IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 December 2024, accepted 12 January 2025, date of publication 15 January 2025, date of current version 22 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3530091

RESEARCH ARTICLE

Enhanced Cognitive Load Detection in Air Traffic Control Operators Using EEG and a Hybrid Deep Learning Approach

YUEYING ZHOU[1,2], (Member, IEEE), JUNJI JIANG[1], LIJUN WANG[1], SHANSHAN LIANG[1], AND HAO LIU[1]

[1]School of Mathematics Science, Liaocheng University, Liaocheng 252000, China
[2]Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Corresponding author: Yueying Zhou (zhouyueying92@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62406131, and in part by the Fundamental Research Funds for the Central Universities under Grant NJ2024029.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Nanjing University of Aeronautics and Astronautics, and performed in line with the Helsinki Declaration.

ABSTRACT The automatic and effective detection of cognitive load for air traffic control (ATC) operators through electroencephalography (EEG) signals provides a covert and objective method for enhancing ATC safety. Nevertheless, the extant paradigm is limited to simple cognitive tasks and lacks real-world scenarios. In this study, a cognitive load-elicited experiment was therefore designed to record the EEG data of eight ATC operators under four distinct simulation scenarios, ascertaining whether they experienced varying degrees of workload. Subsequently, the collected EEG signal was preprocessed. We then used one hybrid deep learning model based on the convolutional layers and a self-attention mechanism to extract the pertinent EEG features. In conjunction with multi-layer perceptron, we decoded cognitive load state into low, high, overload, and special. The experimental results demonstrated that EEG could serve as a reliable measure for predicting ATC load, with an average accuracy of 88.76% and a peak accuracy of 99% at the single-subject level. Additionally, it highlighted the critical role of the frontal regions in decoding cognitive load. This study serves to enhance the efficacy of personalized EEG decoding for ATC operators, furnishing evidence for the feasibility of developing an intelligent load-detecting system.

INDEX TERMS Air traffic control, EEG, load detection, self-attention, CNN, brain mechanisms.

I. INTRODUCTION

C ognitive load detection represents a significant paradigm and typical application of passive brain-computer interface [1], [2], [3], [4], [5]. It serves as the pivotal link in cognitive state perception within the domain of human factors [6] and forms the foundation for a range of application areas, such as cognitive-affective probing, and cognitive modulation [7], [8]. Cognitive load is used to describe the degree of resource consumption or occupancy of the operator's brain for information processing during the execution of human-computer interaction tasks [9]. Prolonged exposure to an extreme load state can result in a range of physiological and psychological health issues, as well as posing significant risks to the safety of operating systems [10], [11], [12], [13], [14], which can potentially lead to human-caused accidents.

Air traffic control (ATC) operators bear the responsibility of monitoring and directing flights through the ATC system and making decisions based on their experience [15], [16], [17], [18]. In the modern work environment, they are confronted with intricate machine interactions, onerous tasks, and the necessity for highly concentrated attention. This has

The associate editor coordinating the review of this manuscript and approving it for publication was Filbert Juwono.

led to a considerable increase in the mental workload of ATC operators. The decisions made by ATC operators not only impact the efficiency of the air traffic system but also have the potential to influence the accident rate. Consequently, it is imperative to develop sophisticated cognitive load monitoring and detection tools to guarantee the security of air traffic operations [17], provide overload warnings, enhance work efficiency and ultimately improve traffic management.

The ATC operators operate in a real-time, continuous environment. However, the prevailing measurements of load states typically rely on data gathered post-test [10] or following mission interruptions. These approaches are insufficient for the real-time evaluation of an operator's cognitive load state. It is therefore imperative that non-intrusive methods be employed. Such methods would permit the implementation of timely preventive measures when high cognitive loads are discerned [1]. Among various non-intrusive techniques, EEG-based cognitive load detection and recognition is particularly noteworthy for its capacity to provide discrete insights into an operator's cognitive state by delivering objective data without distracting them from their primary tasks [1]. Firstly, EEG enables the covert monitoring of the operator's cognitive state without distracting them from the main task [1]. Secondly, it provides objective data and avoids the influence of subjective judgments of the ATC operators on the results. Furthermore, the high temporal resolution of EEG signals allows for the capture of subtle changes in cognitive processes [18]. This facilitates the understanding of cognitive processing strategies and effects of ATC operators under varying loads.

The current research has concentrated on the simulation of cognitive loads in control scenarios, such as tracking and collision prediction tasks [19], in simplified laboratory environments. Although these experiments have rapid response times, they are unable to fully replicate the complexities inherent to real-world application scenarios. Conversely, it is unlikely that ATC operators will be able to perform the simplifying task in real life [20]. Consequently, it is unclear whether the current research can accurately represent the aforementioned difference in workload in real life.

Accordingly, the objective of this study is to develop a more effective method for simulating real-world control situations and eliciting load states of ATC operators, enabling a more precise reflection of the cognitive load change. The overarching framework proposed in this paper is illustrated in Fig. 1, which outlines the cognitive load decoding and mechanism mining of EEG based on ATC operators' participation in simulator training. The framework encompasses several key elements, including the experimental paradigm, EEG acquisition and preprocessing, the application of the deep decoding model, and the subsequent decoding results and brain activation mapping, along with other forms of brain mechanism mining. This will be achieved by capturing EEG data from experienced operators operating in a real control simulator. The experiment aimed to categorize four types of scenarios based on control tasks of varying difficulty levels, low and high as well as overload and abnormal situation handling. This, in turn, facilitates the identification and response to emergencies on time, thereby enhancing operational safety.

Besides, for the EEG-based cognitive states recognition domain, deep learning methods have been widely used to obtain more accurate recognition results than traditional machine learning methods. The deep learning models include convolutional neural network (CNN) models (such as compact CNN [21], [22], [23], 3DCNN [24], 4DRCNN [25]), Transformer-based models [26], [27], [28], [29], and hybrid framework [30]. Among them, a hybrid deep learning model Conformer [30], integrating the CNN and a self-attention mechanism, can capture local and global representations in a unified framework and achieve better recognition performance in decoding emotion and motor imagery tasks.

To achieve the above goals, we designed a cognitive load-elicited experiment to record the EEG data of eight ATC operators working on the simulator under four distinct simulation scenarios, with the explicit aim of ascertaining whether they experienced varying degrees of workload. Subsequently, the preprocessed EEG signal was employed as the model input, and a hybrid deep model was utilized to extract the pertinent EEG features. In conjunction with a multi-layer perceptron, the cognitive load state is decoded under low, high, overload, and special circumstances. The experimental results demonstrated that EEG is an effective and reliable measure for predicting cognitive load in ATC operators, achieving an average accuracy of 88.76% in decoding cognitive load and a peak accuracy of 99% at the single-subject level. The findings also emphasized the importance of the frontal regions in this process. This research strengthens the validity and applicability of load detection results and provides a foundation for developing intelligent closed-loop cognitive load perception system [19]. Such systems could integrate portable physiological signals like EEG into automated decision-making. This could help to inform the future development of adaptive systems in complex working environments [19].

The remainder of the paper is arranged as follows. Section II outlines the load-evoking experimental design for ATC operators, the EEG acquisition procedure, and the hybrid deep Conformer model for load detection. Sections III and IV present the findings of the load detection and offer further discussion, while Section V provides a concluding summary.

II. MATERIALS AND METHODS

The materials and methods of the suggested work are divided into five subsections. The first subsection introduces the ATC subjects involved in the experiment, the second subsection highlights the workload-evoking paradigm and design, the third subsection describes the EEG data collection and preprocessing, the fourth subsection delineates the decoding model of mixed CNN and self-attention, and the final subsection presents the experimental setup.

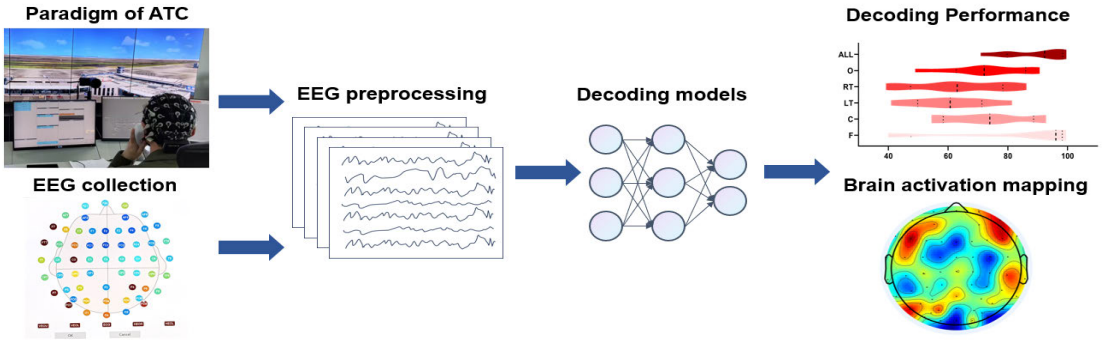


FIGURE 1. EEG-based cognitive load decoding and mechanism mining of ATC system in simulator training.

A. SUBJECTS

Eight male apron controllers (aged 26 to 38 years with a control license) were recruited for the experiment. All participants were male to match the actual gender distribution in the ATC system. The decision was primarily based on the availability of controllers and logistical constraints. Besides, we aim to minimize the inter-subject differences caused by gender. Participants were right-handed, had a bachelor's degree, and were employed in a related job after graduation. They had normal or corrected-to-normal vision. Participants were instructed to abstain from alcohol, and caffeine, and to ensure eight hours of sleep before the experiment. Written informed consent, signed by all participants, was obtained before the experiment.

B. PARADIGM FOR ELICITING LOAD

The ATC simulator is from Nanjing University of Aeronautics and Astronautics. The system can simulate controlled airspace in compliance with Civil Aviation Administration requirements. It facilitates flight planning and enables trainers to engage in experimental projects such as radar recognition, radar guidance, Approach control, conflict detection and resolution during flight, as well as aircraft sorting in radar-controlled Approach, and handling of special flight situations. Wireless communication with the virtual captain creates an environment approximating real-world conditions.

Before the experiment, two student traffic controllers were recruited for a pilot study to select exercises on the control simulator. Based on the NASA-TXL scores of the student participants, the experiment scenarios were categorized into low workload, high workload, overload, and special scenarios. See Table 1 for specific settings of different scenarios.

The EEG load-elicited experiment was divided into four rounds based on varying load levels, with the entire experiment lasting approximately 150-180 minutes. Scenario 1 of the experiment lasts about 30 minutes, while Scenarios 2-4 conclude only 30 minutes during the operation to ensure data consistency.

TABLE 1. Four scenarios that evoke workloads.

Scenarios	Simulator	Evoked Load	RAD	Flights	ARI	Scale
1	PX25-12P02	Low	6/6	12		10.33
2	PX25-17M01	High	9/8	17		28.33
3	TWR401-07-1	Overload	21/8	30		60.67
4	PX25-17M01	Special	9/8	17	✓	78.33

RAD: Ratio of Arrivals to Departures; ARI: Abnormal runway intrusion.

During the experiment, the operators included one apron controller (experimental subject) and one pilot (task completion assistant). All operators were asked to complete a survey questionnaire, gathering details on age, health status, handedness, visual acuity, and other related aspects. We only collected the EEG data and subjective scale of the apron controller. Before the experiment, steps to be completed include: providing training on simulator operation, explaining the experimental procedure, reminding of precautions, completing the information questionnaire, and calibrating the physiological EEG monitoring equipment.

In the experiment, to better assess the workload and fatigue of traffic control officers, the experiments are conducted in the order of Scenarios 4-3-1-2. Actually, in a preliminary pilot experiment, we devised two orders of cognitive load induction. The first order comprised scenarios 1-2-3-4, while the second order included scenarios 4-3-1-2. By collecting and analyzing EEG load data from students undertaking simulated air traffic control tasks under both orders, we were able to ascertain the most effective order for cognitive load induction in the main experiment. Combining power spectral density features and a support vector machine classifier with 10-fold cross-validation, the classification results indicated that the order 2 exhibited better separability. Consequently, the main experiment in our study was conducted using order 2 for eliciting cognitive load.

1) SCENARIO 1 LOW WORKLOAD

The simulator was used to practice PX25-12P02 on the ZWWW map. The practice session was interrupted after

approximately 30 minutes, with a departure-to-arrival flight ratio of 6:6 and a total of 12 flights.

2) SCENARIO 2 HIGH WORKLOAD

Used a simulator to practice PX25-17M01 on the ZWWW map. The exercise was interrupted for about 30 minutes, with a departure/arrival ratio of 9/8 and a total of 17 flights.

3) SCENARIO 3 OVERLOAD

It employed a simulator to practice TWR401-07-1 on the ZWWW map. The practice session was interrupted after approximately 30 minutes, with a departure-to-arrival flight ratio of 21:8 and a total of 30 flights.

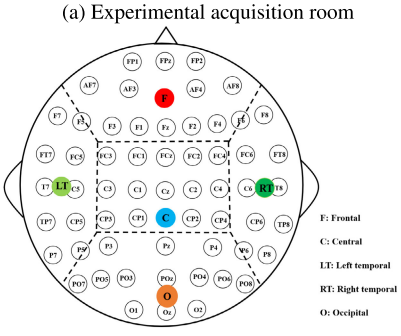
4) SCENARIO 4 SPECIAL CASE

It mirrored that of Scenario 2, with the addition of randomly simulating an abnormal runway intrusion situation from the pilot-in-command position during the exercise.

C. EEG DATA COLLECTION

The experiment took place in a quiet, enclosed civil aviation simulation training laboratory. The subjects sat in comfortable chairs approximately 60 centimeters from the screen display. The stimuli were presented and controlled by the regulatory simulation software. The experimental data collection process is depicted in Fig. 2.

(a) Experimental acquisition room



(b) EEG channel location and partition

FIGURE 2. The experimental data collection process. Here, we divide the 59 EEG channels into five brain regions.

The wireless portable EEG signal acquisition system, NeuSen.W64 (Neuracle, China), was used to record the EEG data from controllers at a 1000 Hz sampling rate. The equipment includes an EEG cap, amplifier, marker

TABLE 2. The number of EEG samples for four scenarios that evoke load.

Sub	Rest	Low	High	Overload	Special
1	123	899	902	902	899
2	185	891	903	901	899
3	150	901	903	1061	906
4	151	903	903	902	904
5	151	909	902	916	893
6	149	908	907	912	918
7	154	904	903	905	901
8	153	902	901	909	931
Sum	1216	7217	7224	7408	7251
Mean	152	902.125	903	926	906.375

synchronization device, and rechargeable battery. According to the 10-20 electrode layout system, EEG signals were recorded from 59 electrodes (refer to Fig. 2(b) for specific electrode placement), with the reference electrode being CPz and the ground electrode being AFz. Throughout the experiment, the impedance of all electrodes was maintained below 5000 ohms.

The preprocessing step of the EEG signal was performed using EEGLAB software [31]. The preprocessing steps applied to the collected EEG data included the following: (1) re-referencing to establish the potential difference between each electrode and the reference electrode; (2) applying a bandpass filter (0.1-70Hz) and a 50Hz notch filter to eliminate power line interference and electromyographic artifacts, thereby preserving the desired frequency range of the EEG signal; (3) segmentation based on predefined labeling information, retaining specific 2-second durations of EEG signal data within each segment, and baseline correction to mitigate data drift effects; (4) artifact removal using independent component analysis (ICA), where components were labeled and subsequently removed using ADJUST [32] and ICLabel [33] plug-in to identify and discard artifacts or electrooculogram components detected in the analysis. Furthermore, we adopt a z-score normalization to reduce the data fluctuation [30], by calculating the mean and standard deviation of training data and applied for testing data.

The final resting state data comprises 1216 samples, while the samples for the different load-induced tasks are 7217, 7224, 7408, and 7251 respectively, with each sample lasting for 2 seconds. Table 2 details the sample sizes for the resting state and load-induced tasks across individual subjects.

D. DECODING MODEL FOR ATC LOAD

Conformer [30] was used to condense local and global representation to decode motor imagery and emotional tasks. It consists of a convolution module, a Transformer Encoder module, and a classifier module. Specifically, taking raw EEG signals as model input, the convolution module learns low-level local features from time and EEG channels by using one-dimensional temporal and spatial convolutional layers. Then, the Transformer Encoder module with multi-head attention is attached to extract long-term

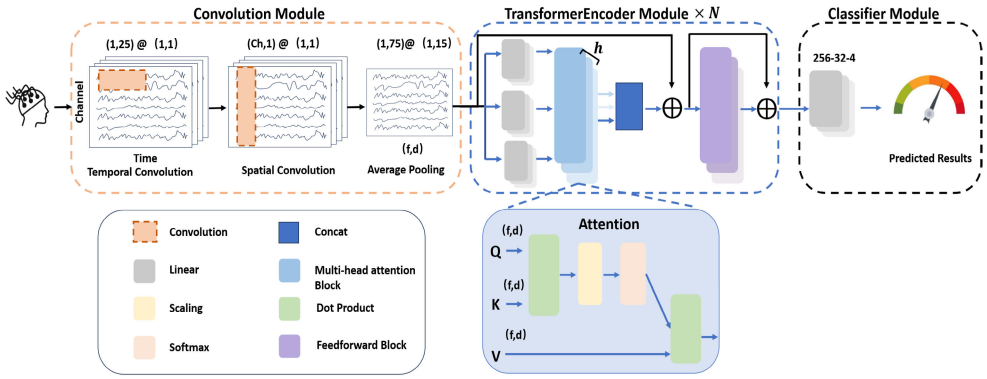


FIGURE 3. The EEG decoding model used in the experiment.

global temporal representation from local characteristics. Finally, fully connected layers are constructed as a classifier, which classifies EEG features and outputs the predicted labels of load states. The general framework and parameter settings are shown in Fig. 3.

1) CONVOLUTION MODULE

The convolution module has been designed under the principles outlined in [21] and [22]. It comprises two one-dimensional spatiotemporal convolution layers, which serve to separate the two-dimensional convolution operator. The model input is shaped as (S, Ch, To), where S is the training number or testing number of EEG samples, and Ch and To are equal to the number of channels and time points of each EEG sample, respectively. The initial layer comprises k kernels of dimensions (1, 25) with a step size of (1, 1), indicating that the operation is conducted in the time aspect. The second layer retains k kernels of size (Ch, 1) with a step size of (1,1), serving as a spatial filter to extract the interactions between various EEG channels. Then, batch normalization is employed to accelerate the training process and mitigate the overfitting issue. The activation function used an exponential linear unit, as recommended in [22]. The third layer comprises average pooling along the time dimension with (1,75) kernel size and (1,15) step size. This pooling layer serves to smooth temporal information, thereby avoiding overfitting and reducing computational complexity. The hyperparameter k is set to 40. Subsequently, the learned features of convolution layers are rearranged by compressing channel dimensions and transposing convolution channel dimensions with temporal dimensions. As such, the feature channel at each time point can be viewed as tokens for subsequent module. The output of the convolution layers is defined as $X \in \mathbb{R}^{f \times d}$, where f is the feature-length, d is the feature-dimension.

2) SELF-ATTENTION MODULE

In this module, self-attention is employed to capture the global temporal dependencies of EEG features, thereby addressing the limitations of the convolution module's restricted receptive field.

The tokens generated by the preceding module are transformed linearly into isomorphic triples, designated as queries (Q), keys (K), and values (V). The Q, K, V are calculated in accordance with learnable weight matrices $W^q \in \mathbb{R}^{d \times d}$, $W^k \in \mathbb{R}^{d \times d}$, $W^v \in \mathbb{R}^{d \times d}$, and obtain $Q = XW^q \in \mathbb{R}^{f \times d}$, $K = XW^k \in \mathbb{R}^{f \times d}$, $V = XW^v \in \mathbb{R}^{f \times d}$. The introduction of a scaling factor serves to prevent the phenomenon of vanishing gradients and to guarantee training stability. Subsequently, the output is processed through a softmax function, which generates the weighting matrix. The weighting matrix is applied to V using a dot product [34]. The aforementioned process of single-head attention is formalized as follows:

$$Att\,(Q,\ K,\ V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V. \tag{1}$$

Given that multi-head self-attention can grasp long-term dependencies of data, we employ it to enlarge representational diversity [28]. The input features are divided into subspaces, each comprising multiple heads. The function of each subspace is to facilitate the learning of attention weights within its respective space. Heads in various subspaces communicate with one another to exchange information regarding the allocation of attention between the subspaces. The tokens are partitioned into h segments, with each segment undergoing independent processing by the self-attention module. The outputs generated by each head are subsequently concatenated to produce the final one.

Assuming we use h heads. As such, the input features are evenly partitioned into h segments, denoted as $x_m \in \mathbb{R}^{f \times \frac{d}{h}}, 1 \leq m \leq h$. Q_m, K_m, $V_m \in \mathbb{R}^{f \times \frac{d}{h}}$ are derived from the linear transformation of the partitioned tokens in the m-th head. We calculate its corresponding Q_m, K_m,V_m by $Q_m = X_m W_m^q$, $K_m = X_m W_m^k$, $V_m = X_m W_m^v$. The self-attention H_m towards subspace m is computed as follows:

$$H_m = Softmax\left(\frac{Q_m K_m^T}{\sqrt{d}}\right)V_m. \tag{2}$$

Multi-head self-attention concatenates all H_m:

$$H_{msa} = Concat\left(H_1, \ldots H_m, \ldots H_h\right). \tag{3}$$

Furthermore, two fully connected (FC) feed-forward layers have been incorporated into the model to enhance its fitting capacity. The dimensions of both the input and output have been maintained throughout this process. The self-attention mechanism has been iterated N times throughout the module.

3) LOAD CLASSIFIER MODULE

At last, a classifier module comprising three FC layers is utilized, resulting in a C-dimensional output vector upon application of the softmax function. The framework employs the cross-entropy as loss function, defined as follows:

$$\mathcal{L} = -\frac{1}{N_b}\sum_{i=1}^{N_b}\sum_{c=1}^{C} y_{ic} log\left(\hat{y}_{ic}\right). \tag{4}$$

where N_b denotes EEG samples number in one batch, y and $\hat{y}$ represent true and predicted labels, respectively, C indicates EEG load categories.

E. EXPERIMENTAL SETUP

The workload classification experiments are run upon the Conformer framework, supported by Pytorch library. We focus on recognizing the subject-dependent workload. The training data and testing data ratio is fixed at 7:3 for each subject. The decoding model was trained with the following parameters: batch size = 8, epoch = 100, Adam optimizer (learning rate = 0.0002), beta 1 = 0.5, and beta 2 = 0.999. We execute self-attention 6 times with 10 heads. By averaging the results of the last 10 epochs, we obtain the recognition score for each subject.

The experiment must acknowledge the cognitive load of a single controller in four distinct scenarios: low, high, overload, and special case. To offer a comprehensive assessment, three metrics are employed in this paper, including accuracy (ACC), macro-averaged F1 score (F1), and Cohen Kappa coefficient (Kappa). Towards i-class, suppose TP_i is true positive, TN_i is true negative, FN_i is false negative, FP_i is false positive. A indicates samples number for one subject.

The detailed definitions of the overall metrics are as follows:

$$ACC = \sum_{i=1}^{C} TP_i/A. \tag{5}$$

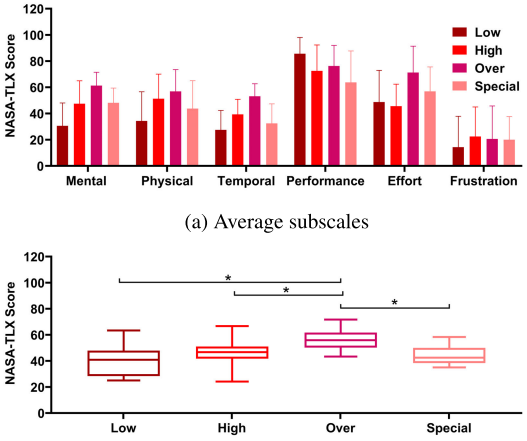
$$F1 = \frac{1}{C}\sum_{i=1}^{C}\left(2 \times pre_i \times sen_i/\left(pre_i + sen_i\right)\right). \tag{6}$$

$$Kappa = \left(ACC - P_e\right)/\left(1 - P_e\right). \tag{7}$$

here, $pre_i = TP_i/\left(TP_i + FP_i\right)$, $sen_i = TP_i/\left(TP_i + FN_i\right)$, P_e is the random probability of producing a consistent outcome by chance.

III. RESULTS

The outcomes of the proposed study are organized into three sections. The first section highlights the analysis of subjective scale data for ATC operators. The second presents experimental results of decoding cognitive loads using a hybrid model. The third explores the influence of different brain regions on decoding performance.



(a) Average subscales

(b) Box-plot Figure

FIGURE 4. The analysis of NASA-TLX scores.

A. SUBJECTIVE ANALYSIS

The ATC load was first evaluated by the NASA-TLX scale [35], [36]. The scale comprises six subscales: mental demand, temporal demand, physical demand, effort, performance, and frustration. Subjects were requested to provide a rating on a scale of 0 to 100 for each subscale. We first give the average subscales under four scenarios in Fig. 4(a). As we can see the mental, physical, and temporal demands increased in line with the rise in task complexity. This resulted in a corresponding decline in task performance, while effort and frustration have no clear tendency.

The overload scenario needs more effort, and both four scenarios have lower frustration scores. We further calculate the final NASA-TLX values for four load-elicited scenarios and display the average NASA-TLX scores of all participants in the box-plot Fig. 4(b) using a one-way analysis of variance with Tukey's Multiple comparison test. Statistical analysis reveals a significant increase in NASA-TLX scores as task difficulty rises, and Low versus over, high versus over, and over versus special, are notably different (p<0.05).

B. EEG DECODING RESULTS

Table 3 presents the accuracy, weighted F1 score, and Kappa results for Conformer on the dataset. As observed, the average decoding performance across subjects is 88.763% accuracy, 88.129% F1 score, and 0.849 Kappa. Additionally, significant variations in decoding performance are evident among subjects. Subjects 1, 2, and 4, 5 demonstrated decoding accuracy exceeding 94%, whereas subjects 3 and

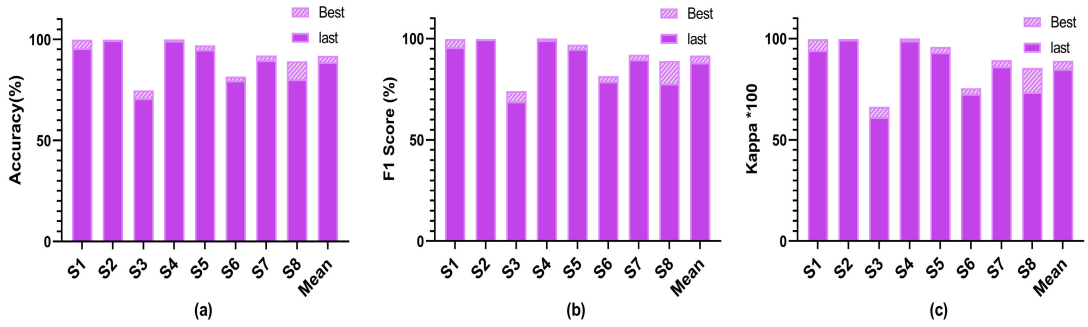


FIGURE 5. The comparison between the average results of the last 10 epochs and best results of (a) accuracy, (b) F1 score, and (c) Kappa results. The average results are expressed in solid dark purple, and the best results are twill.

TABLE 3. The decoding results for ATC task.

Sub	ACC (%)	F1 (%)	Kappa
1	95.716	95.763	0.943
2	99.638	99.638	0.995
3	70.909	69.077	0.613
4	99.261	99.263	0.990
5	94.926	94.923	0.932
6	79.616	78.896	0.727
7	89.723	89.628	0.863
8	80.311	77.847	0.736
Mean	88.763	88.129	0.849
STD	10.621	11.445	0.1417

TABLE 4. The decoding results for online experiment.

Sub	ACC (%)	F1 (%)	Kappa
1	79.873	77.405	0.733
2	84.683	82.844	0.796
3	65.247	61.148	0.541
4	97.993	97.883	0.973
5	84.200	84.163	0.789
6	92.303	92.024	0.897
7	82.087	81.883	0.761
8	68.203	64.421	0.578
Mean	81.824	80.221	0.758
STD	11.015	12.523	0.146

6 showed inferior performance. These findings suggest that the employed decoding model achieves satisfactory results within subjects.

We also record the best performance achieved by each participant and compare it with the average results from the last 10 epochs. As the results show in Fig. 5, the accuracy of participants 2, 4, 5, 6, and 7 matches their respective best performances and the last 10 average values very well. Participants 1, 3, and 8 show more disparities between their best and last 10 epoch results. Generally speaking, the average best performance exceeded 3% in last 10 epochs. Similar trends are observed in the F1 score and Kappa.

We conducted an online experiment on a per-subject basis. To emulate the online process, 2000 samples were randomly selected for training and 600 samples for testing for each subject. At the outset of the experiment, solely the training data were incorporated into the model training and an offline model was constructed. The data were randomly arranged five times to simulate five online tests, and finally the average of the five results was taken as the online result for that subject. The decoding results is listed in Table 4. As we can see, the emulated online results are lower than offline scenarios but also has achievable results. The runtime is 0.016 seconds for decoding each testing sample, which facilitates the real-time decoding performance for ATC task.

C. EFFECT OF BRAIN REGIONS

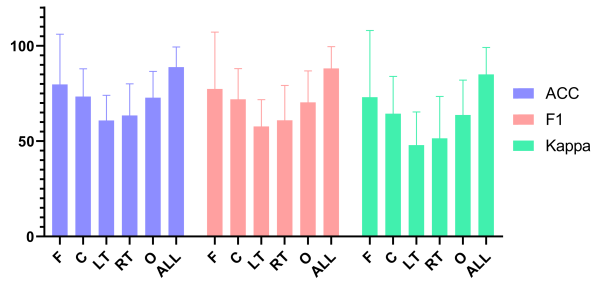


FIGURE 6. Contribution of different brain regions to cognitive load decoding of controllers.

A reduction in the number of channels in an EEG-based cognitive load decoding system is a beneficial approach that not only decreases preparation time and running costs but also improves the system's user-friendliness. This subsection examines the impact of reducing channel numbers on decoding performance. Five distinct channel configurations according to the common brain regions, namely, the frontal lobe, central lobe, left and right temporal lobes, and occipital lobe [7], were employed in the experiments. The detailed division of these brain regions is shown in Fig. 2. During the experiment, there were 14, 14, 8, 8, and 15 electrode channels in the frontal, central, left and right temporal, and occipital lobes of the brain, respectively.

Fig. 6 provides a comprehensive overview of performance metrics (ACC, F1 score, and Kappa) concerning the reduced number of channels. It can be observed that the performance decreases significantly as the channel number is reduced, particularly when using 8 channels from the left and right temporal lobes, respectively. The results demonstrate that comparable results can be obtained from the frontal lobe, followed by the central and occipital lobes, in comparison to using all 59 channels.

IV. DISCUSSION

In the discussion section, we present brain activation mappings across varying levels of cognitive load during the simulated control task. We also include visual analyses of decoding results, comprising confusion matrices for each subject and feature visualization. Finally, we provide an overview of the shortcomings and outlook of this paper.

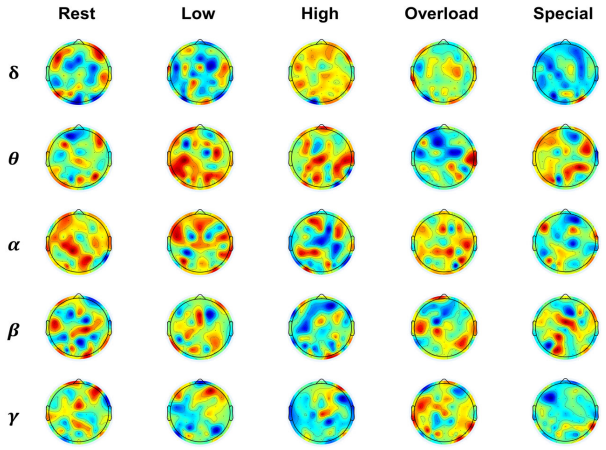


FIGURE 7. PSD distribution of five cognitive states under different frequency bands.

A. BRAIN ACTIVATION MAPPING

The present study employed a psychophysiological approach to investigate the brain activation patterns associated with different cognitive states and frequency bands. The EEG data were analyzed in terms of delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ) bands. The average PSD of 8 subjects for each task state was calculated and normalized to [0,1], resulting in a PSD matrix for each state. These matrices were converted to channels by band format for visualization. The resulting PSD distributions are shown in Fig. 7. We have used the paired t-test to make the statistical analysis between the different load states. The statistical analysis is given in Table 5 with a fixed p-value of $p < 0.05$. The alpha (8-13 Hz) and beta bands (13-30 Hz) have been reported to be associated with changes in workload [4], indicating that the brain is in an engaged and highly conscious state, respectively. In the alpha band, the low vs special and high vs special comparisons yield significant differences. Similarly, in the beta band, the low vs special comparison produces a significant difference. When EEG channels are partitioned as illustrated in Fig. 2(b), as we can see, the PSD feature activation patterns in the frontal and occipital brain regions exhibit distinct patterns between high and low load states. Additionally, in the delta band, the low vs rest1, high, and over comparisons yield significant differences.

TABLE 5. Statistical analysis of brain activity patterns across various workload conditions.

Band	Rest1-L	L-High	L-O	L-Spec	High-O	High-Spec	O-Spec
Delta	*	*	*				
Theta			*				
Alpha				*		*	
Beta				*			
Gamma					*		

Note: * denotes the significant difference. L: Low, O: Over.

B. CONFUSION MATRIX

To comprehensively and intuitively analyze the classification results of the model on the test set, we draw the confusion matrix of each subject, which is used to predict the last epoch of test data. The value at the intersection of the i-th row and j-th column represents the percentage that the real category i is divided into category j by the model. For example, in Fig. 8(a), 75.34% of the low load EEG samples of subject 1 are predicted correctly, and 24.66% of the samples are misclassified to be in the high load category. We can see from the confusion matrix that the high load and overload can be correctly recognized, while low and special cases have a 24%-27% possibility to misclassify in the high load.

For a low load state, subjects 2, 4, 5, and 8 are easy to distinguish, and the recognition accuracy of this category is higher than 92%. For the high load state, except for subjects 3 and 6, the other subjects can be distinguished significantly. In the case of overload, all subjects can be significantly distinguished except for subjects 3 and 8. For special cases, except for subjects 1 and 8, the other subjects can be distinguished significantly.

Based on the above results, it can be found that although different controllers have different performances in recognizing cognitive load state, the mixed depth model used can still construct an effective personalized decoding model. In the future, it is imperative to consider more advanced deep learning methods, to reduce decoding differences of subjects and build a more robust recognition model.

C. FEATURE VISUALIZATION

To visualize the distribution of feature representations learned using the model, we project potential feature representations onto a two-dimensional plane using t-distributed stochastic neighborhood embedding (t-SNE), as illustrated in Fig. 9. It is a nonlinear statistical downscaling method designed to maintain the structure of the data in a low-dimensional space [37]. One subject was randomly selected from the dataset. Following training with and without the conformer model, the feature distribution of the subject is shown in Fig. 9. Fig. 9(a) and (c) represent the original feature

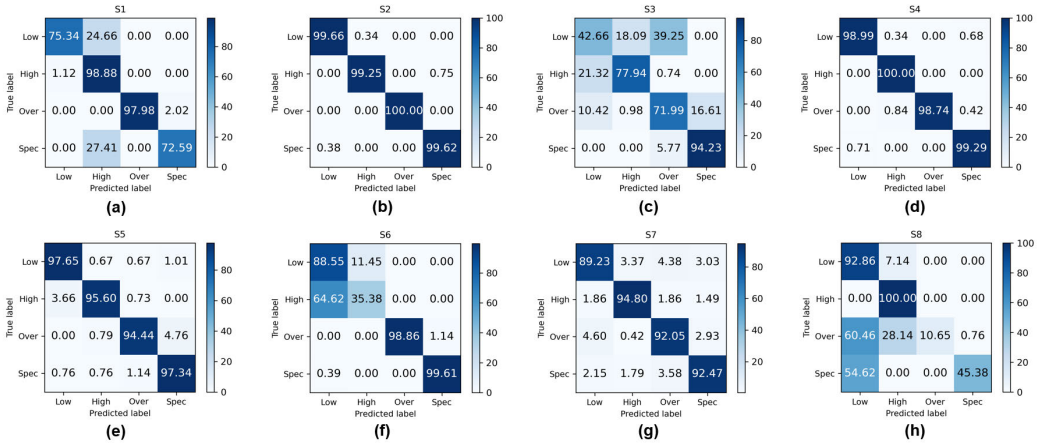


FIGURE 8. Confusion matrix results of each ATC controller, (a) to (h) denotes subjects 1 to 8, respectively.

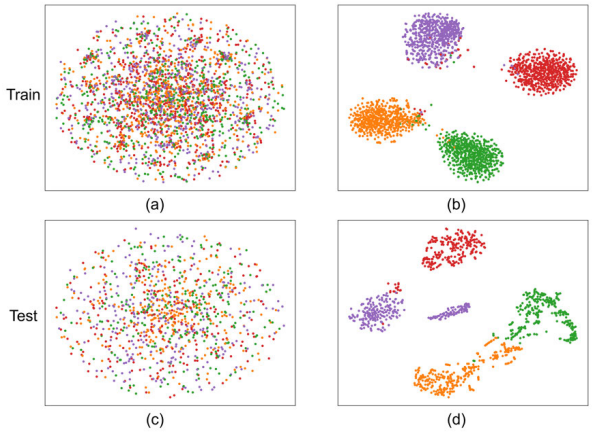


FIGURE 9. Feature visualization by t-SNE demonstrates the impact of introducing deep decoding models for feature learning, (a) and (c) are the original features without decoding, and (b) and (d) are the decoded features.

distribution, while (b) and (d) show the decoded feature distribution. Different colors are used to represent different categories. As illustrated in Fig. 9 (a) and (c), samples belonging to different categories exhibit considerable overlap in both the training and test data, in the absence of the deep model, rendering them challenging to distinguish. In Fig. 9 (b) and (d), the use of decoding model results in a reduction in the distance between samples of the same category, facilitating the formation of clusters. Additionally, the distance between the feature distributions of different categories increases, thereby enhancing the clarity of the class boundaries, which in turn facilitates the classification process.

D. ABLATION OF PREPROCESSING AND STANDARDIZATION

We further explored the effectiveness of the preprocessing steps (e.g., ICA artifact removal) and standardized steps on the decoding results. We have used the EEG data with artifact removal and without artifact removal, standardized and no standardized, to conduct ablation studies using the same workload classification task.

TABLE 6. The ablation experiment.

ICA denoise	Standardize	ACC (%)	F1 (%)	Kappa
x	✓	66.557	63.577	0.555
✓	x	85.393	84.761	0.803
x	x	68.385	66.165	0.579
✓	✓	88.763	88.129	0.849

The ablation experimental results are reported in Table 6, from which we can see that the removal of denoising or standardization operation results the degraded decoding performance. The removal of ICA denoising has a larger impact on the load decoding results, while the removal of standardization operation has less impact on the decoding results. The full model with ICA denoising or standardization obtains the highest classification performance, illustrating the effectiveness of the overall model.

E. LIMITATIONS

This paper has yielded valuable insights into the EEG cognitive load detection. However, there are still a few areas that require further enhancement. Firstly, the number of subjects included in the work was insufficient and lack female subjects, and the EEG collection equipment used was portable. We acknowledge that including female participants would enhance the generalization and applicability of the findings. Future studies could increase the number and gender of subjects to better reflect the gender distribution of the broader ATC workforce, and explore the possibility of using fewer channels for cognitive load decoding [38], [39]. Secondly, the present study employed a block design approach [40], whereby the overall EEG data of the operator was collected over 30 minutes and subsequently segmented following pre-processing. Although such a design can elicit

different levels of loading in subjects, it may limit the temporal dimension of the EEG data and ignore multiple shifts. Future research will aim to extend this work by investigating how cognitive load changes across longer durations and multiple shifts, and how these variations might affect both performance and physiological responses over time. Finally, the current study did not fully explore the effects of temporal and subject differences on the ATC load decoding results [41], [42]. To ameliorate the generalization of the decoding model, future studies might consider these factors to comprehensively assess their effects.

V. CONCLUSION

The performance and efficacy of ATC systems can be ameliorated by keeping the cognitive states of the operator within an ideal scope [19]. This study is designed to elucidate variations in cognitive load among different operators in complex work environments by creating four simulated ATC tasks of varying difficulty levels. EEG data was collected and a hybrid Conformer model was employed for EEG decoding, to validate its capability in constructing personalized decoding models. The findings indicate that EEG is a sensitive and reliable method for predicting operators' cognitive loads, with an average decoding accuracy of 88% and a peak accuracy of 99% at the individual level. Furthermore, the results illustrate the frontal lobe is integral to the decoding of cognitive load, which is implicated in the processing of emotion, personality, movement, intelligence, and speech. This research is designed to advance the intelligent closed-loop system for sensing cognitive load, to inform the layout of prospective adaptive system. The system will integrate portable physiological signals, such as EEG, into automated decision-making processes, intending to mitigate load fluctuations in complex work environments.

REFERENCES

[1] P. Arico, G. Borghini, G. Di Flumeri, N. Sciaraffa, A. Colosimo, and F. Babiloni, "Passive BCI in operational environments: Insights, recent advances, and future trends," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1431–1436, Jul. 2017.

[2] P. Aricò, G. Borghini, G. Di Flumeri, N. Sciaraffa, and F. Babiloni, "Passive BCI beyond the lab: Current trends and future directions," *Physiological Meas.*, vol. 39, no. 8, Aug. 2018, Art. no. 08TR02.

[3] N. Sciaraffa, D. Germano, A. Giorgi, V. Ronca, A. Vozzi, G. Borghini, G. Di Flumeri, F. Babiloni, and P. Aricò, "Mental effort estimation by passive BCI: A cross-subject analysis," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Mexico, Nov. 2021, pp. 906–909.

[4] Y. Zhou, S. Huang, Z. Xu, P. Wang, X. Wu, and D. Zhang, "Cognitive workload recognition using EEG signals and machine learning: A review," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 3, pp. 799–818, Sep. 2022.

[5] Y. Shao, Y. Zhou, P. Gong, Q. Sun, and D. Zhang, "A dual-adversarial model for cross-time and cross-subject cognitive workload decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 2324–2335, 2024.

[6] Y. Ke, H. Qi, F. He, S. Liu, X. Zhao, P. Zhou, L. Zhang, and D. Ming, "An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task," *Frontiers Human Neurosci.*, vol. 8, p. 703, Sep. 2014.

[7] X. Yang, Y. Peng, Y. Han, F. Li, Q. Zhang, S. Wu, and X. Wu, "Interaction between dynamic affection and arithmetic cognitive ability: A practical investigation with EEG measurement," *IEEE Trans. Affect. Comput.*, vol. 15, no. 3, pp. 1427–1438, Jul. 2024, doi: 10.1109/TAFFC.2023.3347391.

[8] D. Lottridge, M. Chignell, and M. Yasumura, "Identifying emotion through implicit and explicit measures: Cultural differences, cognitive load, and immersion," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 199–210, Apr. 2012.

[9] C. D. Wickens, "Multiple resources and performance prediction," *Theor. Issues Ergonom. Sci.*, vol. 3, no. 2, pp. 159–177, Jan. 2002.

[10] G. N. Dimitrakopoulos, I. Kakkos, Z. Dai, J. Lim, J. J. deSouza, A. Bezerianos, and Y. Sun, "Task-independent mental workload classification based upon common multiband EEG cortical connectivity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1940–1949, Nov. 2017.

[11] I. Kakkos, G. N. Dimitrakopoulos, Y. Sun, J. Yuan, G. K. Matsopoulos, A. Bezerianos, and Y. Sun, "EEG fingerprints of task-independent mental workload discrimination," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 10, pp. 3824–3833, 2021.

[12] Y. Zhou, Z. Xu, Y. Niu, P. Wang, X. Wen, X. Wu, and D. Zhang, "Cross-task cognitive workload recognition based on EEG and domain adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 50–60, 2022.

[13] Y. Zhou, P. Wang, P. Gong, F. Wei, X. Wen, X. Wu, and D. Zhang, "Cross-subject cognitive workload recognition based on EEG and deep domain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[14] Y. Yan, L. Ma, Y.-S. Liu, K. Ivanov, J.-H. Wang, J. Xiong, A. Li, Y. He, and L. Wang, "Topological EEG-based functional connectivity analysis for mental workload state recognition," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.

[15] L. L. Di Stasi, M. Marchitto, A. Antolí, T. Baccino, and J. J. Cañas, "Approximation of on-line mental workload index in ATC simulated multitasks," *J. Air Transp. Manage.*, vol. 16, no. 6, pp. 330–333, Nov. 2010.

[16] H. A. Abbass, J. Tang, R. Amin, M. Ellejmi, and S. Kirby, "Augmented cognition using real-time EEG-based adaptive strategies for air traffic control," in *Proc. Hum. Factors rgonom. Soc. Annu. Meeting*. Los Angeles, CA, USA: Sage Publications, 2014, vol. 58, no. 1, pp. 230–234.

[17] Y. Pang, J. Hu, C. S. Lieber, N. J. Cooke, and Y. Liu, "Air traffic controller workload level prediction using conformalized dynamical graph learning," *Adv. Eng. Informat.*, vol. 57, Aug. 2023, Art. no. 102113.

[18] P. Aricò, G. Borghini, G. Di Flumeri, A. Colosimo, S. Bonelli, A. Golfetti, S. Pozzi, J.-P. Imbert, G. Granger, R. Benhacene, and F. Babiloni, "Adaptive automation triggered by EEG-based mental workload index: A passive brain–computer interface application in realistic air traffic control environment," *Frontiers Human Neurosci.*, vol. 10, p. 539, Oct. 2016.

[19] A. R. John, A. K. Singh, T. N. Do, A. Eidels, E. Nalivaiko, A. M. Gavgani, S. Brown, M. Bennett, S. Lal, A. M. Simpson, S. M. Gustin, K. Double, F. R. Walker, S. Kleitman, J. Morley, and C.-T. Lin, "Unraveling the physiological correlates of mental workload variations in tracking and collision prediction tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 770–781, 2022.

[20] K.-J. Chiang, S. Dong, C.-K. Cheng, and T.-P. Jung, "Using EEG signals to assess workload during memory retrieval in a real-world scenario," *J. Neural Eng.*, vol. 20, no. 3, 2023, Art. no. 036010.

[21] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[22] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[23] X. Chen, X. Teng, H. Chen, Y. Pan, and P. Geyer, "Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX," *Biomed. Signal Process. Control*, vol. 87, Jan. 2024, Art. no. 105475.

[24] Y. Wang, Z. Huang, B. McCane, and P. Neo, "EmotioNet: A 3-D convolutional neural network for EEG-based emotion recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7.

[25] F. Shen, G. Dai, G. Lin, J. Zhang, W. Kong, and H. Zeng, "EEG-based emotion recognition using 4D convolutional recurrent neural network," *Cognit. Neurodynamics*, vol. 14, no. 6, pp. 815–828, Dec. 2020.

[26] P. Gong, Z. Jia, P. Wang, Y. Zhou, and D. Zhang, "ASTDF-Net: Attention-based spatial–temporal dual-stream fusion network for EEG-based emotion recognition," in *Proc. 31st ACM Int. Conf. Multimedia*, New York, NY, USA, Oct. 2023, pp. 883–892.

[27] Y. Liu, Y. Zhou, and D. Zhang, "TcT: Temporal and channel transformer for EEG-based emotion recognition," in *Proc. IEEE 35th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2022, pp. 366–371.
[28] G. Fu, Y. Zhou, P. Gong, P. Wang, W. Shao, and D. Zhang, "A temporal-spectral fused and attention-based deep model for automatic sleep staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1008–1018, 2023.
[29] Q. Sun, L. Wang, L. Zhang, P. Gong, Y. Zhou, and D. Zhang, "MSTI-former: A multi-scale spatial–temporal information enhanced transformer for attentive state classification," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, Athens, Greece, May 2024, pp. 1–4.
[30] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG conformer: Convolutional transformer for EEG decoding and visualization," in *Proc. IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, Dec. 2022, pp. 710–719.
[31] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
[32] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, Feb. 2011.
[33] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, pp. 181–197, Sep. 2019.
[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.
[35] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task load Index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Jan. 1988.
[36] S. Rubio, E. Díaz, J. Martín, and J. M. Puente, "Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods," *Appl. Psychol.*, vol. 53, no. 1, pp. 61–86, Jan. 2004.
[37] L. Van der Maaten and G. E. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, Jan. 2008.
[38] R. Zanetti, A. Arza, A. Aminifar, and D. Atienza, "Real-time EEG-based cognitive workload monitoring on wearable devices," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 1, pp. 265–277, Jan. 2022.
[39] P. Gong, P. Wang, Y. Zhou, X. Wen, and D. Zhang, "TFAC-Net: A temporal-frequential attentional convolutional network for driver drowsiness recognition with single-channel EEG," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 7, pp. 7004–7016, Jul. 2024.
[40] R. Li, J. S. Johansen, H. Ahmed, T. V. Ilyevsky, R. B. Wilbur, H. M. Bharadwaj, and J. M. Siskind, "The perils and pitfalls of block design for EEG classification experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 316–333, Jan. 2021.
[41] L. Jin, H. Qu, L. Pang, Z. Zhang, and Z. Lyu, "Identifying stable EEG patterns over time for mental workload recognition using transfer DS-CNN framework," *Biomed. Signal Process. Control*, vol. 89, Mar. 2024, Art. no. 105662.
[42] Y. Wang, M. Han, Y. Peng, R. Zhao, D. Fan, X. Meng, H. Xu, H. Niu, J. Cheng, and T. Liu, "LGNet: Learning local–global EEG representations for cognitive workload classification in simulated flights," *Biomed. Signal Process. Control*, vol. 92, Jun. 2024, Art. no. 106046.

YUEYING ZHOU (Member, IEEE) received the Ph.D. degree in computer science from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2023.

She joined the School of Mathematics Science, Liaocheng University, as a Lecturer, in 2023. Her research interests include brain–computer interface, EEG analysis, and machine learning.

JUNJI JIANG is currently pursuing the M.S. degree with the School of Mathematics Science, Liaocheng University, Liaocheng, China.

His research interests include machine learning and medical image analysis.

LIJUN WANG is currently pursuing the M.S. degree with the School of Mathematics Science, Liaocheng University, Liaocheng, China.

Her research interests include brain–computer interface, EEG analysis, and machine learning.

SHANSHAN LIANG is currently pursuing the M.S. degree with the School of Mathematics Science, Liaocheng University, Liaocheng, China.

Her research interests include brain–computer interface, EEG analysis, and machine learning.

HAO LIU is currently pursuing the M.S. degree with the School of Mathematics Science, Liaocheng University, Liaocheng, China.

His research interests include brain–computer interface, EEG analysis, and machine learning.