



## Subject-specific mental workload classification using EEG and stochastic configuration network (SCN)



Pang Liping<sup>a</sup>, Guo Liang<sup>a</sup>, Zhang Jie<sup>a</sup>, Wanyan Xiaoru<sup>a,\*</sup>, Qu Hongquan<sup>b,\*</sup>, Wang Xin<sup>c</sup>

<sup>a</sup> School of Aviation Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing, 100191, China

<sup>b</sup> College of Information Engineering, North China University of Technology, Beijing, 100144, China

<sup>c</sup> Marine Human Factors Engineering Lab, China Institute of Marine Technology & Economy, Beijing, 100081, China

### ARTICLE INFO

#### Keywords:

Mental workload  
Subject-specific classifier  
Subject-multiple classifier  
Stochastic configuration network  
Multiple subjects

### ABSTRACT

Mental workload assessment of the operators in some safety-critical human-machine systems is an important research topic. In this paper, an experiment was designed to obtain the electroencephalogram (EEG) data under three levels of mental workload. The EEG data of multiple subjects were used for the mental workload classification based on the stochastic configuration network (SCN). The subject-specific classifiers (SSCs) were built by the individual EEG data. The results showed that the range of SSC test accuracy was between 56.5 % and 90.2 % with an average of 75.9 %. The SSC accuracy had a positive correlation with the operating accuracy ( $r = 0.852$ ,  $p < 0.01$ ). For comparison, the subject-multiple classifiers (SMCs) were established with the EEG data of multiple subjects. The results showed that the SSCs had a lower time-consuming and higher prediction accuracy than the SMCs. But the SMCs might embody the trend of statistical performance for a large number of subjects. This study provided an effective modeling method for the classification of mental workload, and it would bring great convenience to the practical application in the future.

### 1. Introduction

With the development of science and technology, the roles of operators tend to be decision-makers and monitors, which means that the physical labor is decreased while the mental labor is increased in a modern human-machine system. Therefore, modern operators are often at a high mental workload. Studies have shown that a high mental workload will cause rapid fatigue, reduced flexibility, stress reaction, frustration and increased errors. Low mental workload will cause the waste of resources, disgust and reduced job performance [1]. So the measurement, assessment and prediction of the mental workload are important topics in ergonomic. Besides, for some safety-critical human-machine systems with a clear division of labor and a small sample of operators, such as the systems for railway drivers, pilots, and astronauts, the study of subject-specific mental workload classification is still very meaningful [2].

The mental workload was first proposed in 1940 with the goal of “optimizing the human-machine system” [3]. Since then various definitions have been proposed [4–8], and their core is all about the relationship between task demand and personal capacity. Among them, people generally agree with the definition of mental workload by

Eggerneier et al. The mental workload is defined by them as the amount of capacity or resource consumed by the operator to complete tasks [5], which depends on task requirements and resource supply that the operator can or will be willing to allocate.

Subjective measurement, performance measurement and psychophysiological measurement are the main methods of mental workload measurement [9]. Many studies have shown that any single indicator or method has its advantages and limitations in mental workload assessment. Therefore, these three methods are recommended to measure mental workload simultaneously in recent years [9]. Among them, subjective measurement is based on the real feelings of operators. Performance measurement is based on the operation of the main task or sub-task by the operators. Physiological measurement is based on the physiological signals of operators, and the typical physiological indexes include electroencephalogram (EEG), event-related brain potential (ERP), electrooculogram (EOG), and electrocardiogram (ECG), etc [10].

With the development of hardware and software, physiological measurement shows more and more potential due to its advantage of objectivity, real-time and little interference to task [11–13]. Therefore, more and more researchers focus on applying algorithms [14] to explore the correspondence between EEG and mental workload levels, and to

\* Corresponding authors.

E-mail addresses: [wanyanxiaoru@buaa.edu.cn](mailto:wanyanxiaoru@buaa.edu.cn) (X. Wanyan), [qhphd\\_ncut@163.com](mailto:qhphd_ncut@163.com) (H. Qu).

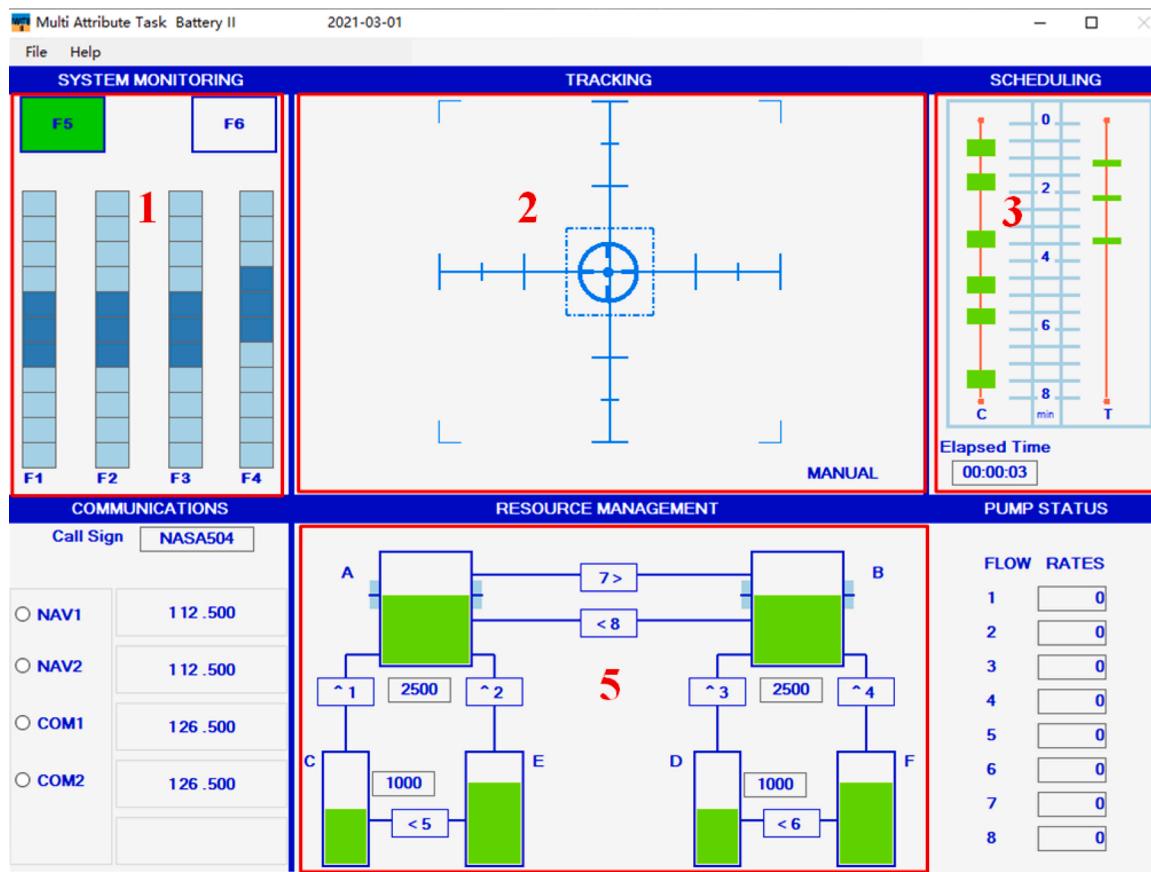


Fig. 1. Interface of MATB II.

establish mental workload classifiers.

Many methods have been proposed to build the classifiers using EEG data, such as Linear Discriminant Analysis (LDA) [15–20], Random Forest (RF) [21,22], Support Vector Machine (SVM) [23–27], and Neural Network (NN) [28–34]. Garrett et al. [35] compared the linear, nonlinear, and feature selection methods for EEG signal classification. The results showed that the nonlinear classification methods were better than the linear ones. In the random forest method, a large amount of storage space and computation time are required for training, and the cumulative errors may occur [36] when the number of decision trees is large. Christensen et al. used EEG data to estimate the operator functional state using LDA, SVM and NN, respectively [37]. The results showed that the classification accuracy of NN is higher than that of SVM and LDA. The NN classifier for mental workload identification has received a lot of attention and achievements in recent years. However, most researchers focus on the NN trained by the back-propagation (BP) algorithm, which is difficult to converge and easy to fall into a local minimum. These deficiencies can be made up by the other type of NN trained by a random algorithm, which randomly generates parameters between the input layer and hidden layer [38]. However, the difficulty in determining hyper-parameters setting is still challenging regardless of NN with BP algorithm or random algorithm. To solve this problem, Wang D et al. proposed the stochastic configuration network (SCN). It inherits the advantage of the random neural network and overcomes the difficulties in the hyper-parameters setting by gradually increasing the hidden nodes under inequality constraints [39–41].

In the existing researches, there were a few studies on the classification of mental workload using EEG, but their accuracy was not high enough. Therefore, the EEG data and the SCN were used in this study to establish the classification model of mental workload. In this study, three mental workload tests were carried out on the MATB II platform

**Table 1**  
Details of four tasks.

Title	Description	Activation frequency during fixed test time		
		LMW	MMW	HMW
System Monitoring	Monitor the scales of F1-F4 in Area 1 and respond with a mouse when the scales are not around the center.	1	12	24
Tracking	In Area 2, keep the target at the grid center by joystick in MANUAL mode and no action is required in AUTO mode.	1	12	24
Scheduling	Monitor scheduling bar in Area 3 and respond to the activated communication with keyboard immediately.	1	12	24
Resource Management	Monitor oil volume in tanks and pump status in Area 4. Click the corresponding oil pump with a mouse when a failure occurs.	1	12	24

for 16 subjects. The NASA-TLX data, performance data and EEG signal were collected and analyzed. The NASA-TLX data could verify the rationality of the experimental design. The EEG data were used to establish the SSCs and SMCs for mental workload based on the SCN. The relationship between the operating performance and the SSC accuracy, and the comparison of the performances between the SSC and the SMC were further discussed based on the classification results. Besides, a comparison between the traditional Artificial Neural Network (ANN) and the SCN was performed to further reveal the superiority of the SCN.



**Fig. 2.** Experimental scene.

## 2. Experiment and data acquisition

### 2.1. Experimental tasks

The MATB II developed by the National Aeronautics and Space Administration (NASA) [42] was used as a platform to perform four types of flight tasks, namely the system monitoring, tracking, scheduling and resource management tasks, respectively [43]. The interface of MATB II is shown in Fig. 1. The above four tasks are located in the areas marked with the numbers 1, 2, 3 and 4, respectively. Details of the four tasks are described in Table 1.

Three mental workload levels, namely low mental workload (LMW), medium mental workload (MMW) and high mental workload (HMW), were defined in this study. The numbers of activated certain tasks during the fixed test time, 16 min, were different for three mental workload tests, as shown in Table 1. Different mental workload levels of the subjects were induced by setting the activation frequency of each task during the fixed test time, which is one of the main methods to study mental workload using MATB II. The activation frequency of 1, 12, and 24 were set to induce LMW, MMW and HMW, separately. These frequencies had been verified in pre-experiment and successfully induced three different levels of mental workload. The operating performance of every subject, the operating accuracy and the reaction time were automatically recorded by the MATB II.

### 2.2. Subjects

Sixteen subjects ( $23.4 \pm 0.8$  years, 15 males and 1 female), labeled as Sub 1 - Sub 16, were selected to participate in the experiment from on-campus postgraduate students with engineering knowledge backgrounds in Beihang University. All volunteers signed the informed consent form. They were right-handed and in good health, with no sleep disorders and normal vision or corrected-to-normal vision. They were required to ensure adequate sleep and a good mental state before the experiment.

### 2.3. Subjective scale

National Aeronautics and Space Administration-Task Load Index (NASA-TLX) was adopted to subjectively evaluate the mental workload in this study [44]. The first part of NASA-TLX intends to obtain scale scores for six factors, namely mental demand, physical demand, temporal demand, effort, own performance and frustration level. The second part intends to obtain weighted scores by paired comparison of the six factors. The sum of the product of the weighted score and the scale score for each factor was converted to the hundred-mark system to obtain the overall score of mental workload [45].

### 2.4. Experimental procedure

The experiments were conducted in a laboratory as shown in Fig. 2.

**Table 2**  
Experimental procedure.

Experimental step	Experimental operation	Time (min)
1. Experiment training	Read the lab guide and train simulation tasks	60
2. Preparation	Wore EEG cap and injected electrode jelly	30
3. Resting test	Carry out the resting test for three minutes	5
4. Condition 1	Carry out the first mental workload test	25
5. Fill in the form and rest	Score in NASA-TLX and rest	15
6. Condition 2	Carry out the second mental workload test	25
7. Fill in the form and rest	Score in NASA-TLX and rest	15
8. Condition 3	Carry out the third mental workload test	25
9. Fill in the form	Score in NASA-TLX	5
10. Fill in the form	Weight in NASA-TLX	5

There were a subject and two experimenters in a laboratory. Experimenter 1 and 2 were responsible for the experimental process and recording the behavior of subjects, separately. The subject was asked to perform the resting and task tests. All subjects were trained to be familiar with the MATB task before the formal experiments. As shown in Table 2, the resting test was carried out firstly, and then three mental workload tests were performed in the order designed by the Latin square method to avoid the practice effect and fatigue effect. The NASA-TLX scale was filled in immediately after each MATB task. The EEG data of subjects were collected during the resting tests and the task tests. Each subject carried on the experiment between 9:00 am and 12:00 am.

### 2.5. EEG data acquisition

Thirty-two EEG electrodes (FP1, FP2, FZ, F3, F4, F7, F8, FCZ, FC3, FC4, FT7, FT8, CZ, C3, C4, T3, T4, A1, A2, CPZ, CP3, CP4, TP7, TP8, PZ, P3, P4, T5, T6, OZ, O1 and O2) were placed on the scalp of the subject according to 10–20 international electrode system [45], as shown in Fig. 3 (a). EEG data were collected by the Neuroscan Neuamps system (Synamps2, Scan 4.3, El Paso, USA). The additional horizontal electro-oculogram (HEOG) and vertical electro-oculogram (VEOG) channels were also recorded, as shown in Fig. 3 (b). All data were referenced online to A1 and later re-referenced off-line to the average of A1 and A2 [34]. The impedances for channels were below  $5\text{ k}\Omega$ , the recording bandwidth was 0.1–200 Hz, and the sampling rate,  $F_s$ , was 1000 Hz. EEG acquisition system is shown in Fig. 4.

## 3. Experimental data

### 3.1. NASA-TLX data

The following conclusions can be obtained by analyzing the NASA-TLX data.

- (1) The overall scores of mental workload gradually increased with the increase of mental workload levels (LMW:  $39.8 \pm 15.0$ ; MMW:  $51.5 \pm 10.0$ ; HMW:  $63.7 \pm 8.2$ ). A significant main effect of three mental workload levels was observed for the overall scores of mental workload ( $F(2, 30) = 35.587, p < 0.001, \eta^2 = 0.703$ ).
- (2) Post hoc Least Significance Difference (LSD) analysis indicated that the overall scores of LMW were significantly lower than that of MMW ( $p < 0.001$ ) and HMW ( $p < 0.001$ ), and the overall scores of MMW were significantly lower than that of LMW ( $p < 0.001$ ).

The above analysis indicated that the overall scores of mental workload varied significantly across LMW, MMW and HMW. Therefore,

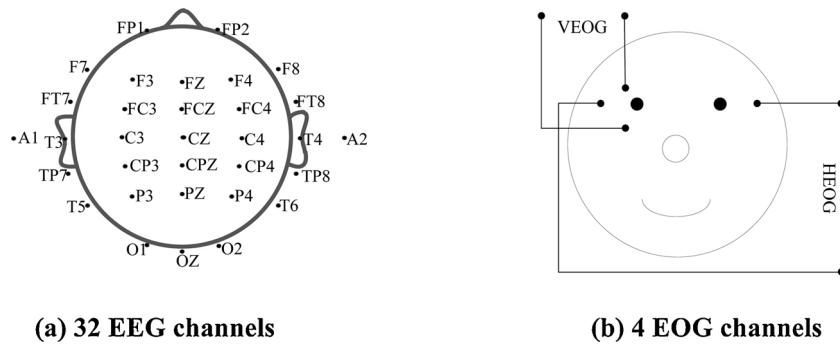


Fig. 3. Channel location.

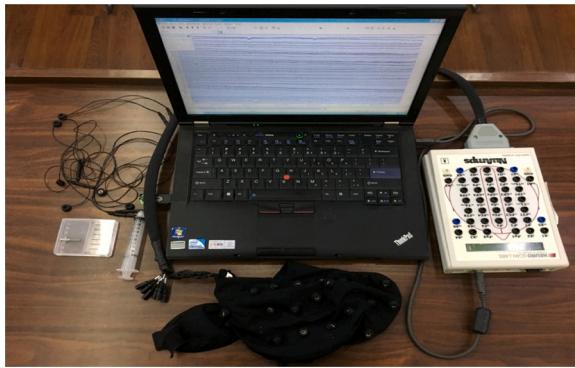


Fig. 4. EEG acquisition system.

the three different mental workload levels could be well induced by the activation frequencies of subtasks designed in this study.

### 3.2. Operating performance

With the increase of the mental workload levels, the average reaction time of all subjects decreased first and then increased (LMW: 3.1 s  $\pm$  0.6 s; MMW: 2.8 s  $\pm$  0.5 s; HMW: 3.1 s  $\pm$  0.5 s), and the average operating accuracy of all subjects increased continuously (LMW: 92.2 %  $\pm$  12.0 %; MMW: 92.5 %  $\pm$  7.3 %; HMW: 92.8 %  $\pm$  4.9 %). Both the reaction time and operating accuracy didn't show any significant differences between the mental workload levels of MATB task (reaction time:  $F(2, 30) = 2.065, p = 0.144, \eta^2P = 0.121$ ; operating accuracy:  $F(2, 30) = 0.055, p = 0.946, \eta^2P = 0.004$ ). This indicated that the elevated mental workload had no obvious effect on the MATB task performance in this study. According to the previous studies [46–48], the reaction time of the MATB task had a U-shaped relationship with the increased mental workload, which indicated that the subjects had higher performance in MMW. The performance first increased to the optimal arousal level and then decreased again due to the increased arousal level. A possible explanation is that the change of task performance is caused by stress with the elevated mental workload [49].

### 3.3. EEG data processing

EEG signal for every mental workload level will be analyzed in the following steps: preprocessing, segmentation and feature extraction, as shown in Fig. 5.

#### (1) Preprocessing

Continuous raw EEG data were cropped firstly to exclude the EEG data outside of the MATB test period. Then EEG data were artifact-corrected, re-referenced and filtered, respectively. It is worth noting

that the temporary artifacts were manually rejected according to the references [50,51], and reference electrodes were reset to A1 and A2. The artifact-corrected EEG data were filtered with a 1 Hz high-pass filter and 30 Hz low-pass filter, respectively.

#### (2) Segmentation

To ensure a sufficient quantity of inputs [34] and the stability [52] of the data for the NN, the middle third of the data ( $t = 300$  s) from each block for each subject was segmented with  $N$  sampling points ( $N = 1024$ ). Adjacent segments overlap by 50 %, that is, 512 sampling points. The sampling rate,  $F_s = 1000$  Hz, so there were  $K$  segments,  $K = 583$ , in each block for each subject. For three blocks, the total number of samples,  $S$ , was 1749.

#### (3) Feature extraction

The power spectral density (PSD) could represent the voltage fluctuations within the neurons in shallow cortical areas [51,53–58]. One-sided PSD of each segment was extracted with Eq. (1) [59]. The energy of  $\alpha$ ,  $\beta$ ,  $\theta$ ,  $\delta$ , namely  $E_\alpha$ ,  $E_\beta$ ,  $E_\theta$  and  $E_\delta$ , was then calculated by the sum of PSD values in the specific frequency band, as shown in Eq. (2) [60].

$$PSD = 2 \times \frac{FFT(signal) \times FFT^*(signal)}{N \times F_s} \quad (1)$$

where  $FFT(signal)$  is the fast Fourier transform (FFT) of the signal;  $FFT^*(signal)$  is the conjugate of  $FFT(signal)$ ;  $N$  is the signal length;  $F_s$  is the sampling rate.

$$\begin{aligned} E_\alpha &= \sum_{freq_\alpha=8}^{14} PSD_\alpha \\ E_\beta &= \sum_{freq_\beta=14}^{30} PSD_\beta \\ E_\theta &= \sum_{freq_\theta=4}^8 PSD_\theta \\ E_\delta &= \sum_{freq_\delta=0.5}^4 PSD_\delta \end{aligned} \quad (2)$$

where  $PSD_\alpha$ ,  $PSD_\beta$ ,  $PSD_\theta$  and  $PSD_\delta$  are the PSD values of  $\alpha$ ,  $\beta$ ,  $\theta$  and  $\delta$ ;  $freq_\alpha$ ,  $freq_\beta$ ,  $freq_\theta$ , and  $freq_\delta$  are the corresponding frequency values.

## 4. SCN classifier and application

### 4.1. SCN principle

The SCN is a stochastic configuration network with an input layer, a hidden layer and an output layer, as shown in Fig. 6. Two strategies are adopted to respond to the difficulties in determining the number of

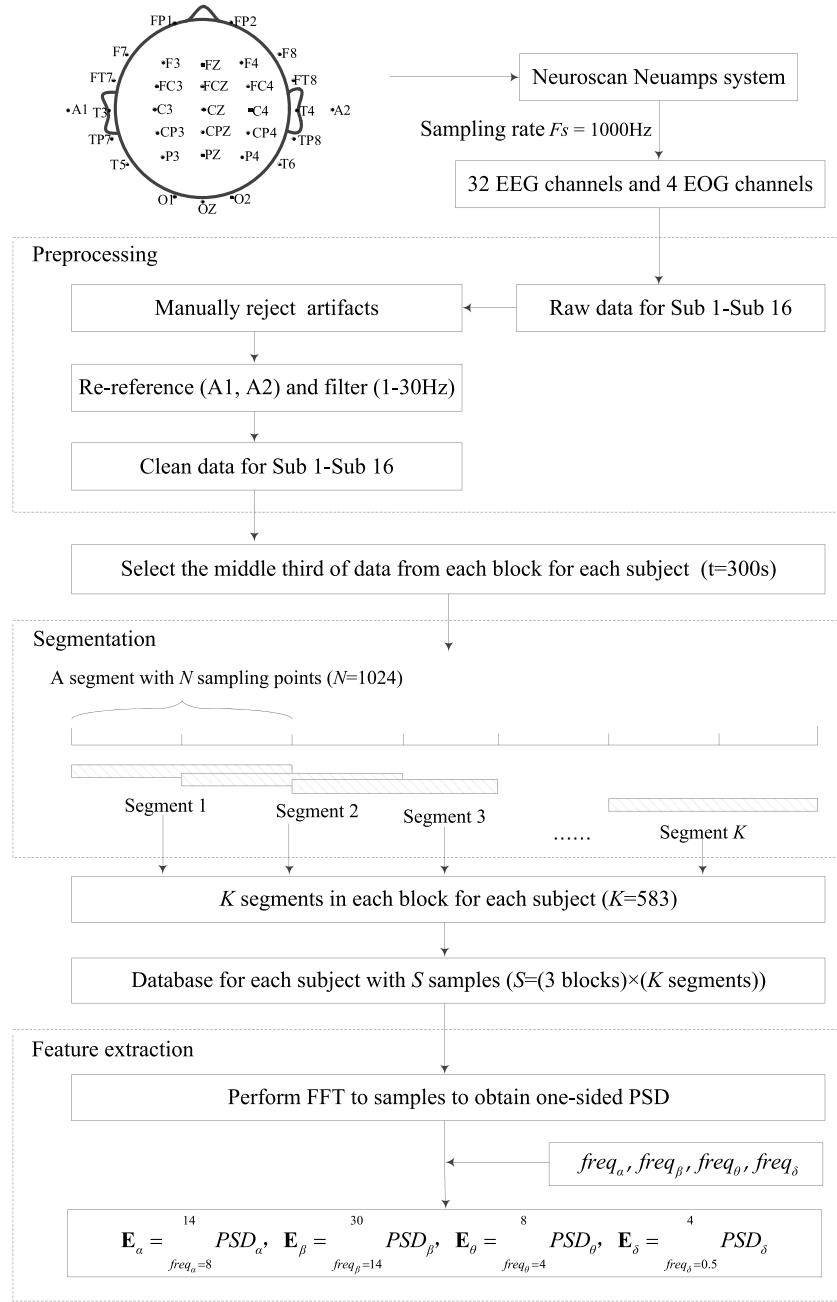


Fig. 5. EEG processing flow.

hidden nodes and the random parameters [39–41].

The number of hidden nodes,  $L$ , is determined by gradually increasing the hidden node under two inequality constraints shown as Eq. (3) and Eq. (4). So the maximum number of hidden nodes,  $L_{\max}$ , and the expected tolerance,  $\varepsilon$ , are defined to determine  $L$ .

$$L \leq L_{\max} \quad (3)$$

$$\|\mathbf{e}_L\|_F \succ \varepsilon \quad (4)$$

where  $\mathbf{e}_L$  refers to the error with  $L$  hidden nodes;  $\|\cdot\|_F$  refers to the  $F$ -norm.

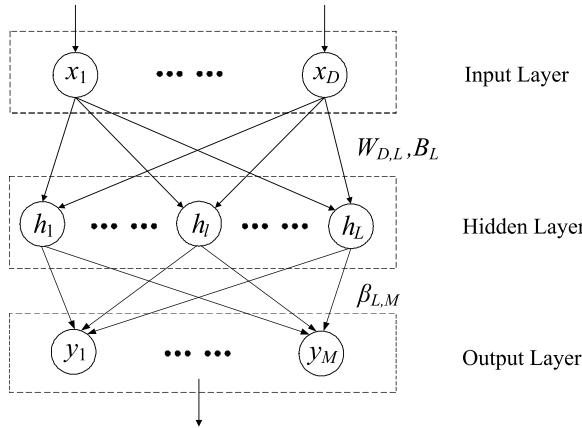
Random parameters are determined by an inequality constraint in the light of a supervisory mechanism, as shown in Eq.(5). Here, we define the following variables: the scope sequence,  $\mathbf{p}$ , for determining weights and biases; the contraction sequence,  $\mathbf{r}$ , for determining the speed of error reduction; and the pool size of candidate nodes,  $T_{\max}$ , for

determining maximum times of random configuration. Besides, the regularization coefficient,  $\lambda$ , is introduced in the original SCN to prevent over-fitting.

$$\xi_L = \frac{\left( (\mathbf{e}_{L-1})^T * \mathbf{h}_L \right)^2}{(\mathbf{h}_L)^T * \mathbf{h}_L} - (1 - r - \mu_L)(\mathbf{e}_{L-1})^T \mathbf{e}_{L-1} > 0 \quad (5)$$

where  $\mu_L$  is a non-negative real number sequence with  $\lim_{L \rightarrow +\infty} \mu_L = 0$  and  $\mu_L \leq (1-r)$ ;  $\mathbf{h}_L$  is the output of hidden layer with  $L$  hidden nodes;  $\xi_L$  is an index defined for selecting the candidate node in the pool.

The principle of the SCN is illustrated by **Algorithm I**.

**Fig. 6.** Structure of the SCN.**Algorithm I**

Given inputs  $\mathbf{X} \in \mathbb{R}^{D \times S}$  and outputs  $\mathbf{Y} \in \mathbb{R}^{M \times S}$ ,  $S$  is the number of samples,  $D$  is the number of input features, and  $M$  is the number of output features;

Set the maximum number of hidden nodes  $L_{\max}$ ;

Set expected tolerance,  $\epsilon$ ;

Set pool size of candidate nodes,  $T_{\max}$ ;

Set scope sequence,  $\mathbf{p} = [p_{\min}; \Delta p; p_{\max}]$  and  $\Delta p$  is not fixed;

Set contraction sequence,  $\mathbf{r} = [r_{\min}; \Delta r; r_{\max}]$  and  $\Delta r$  is not fixed;

Set regularization coefficient  $\lambda$ .

Phase 0: Initialization

1. Initialize  $L=0$ , and  $\mathbf{e}_L = [\mathbf{e}_{L,1}, \dots, \mathbf{e}_{L,m}, \dots, \mathbf{e}_{L,M}]^T = \mathbf{Y}^T$ ; an empty set  $\Omega$ ;

Phase 1: Criterion

2. While  $\|\mathbf{e}_L\|_F > \epsilon$  AND  $L \leq L_{\max}$ , Do

3.  $L=L+1$ ;

Phase 2: Hidden Parameter

4. For  $p \in \mathbf{p}$ , Do

5. For  $r \in \mathbf{r}$ , Do

6. For  $t = 1, 2, \dots, T_{\max}$ , Do

7. Randomly assign  $\mathbf{w}_L^t$  and  $b_L^t$  from  $[-p, p]$ ,  $\mathbf{w}_L^t \in \mathbb{R}^{D \times 1}$ ,  $b_L^t \in \mathbb{R}$

$$8. \quad \Phi_L^t = \frac{1}{1 + \exp\left(-(\mathbf{w}_L^t)^T \mathbf{X} + b_L^t \mathbf{I}\right)}, \mathbf{I} \text{ is the identity matrix};$$

9.  $\mathbf{h}_L^t = [\Phi_L^t]$

$$10. \quad \mathbf{e}_{L,m}^t = \frac{\left((\mathbf{e}_{L-1,m})^T * \mathbf{h}_L^t\right)^2}{\left(\mathbf{h}_L^t\right)^T * \mathbf{h}_L^t} \quad (1 - r) \begin{pmatrix} 1 & & \\ & \ddots & \\ & & L \end{pmatrix} (\mathbf{e}_{L-1,m})^T \mathbf{e}_{L-1,m}$$

11. If  $\min\{\mathbf{e}_{L,1}^t, \dots, \mathbf{e}_{L,m}^t, \dots, \mathbf{e}_{L,M}^t\} > 0$ , then

12.  $\mathbf{e}_L^t = \sum_{m=1}^M \mathbf{e}_{L,m}^t$ , save  $\mathbf{w}_L^t$ ,  $b_L^t$  and corresponding  $\mathbf{e}_L^t$  in  $\Omega$

13. End if (corresponding to Step 11)

14. End for (corresponding to Step 6)

15. If  $\Omega$  is not empty, then

16. Find  $\mathbf{w}_L$ ,  $b_L$  that maximize  $\mathbf{e}_L^t$  in  $\Omega$ ;

17. set  $\mathbf{H}_L = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_L]$

18. Break(go to Step 24)

19. Else

20. Renew  $r$

21. End if (corresponding to Step 15)

22. End for (corresponding to Step 5)

23. Renew  $p$

24. End for (corresponding to Step 4)

Phase 3: Output Weights

$$25. \quad \mathbf{B}_{L,M} = (\mathbf{H}_L^T \mathbf{H}_L + \mathbf{I})^{-1} \mathbf{H}_L^T \mathbf{Y}, \mathbf{I} \text{ is the identity matrix};$$

$$26. \quad \mathbf{Y}_{pre} = \mathbf{B}_{L,M} \mathbf{H}_L$$

$$27. \quad \mathbf{e}_L = \mathbf{Y} - \mathbf{Y}_{pre}$$

28. End while (corresponding to Step 2)

29. Return  $\mathbf{W}_{D,L} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_L], \mathbf{B}_L = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_L], \mathbf{B}_{L,M}$

**4.2. Evaluation**

For multi-class classification, the confusion matrix can describe the

		Predicted class	
		LMW	MMW
HMW	MMW	$U_{11}$	$U_{12}$
	HMW	$V_{11}$	$V_{12}$
MMW	HMW	$U_{21}$	$U_{22}$
	HMW	$V_{21}$	$V_{22}$
HMW	HMW	$U_{31}$	$U_{32}$
	HMW	$V_{31}$	$V_{32}$
		$P_1$	$1 - P_1$
		$P_2$	$1 - P_2$
		$P_3$	$1 - P_3$
		$R_1$	$R_2$
		$1 - R_1$	$1 - R_2$
		$R_3$	$A$
		$1 - R_3$	$1 - A$

Target class

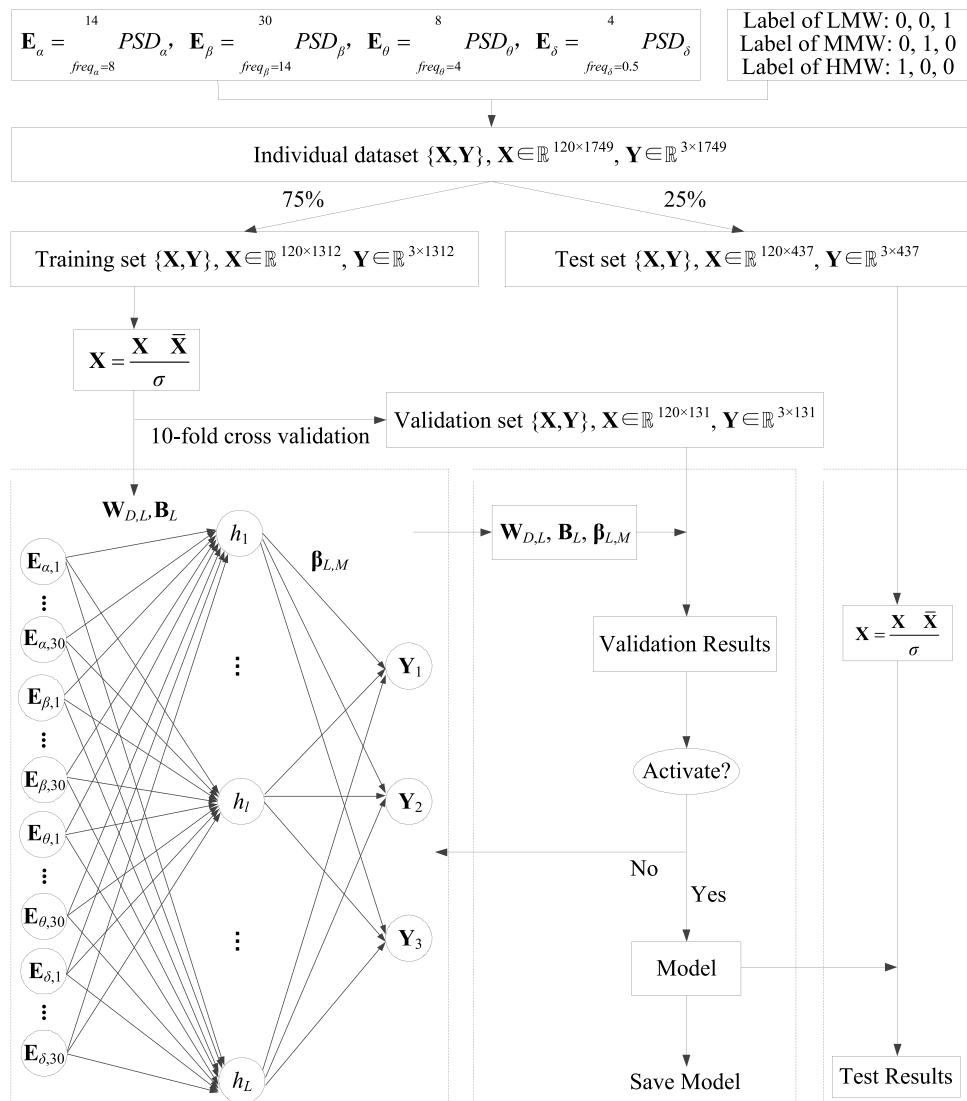
**Fig. 7.** Details of classification results.

classification results in detail. Precision, recall and accuracy derived from the confusion matrix are defined as the model evaluation indexes in this paper. Fig. 7 shows the confusion matrix for our 3-class classification and some evaluation indexes.

- (1) The diagonal element,  $U_{ii}$  ( $i = 1, 2, 3$ ), stands for the number of data correctly classified.  $U_{ij}$  ( $i, j = 1, 2, 3; i \neq j$ ) is the number of the data in the  $j$ -th target class that was misclassified to the  $i$ -th class. In this study, there are three classes, Class  $i$  ( $i = 1, 2, 3$ ), which represent the LMW class, the MMW class and the HMW class, respectively.
- (2)  $V_{ij}$  ( $i, j = 1, 2, 3$ ) is defined as the proportion of  $U_{ij}$  ( $i, j = 1, 2, 3$ ) to the total number of samples, as shown in Eq.(6).

$$V_{ij} = \frac{U_{ij}}{\sum_{i=1}^3 \sum_{j=1}^3 U_{ij}} \times 100\% \quad (6)$$

- (3)  $P_i$  ( $i = 1, 2, 3$ ) is defined as the precision of Class  $i$ . It is the proportion of  $U_{ii}$  ( $i = 1, 2, 3$ ) to the total number of samples whose predicted class is Class  $i$ , as shown in Eq.(7) [61,62].

**Fig. 8.** SCN procedure.

$$P_i = \frac{U_{ii}}{\sum_{j=1}^3 U_{ij}} \times 100\% \quad (7)$$

(4)  $R_j$  ( $j = 1, 2, 3$ ) is defined as the recall of Class  $j$ . It is the proportion of  $U_{jj}$  ( $j = 1, 2, 3$ ) to the total number of samples whose target class is Class  $j$ , as shown in Eq.(8) [61,62].

$$R_j = \frac{U_{jj}}{\sum_{i=1}^3 U_{ij}} \times 100\% \quad (8)$$

(5)  $A$  is defined as the accuracy, as shown in Eq.(9) [62]

$$A = \frac{\sum_{i=1}^3 U_{ii}}{\sum_{i=1}^3 \sum_{j=1}^3 U_{ij}} \times 100\% \quad (9)$$

#### 4.3. SCN method

As shown in Fig. 8, the SCN method consists of three steps. The first step is the preparation of the data sets, as described in Section 4.3.1. Specifically, the individual dataset was obtained from the processed EEG data. And then the samples in the individual dataset were randomly divided into the training set and the test set by a ratio of 75 % to 25 %. Ten-fold cross-validation was used to divide the training set into a training set and a validation set. The validation set was used to optimize the model and determine the appropriate hyperparameters, and the test set was used to evaluate the final model. After that, the data of the training set and the test set were standardized. The second step is the modeling with training sets and validation sets to determine the proper hyper-parameters. The method and its determined hyper-parameters were described in Section 4.3.2. The last step is the modeling evaluation with test sets.

##### 4.3.1. Preparation

###### (1) Input and output

The input of dataset for each subject has a total of  $D$  features and it can be described as  $\mathbf{X} = [\mathbf{X}_\alpha, \mathbf{X}_\beta, \mathbf{X}_\theta, \mathbf{X}_\delta]^T, \mathbf{X} \in \mathbb{R}^{D \times S}$ . Here  $D$  is the product of channels and features, and  $D = 120$ . It is worth noting that  $\mathbf{X}_\alpha = [\mathbf{E}_{\alpha,1}, \mathbf{E}_{\alpha,2}, \dots, \mathbf{E}_{\alpha,30}], \mathbf{X}_\alpha \in \mathbb{R}^{S \times 30}$  and it is similar for  $\mathbf{X}_\beta, \mathbf{X}_\theta, \mathbf{X}_\delta$ .

Samples for LMW, MMW and HMW are separately labeled as “0, 0, 1”, “0, 1, 0” and “1, 0, 0” by one-hot code, so the output of data set for each subject,  $\mathbf{Y}$ , has a total of  $M$  features. Here  $M$  is the number of mental workload blocks and  $M = 3$ . Hence,  $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3]^T, \mathbf{Y} \in \mathbb{R}^{M \times S}$ .

###### (2) Training set, validation set and test set

This paper aims to build the SSC of mental workload using each dataset with  $S$  samples. The samples in the individual dataset were randomly divided into the training set and the test set by a ratio of 75 % to 25 %. Ten-fold cross-validation was used to determine the hyper-parameters of the SCN [63–65].

###### (3) Standardization

Z-score was used to standardize sample data, as shown in Eq.(10).

$$\mathbf{X} = \frac{\mathbf{X} - \bar{\mathbf{X}}}{\sigma} \quad (10)$$

where  $\bar{\mathbf{X}}$  is the average value of  $\mathbf{X}$ ;  $\sigma$  is the standard deviation of  $\mathbf{X}$ .

##### 4.3.2. SSC based on the SCN

For the SSC based on the SCN, the most critical hyper-parameter is the maximum number of hidden nodes,  $L_{\max}$ , which will directly affect the number of iterations of the network [3941]. When  $L_{\max}$  is too small, the network will be under-fitting. When the  $L_{\max}$  is too large, the network will be over-fitting. Both under-fitting and over-fitting will

**Table 3**

Results of accuracies with  $L_{\max}$  varying from 10 to 250 for SSC of Sub 1.

No.	$L_{\max}$	Training accuracy (%)	Validation accuracy (%)
1	10	71.8 ± 0.6	69.7 ± 1.1
2	20	79.1 ± 0.5	76.0 ± 0.9
3	30	82.8 ± 0.3	78.7 ± 0.7
4	40	85.0 ± 0.2	80.3 ± 1.0
5	50	86.5 ± 0.3	81.1 ± 0.6
6	60	87.8 ± 0.2	81.3 ± 0.6
7	70	88.8 ± 0.2	81.4 ± 0.8
8	80	89.8 ± 0.2	81.6 ± 0.7
9	90	90.8 ± 0.2	81.0 ± 0.9
10	100	91.5 ± 0.2	81.2 ± 0.9
11	110	92.3 ± 0.2	81.1 ± 0.9
12	120	93.0 ± 0.2	81.1 ± 0.5
13	130	93.7 ± 0.1	80.8 ± 0.7
14	140	94.3 ± 0.2	80.3 ± 0.9
15	150	94.8 ± 0.2	80.4 ± 1.0
16	160	95.3 ± 0.2	79.9 ± 1.0
17	170	95.8 ± 0.1	79.9 ± 1.1
18	180	96.2 ± 0.2	80.2 ± 0.9
19	190	96.6 ± 0.1	79.8 ± 0.8
20	200	97.0 ± 0.1	79.5 ± 1.0
21	210	97.3 ± 0.1	79.5 ± 0.7
22	220	97.6 ± 0.2	79.0 ± 1.2
23	230	97.9 ± 0.1	79.0 ± 1.0
24	240	98.0 ± 0.1	78.5 ± 0.8
25	250	98.3 ± 0.1	78.4 ± 1.0

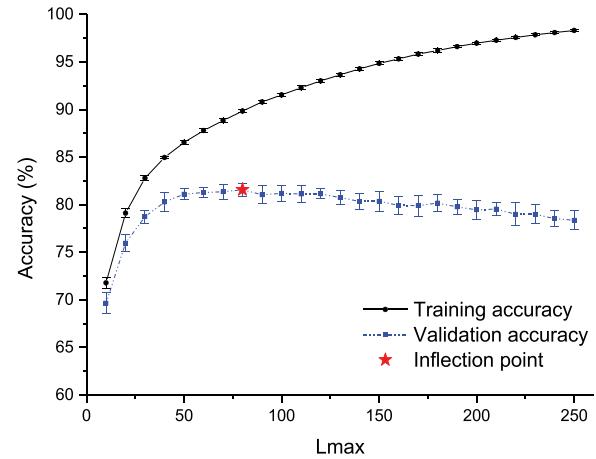
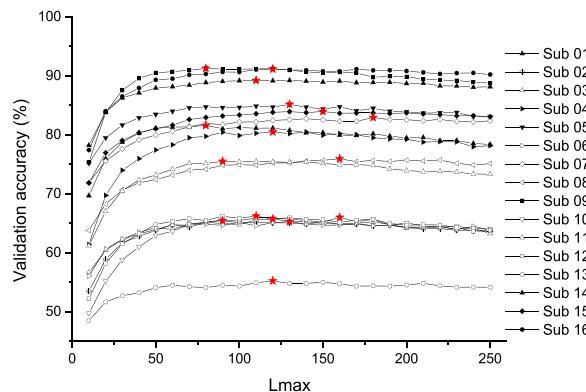


Fig. 9. Change curves of accuracies with  $L_{\max}$  varying from 10 to 250 for SSC of Sub 1.

undermine the effectiveness of the network. So the inflection point, where the accuracies of the validation set firstly rise and then fall with the increase of the number of hidden nodes, is considered as the optimal  $L_{\max}$ .

When the hyper-parameters other than  $L_{\max}$  were adjusted to the appropriate range, their influence on the modeling time of the classifier was more prominent than its accuracy. So after debugging, the pool size of candidate nodes,  $T_{\max}$ , was set as 100. The expected tolerance,  $\epsilon$ , was set as  $10^{-3}$ . The contraction sequence,  $\mathbf{r}$ , was not fixed and it incrementally increased from 0.9 to 1. The scope sequence,  $\mathbf{p}$ , was set as a given set,  $\{0.5, 1, 2, 5, 10, 50, 100, 150, 200\}$ . The regularization coefficient,  $\lambda$ , was set as 0.0005.

$L_{\max}$  was determined by ten-fold cross-validation. Twenty repetitions on each fold are carried out to weaken the random effects of the SCN. Taking the SSC of Sub 1 as an example, the mean value and the standard deviation of accuracies are listed in Table 3 and Fig. 9 with  $L_{\max}$  ranging from 10 to 250. As can be seen, the accuracies of the training set gradually increased when  $L_{\max}$  increased from 10 to 250. The accuracies of the validation set increased first with  $L_{\max}$  ranging from 10 to 80 and



**Fig. 10.** Change curves of accuracies with  $L_{\max}$  varying from 10 to 250 for SSCs.

**Table 4**  
Mean values of accuracies with optimal and suboptimal  $L_{\max}$  for SSCs.

Subject No.	Optimal $L_{\max}$			Suboptimal $L_{\max}$		
	$L_{\max}$	Training accuracy (%)	Validation accuracy (%)	$L_{\max}$	Training accuracy (%)	Validation accuracy (%)
Sub 1	80	89.8	81.6	100	91.5	81.2
Sub 2	90	78.8	65.4	100	80.0	65.1
Sub 3	120	84.1	65.7	100	81.8	64.8
Sub 4	120	92.6	80.5	100	90.9	79.9
Sub 5	130	95.5	85.2	100	93.7	84.6
Sub 6	130	84.0	65.3	100	80.5	65.1
Sub 7	180	96.0	82.9	100	91.5	82.2
Sub 8	160	89.6	75.9	100	85.3	74.9
Sub 9	80	96.1	91.3	100	97.0	91.1
Sub 10	160	88.4	66.0	100	81.4	65.4
Sub 11	90	86.7	75.5	100	87.9	75.4
Sub 12	110	82.9	66.2	100	81.7	65.8
Sub 13	120	76.1	55.2	100	73.2	54.3
Sub 14	110	95.5	89.2	100	95.0	89.2
Sub 15	150	95.4	84.0	100	92.3	83.4
Sub 16	120	97.5	91.1	100	96.7	90.6

then decreased with  $L_{\max}$  ranging from 80 to 250. Therefore, the inflection point, where  $L_{\max}$  is equal to 80, was determined as the optimal  $L_{\max}$  for Sub 1.

The optimal values of  $L_{\max}$  for the SSCs of Sub 2 to Sub 16 were determined by using the same method. They are shown in Fig. 10.

From Figs. 9 and 10, the following conclusions can be observed:

(1) The accuracies of the validation set only decreased slightly when  $L_{\max}$  was greater than the optimal  $L_{\max}$  and decreased significantly when  $L_{\max}$  was far greater than the optimal  $L_{\max}$ . This may be attributed to the introduction of the regularization coefficient to avoid overfitting.

(2) The optimal  $L_{\max}$  for the SSCs of Sub 1 - Sub 16 was different, but it was roughly distributed in the interval [80, 160]. If the optimal  $L_{\max}$  for each subject was updated to the suboptimal  $L_{\max}$  that is close to the optimal  $L_{\max}$ , the impact of this change on the classifier accuracy could be ignored.

The suboptimal  $L_{\max}$  for Sub 1 – Sub 16 was further determined by the following analysis. Table 4 shows the classifier accuracies of the training set and validation set when the suboptimal  $L_{\max}$  was set to 100. As can be seen, the differences of accuracies between the optimal  $L_{\max}$  and the suboptimal  $L_{\max}$  are within  $\pm 1\%$ , which indicated that if the accuracy can be sacrificed to some extent, a general  $L_{\max}$  could be determined for all subjects. This conclusion simplifies the operation of MW classification for multiple subjects and provides great convenience for later practical applications. Therefore,  $L_{\max}$  for the SSCs of Sub 1 - Sub 16 were uniformly set to 100.

**Table 5**  
Classification results of SSCs for Sub 1 - Sub 16.

Subject No.	Training accuracy (%)	Test accuracy (%)	Modeling time (s)
Sub 1	91.3	82.2	3.6
Sub 2	79.5	65.0	3.9
Sub 3	80.7	67.5	3.8
Sub 4	92.0	80.3	3.8
Sub 5	93.1	86.0	3.6
Sub 6	77.4	66.4	3.8
Sub 7	91.5	80.3	3.7
Sub 8	83.6	72.1	4.1
Sub 9	97.0	90.2	3.5
Sub 10	80.3	67.3	3.7
Sub 11	88.8	74.1	3.6
Sub 12	81.9	64.5	3.8
Sub 13	73.1	56.5	3.9
Sub 14	95.7	88.1	3.7
Sub 15	91.7	83.5	3.7
Sub 16	96.5	90.2	3.7
Mean	87.1	75.9	3.7

## 5. Classification results and analysis

The classification results of the SSCs are described in Section 5.1. Further analysis and discussion are given in Section 5.2 – Section 5.3. Besides, the classification accuracies of the two classifiers, the SCN and the traditional ANN, were also compared in this study, as shown in Section 5.4.

### 5.1. Results of SSCs for mental workload

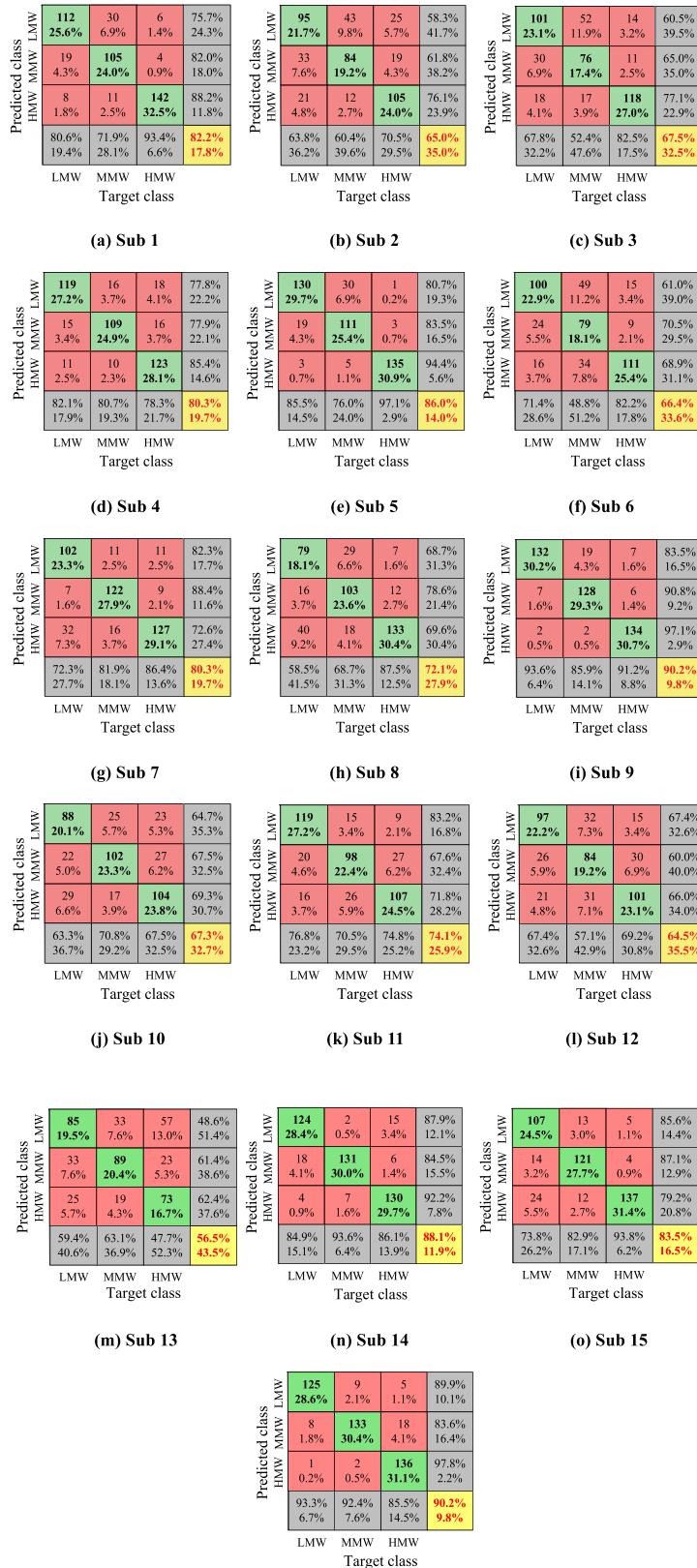
Table 5 shows the classification results of SSCs for Sub 1 - Sub 16. Some conclusions can be drawn from Table 5.

- (1) The range of SSC training accuracies was between 73.1 % and 97.0 % with an average of 87.1 %. The range of SSC test accuracies was between 56.5 % and 90.2 % with an average of 75.9 %, among which the accuracies between 80 % and 90 % constituted the largest proportion. Overall, the SSCs worked well and the classifier accuracies were far greater than the random probability of 33 %.
- (2) The modeling process of SSC was performed on MATLAB R2014a in a computer with an Intel(R) Core(TM) i5-4590 CPU and an installed memory of 8.00 GB. The range of SSC modeling time was between 3.5 s and 4.1 s with an average of 3.7 s. Overall, a rapid classification can be achieved when the hyper-parameters are given in advance.

To explain the classification results clearly, the details of test results for Sub 1 – Sub16 are shown in Fig. 11, which is consistent with Fig. 7.

From Fig. 11, these conclusions can be observed:

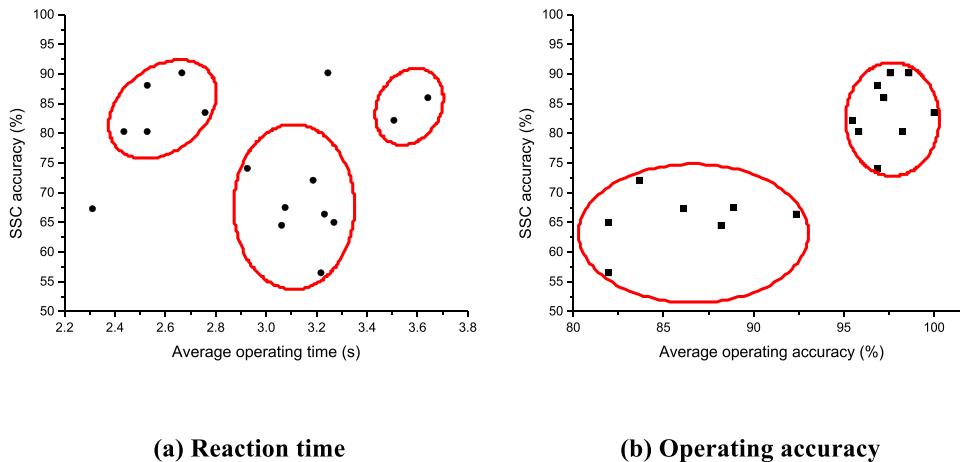
- (1) Taking Sub 1 as an example, the classification precisions of LMW, MMW and HMW were 75.7 %, 82 % and 88.2 %, respectively. The recalls of LMW, MMW and HMW were 80.6 %, 71.9 % and 93.4 %, respectively. Therefore, the SSC of Sub 1 had the best recognition performance for HMW. Similarly, the rule was also true for Sub 2, Sub 3 and Sub 5.
- (2) For the SSC of Sub 11, the classification precisions of LMW, MMW and HMW were 83.2 %, 67.6 % and 71.8 %, respectively. The recalls of LMW, MMW and HMW were 76.8 %, 70.5 % and 74.8 %, respectively. Therefore, the SSC of Sub 11 had the best recognition performance for LMW.
- (3) For the SSCs of other subjects not mentioned above, the highest precision and recall could not simultaneously reach a certain mental workload level. Overall, the recognition performances for LMW, MMW and HMW differed from the SSCs of Sub 1 - Sub 16,

**Fig. 11.** Test details of classification results for Sub 1 – Sub 16.

**Table 6**

Operating performance for Sub 1 - Sub 16.

Subject No.	Reaction time (s)				Operating accuracy (%)			
	LMW	MMW	HMW	Mean	LMW	MMW	HMW	Mean
Sub 1	4.2	3.3	3.1	3.5	100.0	93.8	92.7	95.5
Sub 2	2.8	3.4	3.6	3.3	75.0	79.2	91.7	81.9
Sub 3	3.5	2.7	3.1	3.1	75.0	97.9	93.8	88.9
Sub 4	2.6	2.3	2.4	2.4	100.0	100.0	94.8	98.3
Sub 5	3.5	3.4	4.0	3.6	100.0	100.0	91.7	97.2
Sub 6	2.7	3.7	3.3	3.2	100.0	93.8	83.3	92.4
Sub 7	2.8	2.4	2.4	2.5	100.0	89.6	97.9	95.8
Sub 8	3.9	2.6	3.1	3.2	75.0	85.4	90.6	83.7
Sub 9	3.2	2.9	3.6	3.2	100.0	97.9	94.8	97.6
Sub 10	2.5	1.8	2.7	2.3	75.0	85.4	97.9	86.1
Sub 11	3.5	2.6	2.7	2.9	100.0	97.9	92.7	96.9
Sub 12	3.5	2.8	2.9	3.1	100.0	81.3	83.3	88.2
Sub 13	2.8	3.0	3.9	3.2	75.0	83.3	87.5	81.9
Sub 14	1.9	2.3	3.4	2.5	100.0	95.8	94.8	96.9
Sub 15	3.3	2.7	2.3	2.8	100.0	100.0	100.0	100.0
Sub 16	2.5	2.6	2.9	2.7	100.0	97.9	97.9	98.6



(a) Reaction time

(b) Operating accuracy

Fig. 12. Relationship between operating performance and classifier accuracy.

which may due to the individual differences in the acceptance and adaptability to different mental workload levels.

### 5.2. Operating performance versus SSC accuracy

The impact of operating performance on SSC accuracy is analyzed in this section. Two factors of operating performance are mainly considered, and they are the reaction time and the operating accuracy.

**Table 7**

Classification results of SMCs versus the number of training subjects in SMC.

No.	$L_{\max}$	SMC accuracy (%)	$B_a$ (%)	SMC modeling time (s)	$B_t$ (s)
1	100	90.2	90.2	3.5	3.5
2	160	86.6	90.2	8.8	7.2
3	235	82.1	88.1	17.6	10.9
4	385	81.5	86.0	44.5	14.5
5	425	79.3	83.5	61.6	18.2
6	450	74.8	82.2	75.4	21.8
7	600	69.8	80.3	142.0	25.6
8	625	71.6	80.3	176.2	29.3
9	700	70.2	74.1	232.2	32.9
10	725	72.6	72.1	283.6	37.0
11	775	59.3	67.5	354.1	40.8
12	875	70.0	67.3	473.1	44.5
13	850	65.6	66.4	481.7	48.3
14	800	66.1	65.0	444.5	52.2
15	950	63.4	64.5	681.6	55.9
16	825	57.2	56.5	555.2	59.9

Operating performances for Sub 1 - Sub 16 are shown in Table 6. Fig. 12 describes the relationship between the classifier accuracy and the two factors of operating performance, respectively.

From Fig. 12, these conclusions can be observed:

- (1) In Fig. 12 (a), the relationship between the reaction time and the classifier accuracy is not clear, but the overall trend is that the SSCs are more accurate for subjects with longer or shorter reaction time than for those subjects with moderate reaction time.
- (2) In Fig. 12 (b), the relationship between the operating accuracy and the classifier accuracy is clear, and the overall trend is that SSCs are more accurate for subjects with higher operating accuracy. The classifier accuracy has a positive correlation with the operating accuracy ( $r = 0.852$ ,  $p < 0.01$ ). This may explain the reason why subjects are required to respond to tasks as much as possible in the previous study [34] and may provide a guiding significance for future experiments.

### 5.3. Performance comparison between SSC and SMC

Some researchers prefer to establish classification methods for multiple subjects [66]. The presented SSC in this paper was trained and tested on a single subject. For further comparison, the SMC based on the SCN was built to realize the classification of mental workload for multiple subjects. Fifteen SMCs with different numbers of training subjects, ranging from 2 to 16, were established to explain their accuracies and

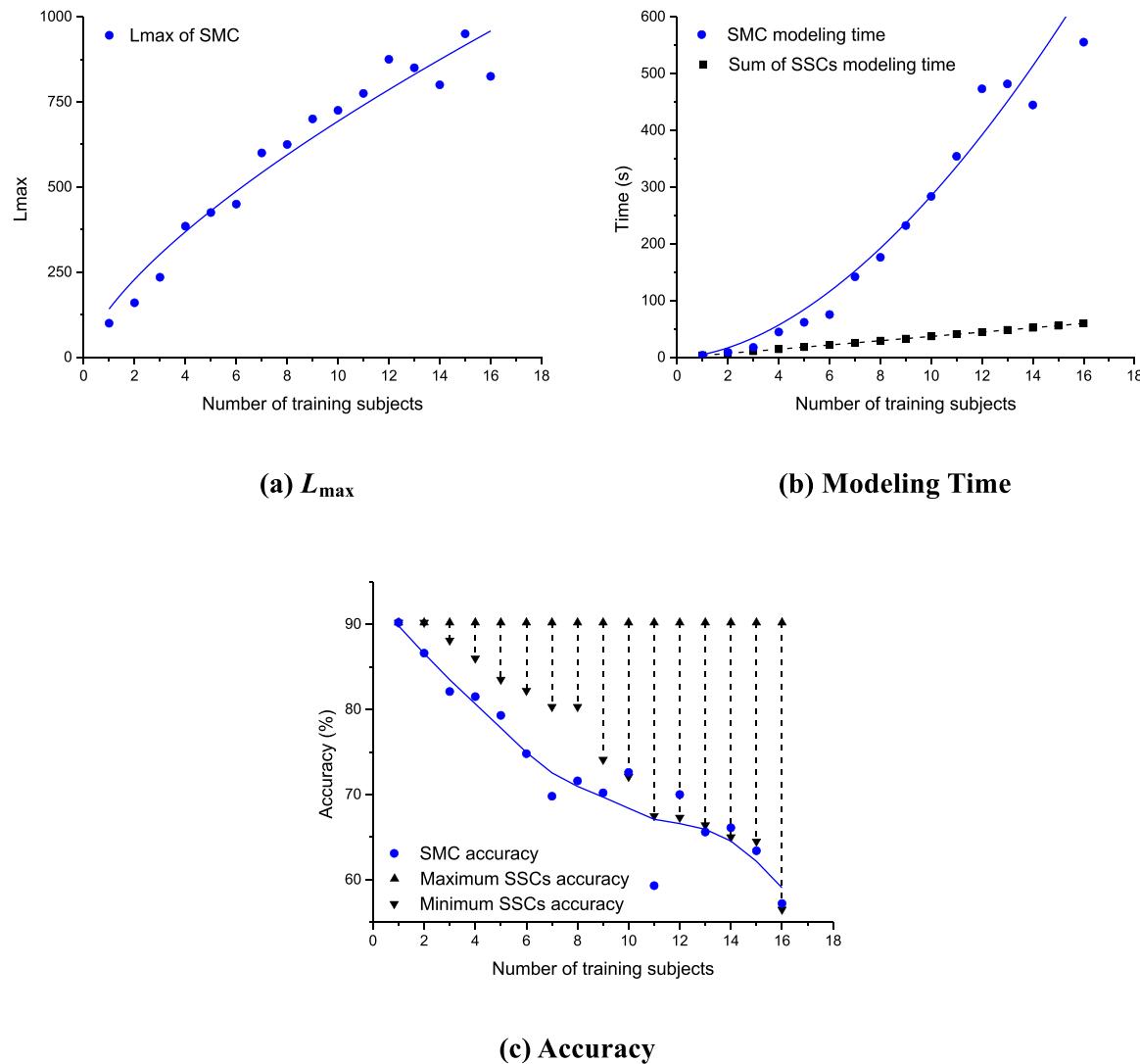


Fig. 13. Comparison between SSC and SMC.

modeling time.

Sub 1 - Sub 16 were rearranged according to the descending order of the SSC test accuracies, and were updated as Sub<sub>M</sub> 1 - Sub<sub>M</sub> 16. The required training subjects of SMCs were sequentially selected according to the order of Sub<sub>M</sub> 1 - Sub<sub>M</sub> 16.

The minimum accuracy of SSCs,  $B_a$ , picked up with Eq. (11), was used to compare with the SMC accuracy. The sum of SSCs modeling time,  $B_t$  in Eq. (12), was used to compare with the SMC modeling time.

$$B_a = \min(A_i), i = 1, 2, \dots, s \quad (11)$$

$$B_t = \sum_{i=1}^s t_i \quad (12)$$

where  $A_i$  is the SSC accuracy of Sub<sub>M</sub>  $i$ ;  $t_i$  is the SSC modeling time of Sub<sub>M</sub>  $i$ ;  $s$  is the number of subjects that will be trained for SMC.

The classification results of SMCs versus the number of training subjects in SMC are illustrated in Table 7.

Fig. 13 shows a clear comparison between the SSCs and the SMCs. Fig. 13 (a) ~ (c) depict the variation trend of  $L_{\max}$ , modeling time and classifier accuracy with different numbers of training subjects, respectively. Besides, Fig. 13 (b) can reflect the difference between the SSCs and the SMC in modeling time when modeling for multiple subjects. Fig. 13 (c) can also reflect the relationship between the minimum SSC

accuracy and the SMC accuracy. Some conclusions can be drawn from Fig. 13.

- (1) In Fig. 13 (a), the  $L_{\max}$  of SMC increased almost linearly from 100 to 825 with the increase of the number of training subjects. This indicated that the network built for multiple subjects was far more complicated than that for an individual, and the structure of the SMC was more complex with more training subjects.
- (2) In Fig. 13 (b),  $B_t$  of SSCs increased linearly and the SMC modeling time increased exponentially with the increase of the number of training subjects. For the SSC, this linear variation of modeling time was caused by conducting multiple SSCs sequentially. For the SMC, the exponential variation of modeling time was strongly related to the linear variation of  $L_{\max}$ , as shown in Fig. 13 (a). This could be explained by the fact that  $L_{\max}$  directly determines the number of iteration and the sigmoid activation function in the SCN. Overall, the above analysis indicated that the sequential SSC modeling for multiple subjects was more favorable in time.
- (3) In Fig. 13 (c), when the number of subjects trained for the SMC was less than or equal to 9, the test accuracies of SMCs were all lower than  $B_a$ . When the number of subjects trained for the SMC was greater than 9, the test accuracies of SMCs were mostly close to  $B_a$ . This indicated that the SSC had an advantage of accuracy over the SMC for a small number of training subjects and the SMC

**Table 8**  
Classification results of the ANN for Sub 1 ~ Sub 16.

Subject No.	Training accuracy (%)	Test accuracy (%)	Modeling time (s)
Sub 1	76.6	77.1	2.0
Sub 2	59.2	56.9	2.1
Sub 3	67.7	59.5	2.3
Sub 4	81.6	77.9	2.1
Sub 5	86.8	84.4	1.8
Sub 6	62.6	60.7	2.2
Sub 7	74.0	71.8	1.9
Sub 8	68.3	63.7	2.5
Sub 9	92.6	88.9	1.9
Sub 10	63.8	61.1	1.8
Sub 11	72.2	71.0	2.1
Sub 12	59.9	55.3	1.9
Sub 13	52.0	50.0	1.5
Sub 14	86.9	80.5	1.8
Sub 15	83.3	78.2	1.5
Sub 16	87.3	84.0	1.4
Mean	73.2	69.6	1.9

might embody the trend of statistical performance for a large number of subjects.

#### 5.4. Performance comparison between the SCN and ANN

To further reveal the advantage of the SCN, it is compared with the traditional ANN. The standard backpropagation algorithm is used to train the ANN. The network weights and biases are randomly initialized at the beginning of the training stage. The stochastic gradient descent method is used to realize the adjustment of weights, biases and the minimization of errors [67]. In this study, the ANN is constructed using the same training set, test set and output variables as the SCN. The number of hidden nodes is the same as the maximum number of hidden nodes in the SCN, which is set to 100. Classification results of the ANN for Sub 1 ~ Sub 16 are illustrated in Table 8.

Fig. 14 shows a clear comparison between the SCN and ANN. Fig. 14 (a) and (b) depict the test accuracy and modeling time for different subjects, respectively.

From Table 8 and Fig. 14, some conclusions can be observed:

(1) The range of ANN training accuracy is between 52.0 % and 92.6 % with an average of 73.2 %. The range of ANN test accuracy is between 50.0 % and 88.9 % with an average of 69.6 %. These accuracies were lower than those of the SCN. The SCN can improve the classification accuracy, to a certain extent.

(2) The modeling processes of the ANN and SCN were performed on the same computer and the same version of the software. The range of

ANN modeling time is between 1.4 s and 2.5 s with an average of 1.9 s. Compared with the traditional ANN, the SCN has approximate modeling time and higher accuracy.

## 6. Conclusion

In this paper, the experiments based on MATB II were carried out to collect EEG data from 16 subjects, and a method of mental workload classification was proposed using EEG data and the SCN. Based on the method, 16 SSCs for individuals and 15 SMCs for different numbers of training subjects were established, respectively. Details were further discussed according to the classification results. Some conclusions can be obtained from our study:

### (1) Classification results of the SSCs

The range of SSC test accuracy was between 56.5 % and 90.2 % with an average of 75.9 %, and the range of SSC modeling time was between 3.5 s and 4.1 s with an average of 3.7 s, which indicated that the SSCs for Sub 1 ~ Sub 16 worked fast and well. Besides, the SSCs for 16 subjects had different recognition performance to LMW, MMW and HMW, which may due to the individual differences in the acceptance and adaptability to different mental workload levels.

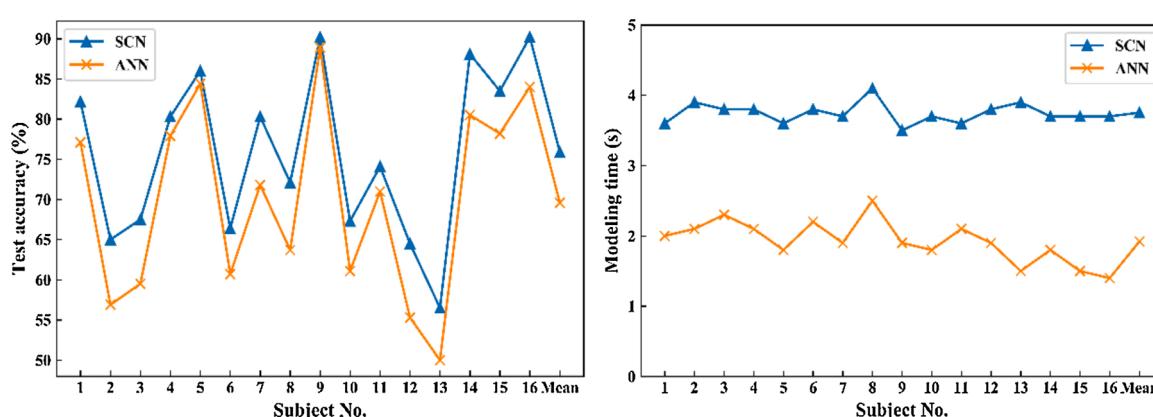
### (2) Relationship between the operating performance and the SSC accuracy

The relationship between the reaction time and the SSC accuracy was not clear. The SSCs might be more accurate for subjects with longer or shorter reaction time than for those subjects with moderate reaction time. Besides, the correlation coefficient between the operating accuracy and the SSC accuracy was 0.852 at the significance level of 0.01, which meant that the SSCs were more accurate for subjects with higher operating accuracy.

### (3) Comparison between the SSC and the SMC

As the number of trained subjects increased, the modeling time of SSCs and SMCs increased linearly and exponentially, respectively. This indicated that the SSC modeling for multiple subjects was more favorable in time. Besides, the SSC had an accuracy advantage over the SMC for a small number of subjects, and the SMC might embody the trend of statistical performance for a large number of subjects.

### (4) Comparison between the SCN and the ANN



(a) Test accuracy

(b) Modeling Time

Fig. 14. Comparison between the SCN and ANN.

Compared with the traditional ANN, the SCN can improve the classification accuracy. This may be because the SCN can overcome the difficulties in the hyper-parameters setting by gradually increasing hidden nodes.

## CRediT authorship contribution statement

**Pang Liping:** Conceptualization-Equal, Data curation-Equal, Funding acquisition-Equal, Investigation-Equal, Methodology-Lead, Project administration-Equal, Supervision-Supporting, Writing-original draft-Equal, Writing-review & editing-Equal. **Guo Liang:** Data curation-Equal, Formal analysis-Equal, Investigation-Lead, Methodology-Equal, Resources-Supporting, Validation-Equal, Visualization-Equal, Writing-original draft-Equal, Writing-review & editing-Equal. **Zhang Jie:** Conceptualization-Equal, Data curation-Lead, Methodology-Equal, Project administration-Supporting, Validation-Equal, Writing-original draft-Equal, Writing-review & editing-Equal. **Wanyan Xiaoru:** Funding acquisition-Equal, Project administration-Lead, Resources-Equal, Software-Equal, Supervision-Equal. **Qu Hongquan:** Project administration-Supporting, Resources-Equal, Software-Equal, Supervision-Equal. **Wang Xin:** Resources-Equal, Software-Equal, Supervision-Equal

## Acknowledgement

This work was supported by the jointly program of National Natural Science Foundation of China and Civil Aviation Administration of China (No. U1733118).

## Declaration of competing interest

None.

## References

- [1] G.F. Wilson, Operator functional state assessment for adaptive automation implementation, in: Biomonitoring for Physiological and Cognitive Performance during Military Operations, 5797, International Society for Optics and Photonics, 2005, pp. 100–104.
- [2] C.J. Lin, M.H. Hsieh, Classification of mental task from EEG data using neural networks based on particle swarm optimization, Neurocomputing 72 (4-6) (2009) 1121–1130.
- [3] E. Bornemann, Untersuchungen über den Grad der geistigen Beanspruchung, Arbeitsphysiologie 12 (2) (1942) 173–191.
- [4] D. Kahneman, Attention and Effort, Prentice-Hall, Englewood Cliffs, NJ, 1973.
- [5] F.T. Eggemeier, G.F. Wilson, A.F. Kramer, et al., Workload assessment in multi-task environments, Multiple-task Performance (1991) 207–216.
- [6] F. Nachreiner, Standards for ergonomics principles relating to the design of work systems and to mental workload, Appl. Ergon. 26 (4) (1995) 259.
- [7] B. Xie, G. Salvendy, Review and reappraisal of modeling and predicting mental workload in single- and multi-task environments, Work Stress 14 (1) (2000) 74–99.
- [8] Wickens, D. Christopher, Multiple resources and performance prediction, Theor. Issues Ergon. Sci. 3 (2) (2002) 159–177.
- [9] Radtintz Thea, Dual frequency head maps: a new method for indexing mental workload continuously during execution of cognitive tasks, Front. Physiol. 8 (2017) 1019.
- [10] K. Ryu, R. Myung, Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic, Int. J. Ind. Ergon. 35 (11) (2005) 991–1009.
- [11] H. Ayaz, P.A. Shewokis, S. Bunce, et al., Optical brain monitoring for operator training and mental workload assessment, Neuroimage 59 (1) (2012) 36–47.
- [12] X.R. Wanyan, D. Zhuang, Y. Lin, et al., Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation, Int. J. Ind. Ergon. 64 (2018) 1–7.
- [13] J.A. de la O Serna, M.R.A. Paternina, A. Zamora-Méndez, et al., EEG-rhythm specific Taylor-Fourier filter bank implemented with B-splines for the detection of epilepsy using EEG signals, IEEE Sens. J. 20 (12) (2020) 6542–6551.
- [14] Steven Lemm, Benjamin Blankertz, Thorsten Dickhaus, et al., Introduction to machine learning for brain imaging, NeuroImage 56 (2) (2011) 387–399.
- [15] J.D. Karch, M.C. Sander, T. Von Oertzen, et al., Using within-subject pattern classification to understand lifespan age differences in oscillatory mechanisms of working memory selection and maintenance, NeuroImage 118 (2015) 538–552.
- [16] Y. Pang, Y. Yuan, K. Wang, Learning optimal spatial filters by discriminant analysis for brain-computer-interface, Neurocomputing 77 (1) (2012) 20–27.
- [17] E. Pacola, V. Quandt, P. Liberalesso, et al., A versatile EEG spike detector with multivariate matrix of features based on the linear discriminant analysis, combined wavelets, and descriptors, Pattern Recognit. Lett. 86 (2017) 31–37.
- [18] N. Brodu, F. Lotte, Anatole Lécuyer, Exploring two novel features for EEG-based brain-computer interfaces: multifractal cumulants and predictive complexity, Neurocomputing 79 (none) (2012) 87–94.
- [19] Shubham Dodia, Damodar Reddy Edla, Annushree Bablani, et al., An efficient EEG based deceit identification test using wavelet packet transform and linear discriminant analysis, NeuroImage 314 (2019) 31–40.
- [20] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R.B. Pachori, H. Wang, A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI, IEEE Trans. Instrum. Meas. 70 (2021) 1–9. February article sequence number: 4002709.
- [21] A. Bhattacharyya, R.K. Tripathy, L. Garg, R.B. Pachori, A novel multivariate-multiscale approach for computing EEG spectral and temporal complexity for human emotion recognition, IEEE Sens. J. 21 (February 3) (2021) 3579–3591.
- [22] A. Ullal, R.B. Pachori, Variational mode decomposition based automated diagnosis method for epilepsy using EEG signals, in: S. Day, S.K. Pani, J. Rodrigues, B. Majhi (Eds.), Deep Learning, Machine Learning and IoT in Biomedical and Health Informatics Techniques and Applications, Biomedical Engineering, CRC Press, 2020.
- [23] X.W. Wang, D. Nie, B.L. Lu, Emotional state classification from EEG data using machine learning approach, Neurocomputing 129 (2014) 94–106.
- [24] Z. Sun, Y. Qiao, B.P.F. Lelieveldt, et al., Integrating spatial-anatomical regularization and structure sparsity into SVM: improving interpretation of Alzheimer's disease classification, NeuroImage 178 (2018) 445–460.
- [25] B. Swiderski, S. Osowski, A. Cichocki, et al., Single-class SVM and directed transfer function approach to the localization of the region containing epileptic focus, Neurocomputing 72 (7-9) (2009) 1575–1583.
- [26] Barjinder Kaur, Dinesh Singh, Partha Pratim Roy, EEG based emotion classification mechanism in BCI, Procedia Comput. Sci. 132 (2018) 752–758.
- [27] P.éicles B.C. Miranda, Ricardo B.C. Prudêncio, de Carvalho, P.L.F. André, et al., A hybrid meta-learning architecture for multi-objective optimization of SVM parameters, Neurocomputing 143 (2014) 27–43.
- [28] C. Ieracitano, N. Mammone, A. Bramanti, et al., A convolutional neural network approach for classification of dementia stages based on 2D-Spectral representation of EEG recordings, Neurocomputing 323 (2019) 96–107.
- [29] R. Kottaimalai, M.P. Rajasekaran, V. Selvam, et al., EEG signal classification using principal component analysis with neural network in brain computer interface applications. 2013 IEEE International Conference on Emerging Trends in Computing Communication and Nanotechnology (ICECCN), IEEE, 2013, pp. 227–231.
- [30] S.K. Agarwal, S. Shah, R. Kumar, Classification of mental tasks from EEG data using backtracking search optimization based neural classifier, Neurocomputing 166 (2015) 397–403.
- [31] Z. Wang, R.M. Hope, Z. Wang, et al., An EEG workload classifier for multiple subjects, 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 6534–6537.
- [32] L. Mingfei, W. Wei, G. Zhenghui, et al., Deep learning based on Batch Normalization for P300 signal detection, Neurocomputing 275 (2018) 288–297.
- [33] Y. Ke, H. Qi, F. He, et al., An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task, Front. Hum. Neurosci. 8 (2014) 703.
- [34] C.L. Baldwin, B.N. Penaranda, Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification, NeuroImage 59 (1) (2012) 48–56.
- [35] L.D. Garrett, B.W. Anderson, M.H. Thaut, et al., Comparison of linear, nonlinear, and feature selection methods for EEG signal classification, IEEE Trans. Neural Syst. Rehabil. Eng. 11 (2) (2013) 141–144.
- [36] K. Cai, B.P. Alalibo, W. Cao, et al., Hybrid approach for detecting and classifying power quality disturbances based on the variational mode decomposition and deep stochastic configuration network, Energies 11 (11) (2018) 3040.
- [37] J.C. Christensen, J.R. Estep, G.F. Wilson, et al., The effects of day-to-day variability of physiological data on operator functional state classification, NeuroImage 59 (1) (2012) 57–63.
- [38] W. Cao, X. Wang, Z. Ming, et al., A review on neural networks with random weights, Neurocomputing 275 (2017) 278–287.
- [39] D. Wang, M. Li, Stochastic configuration networks: fundamentals and algorithms, IEEE Trans. Cybern. 47 (10) (2017) 3466–3479.
- [40] D. Wang, C. Cui, Stochastic configuration networks ensemble for large-scale data analytics, Inf. Sci. 417 (2017) 55–71.
- [41] D. Wang, M. Li, Robust stochastic configuration networks with kernel density estimation for uncertain data regression, Inf. Sci. 412-413 (2017) 210–222.
- [42] J.R. Comstock, R.J. Arnegard, The Multi-attribute Task Battery for Human Operator Workload and Strategic Behavior Research, 1992.
- [43] C. Feng, X. Wan, S. Liu, Influence of different attention allocation strategies under workloads on situation awareness, Acta Astronaut. Sin 31 (2020), 123307.
- [44] S.G. Hart, L.E. Staveland, Development of NASA-TLX (Task Load Index): results of empirical and theoretical research, Adv. Psychol. 52 (1988) 139–183. North Holland.
- [45] D.M. Greer, J. Yang, P.D. Scripko, et al., Clinical examination for prognostication in comatose cardiac arrest patients, Resuscitation 84 (11) (2013) 1546–1551.
- [46] J. Zhang, L. Pang, X. Cao, et al., The effects of elevated carbon dioxide concentration and mental workload on task performance in an enclosed environmental chamber, Build. Environ. 178 (2020), 106938.

- [47] D. De Waard, K.A. Brookhuis, The Measurement of Drivers' Mental Workload, 1996.
- [48] J. Rueb, M. Vidulich, J. Hassoun, in: Establishing Workload Acceptability: an Evaluation of a Proposed KC-135 Cockpit Redesign, Proceedings of the Human Factors Society 36th Annual Meeting, Santa Monica, 1992, pp. 17–21.
- [49] X.R. Wanyan, D.M. Zhuang, Measurement and Application of Mental Workload in Pilots, Science Press, Beijing, 2014, pp. 12–14, in Chinese.
- [50] L. Pion-Tonachini, K. Kreutz-Delgado, S. Makeig, ICLabel: An automated electroencephalographic independent component classifier, dataset, and website, *NeuroImage* 198 (2019) 181–197.
- [51] E.V. Chemerisova, M.S. Atanov, I.N. Mikheev, et al., Classification of verbal and mathematical mental operations based on the power spectral density of eeg, *Psychology* 15 (2) (2018) 268–278.
- [52] G.F. Wilson, C.A. Russell, Operator functional state classification using multiple psychophysiological features in an air traffic control task, *Hum. Factors J. Hum. Factors Ergon. Soc.* 45 (3) (2003) 381–389.
- [53] H. Qu, Y. Shan, Y. Liu, et al., Mental workload classification method based on EEG independent component features, *Appl. Sci.* 10 (9) (2020) 3036.
- [54] Z. Yin, J. Zhang, Cross-session classification of mental workload levels using EEG and an adaptive deep learning model, *Biomed. Signal Process. Control* 33 (2017) 30–47.
- [55] V. Menon, S.M. Rivera, C.D. White, G.H. Glover, A.L. Reiss, Dissociating pre-frontal and parietal cortex activation during arithmetic processing, *Neuroimage* 12 (4) (2000) 357–365.
- [56] A. Gundel, G. Wilson, Topographical changes in the ongoing EEG related to the difficulty of mental tasks, *Brain Topogr.* 5 (1) (1992) 17–25.
- [57] M. Akin, M. Kurt, N. Sezgin, M. Bayram, Estimating vigilance level by using EEG and EMG signals, *Neural Comput. Appl.* 17 (3) (2008) 227–236.
- [58] S.W. Chuang, L.W. Ko, Y.P. Lin, R.S. Huang, T.P. Jung, C.T. Lin, Co-modulatory spectral changes in independent brain processes are correlated with task performance, *Neuroimage* 62 (3) (2012) 1469–1477.
- [59] R.N. Youngworth, B.B. Gallagher, B.L. Stamper, An overview of power spectral density (PSD) calculations, in: Optical Manufacturing and Testing VI, 5869, International Society for Optics and Photonics, 2005, 58690U.
- [60] Liu N H, C.Y. Chiang, H.C. Chu, Recognizing the degree of human attention using EEG signals from mobile sensors, *Sensors* 13 (8) (2013) 10273–10286.
- [61] G. Cheng, J. Han, P. Zhou, et al., Multi-class geospatial object detection and geographic image classification based on collection of part detectors, *ISPRS J. Photogramm. Remote. Sens.* 98 (1) (2014) 119–132.
- [62] Z. Yin, J. Zhang, Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights, *Neurocomputing* 260 (2017) 349–366.
- [63] J.D. Rodriguez, A. Perez, J.A. Lozano, Sensitivity analysis of k-Fold cross validation in prediction error estimation, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (3) (2010) 569–575.
- [64] K. Polat, Salih. Güneş, Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals, *Expert Syst. Appl.* 34 (3) (2008) 2039–2048.
- [65] K. Zeng, G. Ouyang, H. Chen, et al., Characterizing dynamics of absence seizure EEG with spatial-temporal permutation entropy, *Neurocomputing* 275 (2017) 577–585.
- [66] G.F. Wilson, C.A. Russell, Real-time assessment of mental workload using psychophysiological measures and artificial neural networks, *Hum. Factors* 45 (4) (2003) 635–644.
- [67] H. Qu, S. Fu, L. Pang, et al., Rapid temperature prediction method for electronic equipment cabin, *Appl. Therm. Eng.* 138 (2018) 83–93.