2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)

LSTM Model with Self-Attention Mechanism for EEG Based Cross-Subject Fatigue Detection

Haoheng Kuang*
College of Automation Science and Engineering
South China University of Technology
Guangzhou, China
aukhheng@mail.scut.edu.cn

Jun Qu
South China Brain-computer Interface Technology Co., Ltd
Guangzhou, China
qujun@ihnnk.com

***Abstract*—The Brain-Computer Interface (BCI) system based on electroencephalography (EEG) is proven to detect human mental fatigue. In recent approaches, however, the procedure of EEG data requires a lot of feature engineering and is challenging to achieve ideal recognition accuracy in cross-subject scenarios. This paper presents a novel deep learning model towards remarkably accurate based on the self-attention-based Long Short-Term Memory (LSTM) model. Our study shows that LSTM can find the relevant features of each frequency band between acquisition channels, rather than independently concatenate high-dimensional EEG data into a feature vector; the self-attention mechanism can select the information that is more critical to the current mission from high-dimensional data. In the experiment, the public dataset we selected was labelled with two fatigue levels, and the sample balance was achieved by randomly deleting most samples. Our result shows that our model achieves a 78.84% accuracy rate and outperforms the other methods in a cross-subject situation for fatigue detection. Specifically, self-attention based LSTM improves the accuracy higher than EEG-Net by 19.84% and subject matching by 4.52%.**

Keywords-EEG; LSTM; Self-Attention; Transfer Learning; Cross-subject fatigue recognition

I. INTRODUCTION

In recent years, with the rapid development of the transportation industry, nasty road traffic safety accidents have also increased. According to statistics, fatigued driving is one of the leading causes of traffic accidents[1]. The detection and warning of a driver's mental fatigue can solve this problem. Studies over the past two decades have provided many fatigue detection methods, and these methods can be divided into two categories, based on human behaviour characteristics and human physiological characteristics [2]. For these human behaviour characteristics methods, some studies have proposed that the camera is used to record the state of the eyes. If the eyes are found closed for 5 consecutive frames, the system will conclude that the driver is in a fatigue state and send a warning signal [3]; Gu H et al. [4] also found that it is effective to design an automated face reader for fatigue detection. For these physiological characteristics, Huang K C et al. [5] proposed a fatigue detection system based on EEG, and some studies combine electroencephalography (EEG), electrooculography (EOG) and electromyography (EMG) for fatigue detection [6].

In the above methods, EEG signal is considered the "gold standard" to reflect people's mental state. It is the overall performance of brain nerve cell activity and can reflect the state of the brain in real-time. Many studies have proved the effectiveness of EEG signals in fatigue detection. Gharagozlou et al.[7] found that the power of the α wave increases when people are in a state of fatigue. Wang et al. [8] found that using the wavelet entropy of the EEG signal as a classification feature can achieve an accuracy of 90.7% in fatigue detection. However, the EEG signal is non-stable and non-linear. The above methods can only achieve good results in intra-subject conditions but not in cross-subjects or even the same subject in different trials. In order to solve this problem, most research on cross-subject classification has been carried out. Lawhern V J et al.[9] proposed a compact convolutional neural network for EEG (EEGNet), and the study showed the interpretation of its learned features; Yeo M V M et al. proposed the traditional machine learning method Support Vector Machine (SVM) in cross-subject fatigue detection[10]; A framework called subject match [11] use the multi-source domain alignment layers to collect source domain statistics and train the model, this framework also achieves a state-of-art accuracy for the public driving dataset [12].

In the above method, transfer learning plays an important role in the cross-subject situation. The key idea of transfer learning is to transfer knowledge from the source domain to the target domain. Domain adaption is a mainstream technique and has been applied in EEG processing. Lan Z et al. [13] have come up with a technique called Transferable Component Analysis (TCA). TCA mitigate the distribution mismatch by minimizing the maximum mean discrepancy (MMD) in a reproducing kernel Hilbert space. When the train data (source domain) and the test data (target domain) tend to distribute similarity, they get the higher accuracy of the trained model. Another transfer learning technique is called Maximum Independence Domain Adaption (MIDA) [14]. MIDA seeks to maximize the independence between train data and test data. With the help of the above transfer learning techniques, the latest research [11] achieves 74.32% accuracy in fatigue detection.

In this paper, we proposed a novel deep learning model for fatigue detection. The model uses the Power Spectral Density feature of EEG signal as input, then uses self-attention based LSTM to calculate the probability of fatigue state. LSTM with

978-1-6654-0605-5/21/$31.00 ©2021 IEEE

attention mechanism method has also been applied in motor imagery (MI) [15] and shows that it is effective to use this approach to solve problems related to EEG signal. This research [15] uses raw EEG signals as input, which is different from our method. The transfer learning techniques we use is based on the feature of the deep network model, which can learn the parameters of specific layers in the network. The deep network can automatically transfer the difference between the source and target domains with a tiny test data. The classification accuracy results of our self-attention based LSTM method is higher than the previous researches, which used different model and TCA or MIDA transfer learning techniques.

In this paper, we evaluate the different methods through the same public dataset and Power Spectral Density (PSD) feature. The arrangement of our paper is as follow. The public driving dataset and data preparation are described in Section II; The proposed self-attention based LSTM model and transfer learning method are present in Section III. Section IV shows the performance of different methods. Finally, discuss our methodology of fatigue detection in Section V.

II. Materials

A. Dataset description

In order to compare the performance of EEG based fatigue detection methods, it is necessary to analyze the data under unified experimental conditions. Our paper used a public driving dataset from Cao et al. [16], as shown in Figure 1. This dataset collected 32 channels of EEG data from 62 groups of 27 subjects. The subjects drove the simulated vehicle as instructed in VR immersive simulator. During driving, the background program randomly induces lane-departure events and causes the vehicle to drift. In the experiment, subjects' reaction time to drift is an important index to evaluate the degree of mental fatigue.

In this experimental system, EEG data is sent to the synchronous trigger by the Neuroscan EEG acquisition system, and the time nodes of vehicle lane departure are recorded in log files by computer. EEGLAB is then used in Matlab to import and align the lane-departure events to the EEG data in the synchronous trigger. During the entire 90-minute experiment, EEG data were recorded using 30 EEG electrodes and 2 reference electrodes. EEG electrodes were placed in a modified 10-20 system with a contact impedance less than 5KΩ between the electrodes and the skin. Finally, the collected EEG data were digitized at 500Hz, and stored with annotations in ".set" and ".fdt" files that Matlab or Python can read.

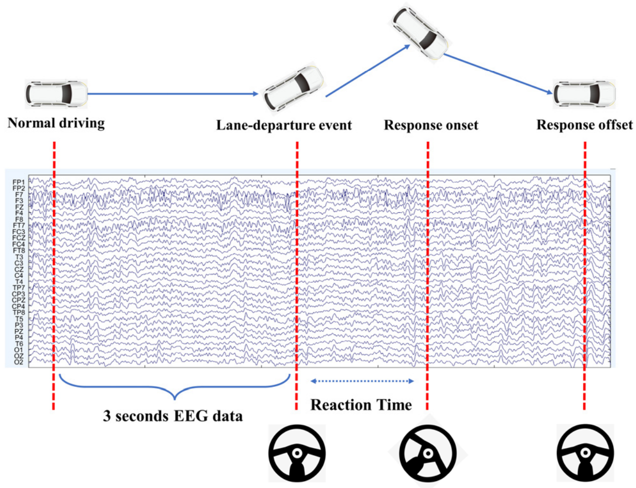

Figure 1. Illustrating the experimental procedures in the public datasets.[16]

B. Data preparation

EEG signal is a kind of bioelectrical signal that can directly reflect the state of the brain. It has the characteristics of low frequency and weak, and its amplitude is generally only $10-50\ \mu V$, the maximum is $100\mu V$, and the frequency is mostly between 0.5 and 50Hz. EEG signals are usually accompanied by strong interference and a low signal-to-noise ratio (SNR). The collected EEG signals often contain complex interference components, such as electrode and skin contact noise, ECG, EOG and myoelectric artifacts. The paper [16] described raw EEG signals are filtered by 1 Hz high-pass and 50 Hz low-pass finite impulse response (FIR) filters. Then manually removed the apparent eye blinks that contaminate EEG signals by visual examination. Eye and muscle artifacts were removed by EEGLAB's Automatic Artifact Removal (AAR) plug-in. The final sampling frequency of this pre-processed version of the dataset is 500Hz.

In our experiment, we down-sampled the raw EEG data to 128Hz. Then we intercepted the EEG signal 3 seconds before the lane-departure events (shown as Figure 1), and the Reaction Time (RT) after lane-departure events was used as an indicator of fatigue. Furthermore, the EEG data of each subject were filtered in the spatial and temporal order. Surface Laplacian is used for spatial filtering. Surface Laplacian spatial filtering enhances the spatial resolution of EEG signals by filtering out spatial features between several adjacent electrodes. The specific implementation is shown in Formula (1) (2). In the time domain, every channel of EEG signal is filtered using a 6-order Butterworth bandpass filter with a frequency range of 1HZ-30Hz. Finally, Each sample we got is a 30×384 matrix with the corresponding RT.

$$V_{LAP(i)} = V_{(i)} - \sum_{j \in S(i)} w_{i,j} V_{(j)} \tag{1}$$

$$w_{i,j} = \frac{1/d_{i,j}}{\sum_{j \in S(i)} 1/d_{i,j}} \tag{2}$$

Referring to the paper [17], the RT corresponding to a single lane-departure event is called Local RT, and the average value of Local RT of all periods within 90 seconds before lane-departure event is denoted as Global RT. All Local RT in the experiment was sorted, and the 5th percentile of local RT was taken as Alert RT. The events with both local RT and Global RT less than 1.5 times Alert RT were recorded as the awake state, and the events with greater than 2.5 times Alert RT were recorded as the fatigue state. These are the criteria for judging whether the subject is fatigue state or not.

In this dataset, some subjects did not show an apparent fatigue state during the experiment. According to the above criteria, we screened the experimental data of 11 subjects. In order to ensure sample balance, When the difference in the number of two types of samples is too large, we randomly deleted most samples. Finally, we obtained the number of each qualified subject, as shown in Table 1.

TABLE I. NUMBER OF EVENTS PER QUALIFIED SUBJECT

Subject Id	Number of Events	
	Fatigue State	*Awake State*
1	96	94
5	66	66
22	75	75
31	74	118
35	101	85
41	116	83
42	51	51
43	70	70
44	72	72
45	54	54
53	131	113
Total	906	881

C. Feature extraction

The frequency-domain characteristics of EEG signals are prominent. The Short-Time Fourier Transform (STFT) with Hamming window was applied to the EEG data of each event, and the Power Spectral Density (PSD) was obtained at δ (1-4Hz), θ (4-8Hz), α (8-12Hz), and β (12-30Hz) spectral bands. Finally, feature data obtained is a 30×4 matrix with the corresponding RT.

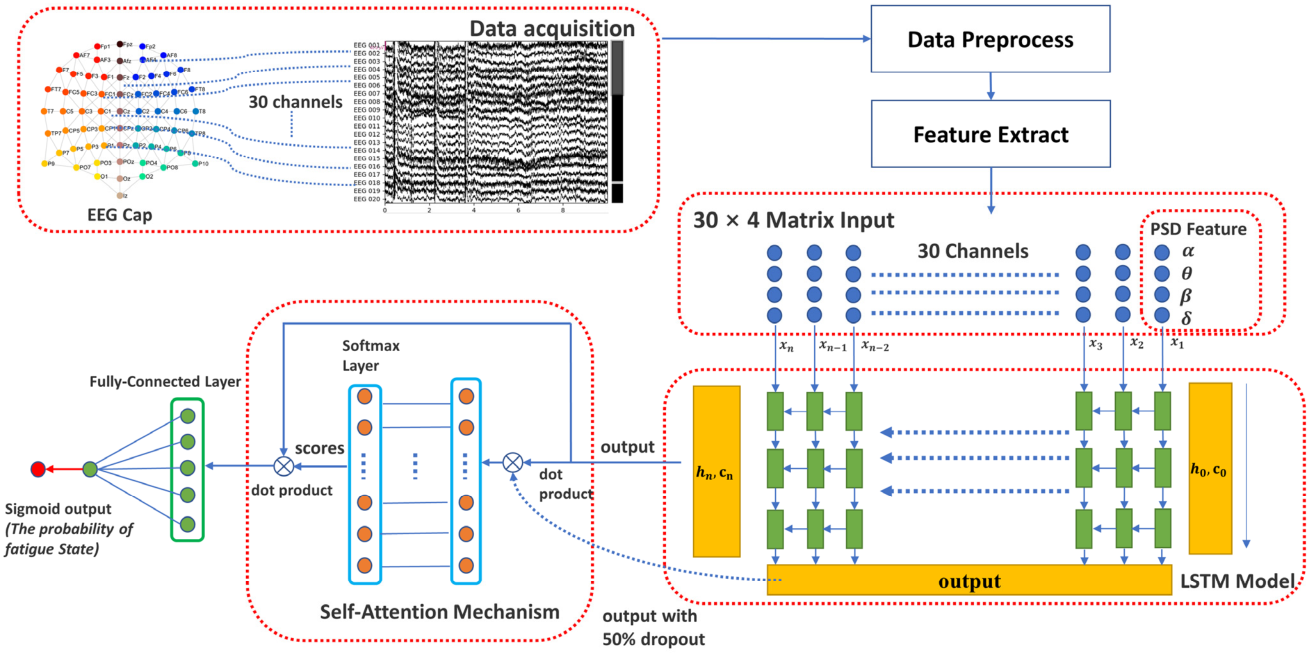

Figure 2. EEG based Self-Attention-LSTM fatigue detection model

III. METHODOLOGY

A. LSTM Model

Long Short Term Memory network (LSTM) is a special kind of Recurrent Neural Network (RNN); it is very good at processing sequence data. EEG signals can be regarded as both time sequence and spatial sequence. In the time domain, EEG data are similar to time-varying sound waves. Spatially, due to its multi-channel nature, different EEG channels transmit specific and relevant information. For example, the EEG signals of the forehead and back neck often express different meanings. Therefore, it is a feasible method to extract deep features by the LSTM model.

As mentioned in paper [18], the formula of LSTM forward propagation is as follows (3-8). The key of LSTM is the value stored in its memory cell. LSTM deletes and adds information to the memory cell through a structure called "gate". LSTM has 3 kinds of gates: input gate, forgetting gate and output gate. In formula (3) (4) (6), The parameter i_t, f_t, o_t is the degree of the 3 gates respectively. The parameter g_t is the conventional RNN operation for input. In the formula, we can see that there are 2 outputs of LSTM: cell state c_t and hidden state h_t. The parameter c_t is the product of input and forgetting gate; it is the content of the current LSTM cell itself. The parameter h_t is obtained through the output gate, and it will be passed to the next LSTM cell.

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \tag{3}$$

$$f_t = \sigma\left(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}\right) \tag{4}$$

$$g_t = \tanh\left(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}\right) \tag{5}$$

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \tag{6}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{7}$$

$$h_t = o_t \odot \tanh(c_t) \tag{8}$$

In our experiment, as shown in Figure 2, the input of LSTM is the PSD feature of raw EEG Data. The input x_i means the extracted feature from $\alpha, \beta, \theta, \delta$ waves. We use the channel from 1-30 as the sequence of LSTM, and it is an easy way to find the relationship between different channels. The LSTM model is implemented by using the deep learning framework Pytorch. The hyperparameter of LSTM can affect the classification performance of the model, and we set the hidden layer size to 4, the dropout of LSTM cell to 50%, and the number of LSTM layers to 1 after the experiment. The reason why we did not choose a deeper LSTM layer is to prevent overfitting.

B. Self-Attention Mechanism

The self-attention mechanism[19] uses the attention mechanism[20] to "dynamically" generate the weights of different connections to deal with the variable-length information sequence. In our method, the input of the self-attention mechanism is the output of the LSTM model. We define the input vector as $X = [x_1, \cdots, x_N]$, x_N represents the Nth input information. Then it is obtained as query vector Q (9), key vector K (10) and value vector V (11) by a linear transformation.

$$Q = W_Q X \tag{9}$$

$$K = W_K X \tag{10}$$

$$V = W_V X \tag{11}$$

From the above formula, Q in self-attention mechanism is the transformation of the input X itself, while in traditional attention mechanism, Q comes from the outside.

The self-attention mechanism uses the Dot Product model, the output vector can be written as formula (12). After the key vector K is transposed, it is dot product with the query vector Q. After d_k scaling, the attention score on each output channel is obtained through the softmax function. The results obtained are dot product with the value vector V to obtain the output H of the attention model.

$$H = V\text{softmax}\left(\frac{K^{\mathrm{T}}Q}{\sqrt{d_k}}\right) \tag{12}$$

As shown in figure 2, the parameter in the self-attention mechanism is $K = V = X$, and the value vector Q is the output X of LSTM with 50% dropout.

Finally, we pass the output vector H through a Fully Connected layer with a sigmoid activate function $S(x) = 1/(1 + e^{-x})$. The output value of the sigmoid is between 0-1, which means the probability that the current sample belongs to fatigue state or awake state.

C. Transfer Learning

In the experiment, we verified the effectiveness of the model on 11 subjects. The deep learning model can learn a wide range of features in the training data, but due to the instability of the EEG signal, we choose transfer learning to enhance the accuracy of the cross-subject model. We divide 11 subjects into training subjects and test subjects, select about 20% from the test subjects, and ensure that the samples are balanced. Transfer learning is carried out in the following ways.

As shown in Figure 3, the implementation steps are as follows:

- *Step 1.* we pre-train a source model on the source dataset (training subjects).
- *Step 2.* Create a new model as the target model. It duplicates all model designs and their parameters on the source model except the output layer. We assume that these model parameters contain the knowledge learned from the source training subjects, which is also applicable to the test subjects. We also assume that the output layer of the source model is closely related to the label of the training subjects, so it is not used in the target model.
- *Step 3.* Randomly initialize the model parameters of the output layer and set a minimal learning rate for hidden layers (Fine-tuning)

- *Step 4.* Train the target model on the training subjects. We will retrain the output layer, and the parameters of other layers are obtained by fine-tuning based on the parameters of the source model.
- *Step 5.* In our experiment, we fine-tune the parameters of the LSTM model and attention mechanism, use the learning rate of $1e^{-6}$, retrain the output layer of the fully connected layer, and set the learning rate to $1e^{-3}$.

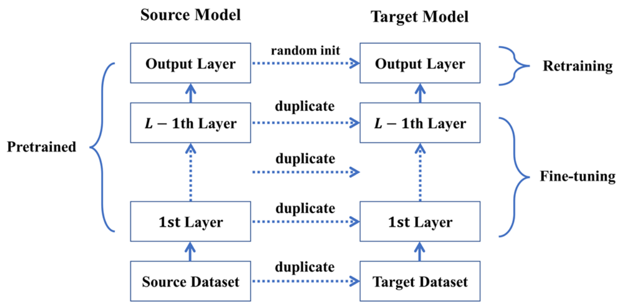

Figure 3. Schematic diagram of Transfer Learning

IV. Result

In the experiment, we use leave-one subject out cross-validation to test the classification accuracy of the model. Under the same experimental conditions as described in [11], we compare the classification accuracy of our attention based LSTM model with EEGNet, SVM and Subject Matching. These comparison methods also use transfer learning or similarity transformation on the test data. The results are shown in Table 2.

From the classification accuracy of different models on 11 subjects, the attention-based LSTM model achieves the highest classification accuracy of 78.84%. Compared with EEGNet, which also uses the deep learning model, our model improves the classification accuracy by 19.84%. Compared with the traditional machine learning method SVM, the classification accuracy of the attention-based LSTM model improves by 7.17%. Compared with subject matching, our model improves the accuracy by 4.52%. In addition, the standard deviation (STD) of classification accuracy of our model is also outstanding. Our method achieves a standard deviation of 6.53, which is lower than subject matching and SVM. Although the standard deviation is not as good as EEGNet, it has a significant advantage in accuracy. The result shows that our model has better generalization ability in across-subjects situations.

TABLE II. Classification accuracies across 11 subjects

Subject Id	**Methods**			
	EEGNet	***SVM***	***Subject matching***	***Attention Based LSTM***
1	57.08	79.26	78.72	**80.92**
5	59.16	71.97	68.18	**79.25**
22	59.07	66.67	**79.33**	75.83
31	56.04	65.54	**68.24**	66.39
35	57.57	82.59	**85.27**	80.53
41	55.31	74.10	**83.73**	83.12
42	58.01	60.78	64.71	**76.83**
43	54.16	65.15	57.20	**82.14**
44	59.12	88.22	78.03	**92.24**
45	73.72	71.30	**82.41**	77.59
53	59.51	62.83	71.68	**72.45**
Mean	59.00	71.67	74.32	**78.84**
Std	**5.18**	8.31	8.94	6.53

V. Discussion

As we expected, we have achieved good results by using the attention-based LSTM model. However, we believe that the EEGNet of the deep learning model does not achieve good results because EEGNet widely learns the PSD feature of each channel without integrating the information from every channel. In our proposed model, the LSTM model correlates the PSD features between each channel, and the attention mechanism evaluates the importance of the features hidden in each LSTM cell. Not every feature has the same importance for high-dimensional data, it is challenging to select the critical features manually, and our model solves this problem.

From the perspective of cross-subject transfer learning, Transfer learning methods in related work often need projection mapping of training data and test data. Our model can automatically find the appropriate data projection by retraining the output Fully Connected layer, reducing computational difficulty and complexity.

In terms of interpretability, deep learning models are usually complex to explain the parameters in the hidden layer. Interpretability is not the focus of this paper. However, research is also needed to determine the performance of the attention-based LSTM model under a larger amount of training data and test data.

VI. Conclusion

In this paper, we propose a novel deep learning model for fatigue detection based on EEG signals. This study has shown that the self-attention based LSTM model can mine the deep information of power-spectral-density of multi-channel EEG signal. The results indicate that the attention-based LSTM model has 78.84% accuracy in fatigue state classification, which is better than other similar classification models. The transfer learning method can be directly used on the model rather than looking for similarity mapping between raw EEG signals. The findings will be of interest to the detection of fatigue driving, and we only need to use the data of training subjects to pre-train a source model. Then, a small amount of test subject data is needed, and a high accuracy fatigue detection model can be obtained through transfer learning. The study certainly adds a very effective and straightforward way to detect fatigue from EEG signals.

ACKNOWLEDGMENT

This work was supported in part by the Key Research and Development Program of Guangdong Province, China, under Grant 2018B030339001 and in part by the Key Realm Research and Development Program of Guangzhou, China, under Grant 202007030007.

REFERENCES

[1] S. K. L. Lal and A. Craig, 'A critical review of the psychophysiology of driver fatigue', Biological Psychology, vol. 55, no. 3, pp. 173–194, Feb.2001, doi: 10.1016/S0301-0511(00)00085-5.

[2] Hartley L, Horberry T, Mabbott N, et al. Review of fatigue detection and prediction technologies[J]. National Road Transport Commission, 2000: 1-41.

[3] Devi M S, Bajaj P R. Driver fatigue detection based on eye tracking[C]//2008 First International Conference on Emerging Trends in Engineering and Technology. IEEE, 2008: 649-652.

[4] Gu H, Ji Q. An automated face reader for fatigue detection[C]//Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. IEEE, 2004: 111-116.

[5] Huang K C, Huang T Y, Chuang C H, et al. An EEG-based fatigue detection and mitigation system[J]. International journal of neural systems, 2016, 26(04): 1650018.

[6] M. Doudou, A. Bouabdallah, and V. Berge-Cherfaoui, 'Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges', Int. J. ITS Res., vol. 18, no. 2, pp. 297–319, May 2020, doi: 10.1007/s13177-019-00199-w.

[7] F. Gharagozlou et al., 'Detecting Driver Mental Fatigue Based on EEGAlpha Power Changes during Simulated Driving', p. 9.

[8] Q. Wang, Y. Li, and X. Liu, 'Analysis of Feature Fatigue EEG SignalsBased on Wavelet Entropy', Int. J. Patt. Recogn. Artif. Intell., vol. 32, no.08, p. 1854023, Aug. 2018, doi: 10.1142/S021800141854023X.

[9] Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces[J]. Journal of neural engineering, 2018, 15(5): 056013.

[10] Yeo M V M, Li X, Shen K, et al. Can SVM be used for automatic EEG detection of drowsiness during car driving?[J]. Safety Science, 2009, 47(1): 115-124.

[11] Cao Z, Chuang C H, King J K, et al. Multi-channel EEG recordings during a sustained-attention driving task[J]. Scientific data, 2019, 6(1): 1-8.

[12] Li R, Wang L, Sourina O. Subject matching for cross-subject EEG-based recognition of driver states related to situation awareness[J]. Methods, 2021.

[13] Lan Z, Sourina O, Wang L, et al. Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets[J]. IEEE Transactions on Cognitive and Developmental Systems, 2018, 11(1): 85-94

[14] F. Qiao, L. Zhao, X. Peng, Learning to learn single domain generalization, in: inProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020, pp. 12556–12565.

[15] Hou Y, Jia S, Zhang S, et al. Deep feature mining via attention-based BiLSTM-GCN for human motor imagery recognition[J]. arXiv preprint arXiv:2005.00777, 2020.

[16] Cao Z, Chuang C H, King J K, et al. Multi-channel EEG recordings during a sustained-attention driving task[J]. Scientific data, 2019, 6(1): 1-8.

[17] Liu Y, Lan Z, Cui J, et al. Eeg-based cross-subject mental fatigue recognition[C]//2019 International Conference on Cyberworlds (CW). IEEE, 2019: 247-252.

[18] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[19] Tan Z, Wang M, Xie J, et al. Deep semantic role labeling with self-attention[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1).

[20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.