EEG-Based Spatio–Temporal Convolutional Neural Network for Driver Fatigue Evaluation

Zhongke Gao, Xinmin Wang, Yuxuan Yang, Chaoxu Mu, Qing Cai, Weidong Dang, and Siyang Zuo

***Abstract*— Driver fatigue evaluation is of great importance for traffic safety and many intricate factors would exacerbate the difficulty. In this paper, based on the spatial–temporal structure of multichannel electroencephalogram (EEG) signals, we develop a novel EEG-based spatial–temporal convolutional neural network (ESTCNN) to detect driver fatigue. First, we introduce the core block to extract temporal dependencies from EEG signals. Then, we employ dense layers to fuse spatial features and realize classification. The developed network could automatically learn valid features from EEG signals, which outperforms the classical two-step machine learning algorithms. Importantly, we carry out fatigue driving experiments to collect EEG signals from eight subjects being alert and fatigue states. Using 2800 samples under within-subject splitting, we compare the effectiveness of ESTCNN with eight competitive methods. The results indicate that ESTCNN fulfills a better classification accuracy of 97.37% than these compared methods. Furthermore, the spatial–temporal structure of this framework advantages in computational efficiency and reference time, which allows further implementations in the brain–computer interface online systems.**

***Index Terms*— Brain–computer interface (BCI), convolutional neural network (CNN), deep learning (DL), electroencephalogram (EEG), fatigue driving, spatio–temporal data.**

I. Introduction

MENTAL fatigue refers to a complicated physiological and psychological condition accompanied with lessened alertness and decremental integrated performance. According to individual motivation, it is defined as a sensation of weariness with one's unwillingness to carry on performing in the task and may lead to reduced work efficiency and increased accident possibility [1]. As reported by *World Health Organization* [2], about 127 000 people were killed by road accidents annually, and about one-third of the victims were aged from 15 to 29 years. Among the causes of the accidents, fatigue driving devotes to a conservative estimate of above 10 000 deaths, which is a great threat to road safety. These observations indicate that fatigue working has caused great troubles in the transportation industry, demanding for detecting techniques to precisely recognize fatigue states [3].

Up until now, several types of human clues have been used for fatigue detection, including facial expressions [4], speech signals, and some physiological indexes such as electroencephalogram (EEG) data and dermal resistance [5]. However, the nonphysiological sources such as face expressions vary from different living habits and cultural backgrounds, thus may not reliable to a certain extent. In addition, some physiological indexes such as dermal resistance could be affected by temperature and humidity. On consideration, EEG signal is the chief source among these, which contains a great deal of physiological information of a working brain. It can truly reflect the pathological and mental states of a human body due to its good temporal resolution and information richness [6]. Moreover, with the rapid development of wearable EEG devices and dry electrode techniques [7], various tasks can be easily implemented with EEG online systems and are promising for practical applications. In view of these foundations, we focus on driver fatigue evaluation based on continuously recording multichannel EEG signals.

It is noticed that the EEG signal is extremely weak with low signal-to-noise ratios, which makes it fairly difficult to develop computational algorithms for fatigue detection. Against these odds, considerable studies have been conducted on fatigue evaluation tasks. Subjects may be suffered from an aggravated fatigue reaction in a monotonous and cumulative process [8], which has an effect on physiological indicators in turn. A suitable fatigue threshold could be established according to individual feedbacks and derivable indicators, and thus fatigue detection can be turned into a classification exploration. Many novel methods have been proposed to extract valuable information from EEG signals, e.g., time–frequency analysis [9], complex network methods [10]–[12], and nonlinear analysis [13]. Particularly, several studies have given great ideas to distinguish subjects' alert and fatigue states [14]. In [15], four frequency features were extracted across four principal frequency bands from EEG signals and then fed into the support vector machine (SVM), which achieved an excellent classification effect on fatigue detection. In [16], an EEG-based system was presented to effectively detect drivers' fatigue states by calculating four entropies as features and analyzing the effects of multiple entropy fusion. To enhance the performance, the feature extraction processes

Manuscript received November 30, 2017; revised July 11, 2018 and October 26, 2018; accepted December 2, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61873181, Grant 61473203, and Grant 61773284 and in part by the Natural Science Foundation of Tianjin, China under Grant 16JCYBJC18200. *(Corresponding author: Zhongke Gao.)*

Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, and W. Dang are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: zhongkegao@tju.edu.cn).

S. Zuo is with the Key Laboratory of Mechanism Theory and Equipment Design, Ministry of Education, Tianjin University, Tianjin 300072, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2018.2886414

cannot ignore the interference of electrooculogram (EOG) and other physiological signal noises. In [17], a multimodal approach was proposed to estimate driver fatigue by combining EEG and forehead EOG with adopted electrode placement, which targeted at integrating the temporal dependencies of both and received an improved performance. In [18], an independent component was used for source separation from EEG signals, then coupled with autoregressive modeling and Bayesian neural network, which worked effectively for fatigue identification. Overall, it is challenging to design a single framework to overcome these problems and be available for wide varieties of users.

According to international standards [19], EEG signals are recorded from multiple active electrodes attached to cerebral cortex with a fixed spatial arrangement. Among the EEG classification methods mentioned above, the feature extraction process of time–frequency analysis may neglect the valuable information of electrode correlations from EEG signals. It can be concretely interpreted as a spatial dimension, and some methods based on the common spatial patterns (CSPs) method were proposed to manage with spatial information. The CSP method aims to find a set of spatial filters that can maximize the between-class distance. Then, the computed relative energy of the filtered channels is fed into linear classifiers for classification. A regularized CSPs algorithm was presented as feature extraction in [20], which significantly outperformed traditional CSPs in motor imagery classification. Several successful attempts have been conducted to improve the CSP methods. The filter bank CSP method in [21] was extended to help improve the performance of signal decomposition. In applied contexts, summarily, spatial–temporal filtering methods have an outstanding capacity to deal with EEG signal analysis. Even though some CSP-based methods can gradually increase the performances, they still have room for improvement in mapping relationships, especially the concern on temporal dependencies.

Recently, deep learning (DL) techniques have shown distinctive capabilities in object detection, speech recognition, and time series classification [22], [23]. These models possess high computational efficiency and low model complexity. In addition, many advanced methods based on DL techniques have been explored and successfully converted into practical applications. Among these previous studies, DL methods have also shown effective performances in the analysis of EEG classification tasks. The DL framework should first reduce the dimension of EEG signals and then transform them into new representations without any significant information loss. Some studies have applied convolutional neural networks (CNNs) into many kinds of EEG recognition tasks, such as emotion recognition [24], P300 detection [25], effective radiated power classification [26], and memory performance prediction [27]. In [28], a novel channel-wise CNN was proposed to evaluate fatigue states on raw EEG data and ICA-transformed data, and it received an improved performance. These studies convey solid and meaningful design solutions for EEG signals. Furthermore, a detailed survey was presented in [29], which reviewed how to design and train CNNs without handcrafted features for EEG-based brain mapping.

In another work [30], by using fast Fourier transform (FFT), EEG sequences were converted into 2-D meshlike hierarchies and then fed into the convolutional recurrent neural network for human intended movement identification. The results confirmed the effectiveness of FFT-based DL methods in EEG recognition tasks. Recently, by modifying the filter-bank CSPs methods, EEG signals were turned into new temporal representations and a CNN architecture was introduced for motor imagery EEG signal classification. The framework outperformed the existing results on the motor imagery data set [31]. These two studies converted EEG signals into new characteristics by using feature extraction methods, which can be fed into DL frameworks with more concrete information. During the analysis procedure, they both focused on the treatments of spatial relations and temporal dependencies from EEG signals.

In this paper, we develop a novel framework, namely as EEG-based spatial–temporal CNN (ESTCNN), for driver fatigue detection. It attaches importance to temporal dependencies learning for each electrode and also strengthens spatial information extraction. The developed framework has an effective performance on EEG signal classification tasks. First, we introduce the core block which has some advantages on temporal dependencies extraction and then combine it with dense layers to meet the spatial–temporal information of EEG signals. Second, it constantly reduces the data dimension in the inference procedure, which gives rise to computational efficiency and reference response. The proposed framework has achieved the best performance on the fatigue data set compared with eight competitive methods.

To gain insight into fatigue detection, this paper is organized as follows. First, we introduce the core block and present the ESTCNN framework with the learning process and implementation details. Second, we present a systematic introduction of the fatigue experiment and the acquisition and preprocessing of EEG signals. Third, we present the overall performance of the model and compare it with eight competitive methods. Finally, we present conclusions with an outlook of future applications on broader brain–computer interface (BCI) systems.

II. Methods

To build a proper network, it is important to consider spatial relations and temporal dependencies using EEG signals. Due to its inherent distribution, the rational information along different dimensions could greatly enhance the model performance and make the model more explanatory. We first introduce the core block of the ESTCNN framework and then present the concrete model architecture and learning procedures. Finally, we give the implementation details.

A. Core Block

Several DL methods, such as convolutional layer and recurrent layer, have been proposed to deal with the dynamic properties of EEG signals for excavating effective information of partial snippets. Temporal convolution can be used as a feature extraction module to process time series, and when compared with recurrent models, it has better computational

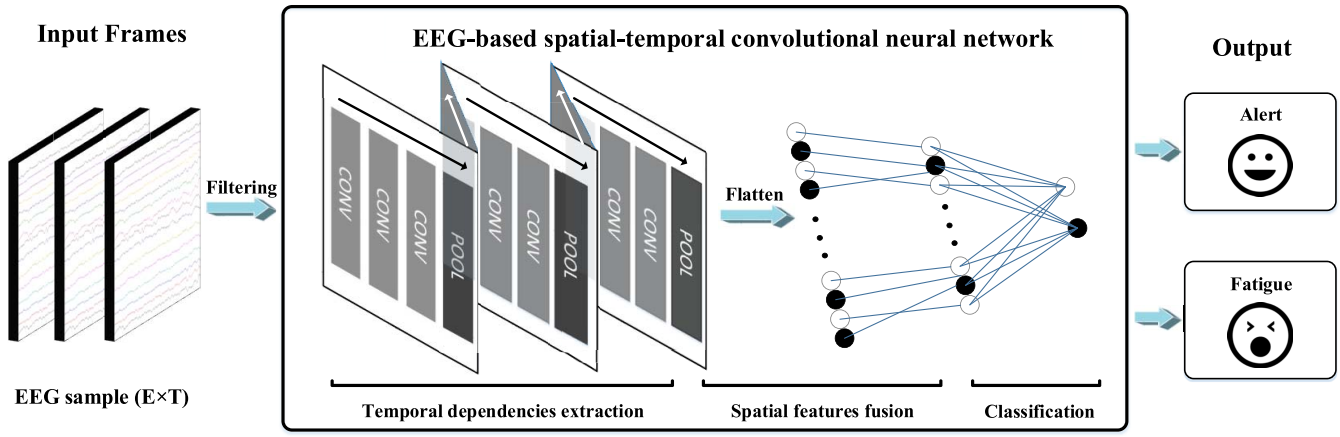

Fig. 1. High-level pipeline of the ESTCNN framework.

efficiency on time series classification tasks [32]. We get inspired and propose this core block for two reasons in the task. First, convolution layers in the core block could process information in a hierarchical form, so be easier to capture high-level features from EEG signals. Second, we also add pooling layer to balance train accuracy and generalization. The core block consists of three convolutional blocks and a pooling layer. Each convolutional block orderly consists of a 1×3 convolution, a rectified linear activation [33], and a batch normalization (BN) [34], referring to [35]. We use valid padding for convolution so that the output through three convolutions has six fewer dimensions than the input. In addition, three convolutions in each core block share the same hyperparameters.

From the view of feature level, the core block can be explained as follows. Each convolution layer in the core block performs weighted feature combiners with nonlinear activation. Then, the extracted features are cross channel convoluted repeatedly in the next convolution layer. It contributes to extract the high-level information across the temporal dimension. In addition, a pooling layer is followed to ease overfitting. Here, we take this convolution structure as core block to make better predictions. This change makes it straight to understand the process on temporal dimension.

B. Model Architecture

Fig. 1 shows the high-level pipeline of the proposed method for EEG-based fatigue evaluation. The framework could learn effective information through the combination of the core blocks and dense layers.

Let matrix $X \in R^{E \cdot T}$ serve as input samples (for 30×100 volumes) with paired labels $Y \in R$, where E denotes recorded electrodes, T denotes sampling points, and Y varies from 0 to 1 as the associated label prediction. The proposed framework contains convolutional layers, pooling layers, and dense layers.

In the experiment, we reach a network with depth selected to 14 layers, which consists of three core blocks, two dense layers, and a softmax layer. The details of network structure are presented in Table I. In this framework, three core blocks

TABLE I

DETAILS OF THE ESTCNN FRAMEWORK. THE SYMBOL [] CONTAINS THE KERNEL SIZE, THE NUMBER OF FEATURE MAPS, AND THE TYPE OF LAYER, RESPECTIVELY

Layers	Output size	ESTCNN
Input	30×100	–
Core block 1	30×94	[1×3, map 16, conv]×3
	30×47	1×2, maxpool, stride2
Core block 2	30×41	[1×3, map 32, conv]×3
	30×20	1×2, maxpool, stride2
Core block 3	30×14	[1×3, map 64, conv]×3
	30×2	1×7, averagepool, stride7
Classification	50	Dense layer
	2	Softmax

are initially performed to handle input samples. Each pooling layer goes only in temporal dimension without overlap, where the first two are set as max pooling with kernel size 2, and the third is set as average pooling with kernel size 7. Two dense layers, behind the third core block, are followed with a Softmax classifier. In addition, no normalization is adopted in the data preprocessing process. The node numbers of two dense layers are 50 and 2, respectively. In the convolutional layers, kernel size is set as 3 with default stride and $16x$ filters, where x is initialized to 1 and doubled every core block.

C. Model Learning

A unit in the CNN is denoted by $x_{l,k,(m,n)}$, where l is the layer, k is the feature map, and (m, n) is the position of the unit in this feature map. Likewise, $\sigma_{l,k,(m,n)}$ is denoted as the scalar product between a group of input neurons. Then, $x_{l,k,(m,n)}$ can be obtained as

$$x_{l,k,(m,n)} = f(\sigma_{l,k,(m,n)}) \tag{1}$$

where f is the rectified linear units function [33] used for whole network layers.

Notably, each neuron of feature maps in the convolutional layer shares the same set of weights, which aims to decrease the number of weight parameters. They are attached to a subset

of the neurons of the former layer, which depends on the exact position of this neuron. Or rather, the neuron weights are trained independently to their corresponding receptive fields. In ESTCNN structure, the computation process of three core blocks is quite similar, so we introduce the first block and other reasoning parts. Let layer n denote L_n, then the information transmission process could be described as follows.

1) For Layer L_1

$$\sigma_{1,k,(m,n)} = \omega_{1,k,0} + \sum_{i=1}^{A} I_{m+i-1,n} \cdot \omega_{1,k,i} \tag{2}$$

where $\omega_{1,k,0}$ is a threshold and $\omega_{1,k,j}$ denotes a set of weights with $1 \le i \le A$ ($A = 3$). Here, m corresponds to the convolutional kernel used in this framework, which amounts for temporal filters. With $C + 1$ weights for each map, this layer is designed to extract more valid temporal features among all electrodes.

2) For Layer L_2

$$\sigma_{2,k,(m,n)} = \omega_{2,k,0} + \sum_{i=1}^{N_1} \sum_{j=1}^{B} X_{1,i,(m+j-1,n)} \cdot \omega_{2,k,i} \tag{3}$$

where N_1 denotes the number of feature maps in layer L_1 and B denotes the kernel size of L_2. This layer is employed to further process the temporal information extract from layer L_1 by cross channel.

3) For Layer L_3

$$\sigma_{3,k,(m,n)} = \omega_{3,k,0} + \sum_{i=1}^{N_2} \sum_{j=1}^{C} X_{2,i,(m+j-1,n)} \cdot \omega_{3,k,i} \tag{4}$$

where N_2 denotes the number of feature maps in layer L_2 and C denotes the kernel size of L_3. This layer is employed to extract the high-level information of temporal dependencies based on the feature maps from layer L_2.

4) For Layer L_4

$$\sigma_{4,k,(m,n)} = \max(x_{3,k,(i,n)}, x_{3,k,(i+1,n)}). \tag{5}$$

No parameter is occupied in this layer and k is fixed. The pooling layer reduces the dimension of feature maps by half, which aims to ease overfitting. The second and third core blocks (L_5–L_8 and L_9–L_{12}) follow the same rules of the first core block (L_1–L_4) and can be deduced from it.

5) For Layer L_{13}

$$\sigma_{13,n} = \omega_{13,0,n} + \sum_{i=1}^{N_{12}} \sum_{j=1}^{D} X_{12,i,j} \cdot \omega_{13,n} \tag{6}$$

where $\omega_{13,0,n}$ is a threshold, N_{12} denotes the number of feature maps and each has D neurons in layer L_{12}. In addition, L_{13} has n neurons and is fully connected to the flattened L_{12}. This layer plays a role of making channel feature combinations.

6) For Layer L_{14}

$$\sigma_{14,n} = \omega_{14,0,n} + \sum_{i=1}^{N_{13}} X_{13,i} \cdot \omega_{14,n} \tag{7}$$

where $\omega_{14,0,n}$ is a threshold and N_{13} denotes the neuron number in layer L_{13}. L_{14} is fully connected to L_{13}. The layer aims to select the valid spatial information for classification.

D. Implementation

We take the robust normalization strategy in the fancy 14-layer network, adopting the CNN architecture from electrocardiogram arrhythmias detection in [36]. Hence, BN layers are repeatedly employed to optimize such a deep network manageably, and the stacking structure of the network is used to enlarge the receptive field of high-level convolution kernels as well as the nonlinearity of model fitting. We use the classical backpropagation as a learning algorithm to tune up the thresholds and weights of the network [37], which is reflected by the promotion of model accuracy on the validation set. The cross-entropy objective function is employed as a loss function to estimate the model performance.

During the process of training the models, the weight of the convolutional layers is initialized with glorot_normal initializer mentioned in [38]. We adopt the stochastic gradient descent [39] optimizer with the learning rate 0.001 and other default parameters. During the optimization for backpropagation, we save the optimal model evaluated on the validation data set. The model is implemented on a workstation with an Intel CPU (i7-6850k, 3.6 GHz) and an NVIDIA GPU (Titan Xp) using the Keras library,[1] which is extended from Google Tensorflow.[2] The typical duration for model training is about 5 min.

III. Experiment

For the purpose of investigating the effective representations of fatigue states, we elaborately designed a fatigue driving experiment for fatigue evaluation to collect EEG data, which could produce valuable original data sets. In our experiments, we took full consideration about the time factor and subjects' cooperative attitudes. In addition, we arranged the driving test environment as real life as possible, which aimed to elicit the strong physiological changes for subjects. In this section, we introduce the subject information, experiment protocol, and data acquisition and preprocessing, respectively.

A. Subjects

In the study, eight right-handed undergraduates (five males and three females; mean: 22.73, standard deviation (STD): 1.69) aged from 19 to 26 voluntarily performed the experiment. None of them had any disorders related to psychiatric. Subjects were required to refrain from antifatigue drinks or drowsiness-causing medications for 2 days before the experiment. Concurrently, they needed to keep reasonable rest with sleep durations of more than 7 h per night. They should comply with these regulations so as to join the study. As all subjects had no exposure to driving simulators, they were asked to practice driving until they get skilled. Subjects were

[1] https://keras.io/

[2] https://www.tensorflow.org/

Fig. 2. Experimental settings. (a) Experimental scene from a researcher's perspective. (b) Simulated driving system. (c) Forehead EOG placement from Neuroscan system.

advised to stop driving at any moment during the experiment when any discomfort appeared.

B. Experiment Protocol 配有脚踏板、方向盘和离合器

Driving on a real highway and simultaneously conducting another task was highly perilous for subjects and other drivers. Therefore, the study was conducted in the Laboratory of Complex Networks and Intelligent Systems, Tianjin University. We used the driving simulator, $PGFD001$, equipped with a pedal, a steering wheel, and a clutch. In the virtual driving software $3D\ Instructor2$, we used the general car $Phaeton2.0L$ with automatic shifting by default. The simulated environment was a monotonous expressway with few bends, sunny day, and bare roadside scenery. Furthermore, we added a webcam $360D618$, a projector, and a stereo cabinet for better feel. The experimental settings are shown in Fig. 2.

To heighten the fatigue sensations of subjects, the trials began during 14:00–15:30, which was proven as an easy-trapped period for fatigue. Each trial lasted about 90 min. Full-course scalp EEG signals of subjects were collected in an isolated and silent room. In addition, we also monitored subjects' facial states via a front-facing camera to verify fatigue levels. According to Karolinska Sleepiness Scale that assesses from 1 (extremely alert) to 9 (very sleepy) [40], drivers' fatigue states are divided into alert, mild fatigue, and fatigue for evaluation. Before the experiment, there were 10 min for scene setup and about 20 min for driving practice. First, after a 3-min survey, the subject would keep driving until he/she report his/her mild fatigue, which generally lasted about 30 min as the alert state. Then, after a 10 min continuously driving for transition, the subject went through another 30-min driving, which was used as a fatigue state. After this stage, each subject was asked to answer some questions and report postexperiment comments. The recorded duration for each subject was nearly 90 min, which varied with individual differences. The timeline of the experiment is shown in Fig. 3. Due to that, the experiment procedure was a little tiresome, repetitive but mental engaged; the fatigue level of subjects became increasingly stronger during the experiment. As the vital physiological measurements of fatigue states, their subjective assessments and actual behaviors revealed the increasingly suffered fatigue.

C. Data Acquisition and Preprocessing

As a measurement of driver fatigue states, the EEG recording device was equipped with the Neuroscan system with 40 electrodes at a sampling frequency of 1000 Hz, which were arranged in accordance of the standard international 10/20 system. Before the acquisition, the skin impedance of EEG electrodes was adjusted below 5 kΩ by injecting conductive gel. Among these 40 electrodes, except for 4 electrodes as the interior structure, 2 were defined as reference electrodes and 4 (placed across horizontal and vertical directions) were used for monitoring eye movements. All subjects were advised to possibly restrict unnecessary body movements, maintain a constant speed, and avoid car collision during data collection.

Raw EEG signals were preprocessed with EEGLAB toolbox [41]. We reduced the sampling rate to 100 Hz and performed a bandpass filter of 1–50 Hz on the EEG signals to remove artifacts since the power frequency interference is above 50 Hz and some useless physiological noises are below 1 Hz. After that, there were 30 channels for the preprocessed EEG signals. We removed the 10-min transition period and obtained two pieces of about 30-min signals, labeled with alert and fatigue categories. Then, we divided the signals into samples of 1-s epochs without overlap and intentionally selected larger class samples to make classes balanced. There are 2800 samples per subject, of which the number of each category is 1400.

IV. Result

In this section, to validate the performance of the proposed framework, we first give the overall performance evaluated with 10-fold cross validation. Then, we make a comparison with other common structures and competitive works to prove the temporal superiority of our proposed structure. Finally, we analyze the impact of the ESTCNN framework on the spatial–temporal information with some brief discussions.

A. Overall Performance

The proposed ESTCNN is trained to detect fatigue states for each subject. The individual performances are obtained on eight subjects by 10-fold cross validation. For each subject, 90% of the samples are randomly selected for training and the remaining 10% are reserved for validation. We show the recognition effects with classification accuracy and STD on the validation sets. Fig. 4 presents the performance of the ESTCNN framework on the fatigue data set.

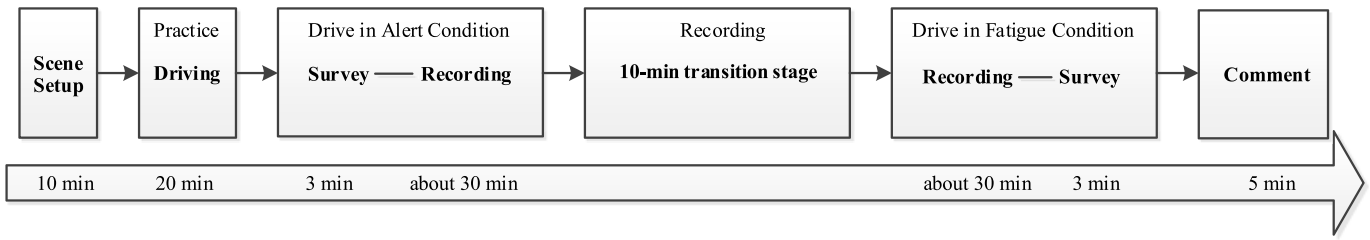

Fig. 3. Timeline of the experiment.

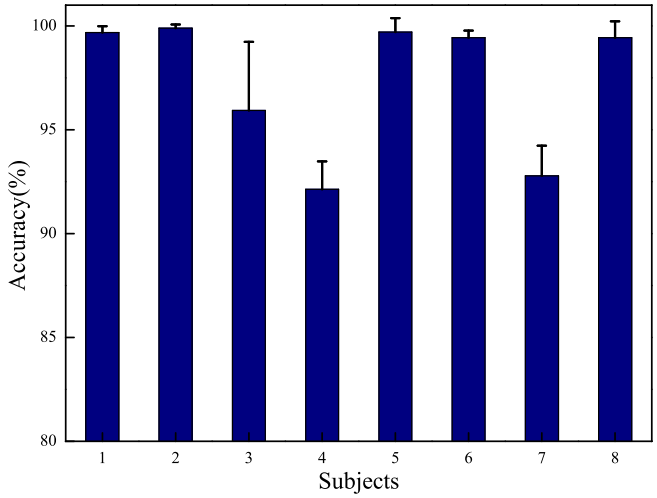

Fig. 4. Overall performance of the ESTCNN framework.

From Fig. 4, we find that the ESTCNN model is stably effective on the whole data set and all the accuracies surpass 92%. The average accuracy achieves 97.37% with STD 3.30%. Among them, the accuracy of five subjects is over the mean accuracy. Except for some indeed uncontrollable factors, the accuracy difference is possibly caused by subjects' physical conditions. By examining the recorded experiment videos, we found that S4 showed slight fatigue at the beginning of the experiment, but reported it after 30 min of the experiment. Just the reverse, S7 almost stayed alert throughout the experiment but reported slight fatigue half an hour before the end of the experiment. These two subjects did not well report the degree of subjective fatigue, which accounts for the lower accuracies of the two subjects. It suggests that, during carrying out the experiments, we should inform the subjects of reporting actual individual situations, which reflects that the proposed method has the robust ability to learn effective information from two categories of samples for recognition.

B. Method Comparison

From the overall performance, we find that our proposed method has effective performances, which should be due to the spatial–temporal structure of the framework. In order to deeply explore the spatial–temporal ability of the method on recognition tasks, we make a comparison with several commonly used structures and other existing works.

Actually, convolutional and recurrent layers are often used to extract the temporal dependencies in deep networks [31], [42]. Therefore, we give three baselines to verify the spatial–temporal extraction capabilities of ESTCNN. We also select other advanced EEG-based signal classification methods, including some feature-based methods, to check the performance of the proposed model. Here, we briefly introduce some of the details for these compared models.

1) *PSD-SVM:* It extracted the EEG power spectrum density features and used SVM classifier to determine the fatigue level.
2) *CSP-SVM:* It fed the relative energy of the filtered channels from CSPs methods into SVM classifiers, which was used to test the results in [28].
3) *LSTM:* It used the deep long short-term memory (LSTM) architecture for binary classification in [42], which consisted of two LSTM layers and a sigmoid activation function.
4) *CNN-A:* It replaced each core block with a 1×7 convolution layer and a max-pooling layer, then added a 1×3 convolutional layer and a global average pooling layer at the end.
5) *CNN-B:* It replaced each 1×7 convolution layer with three 1×3 convolution layers in CNN-A model. Other hyperparameters remained consistent.
6) *CNN-C:* It developed a four-layer CNN for spatial feature fusion and temporal feature extraction on steady-state visual evoked potential classification in [43].
7) *CNN-D:* It proposed a novel channel-wise CNN with raw EEG data on driver's cognitive performance prediction tasks in [28].
8) *FFT-CNN:* It transformed EEG signals into multispectral images and trained with a combined CNN and LSTM framework for classification in [30].

For an effective comparison, we select the eight above-mentioned methods, which are some baselines and typical works from different perspectives. There are some conventional methods for EEG analysis, such as PSD-SVM in [44]. It extracted power spectrum density features to decode time series while neglecting the spatial information of EEG signals. CSP-SVM considered the spatial patterns to discover discriminative information. DL methods brought some successful attempts to improve recognition performance taking full consideration for spatial–temporal properties. Therefore, we give three baselines to verify the superiority of our proposed framework with details

TABLE II
STRUCTURES OF THREE BASELINE METHODS. THE USAGE OF SYMBOL [] IS SAME AS ABOVE

CNN-A	CNN-B	LSTM
Input sample(30×100)		
[1×7, map 16, conv]	[1×3, map 16, conv]×3	LSTM
1×2, maxpool, stride2	1×2, maxpool, stride2	(units 100, dropout 0.5)
[1×7, map 32, conv]	[1×3, map 32, conv]×3	LSTM
1×2, maxpool, stride2	1×2, maxpool, stride2	(units 50, dropout 0.5)
[1×7, map 64, conv]	[1×3, map 64, conv]×3	–
1×2, maxpool, stride2	1×2, maxpool, stride2	
[1×3, map 2, conv]		Fc-50
Global avepool		Fc-2
Softmax		

shown in Table II, including LSTM, CNN-A, and CNN-B. We also choose five competitive works for further comparison, including PSD-SVM, CSP-SVM, CNN-C, CNN-D, and FFT-CNN. From the above, we select two feature-based methods and six DL frameworks to test and evaluate our developed algorithm.

According to the structure parameters presented in the original papers, we reproduce these methods to analyze our EEG data set. Note that we use the open-source code released from the FFT-CNN method. These above-mentioned are all designed for EEG signal analysis in the original papers. We apply these methods to all eight subjects with 10-fold cross validation and receive the mean accuracy of each method. Fig. 5 shows the recognition accuracies of the compared methods.

In order to investigate the importance of extracting temporal information, we take three baseline methods LSTM, CNN-A, and CNN-B for comparison. It is proven that the receptive field of a 5×5 kernel is equal to the receptive field of two stacked 3×3 kernels [45]. We replace the large 7×7 kernel with three small 3×3 kernels. The small kernel has advantages in many aspects, such as increasing the nonlinearity of model fitting and reducing the occupied memory of model. Thus, we build the third baseline method CNN-B. From Fig. 5, LSTM and CNN-A perform well and achieve 91.34% and 87.34%, respectively. It indicates that convolutional and recurrent layers are capable of extracting temporal information from EEG signals. By replacing large kernels with small kernels, CNN-B contributes to a slight increase in its performance, which reaches 92.73%.

The classification accuracy of all three baseline methods exceeds 87% with some differences, in which LSTM outperforms CNN-A by 4%. After increasing the model nonlinearity, CNN-B has a better performance than LSTM. Note that we employ global average pooling layer as the final layer to average every channel into a value in the CNN baselines. In the model modification, we reduce the degree of signal compression by using two values to represent each channel and receive a considerable improvement. The results indicate that two representative values are more robust to retain effective information across spatial dimensions. As the adjusted model, our ESTCNN provides an accuracy of 97.37%, which surpasses

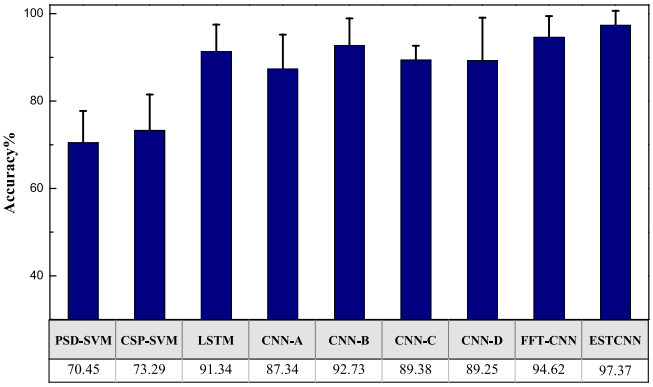

Fig. 5. Recognition accuracies of the compared methods.

CNN-B by over 4%. It suggests that the overall structure is beneficial for EEG-based fatigue evaluation, of which the core block could enhance temporal information extraction.

Some DL and non-DL methods have been conducted to explore different kinds of EEG analysis tasks. We employ the 10-fold cross validation to estimate the classification effects among ESTCNN and other competitive methods. Frequency features are often used in the above-mentioned studies, including PSD-SVM and FFT-CNN. Meanwhile, CSP method is also used to investigate stable patterns, so we take CSP-SVM into comparison. PSD-SVM gets the mean accuracy of 70.45% while CSP-SVM has a better performance of 73.29%. However, these two methods are worse than other DL-related methods, which reflects the robust capacity of DL methods on learning representations. The classification accuracy of these DL methods varies between 87% and 97%. Among the methods using the EEG signals as input, CNN-B and ESTCNN work well in the whole data set. Moreover, some prior knowledge-based DL methods also benefit a lot from prior information. FFT-CNN combined with FFT has a performance of over 94%. Note that in terms of STD, the STD of these eight methods varies from 9.84% to 3.28%. CNN-D receives the maximum variance of 9.84% and ESTCNN has the STD of 3.3%.

Among these eight methods, ESTCNN shows the best performance for driving fatigue evaluation tasks with considerable advantages compared with other methods. This suggests that our ESTCNN framework can robustly capture effective information using EEG signals from eight subjects, due to its good abilities for temporal dependencies extraction and spatial features fusion. Overall, the ESTCNN framework delivers an excellent performance on the fatigue data set.

C. Discussion

The above-mentioned comparison analysis shows that our ESTCNN framework performs best than other methods under the same validation scheme. As the ESTCNN can be taken as an integration of the core blocks and dense layers, it is novel to investigate the importance of the core block for extracting temporal dependencies and dense layer for spatial feature fusion. Comparing LSTM and CNN-A with ESTCNN, the core block in the ESTCNN framework performs better than conventional DL layers, while these layers or blocks

are targeted for extracting temporal dependencies. Meanwhile, comparing CNN-B with ESTCNN, the accuracy of ESTCNN is about 5% higher than CNN-B with lower deviations. These improvements of performance suggest the effectiveness of our proposed framework which learns both spatial and temporal information using EEG signals.

In the model evaluation, two typical indicators are the validate accuracy and standard variance. In comparison, our proposed ESTCNN method has considerable advantages in accuracy and variance. By applying the core blocks and dense layers to reduce temporal dimensions, the network is designed to learn representations from multichannel EEG signals. Due to the data-driven nature of the model, the developed ESTCNN model keeps enough flexibility to address many kinds of EEG-based recognition tasks. Compared to the above-mentioned approaches, ESTCNN model requires less preprocessing on the multichannel data and is more convenient to be implemented in the BCI online system.

V. Conclusion

Fatigued driving is a social problem that requires more attentions, and DL methods could promote the development of computational models for detecting task. In summary, we have developed a spatial–temporal CNN to detect driver fatigue from EEG signals. The proposed method shows significant improvements in the model performance and can learn more robust representations from EEG signals. The vital procedures are in two parts: first, we introduce the core block to deal with the information on the temporal dimension. Second, we utilize the dense layer to fuse the spatial features among the electrodes. We investigate the importance of spatial information and temporal dependencies by giving three baseline methods and five competitive studies. The results show that the performance is greatly improved due to the involvement of spatial–temporal information in EEG-based classification tasks.

It would be a great potential to extend the proposed method to numerous areas, such as multisource information fusion tasks. Further works would focus on the combination with feature-based methods to improve the model performance. Look forward, with the effectiveness and generality of the ESTCNN model, we expect it to be useful for broader applications in BCI systems.

References

[1] M. Tops and M. A. S. Boksem, "Absorbed in the task: Personality measures predict engagement during task performance as tracked by error negativity and asymmetrical frontal activity," *Cogn. Affect. Behav. Neurosci.*, vol. 10, no. 4, pp. 441–453, Dec. 2010.

[2] F. Racioppi, L. Eriksson, C. Tingvall, and A. Villaveces, *Preventing Road Traffic Injury: A Public Health Perspective for Europe*. Geneva, Switzerland: World Health Organ, 2004.

[3] D. F. Dinges, "An overview of sleepiness and accidents," *J. Sleep Res.*, vol. 4, no. 2, pp. 4–14, Dec. 1995.

[4] M. Karchani *et al.*, "Presenting a model for dynamic facial expression changes in detecting drivers' drowsiness," *Electron. Phys.*, vol. 7, no. 2, pp. 1073–1077, Apr./Jun. 2015.

[5] M. Fallahi, M. Motamedzade, R. Heidarimoghadam, A. R. Soltanian, and S. Miyake, "Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study," *Appl. Ergonom.*, vol. 52, pp. 95–103, Jan. 2016.

[6] J. A. Horne and S. D. Baulk, "Awareness of sleepiness when driving," *PsychoPhysiology*, vol. 41, no. 1, pp. 161–165, Jan. 2004.

[7] Y. M. Chi, Y.-T. Wang, Y. J. Wang, C. Maier, T.-P. Jung, and G. Cauwenberghs, "Dry and noncontact eeg sensors for mobile brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 2, pp. 228–235, Mar. 2012.

[8] S. K. L. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, Feb. 2001.

[9] C.-T. Lin *et al.*, "Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2469–2476, Nov. 2006.

[10] Z.-K. Gao, Q. Cai, Y.-X. Yang, N. Dong, and S.-S. Zhang, "Visibility graph from adaptive optimal kernel time-frequency representation for classification of epileptiform EEG," *Int. J. Neural Syst.*, vol. 27, no. 4, p. 1750005, Jun. 2017.

[11] Z. K. Gao, M. Small, and J. Kurths, "Complex network analysis of time series," *Europhys. Lett.*, vol. 116, no. 5, p. 50001, Jan. 2017.

[12] Z. K. Gao *et al.*, "An adaptive optimal-Kernel time-frequency representation-based complex network method for characterizing fatigued behavior using the SSVEP-based BCI system," *Knowl.-Based Syst.*, vol. 152, pp. 163–171, Jul. 2018.

[13] L.-L. Chen, Y. Zhao, J. Zhang, and J.-Z. Zou, "Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7344–7355, Nov. 2015.

[14] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 58–75, Jul. 2014.

[15] V. M. Y. Mervyn, X. Li, K. Shen, and E. P. V. Wilder-Smith, "Can SVM be used for automatic EEG detection of drowsiness during car driving?" *Saf. Sci.*, vol. 47, no. 1, pp. 115–124, Jan. 2009.

[16] J. Min, P. Wang, and J. Hu, "Driver fatigue detection through multiple entropy fusion analysis in an EEG-based system," *PLoS ONE*, vol. 12, no. 12, p. e0188756, Dec. 2017.

[17] W. L. Zheng and B.-L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *J. Neural Eng.*, vol. 14, no. 2, p. 026017, Feb. 2017.

[18] R. Chai *et al.*, "Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 715–724, May 2017.

[19] G. H. Klem, H. O. Lüeders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 52, no. 3, pp. 3–6, 1999.

[20] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2010.

[21] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Hong Kong, Jun. 2008, pp. 2390–2397.

[22] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[23] S. Lin and G. C. Runger, "GCRNN: Group-constrained convolutional recurrent neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4709–4718, Oct. 2018.

[24] Y. X. Yang *et al.*, "A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG," *Chaos*, vol. 28, no. 8, p. 085724, Aug. 2018.

[25] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[26] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, "Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 11, pp. 2030–2042, Nov. 2014.

[27] X. Sun, C. Qian, Z. Chen, Z. Wu, B. Luo, and G. Pan, "Remembered or forgotten?—An EEG-based computational prediction approach," *PLoS ONE*, vol. 11, no. 12, p. e0167497, Dec. 2016.

[28] M. Hajinoroozi, Z. Mao, T.-P. Jung, C.-T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Process., Image Commun.*, vol. 47, pp. 549–555, Sep. 2016.

[29] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[30] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," in *Proc. (ICLR)*, San Juan, Puerto Rico, Feb. 2016, pp. 1–15.
[31] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.
[32] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 47–54.
[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, Haifa, Israel, 2000, pp. 807–814.
[34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, Mar. 2015, pp. 448–456.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zürich, Switzerland, Oct. 2016, pp. 630–645.
[36] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng. (2017). "Cardiologist-level arrhythmia detection with convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1707.01836
[37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.
[38] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, Sardinia, Italy, Mar. 2010, pp. 249–256.
[39] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. COMPSTAT*, Paris, France, 2010, pp. 177–186.
[40] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci.*, vol. 52, nos. 1–2, pp. 29–37, Jul. 1990.
[41] 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.
[42] R. G. Hefron, B. J. Borghetti, J. C. Christensen, and C. M. S. Kabban, "Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation," *Pattern Recognit. Lett.*, vol. 94, pp. 96–104, Jul. 2017.
[43] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS ONE*, vol. 12, no. 2, p. e0172578, Feb. 2017.
[44] X. Zhang *et al.*, "Design of a fatigue detection system for high-speed trains based on driver vigilance using a wireless wearable EEG," *Sensors*, vol. 17, no. 3, p. 486, Mar. 2017.
[45] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

Zhongke Gao received the M.Sc. and Ph.D. degrees from Tianjin University, Tianjin, China, in 2007 and 2010, respectively.

Since 2016, he has been a Full Professor with the School of Electrical and Information Engineering, Tianjin University, where he is currently the Director of the Laboratory of Complex Networks and Intelligent Systems. His current research interests include deep learning, EEG analysis, complex networks, brain–computer interface, and wearable intelligent devices.

Xinmin Wang received the bachelor's degree in automation from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in 2017, where he is currently pursuing the master's degree in control science and engineering with the School of Electrical and Information Engineering.

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

Yuxuan Yang received the bachelor's degree in automation from Anhui University, Hefei, China, in 2014, the master's degree in automation from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in 2017, where she is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering.

Her current research interests include brain–computer interface, machine learning, and complex networks.

Chaoxu Mu received the B.S. degree from the Harbin Institute of Technology, Harbin, China, in 2006, the M.S. degree from Hohai University, Nanjing, China, in 2009, and the Ph.D. degree in control science and engineering from Southeast University, Nanjing, in 2012.

Her current research interests include computational intelligence, nonlinear system control and optimization, adaptive and learning systems, smart grid, and machine learning.

Qing Cai received the B.Sc. degree in automation from Liaoning Shihua University, Liaoning, China, in 2014. She is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China.

Her current research interests include EEG, fMRI, brain network, and complex networks.

Weidong Dang received the B.Sc. degree in automation from Tianjin University, Tianjin, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering.

His current research interests include sensor design, multisource information fusion, measurement science and technology, multiphase flow, and complex networks.

Siyang Zuo received the M.Eng. and Ph.D. degrees in information science and technology from the University of Tokyo, Tokyo, Japan, in 2009 and 2013, respectively.

He is currently a Professor with the Key Laboratory of Mechanism Theory and Equipment Design, Ministry of Education, School of Mechanical Engineering, Tianjin University, Tianjin, China. His current research interests include medical robotics and imaging techniques.