitation networks and the latter from the correlation coefficient (Fig. 4).

3.3.1 Adaptive feature recalibration method

The AFR method aims to improve feature representation ability by adaptively selecting the discriminative features through the squeeze-and-excitation (SE) network [29].

Our AFR CNN model contains three SE blocks, which are inserted into three normal convolution layers. The specific structure is illustrated in Fig. 5, and the parameters of each CNN layer are described in Table 1. A special classification task is set to evaluate the quality of the weight coefficients generated by the first SE block. In particular, we choose a depthwise convolution (DC) layer instead of a generic convolution (GC) for the first SE block since a generic convolution mixes the channel information, which is not conducive to channel interpretation [28].

For each SE block, the "squeeze" operation first aggregates feature maps across the time dimension by carrying out global average pooling, which shrinks feature map $F \in R^{N \times d}$ to $S \in R^{N \times 1}$, where N denotes the total feature size and d denotes the data length. Then, an "excitation"
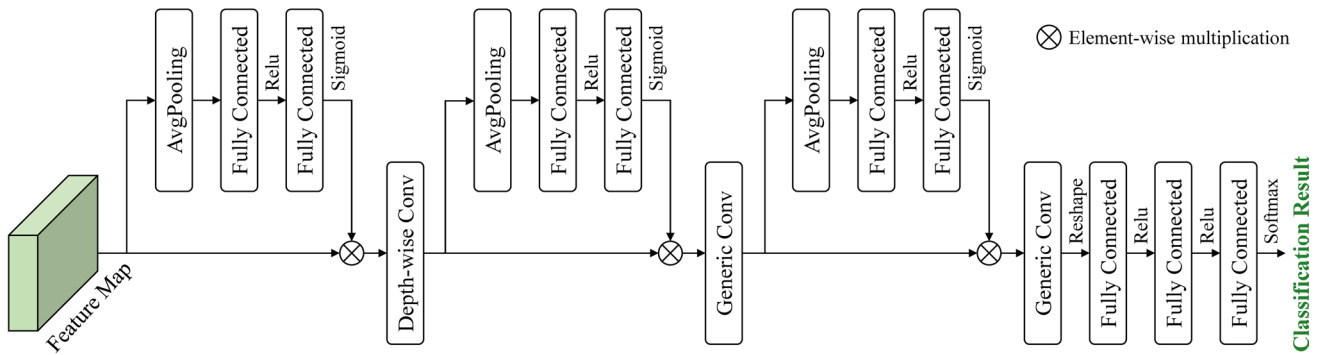
Fig. 5 The structure of the AFR convolutional neural network (batch normalization after each convolutional block is omitted in the figure)

Table 1 The parameters of each CNN layer

Layer	Param	Output shape (batch, channel, signal)
Input		$38 \times 123 \times 90$
Depthwise conv	Filter = 123, ksize = 7, stride = 2	$38 \times 123 \times 44$
Generic conv	Filter = 48, ksize = 5, stride = 2	$38 \times 48 \times 21$
Generic conv	Filter = 8, ksize = 3, stride = 2	$38 \times 8 \times 10$
Fully connected	Weight = $38 \times 80 \times 32$	$38 \times 32 \times 1$
Fully connected	Weight = $38 \times 32 \times 16$	$38 \times 16 \times 1$
Fully connected	Weight = $38 \times 16 \times 2$	$38 \times 2 \times 1$

operation is followed, which captures the featurewise dependencies and outputs per-feature modulation weights by two fully connected layers. In particular, the first fully connected layer is followed by a ReLU activation function to perform dimensionality reduction. The reduction factor needs to be carefully adjusted within the training process. The second fully connected layer is followed by a smoothing sigmoid activation function to increase the dimensionality. The output of the SE operation can be formulated as:

$$e = \sigma\left(W_2 \delta\left(W_1(S)\right)\right) \tag{12}$$

where W_1 and W_2 refer to the two fully connected layers, σ is the sigmoid, and δ represents the ReLU activation function.

We implement our model with the PyTorch framework. The binary cross-entropy was chosen as a loss function since our training task can be regarded as a binary task. Its mathematical representation is as follows.

$$L = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \bullet \log\left(p_i\right) + \left(1 - y_i\right) \bullet \log\left(1 - p_i\right) \right] \tag{13}$$

where y_i refers to the class label, p_i denotes the probability that the class label is consistent with the class label predicted by the model, and N represents the total number of samples. Additionally, ADAM was introduced as the optimizer with a learning rate of $1e-2$ to speed up the correction of the loss function. The batch size was experimentally selected as 38, and the model was trained for 150 epochs. During the training, the dataset was split into 80% for training and the remaining 20% for validation. We adopted a fivefold cross-validation and took the average accuracy as our result.

After approximately 50 iterations, the accuracy is close to 100% with the loss down to zero. Because of available limited labelled training samples, the neural network may suffer from overfitting. However, our primary goal is to select discriminative features for a given task, regardless of its generalization ability on other tasks. In contrast, a degree of overfitting will help the prominent features to be more obvious [53].

3.3.2 Correlation-based feature selection method

The correlation-based feature selection (CFS) method is a supervised method of feature selection, through which each feature obtains a correlation coefficient, determining how highly the feature correlates with the target class [54, 55]. We calculated the correlation coefficient between the features and the label of each feature to express each feature's correlation with cognitive load. A large correlation coefficient value indicates a strong correlation between the feature and the target class, which can be used as the basis for selecting features. Features most relevant to cognition load can be identified by ranking their coefficients in descending order. The correlation coefficient between each feature and cognitive load is calculated as follows:

$$R_{r_i y} = \frac{C_{r_i y}}{\sqrt{C_{r_i} C_y}} \qquad (14)$$

In this index, C stands for "covariance" and R stands for "correlation coefficient."

3.4 Classification

To demonstrate the effectiveness of the AFR method, a uniform classifier needed to be predefined to compare the classification performance of the distinguishing features selected by the two methods. According to previous research [56], SVMs usually result in better generalization than softmax due to their maximum margin property [57]. In addition, the radial basis function (RBF) can map raw data to infinite-dimensional space, which usually leads to better performance and is widely used in practical applications [58]. Furthermore, the original SVM implementation solves binary class/nonclass separation problems, which is in line with our goal [56]. Thus, RBF-SVM was adopted as a reference to evaluate the performance of the introduced method in our research.

The SVM approach was first introduced by Vapnik as a potential alternative to conventional artificial neural networks [59–61]. When SVM is used, a set of training samples needs to be given, where each sample contains several features and is accompanied by a property label clarifying to which of the two categories it belongs. SVM is realized in the following way: a dataset with labels is given, with the task of finding a linear plane $w^T x_i + b$ meeting the following condition:

$$y_i\left(w^T x_i + b\right) \geq 1, i = 1, 2, \dots, N \qquad (15)$$

where x_i denotes the feature vector and y_i stands for its corresponding label.

In addition, the RBF-SVM model was trained using fivefold cross-validation on the datasets in this study.

Table 2 Means and standard deviations of all variables by conditions

	NED group		PED group	
	M	SD	M	SD
Delta power	35.26	18.9	38.00	15.85
Theta power	10.70	5.08	16.12	5.83
Alpha power	11.72	4.36	16.37	8.60
Beta power	38.11	18.92	26.46	13.02
Gamma power	4.58	2.39	3.11	1.90
Sample entropy	0.66	0.044	0.60	0.43
Approximate entropy	1.44	0.10	1.19	0.093
Wavelet entropy	2.14	0.018	2.13	0.019
Left pupil diameter	3.71	0.42	3.29	0.35
Right pupil diameter	3.72	0.46	3.37	0.31
Fixation duration	314.00	137.50	218.12	103.23
Perceived task difficulty	5.71	1.76	4.28	1.23
Learning performance	3.10	1.73	5.09	2.07

4 Results

4.1 Statistical analysis of physiological features

To verify the effectiveness of applying physiological measures to distinguish two cognitive states, one-way analysis of variance (ANOVA) and Mann–Whitney U tests were conducted to make comparisons between the NED group ($n = 19$) and the PED group ($n = 19$). The descriptive statistics of all variables in each group are presented in Table 2.

Figure 6 shows the weight distribution of five frequency band powers in different brain regions. One-way ANOVA revealed that there was a main effect with the dependent variable theta band ($F(1, 37) = 9.36, p = 0.004, \eta_p^2 = .206$) and alpha band power across the two groups ($F(1, 37) = 4.43, p = .042, \eta_p^2 = .110$). Compared to the NED group, subjects had increased theta

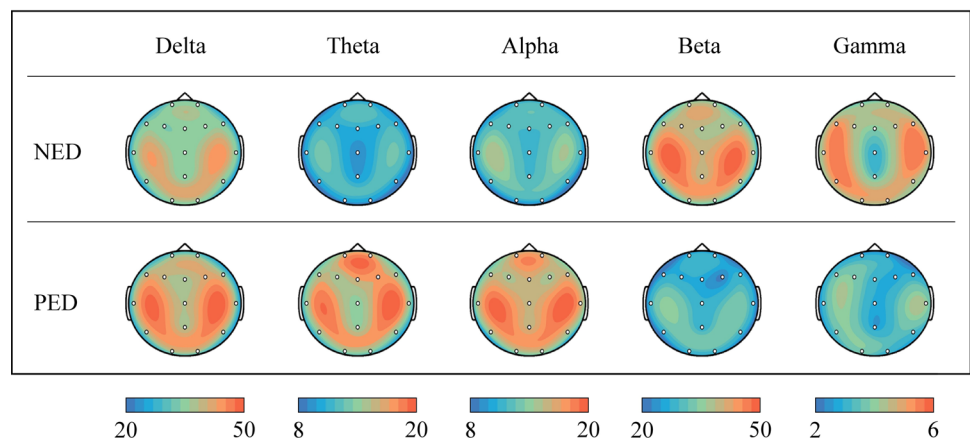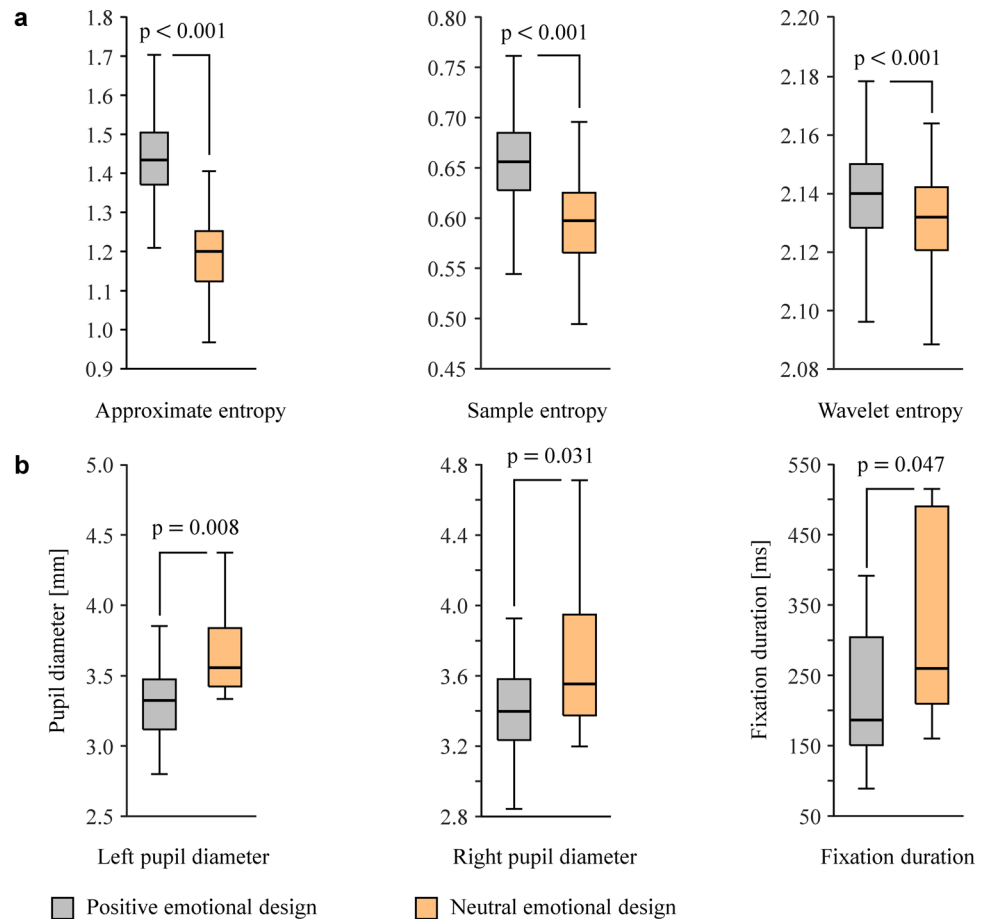**Fig. 6** The weight distribution of five frequency band powers in different brain regions

Fig. 7 Boxplot data for participants across the two groups: **a** spectral entropy (approximate entropy, sample entropy, and wavelet entropy). **b** Eye-tracking (left and right pupil diameters, and fixation duration)

and alpha activity while learning the PED online courses, indicating that the participants in the PED group had a lower cognitive load level than those in the NED group [62].

Boxplots of approximate, sample, and wavelet entropy in the two experimental groups are shown in Fig. 7a. One-way repeated-measures ANOVA revealed a significant main effect across two groups on approximate entropy $(F(1, 37) = 624.03, p < .001, \eta_p^2 = .617)$, sample entropy $(F(1, 37) = 191.58, p < .001, \eta_p^2 = .331)$, and wavelet entropy $(F(1, 37) = 27.31, p < .001, \eta_p^2 = .069)$. For subjects who watched the online course adopting a positive emotional design, the mean spectral entropy (i.e., approximate, sample, and wavelet entropy) was higher than those who watched the online course with a neural emotion design, suggesting that the PED online course can reduce the cognitive load on learners [63].

Figure 7b shows the boxplot of eye-tracking features across the two conditions. The one-way ANOVA revealed a main effect of the two online courses on the left pupil diameter $(F(1, 37) = 8.28, p = .008, \eta_p^2 = .24)$, right pupil

diameter $(F(1, 37) = 5.19, p = .031, \eta_p^2 = .13)$ and fixation duration $(F(1, 37) = 4.35, p = .047, \eta_p^2 = .143)$. Subjects in the PED group had smaller pupil diameter and shorter fixation duration than those in the NED group. These results reveal that learning material employing neutral emotional design had a higher cognitive load assumption when compared to PED online courses [20, 64].

In addition, subjective ratings of cognitive load and objective scores of the learning performance test collected after the experiment were also analysed to confirm the results drawn from the physiological measures. For the ANOVA with the dependent variable subjective ratings, learners who watched NED online courses perceived higher difficulty than those who viewed PED online courses, $F(1, 38) = 7.83, p = .009, \eta_p^2 = .19$. There was also a significant learning performance effect, $F(1, 38) = 5.65, p = .028, \eta_p^2 = .24$, indicating that the PED group can better understand and retain the information presented in the courses. These results support our hypothesis and are consistent with the conclusions of Mayer and Estrella [32] and Um et al. [33].

Table 3 Means and standard deviations of classification accuracies achieved through the AFR and CFS methods

Feature dimension	AFR		CFS	
	M	SD	M	SD
10	81.06	0.35	78.72	0.12
20	90.36	0.15	88.4	0.29
30	92.92	0.23	91.92	0.40
40	95.32	0.18	93.32	0.15
50	95.2	0.20	93.02	0.19
60	95.56	0.16	92.8	0.11
70	95.12	0.29	92.48	0.16
80	94.76	0.40	92.7	0.18
90	94.1	0.37	92.82	0.12
100	94.58	0.29	92.7	0.17
110	94	0.21	93.24	0.21
120	93.34	0.21	93.28	0.29
All	93.44	0.21	93.4	0.19

4.2 Recognition performance comparison

To demonstrate the advantage of the introduced AFR method, we evaluated the performance of the distinguishing features selected by the AFR method and used the CFS method to make an overall comparison. In the experiments, classification accuracy (ACC) was adopted as the evaluation criterion, and all these methods were retested on the datasets obtained from 38 valid participants in this study.

We first calculated the average weight distribution between features and the class labels for all EEG and eye-tracking features and sorted them in descending order. Then, a useful tensor was declared, and the features corresponding to the top 10 weight distributions were initialized as the initial useful tensor. Every time an extra 10 features were added to the useful tensor, an RBF-SVM was employed. The changes in classification accuracy are depicted in Table 3.

In Table 3, we can learn that the AFR method can achieve relatively high average accuracies compared to the CFS method. For the top 110 dimensions, the results achieved by the AFR method are 2.34%, 1.96%, 1.00%, 2.00%, 2.18%, 2.76%, 2.64%, 2.06%, 1.28%, 1.88%, and 0.76% higher than those of the CFS method in average accuracy. Specifically, when the number of dimensions reaches 60, our AFR method reaches the highest average accuracy of 95.56%. Moreover, when the classification accuracy reaches the maximum value, the feature dimension decreases to 60, which is much smaller than the original full dimension. However, all features need to be added to the useful tensor for the CFS method to achieve the highest average accuracy (93.5% with a full feature dimension of 123).
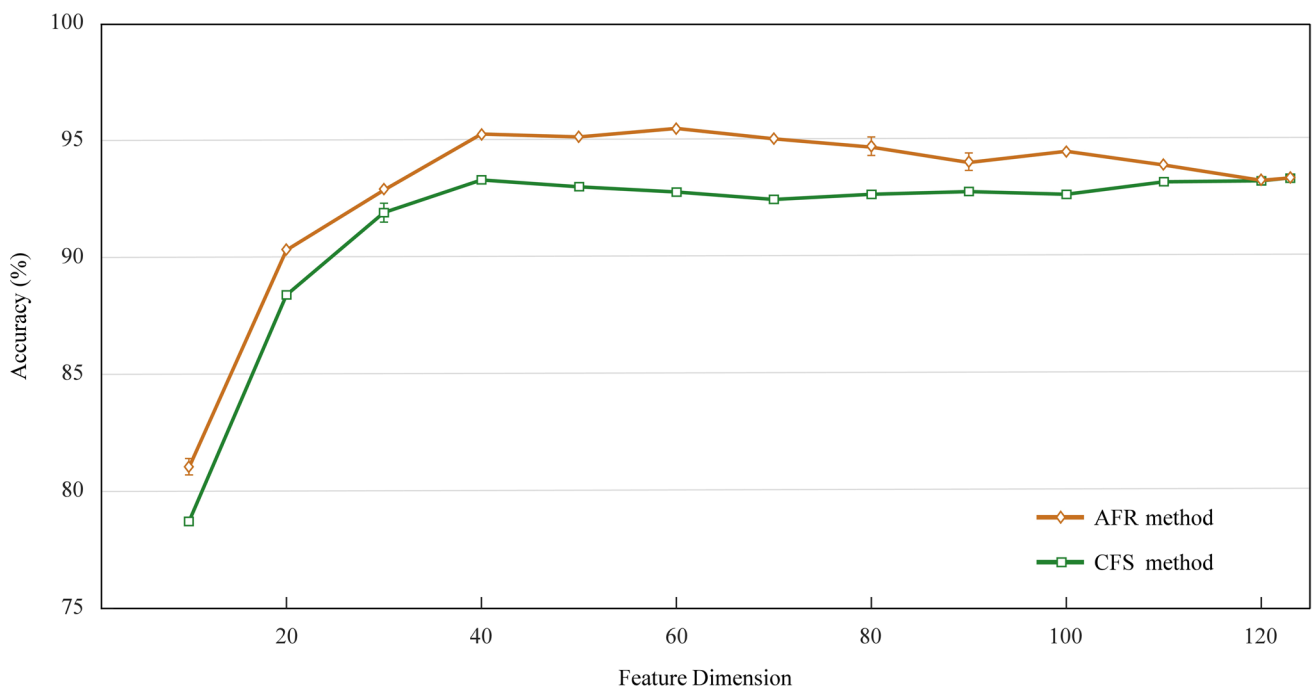

Fig. 8 The classification performance of the selected feature combination using the AFR and CFS methods (a standard error of less than 0.3 was omitted in the figure)
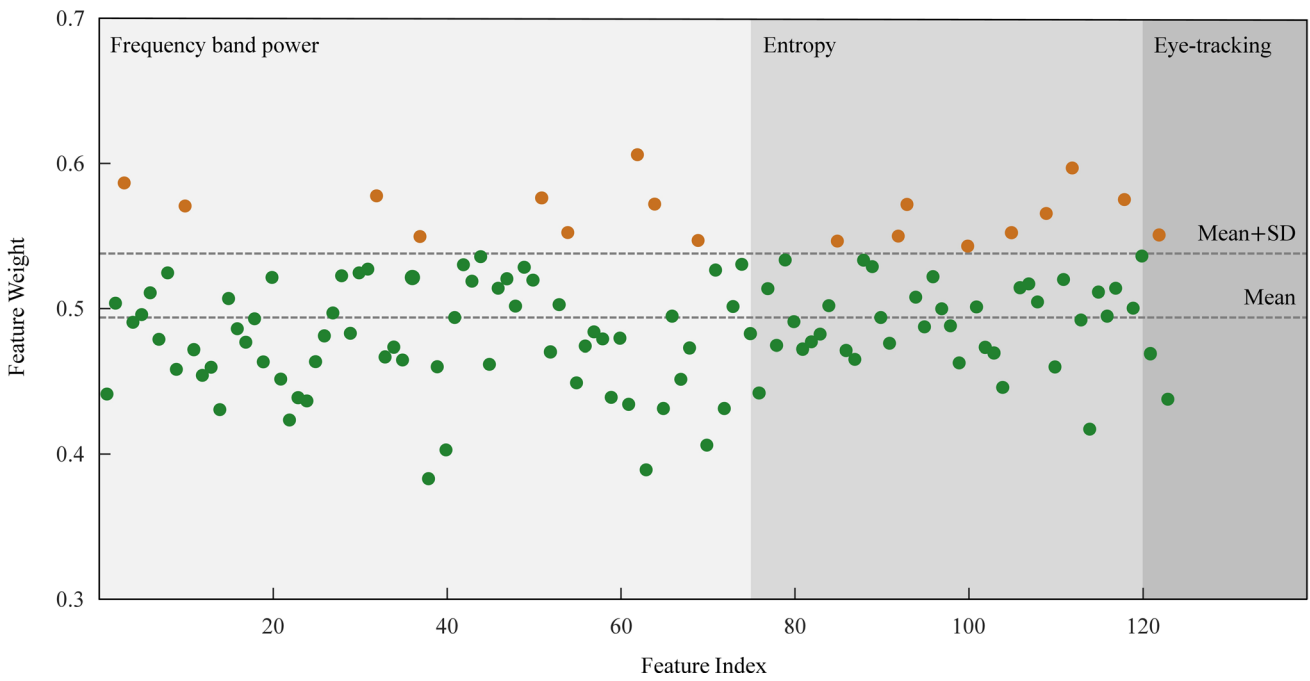
Fig. 9 The feature weights obtained by the AFR method ("mean weight" and "mean + standard deviation weight" are marked in the figure)

5 Discussion

In this study, we combined EEG and eye-tracking measures to evaluate cognitive load in distance learning. As expected, the statistical analysis results in "Sect. 4.1" indicate the feasibility and validity of using physiological measures to monitor cognitive load in remote learning. The learners with lower cognitive load had higher theta and alpha band power,
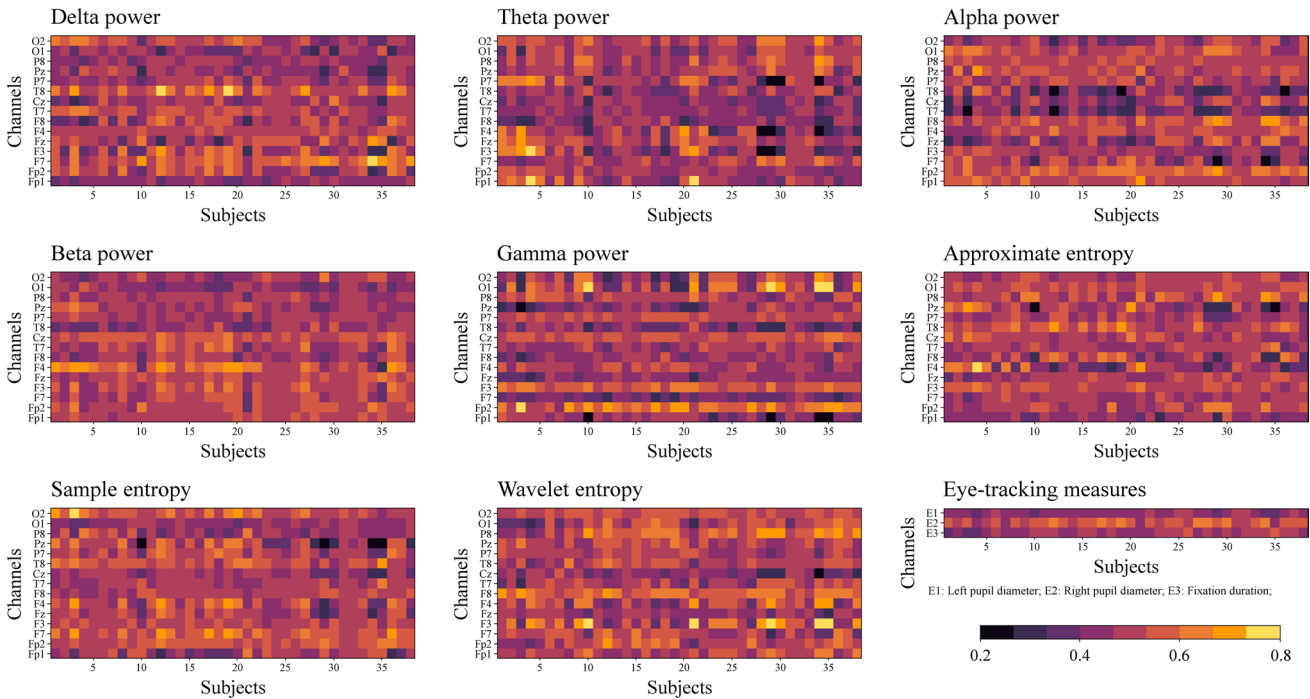

Fig. 10 The average feature weight on each subject of different kinds of features

higher approximate, sample, and wavelet entropy, smaller pupil diameter, and shorter fixation when watching online courses. The results in "Sect. 4.2" demonstrate that our AFR method is better in maintaining a good compromise between performance and computational cost compared to the CFS method, as well as a 2.06% improvement in accuracy and a 51.21% reduction in feature dimension. Therefore, we demonstrated that the application of physiological measures and the AFR method is beneficial to enhancing the feasibility of cognitive load recognition in distance learning.

There are three possible explanations for the excellent performance of the AFR method. First, this might be because the AFR method is highly sensitive to irrelative noise, which is consistent with previous findings [65]. Figure 8 shows the visualization result of the classification performance with discriminative features obtained by the AFR and CFS methods. As shown in Table 3, for the AFR method, the classification accuracy achieved by 60 dimensions (95.56%) is better than that achieved by full feature dimensions (93.44%), which demonstrates the ability of the AFR method to effectively remove interference information. In addition, the accuracy of the AFR method presents an overall downwards trend after the highest point in dimension 60 (see Fig. 8), indicating that fusing irrelevant interfering features reduces the recognition accuracy. In contrast, for the CFS method, all features need to be added to the useful tensor to achieve the highest average accuracy. As shown in Fig. 8, the CFS method accuracy shows a fluctuating upwards trend after 40 dimensions, indicating the poor ability of the CFS method to identify redundant or irrelevant information.

Second, we plotted the average weights of all subjects for each feature and highlighted features with average weight values larger than "mean + standard deviation" (see Fig. 9). It can be seen that the highlighted physiological features were not limited to a particular type of feature, which were in turn scattered in all types of features, indicating that the AFR method can help us combine the robust advantages of multiple and multimodal physiological features to significantly enhance the cognitive load recognition performance. These findings were consistent with previous studies showing that combining multimodal physiological features could enhance recognition [66].

The last possible explanation might lie in the strength of the multiresolution convolutional neural network, which makes the AFR method not only powerful but also superior to statistical and machine learning algorithms [65]. AFR explicitly models the interdependencies between the features and efficiently captures features that are most salient for a given task through a residual squeeze-and-excitation block, which helps the lower layers of the network to exploit more contextual information outside its local receptive field [67].

The application of physiological measures and the AFR method opens the way for cognitive load recognition during the online teaching of engineering courses and the results will hopefully serve as helpful feedback for teachers to adjust teaching strategies immediately and reasonably. Additionally, this technology also provides evidence for evaluating online courses. In this case, the learners who received the PED online course had higher theta and alpha band power, higher approximate, sample, and wavelet entropy, smaller pupil diameter, and shorter fixation duration, indicating that online courses employing positive emotional design consume less cognitive load than those using neural emotional design, thereby leading to better learning performance [62].

However, one limitation of this study is that we ignored individual-specific bias. The weight distributions of individual features for each subject are visualized in Fig. 10. Although the results for most subjects are generally consistent with Fig. 9, the discriminative features showed differences among different subjects. This is an important issue for future research since developing a suitable algorithmic model to address individual-specific bias may further enhance the performance of cognitive load recognition.

The future development of this research will focus on combining the channel attention mechanism with the spatial attention mechanism. The channel attention mechanism diverts attention to the most critical features, while the spatial attention mechanism diverts attention to the most relevant parts. The combined use of the two mechanisms may further enhance feature learning, leading to higher classification performance and lower total computational cost by emphasizing important information while suppressing noise.

6 Conclusion

The major takeaway of this study was to introduce physiological measures and the AFR method for higher feasibility in cognitive load recognition. We utilized multimodal physiological signals as tools to monitor the learners' cognitive loads and adopted the AFR algorithm to further enhance the feasibility of cognitive load recognition in remote learning. The results demonstrated that physiological measures can significantly distinguish different cognitive load levels. Additionally, the introduced AFR algorithm can be effectively used to capture discriminative features, thus achieving good performance in terms of accuracy and computational cost. Continuously monitoring learners' cognitive states provides not only instantaneous recognition of cognitive overload or underload but also valuable feedback to distance-learning system designers and administrators so

they may take appropriate countermeasures to improve the overall learning effect.

References

1. Paas F, Renkl A, Sweller J (2004) Cognitive load theory: instructional implications of the interaction between information structures and cognitive architecture. Instr Sci 32:1–8
2. Sweller J (2010) Element interactivity and intrinsic, extraneous, and germane cognitive load. Educ Psychol Rev 22:123–138. https://doi.org/10.1007/s10648-010-9128-5
3. Sweller J, van Merrienboer JJG, Paas FGWC (1998) Cognitive architecture and instructional design. Educ Psychol Rev 10:251–296. https://doi.org/10.1023/A:1022193728205
4. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM (2003) Cognitive load measurement as a means to advance cognitive load theory. Educ Psychol 38:63–71. https://doi.org/10.1207/S15326985EP3801_8
5. Baddeley A (1992) Working memory: the interface between memory and cognition. J Cogn Neurosci 4:281–288. https://doi.org/10.1162/jocn.1992.4.3.281
6. Frey J, Mühl C, Lotte F, Hatchet M (2013) Review of the use of electroencephalography as an evaluation method for human-computer interaction[J]. arXiv preprint arXiv:1311.2222. https://doi.org/10.48550/arXiv.1311.2222
7. Tran TQ, Boring RL, Dudenhoeffer DD, Hallbert BP, Keller MD, Anderson TM (2007) Advantages and disadvantages of physiological assessment for next generation control room design. In: 2007 IEEE 8th human factors and power plants and HPRCT 13th Annual Meeting. pp. 259–263. https://doi.org/10.1109/HFPP.2007.4413216
8. Gevins A, Smith ME (2000) Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. Cereb Cortex 10:829–839. https://doi.org/10.1093/cercor/10.9.829
9. Murata A (2005) An attempt to evaluate mental workload using wavelet transform of EEG. Hum Factors 47:498–508. https://doi.org/10.1518/001872005774860096
10. Başar E, Başar-Eroglu C, Karakaş S, Schürmann M (2001) Gamma, alpha, delta, and theta oscillations govern cognitive processes. Int J Psychophysiol 39:241–248. https://doi.org/10.1016/S0167-8760(00)00145-8
11. Gevins A (1997) High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. Cereb Cortex 7:374–385. https://doi.org/10.1093/cercor/7.4.374
12. Herweg NA, Solomon EA, Kahana MJ (2020) Theta oscillations in human memory. Trends Cogn Sci 24:208–227. https://doi.org/10.1016/j.tics.2019.12.006
13. Sauseng P, Griesmayr B, Freunberger R, Klimesch W (2010) Control mechanisms in working memory: a possible function of EEG theta oscillations. Neurosci Biobehav Rev 34:1015–1022. https://doi.org/10.1016/j.neubiorev.2009.12.006
14. Schacter DL (1977) EEG theta waves and psychological phenomena: a review and analysis. Biol Psychol 5:47–82. https://doi.org/10.1016/0301-0511(77)90028-X
15. Sterman MB, Mann CA, Kaiser DA, Suyenobu BY (1994) Multiband topographic EEG analysis of a simulated visuomotor aviation task. Int J Psychophysiol 16:49–56. https://doi.org/10.1016/0167-8760(94)90041-8
16. Chen W, Zhuang J, Yu W, Wang Z (2009) Measuring complexity using FuzzyEn, ApEn, and SampEn. Med Eng Phys 31:61–68. https://doi.org/10.1016/j.medengphy.2008.04.005
17. Gribkov D, Gribkova V (2000) Learning dynamics from nonstationary time series: analysis of electroencephalograms. Phys Rev E 61:6538–6545. https://doi.org/10.1103/PhysRevE.61.6538
18. Pincus SM (1991) Approximate entropy as a measure of system complexity. Proc Nat Acad Sci 88(6):2297–2301. https://doi.org/10.1073/pnas.88.6.2297
19. Skaramagkas V, Giannakakis G, Ktistakis E, Manousos D, Karatzanis I, Tachos N, Tripoliti EE, Marias K, Fotiadis DI, Tsiknakis M (2021) Review of eye tracking metrics involved in emotional and cognitive processes. IEEE Rev Biomed Eng 1–1. https://doi.org/10.1109/RBME.2021.3066072
20. Kahneman D, Beatty J (1966) Pupil diameter and load on memory. Science. https://doi.org/10.1126/science.154.3756.1583
21. van der Wel P, van Steenbergen H (2018) Pupil dilation as an index of effort in cognitive control tasks: a review. Psychon Bull Rev 25:2005–2015. https://doi.org/10.3758/s13423-018-1432-y
22. Borys, M., Tokovarov, M., Wawrzyk, M., Wesolowska, K., Plechawska-Wojcik, M., Dmytruk, R., Kaczorowska, M., 2017. An analysis of eye-tracking and electroencephalography data for cognitive load measurement during arithmetic tasks, in: 2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE). IEEE, Bucharest, Romania, pp. 287–292. https://doi.org/10.1109/ATEE.2017.7905130
23. Scharinger C, Schüler A, Gerjets P (2020) Using eye-tracking and EEG to study the mental processing demands during learning of text-picture combinations. Int J Psychophysiol 158:201–214. https://doi.org/10.1016/j.ijpsycho.2020.09.014
24. Pandey P, Miyapuram KP (2021) Brain2depth: lightweight CNN model for classification of cognitive states from EEG recordings. In: Annual conference on medical image understanding and analysis. Springer, Cham, pp 394–407. https://doi.org/10.1007/978-3-030-80432-9_30
25. Michalak K, Kwaśnicka H (2006) Correlation-based feature selection strategy in classification problems. Int J Appl Math Comput Sci 16:503–511
26. Wang X-W, Nie D, Lu B-L (2014) Emotional state classification from EEG data using machine learning approach. Neurocomputing 129:94–106. https://doi.org/10.1016/j.neucom.2013.06.046
27. Nie D, Wang X-W, Shi L-C, Lu B-L (2011) EEG-based emotion recognition during watching movies, in: 2011 5th International IEEE/EMBS Conference on Neural Engineering. IEEE, Cancun, pp 667–670. https://doi.org/10.1109/NER.2011.5910636
28. Guo M-H, Xu T-X, Liu J-J, Liu Z-N, Jiang P-T, Mu T-J, Zhang S-H, Martin RR, Cheng M-M, Hu S-M (2022) Attention mechanisms in computer vision: a survey. Comput Vis Media 8:331–368. https://doi.org/10.1007/s41095-022-0271-y
29. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Salt Lake City, UT, pp 7132–7141
30. Mursalin M, Zhang Y, Chen Y, Chawla NV (2017) Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing 241:204–214. https://doi.org/10.1016/j.neucom.2017.02.053
31. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC (1973) Quantification of sleepiness: a new approach. Psychophysiology 10:431–436. https://doi.org/10.1111/j.1469-8986.1973.tb00801.x

32. Mayer RE, Estrella G (2014) Benefits of emotional design in multimedia instruction. Learn Instr 33:12–18. https://doi.org/10.1016/j.learninstruc.2014.02.004

33. Um E "Rachel," Plass JL, Hayward EO, Homer BD (2012) Emotional design in multimedia learning. J. Educ. Psychol. 104:485–498. https://doi.org/10.1037/a0026609

34. Kalyuga S, Chandler P, Sweller J (2000) Incorporating learner experience into the design of multimedia instruction. J Educ Psychol 92:126

35. Kalyuga S, Sweller J (2005) Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. Educ Technol Res Dev 53:83–93

36. Tuovinen JE, Paas F (2004) Exploring multidimensional approaches to the efficiency of instructional conditions. Instr Sci 32:133–152

37. Herwig U, Satrapi P, Schönfeldt-Lecuona C (2003) Using the international 10–20 EEG system for positioning of transcranial magnetic stimulation. Brain Topogr 16:95–99. https://doi.org/10.1023/B:BRAT.0000006333.93597.9d

38. Duncan CC, Barry RJ, Connolly JF, Fischer C, Michie PT, Näätänen R, Polich J, Reinvang I, Van Petten C (2009) Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clin Neurophysiol 120:1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045

39. Gratton G, Coles MGH, Donchin E (1983) A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 55:468–484. https://doi.org/10.1016/0013-4694(83)90135-9

40. Pesonen M, Hämäläinen H, Krause CM (2007) Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. Brain Res 1138:171–177. https://doi.org/10.1016/j.brainres.2006.12.076

41. Aftanas LI, Lotova NV, Koshkarov VI, Pokrovskaja VL, Popov SA, Makhnev VP (1997) Non-linear analysis of emotion EEG: calculation of Kolmogorov entropy and the principal Lyapunov exponent. Neurosci Lett 226:13–16. https://doi.org/10.1016/S0304-3940(97)00232-2

42. Zhang C, Wang H, Fu R (2014) Automated detection of driver fatigue based on entropy and complexity measures. IEEE Trans Intell Transp Syst 15:168–177. https://doi.org/10.1109/TITS.2013.2275192

43. Al-Angari HM, Sahakian AV (2007) Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome. IEEE Trans Biomed Eng 54:1900–1904. https://doi.org/10.1109/TBME.2006.889772

44. Mazher M, Abd Aziz A, Malik AS, Ullah Amin H (2017) An EEG-based cognitive load assessment in multimedia learning using feature extraction and partial directed coherence. IEEE Access 5:14819–14829. https://doi.org/10.1109/ACCESS.2017.2731784

45. Richman JS, Lake DE, Moorman JR (2004) Sample entropy. In: Methods in enzymology, vol 384. Academic Press, pp 172–184. https://doi.org/10.1016/S0076-6879(04)84011-4

46. Yentes JM, Hunt N, Schmid KK, Kaipust JP, McGrath D, Stergiou N (2013) The appropriate use of approximate entropy and sample entropy with short data sets. Ann Biomed Eng 41:349–365. https://doi.org/10.1007/s10439-012-0668-3

47. Kukker A, Sharma R (2019) EEG based epilepsy seizure analysis and classification methods: an overview. Int J Comput Sci Eng 7(8):328–346. https://doi.org/10.26438/ijcse/v7i8.328346

48. Ocak H (2009) Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. Expert Syst Appl 36:2027–2036. https://doi.org/10.1016/j.eswa.2007.12.065

49. Subasi A (2007) EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst Appl 32:1084–1093. https://doi.org/10.1016/j.eswa.2006.02.005

50. Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. IEEE Trans Inf Theory 36:961–1005. https://doi.org/10.1109/18.57199

51. Zhang J, Yin Z, Chen P, Nichele S (2020) Emotion recognition using multi-modal data and machine learning techniques: a tutorial and review. Inf Fusion 59:103–126. https://doi.org/10.1016/j.inffus.2020.01.011

52. Kira K, Rendell LA (1992) A practical approach to feature selection. In: Sleeman D, Edwards P (eds) Machine learning proceedings 1992. Morgan Kaufmann, San Francisco, pp 249–256

53. Hu J, Shen L, Albanie S, Sun G, Wu E (2020) Squeeze-and-Excitation Networks. IEEE Trans Pattern Anal Mach Intell 42(8):2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

54. Hall MA (2000) Correlation-based feature selection of discrete and numeric class machine learning. (Working paper 00/08). University of Waikato, Department of Computer Science, Hamilton, New Zealand

55. Hall MA (1999) Correlation-based feature selection for machine learning. The University of Waikato

56. Noble WS (2006) What is a support vector machine? Nat Biotechnol 24:1565–1567. https://doi.org/10.1038/nbt1206-1565

57. Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92. Association for Computing Machinery, New York, NY, USA, pp. 144–152. https://doi.org/10.1145/130385.130401

58. Zhao Z-D, Lou Y-Y, Ni J-H, Zhang J (2009) RBF-SVM and its application on reliability evaluation of electric power system communication network. In: 2009 International Conference on Machine Learning and Cybernetics. pp 1188–1193

59. Vapnik V (1999) The nature of statistical learning theory. Springer Science & Business Media

60. Vapnik V (1998) The support vector method of function estimation. In: Suykens JAK, Vandewalle J (eds) Nonlinear modeling: advanced black-box techniques. Springer, Boston, pp 55–85

61. Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10:988–999. https://doi.org/10.1109/72.788640

62. Zammouri A, Ait Moussa A, Mebrouk Y (2018) Brain-computer interface for workload estimation: assessment of mental efforts in learning processes. Expert Syst Appl 112:138–147. https://doi.org/10.1016/j.eswa.2018.06.027

63. Plass JL, Heidig S, Hayward EO, Homer BD, Um E (2014) Emotional design in multimedia learning: effects of shape and color on affect and learning. Learn Instr 29:128–140. https://doi.org/10.1016/j.learninstruc.2013.02.006

64. Kruger JL, Doherty S (2016) Measuring cognitive load in the presence of educational video: towards a multimodal methodology. Aust J Educ Technol 32(6). https://doi.org/10.14742/ajet.3084

65. Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PT, Mu TJ, Zhang SH, Martin RR, Cheng MM, Hu SM (2021) Attention mechanisms in computer vision: a survey. Comput Vis Media 8(3):331–368. https://doi.org/10.1007/s41095-022-0271-y

66. Lu Y, Zheng WL, Li B, Lu BL (2015) Combining eye movements and EEG to enhance emotion recognition. In: Twenty-Fourth International Joint Conference on Artificial Intelligence. pp 1170–1176

67. Eldele E, Chen Z, Liu C, Wu M, Kwoh C-K, Li X, Guan C (2021) An attention-based deep learning approach for sleep stage classification with single-channel EEG. IEEE Trans Neural Syst Rehabil Eng 29:809–818. https://doi.org/10.1109/TNSRE.2021.3076234

Chennan Wu is currently pursuing the bachelor degree in the School of Information and Electronic Engineering, Zhejiang University of Science and Technolog, Hangzhou, China. His research interests include cognitive science, physiological signal processing, and machine learning.

Yang Liu is currently an associate professor of the School of Information and Electronic Engineering, Zhejiang University of Science and Technology. She received her Ph.D. degree from the College of Computer Science and Technology, Zhejiang University, China, in 2012, and her master degree from the Department of Education Technology, Central China Normal University, China, in 2003. Her research interests include e-learning, cognitive science, and artificial intelligence.

Xiang Guo is currently a lecturer of the School of Information and Electronic Engineering, Zhejiang University of Science and Technology. He received his Ph.D. from the college of Information Science & Electronic Engineering, Zhejiang University, China, in 2019. His research interests include signal processing and artificial intelligence.

Tianshui Zhu is currently a BS student of the School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, China. His research interests include emotional regulation process, biomedical signal processing, and deep learning.

Zongliang Bao is currently a senior experimentalist of the School of Information and Electronic Engineering, Zhejiang University of Science and Technology. He received his master degree from the Zhejiang University of Technology, China. His research interests include educational technology and design thinking.