Detection of Train Driver Fatigue and Distraction Based on Forehead EEG: A Time-Series Ensemble Learning Method

Chaojie Fan, *Student Member, IEEE*, Yong Peng, *Member, IEEE*, Shuangling Peng, Honghao Zhang, *Member, IEEE*, Yuankai Wu, *Member, IEEE*, and Sam Kwong, *Fellow, IEEE*

***Abstract*— Train driver fatigue and distraction are the main reasons for railway accidents. One of the new technologies to monitor drivers is by using the EEG signals, which provides vital signs monitoring of fatigue and distraction. However, monitoring systems involving full-head scalp EEG are time-consuming and uncomfortable for the driver. The aim of this study was to evaluate the suitability of recently introduced forehead EEG for train driver fatigue and distraction detection. We first constructed a unique dataset with experienced train drivers driving in a simulated train driving environment. The EEG signals were collected from an EEG recording device placed on the driver's forehead, and numerous features including energy, entropy, rhythmic energy ratio and frontal asymmetry ratio were extracted from the EEG signals. Therefore, a time-series ensemble learning method was proposed to perform fatigue and distraction detection based on the extracted feature. The proposed method outperforms other popular machine learning algorithms including Support Vector Machine(SVM), K-Nearest Neighbor(KNN), Decision Tree(DT), and Long short-term memory(LSTM). The proposed method is stable and convenient to meet the real-time requirement of train driver monitoring.**

***Index Terms*— Train driver, fatigue and distraction, forehead EEG, time-series ensemble learning method.**

Manuscript received 30 January 2020; revised 26 August 2020, 7 December 2020, 27 February 2021, and 14 July 2021; accepted 1 November 2021. Date of publication 18 November 2021; date of current version 9 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 52075553 and Grant 52105523, in part by the Natural Science Foundation of Hunan under Grant 2020JJ7030, in part by the Fundamental Research Funds for the Central Universities of Central South University under Grant 2019zzts553, and in part by the Postgraduate Scientific Research Innovation Project of Hunan Province under Grant CX20210099. The Associate Editor for this article was C. Wu. *(Corresponding authors: Yong Peng; Sam Kwong.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Xiangya No.2 Hospital of Central South University Institutional Review Board under Application No. S115.

Chaojie Fan is with the Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410000, China, and also with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: fcjgszx@csu.edu.cn).

Yong Peng is with the Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410000, China (e-mail: yongpeng@csu.edu.cn).

Shuangling Peng is with the School of Rolling stock, Guangzhou Railway Polytechnic, Guangzhou 510000, China (e-mail: pengshuangling@gtxy.edu.cn).

Honghao Zhang is with the School of Mechanical Engineering, Shandong University, Jinan 250061, China (e-mail:honghao_zhang@sdu.edu.cn).

Yuankai Wu is with the Department of Civil Engineering, McGill University, Montréal, QC H3A 0G4, Canada (e-mail: kaimaogege@gmail.com).

Sam Kwong is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: cssamk@cityu.edu.hk).

Digital Object Identifier 10.1109/TITS.2021.3125737

I. Introduction

INATTENTION of train driver, including fatigue and distraction, has been one of the main causes of railway traffic accidents. The inattention of train driver directly causes the "signal passed at danger" (SPAD), which is the largest safety breach in railway [1]. European researchers collected railway traffic accident data of 24 different countries from 1997 to 2011 and analyzed the factors responsible for accidents [2]. Factors 'Fatigue' and 'Distraction' accounted for 10.4% and 23.1% of total accidents, respectively. It is worth mentioning that the proportion of serious accidents caused by fatigue, 16.5%, which is twice that of less serious incidents.

Nowadays, Automatic Train Operation (ATO) is an exciting technology. In order to improve the capacity of the railway system, the European Railway Agency suggested focus on the implementation of Grades of Automation-2(GoA-2) particularly [3]. The GoA-2 means semi-automatic train operation (STO). Though the speed control is automated the train drivers are still responsible for safety and punctuality, so they can take over control at any time [3]. The train drivers are required to be more concentrated and keep look-out status to keep safety. However, due to the GoA-2 semi-automatic train operation system, the participation of train drivers decreases. They are in a monotonous monitoring state for a long time, and need to spend more energy to stay alert, which is easily to cause fatigue and distraction [3].

To sum up, whether now or in the future, train driver fatigue and distraction are the main factors that threaten the safety of railway. Therefore, the detection of train driver fatigue and distraction in real-time is necessary. In this study, a time-series ensemble learning method using forehead EEG is proposed to detect train driver fatigue and distraction continuously. The detection model is trained to identify three driving status alert, fatigue, and distraction.

A. Related Work

In the field of transportation engineering, the detection methods can be roughly divided into four categories:

1) Vehicular Behavior: Vehicle-data-based detection technology is realized by detecting the driver's operation behaviors on the steering wheel and the pedal, along with the speed, acceleration, and lane offset of the vehicle [4]–[6].

In [7], depending on vehicle position, velocity, and acceleration throttle and brake pedal positions, an support vector machines (SVM) model is used to classify distracted driving and normal driving. However, vehicle-based detection technology has certain limitations with respect to the train drivers. Because railway tracks are fixed and railway automation is extensive, the main duty of train driver is to keep alert to prepare for possible danger. [8].

2) Driver Behavior: According to related video information, many driver behavior features such as ocular activities, mouth activities, and head pose are used to estimate the level of driver fatigue and distraction [4], [5], [9]–[11]. However, facial expressions and driver bahavior change minimally when cognitive distraction occurs, which is considered an intrinsic state [12].

3) Periphery Physiological Signals: The periphery physiological signals include ECG signals, SEMG signals, EDA signals and so on. Recent research has proposed various methods to extract features from the physiological signals for fatigue and distraction detection [4], [5]. In [13], the author extracts physiological signal features including interbeat interval (IBI) and heart rate variability (HRV) from ECG, as well as the interval between skin conductance responses (SCRs) from EDA. To some extent, these physiological signal features can capture cognitive distractions, but the detecting stability and accuracy is not enough.

4) Electroencephalogram (EEG): Due to the superior performance of detecting brain information, EEG is called the gold standard for detecting driver fatigue and distraction [4]. Wu *et al.* [14] used EEG signal and deep learning model to detect fatigue status of pilot, and got good performance. In [15], four EEG band waves (including delta, theta, alpha and beta) are evaluated for 52 subjects during monotonic driving. The results show the ratios of the slow waves and fast waves increasing over driving time. For distraction detection, in [16], in simulated driving experiments, the EEG theta and beta power increase briefly after unexpected car deviations and the onset of mathematical questions, which proves that driving distraction can also be estimated from EEG signals. In summary, EEG had the highest and most robust performance for detecting driver fatigue and distraction. However, conventional wet EEG acquisition technology requires the coating of gel in a gel-based cap, which is inconvenient for application and stains the driver's hair. The dry EEG acquisition technology can acquire EEG signals rapidly and reliably without gel by using the dry electrodes [17]. It can't be neglected that using the dry cap is uncomfortable due to the dry electrode must be firmly attached to the scalp. Therefore, it cannot meet the requirements of long-term detection.

However, as aforementioned, vehicular behavior, driver behavior, and periphery physiological signals can't meet the requirement of detecting train driver fatigue and distraction stably and accurately. Although the EEG can fit the bill, the clumsy EEG cap is too troublesome to fill the practical usage. The forehead is the only hairless area of human head, so that forehead EEG can be collected by simple electrodes which can be put in the headband, even glasses [18]. Whereas, there is no research exploring whether the EEG signals from forehead area provide enough information to detect fatigue and distraction simultaneously. To fill the gaps mentioned above, this work proposes a practical and robust train driver fatigue and distraction detection method using forehead EEG signals.

B. Contribution

The contributions of the paper can be summarized as follows:

First, to the best of the authors' knowledge, this is the first study which proposed a stable and feasible solution for simultaneously detecting the fatigue and distraction of train drivers.

Second, a novel "future feasible solution" approach is proposed to identify train driver fatigue and distraction by using forehead EEG. Our method is low-cost and is easier to be used in real-world applications.

Third, the proposed time-series ensemble method can improve the detection performance significantly with less time consumption. The proposed algorithm only trains base model once and directly combines classifier results from base model in temporal dimension.

C. Paper Organization

The organization of this paper is as follows: Section II explains in detail the experimental design. Section III discusses the proposed method. Section IV describes the results. Section V discusses the results. Finally, Section VI concludes the paper.

II. Design of the Experiment

A. Subjects

A total of 7 professional train drivers from the China Railway Guangzhou Railway Group Co., Ltd. were recruited to conduct simulated driving experiments. All of the subjects were right-handed men with ages ranging from 23 to 39, which basically accorded with the age distribution of train drivers in China [19]. The subjects were healthy and had no history of genetic disease. They did not receive medical treatment for disease before participating in the experiment. Drugs and alcohol were prohibited during the 24 hours before the experiment. Moreover, except for the fatigue task, the subjects were instructed to ensure sufficient rest before the experiment. This Project was approved by Xiangya No.2 Hospital of Central South University Institutional Review Board.

B. Apparatus

The experimental equipment primarily includes a simulated train driving system and a biological multichannel recorder. The simulated train driving system was developed by YUNDA, which is the designated supplier of CHINA RAILWAY. This stimulated system had been validated by CHINA RAILWAY [20]. The simulated train driving system includes train console, driver's seat, driver controller (traction/brake/reverse handle), control panel, LKJ display, and CIR display. These components were included in every actual train operating systems except for PIS phones and driver pedals. Among them, the high-definition display was used to simulate track and environmental information to create a realistic driving environment for the subjects. The biological multichannel recorder used in the experiment was the MP150 model provided by BIOPAC Company. The host had 16 analog data acquisition channels, 16 digital input and output channels, 16-bit A/D conversion rate, and 0.0005 Hz to 400 KHz sampling rate. The BioNomadix EEG wireless transmitters transmitted the acquired EEG signals.

C. Procedure

The primary task was the simulated train driving task on the designated line. The experimental line is the northbound section of the Beijing-Guangzhou line, and the task is initiated at Hengyang station. The train runs under the supervision of the train operation monitoring system (LKJ) and meets the requirements of Chinese railway regulations. The full speed of the train was maintained at 120 km/h. Fig. 2 "experiment" part shows that the simulated train driving task was completed by the subjects. All experiments started at 9:00 in order to avoid the influence of circadian rhythms. Before the experiments, the subjects practiced driving for approximately 5 minutes on the simulated train driving system, although they had used the system for training previously. The Stanford Sleepiness Scale (SSS) [21] was used to ascertain self-reported sleepiness before the experiments. As required, each subject needs to finish three tasks.

D. Tasks

Tasks were designed to simulate the primary fatigue and distraction types of the train driver, which are different from those of other professional drivers. The corresponding analysis and task designing are as follows:

1) Fatigue Task: Fatigue is a complex condition characterized by lack of alertness and drowsiness. Researchers generally divided driving fatigue into physical fatigue and mental fatigue [4]. The former is caused by heavy manual labor, while the latter is caused by long work durations or sleep deprivation. The characteristics of train driving includes high automaticity, sample driving operations, and monotonous train-driving scenarios, and high demand for concentration [1]. Therefore, mental fatigue is the main driving fatigue state among train drivers.

In this study, the mental fatigue of our subjects was caused by long-time driving without enough rest. The Chinese train drivers must perform duty during night shifts, which requires driving trains for 3 hours or more at midnight. Usually, they are forced to rest. For experiments required, the fatigue task subjects need to finish 30-minute simulated driving after the night shifts without rest. Furthermore, to ensure that all of the subjects were fatigued, the subjects were required to report the scores of SSS every two minutes. The fatigue task started only when the self-report SSS scores reached four or more.

2) Distraction Task: Driver distraction was defined as 'the diversion of attention away from activities critical for safe driving towards a competing activity' [22]. Distractions can be divided into the following three categories according to source and demand: Visual distractions (e.g., looking around); Manual distractions (e.g., raising a hand from steering); Cognitive distractions (e.g., thinking about something not related to driving) [23]. In this study, auditory distraction is considered as a cognitive distraction, as defined in the literature [12]. Train drivers are not allowed to take any non-work related objects, such as telephones, books, or food, which are the main factors that cause manual distractions during driving work. Moreover, due to the existence of the real-time video monitoring system, the train drivers are required to maintain the lookout position and driving attitude at all times. Therefore, the management system greatly reduces the possibility of visual distraction and manual distraction. However, the strict rules cannot completely eliminate driver cognitive distraction phenomena. In an Australian incident investigation report [24], the driver had been thinking about things which were not related to driving, such as his upcoming wedding, when the accident happened. Cognitive distraction is still inevitable in the group of train drivers.

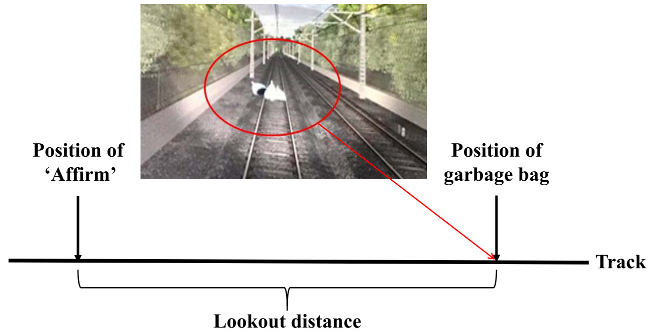


Fig. 1. Definition of lookout distance.

In this study, the auditory math task was designed to make subjects cognitively distracted during driving, without any visual and manual distraction [25]. In this task, the subjects heard mathematical questions continuously and provided the right answers immediately. The time of the distraction task was also 30 minutes.

3) Alert Task: Subjects completed 30 minutes of the simulated train driving task in an alert state with sufficient rest before the experiment.

E. Driving Performance

The lookout distance(LD) is applied to measure the driving performances of train drivers under different tasks and to prove the effectiveness of our experimental task division. It refers to the longest sight distance in the train simulation driving task. During the train simulation driving task, garbage bag obstacles appear randomly. The driver is required to press the 'Affirm' button on the LKJ system for the first time when an obstacle is found. The system records the 'Affirm' time and position. The distance between the 'Affirm' position and the actual position of the obstacle was defined as LD. The random garbage bag obstacle in the task and the definition of lookout distance was shown in Fig.1.

F. EEG Collection

The raw EEG data was collected from the forehead, which is the only hairless site on the scalp. Therefore, EEG electrodes

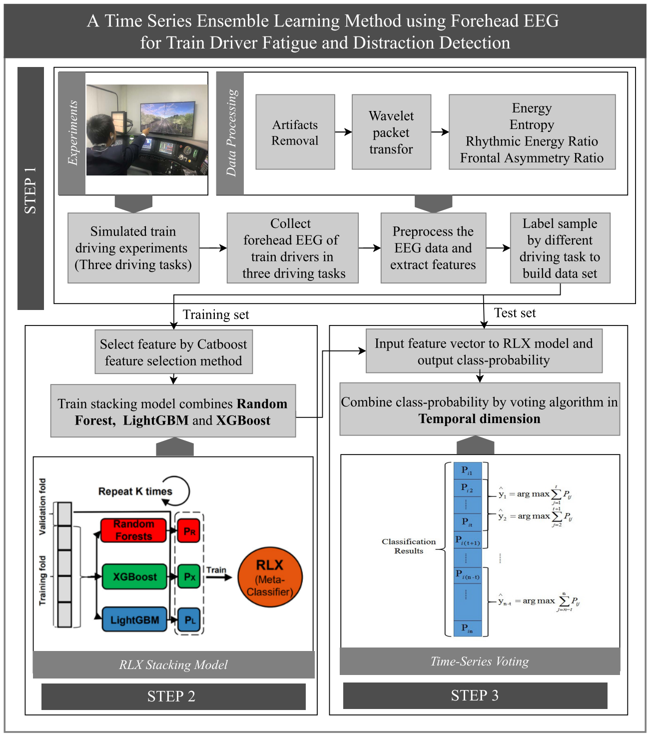


Fig. 2. Detailed flowchart of the proposed detection method.

were placed according to the International 10–20 System: Fp1, Fp2. In this study, A1–A2 was chosen as the reference electrode.

III. Method

The detailed procedure of this study can be summarized in three steps, which is described by the detailed flowchart shown in Fig. 2. The framework consists of three steps. Step 1 is to build train driver fatigue and distraction data set. Step 2 is to select feature and train the RLX model. And step 3 involves the time-series voting method.

Two EEG electrodes were placed on subjects' forehead to collect forehead EEG. They need to finish simulated train driving experiments with three different tasks. After removing artifacts and extracting rhythm waves, the EEG signals were divided into 0.5s-long samples. For each sample, four types of features were extracted: 1) energy; 2) entropy; 3) rhythmic energy ratio; 4) frontal asymmetry ratio, and the total number of features reached 97. The samples which were collected in alert, fatigue and distraction task were labeled as 'alert', 'fatigue' and 'distraction' respectively. Each sample from the data set has 97 features and one label. Furthermore, the dataset was split into training and test datasets. In the training phase (Step 2), we used CatBoost and Prediction Values Change (PVC) method to evaluate the feature importance, and excluded the useless features from the feature set. And then Random Forest, LightGBM, and XGBoost were combined by StackingCV to train base model named RLX model. Finally, in Step 3, the feature vector of each sample was input to the RLX model, and got the category probability value of each sample. In the temporal dimension, the category probability value of continuous samples was combined with the voting algorithm to obtain the final detection result. The followings are details.

A. Data Preprocessing and Feature Extraction

The raw EEG signal was processed through the bandpass filters, which were set as 0.5-60 Hz, and processed to the 50 Hz notch filter. The artifacts were then removed by adaptive filtering of the EEG signal. Finally, downsampling to 128 Hz was performed from the sample recording rate of 500 Hz used during the hardware phase.

Sliding windows with length of 0.5 s were employed to analyze experimental data using the wavelet packet transform to extract five rhythm waves from EEG data. In addition, four types of features were extracted: 1) energy; 2) entropy (containing Shannon entropy, approximate entropy, sample entropy, and combined entropy); 3) rhythmic energy ratio; 4) frontal asymmetry ratio (feature ratio of Fp1/Fp2). The total number of features reached 97.

1) Wavelet Packet Transform: Wavelet packet transform (WPT) provides multiresolution and time-frequency analysis for nonstationary EEG data. In the WPT, we acquire new lower resolution approximation spaces plus detailed spaces by decomposing both the approximation space and detail space. Let $\psi(t)$ and $\phi(t)$ be the corresponding mother wavelet function and the scaling function in the WT, respectively, and $\psi^0(t) = \psi(t)$, $\phi^0(t) = \phi(t)$. We can then construct the following wavelet basis according to Equation (1) [26]:

$$\psi_{j,k}^{2i} = \frac{1}{\sqrt{2}}\psi^{2i}\left(\frac{2^j k - t}{2^j}\right) = \sum_n h(n)\psi_{j-1,2k-n}^{i}(t) \tag{1}$$

$$\psi_{j,k}^{2i+1} = \frac{1}{\sqrt{2}}\psi^{2i+1}\left(\frac{2^j k - t}{2^j}\right) = \sum_n g(n)\psi_{j-1,2k-n}^{i}(t) \tag{2}$$

where i is the node number, j is the level of decomposition, and $h(n)$ and $g(n) = (-1)^{1-n}h(1-n)$ are a pair of quadrature mirror filters. The wavelet transform coefficients of a given function at the jth level and kth point are computed via the following recursion equations [27]:

$$d_j^{2i}(k) = \int f(t)\psi_{j,k}^{2i}(t)dt = \sum_n h(n)d_{j-1}^{i}(2k-n) \tag{3}$$

$$d_j^{2i+1}(k) = \int f(t)\psi_{j,k}^{2i+1}(t)dt = \sum_n g(n)d_{j-1}^{i}(2k-n) \tag{4}$$

Letting the original signal has 2^N sample points, the complete reconstruction of $f(t)$ can be expressed as follows:

$$f(t) = \sum_{i=o}^{2^{j-1}-1}\sum_{k=o}^{2^{N-1}-1} d_j^{2i}(k)\psi_{j,k}^{2i}(t) + \sum_{i=o}^{2^{j-1}-1}\sum_{k=o}^{2^{N-1}-1} d_j^{2i+1}(k)\psi_{j,k}^{2i+1}(t) \tag{5}$$

According to the frequency band of the desired signal, the corresponding wavelet decomposition was selected, and Equation (5) was used to reconstruct the desired signal. The EEG data was decomposed into delta (0-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-32 Hz) waves by 6-layer wavelet packet.

2) Energy: Because the EEG rhythms of human beings are different when they are in inattentive or normal states, the energies of different EEG rhythms were essential features for driver fatigue and distraction detection [28]. The energies of the different EEG rhythms in the sliding window were defined according to Equation (6):

$$E_i = \sum_{i=1}^{n} [X_i]^2 \tag{6}$$

where X_i is the amplitude.

3) Shannon Entropy: Shannon entropy (SE) is used to measure the uncertainty of systems [29]. The higher the uncertainty of EEG signals are, the higher the Shannon entropy is. The Shannon entropy H_n of variable X can be found according to Equation (7):

$$H(X) = -\sum_{i=0}^{N-1} p_i log_2 p_i \tag{7}$$

where p_i is the probability of X.

4) Approximate Entropy: Approximate entropy (AE) is a nonlinear dynamic parameter that measures the complexity of a sequence [30]. Approximate entropy offers the advantages of strong anti-noise and anti-interference abilities, which was applied to unsteady random signals such as the EEG signals.

$$\begin{aligned} ApEn(m, r, N) = {} & \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C_i^m(r) \\ & - \frac{1}{N-m} \sum_{i=1}^{N-m} \log C_i^{m+1}(r) \end{aligned} \tag{8}$$

$$C_i^m(r) = \frac{B_i}{N-m+1} \tag{9}$$

where B_i is the number satisfying $d\,|X(i), X(j)| \leq r$, and $d\,|X(i), X(j)|$ is the distance between $X(i)$ and $X(j)$. m is the embedded dimension, N is the number of data points in the phase space, and r is the tolerance parameter. This paper takes $m = 2$ and $r = 0.2 * SD$.

$$X(i) = [x(i), x(i+1), \cdots, x(i+m-1)], X(i) \in R^m \tag{10}$$

$$X(j) = [x(j), x(j+1), \cdots, x(j+m-1)], X(j) \in R^m \tag{11}$$

5) Sample Entropy: Sample entropy is a new time series complexity measure which is different from Shannon entropy and approximate entropy [28]. It can be represented by $SampEn(m, r, N)$, where N is the length, r is the tolerance parameter, and m or $m+1$ is the dimension. The purpose of sample entropy is to reduce the error of the approximate entropy, which is more closely related to the known random signal than the approximate entropy.

$$SamEn(m, r, N) = -\ln \frac{B^{m+1}(r)}{B^m(r)} \tag{12}$$

$$B^m(r) = \frac{1}{N-mr} \sum_{i-1}^{N-m} C_i^m(r) \tag{13}$$

6) Combined Entropy: Some studies have shown that combined entropy applied to driver fatigue and distraction detection [31]. The combined entropy (CE) is the sum of Shannon entropy, Approximate entropy, and Sample entropy. $CE = SE + ApEn + SampEn$.

7) Rhythmic Energy Ratio: The rhythmic energy ratio is the ratio of different EEG rhythm wave energies. Some studies found that the energy ratio can be used to detect driver fatigue and distraction [32]. In this study, we calculated 4 ratios of energy as follow: E_α/E_θ, $/E_\beta/E_\delta$, E_α/E_β, E_θ/E_β, $(E_\theta + E_\alpha)/E_\beta$ and $(E_\theta + E_\alpha)/(E_\alpha + E_\beta)$. In addition, special energy ratios called related energy were considered, including E_δ/E, E_θ/E, E_α/E, E_β/E, and E_γ/E. The E represents the total energy of five rhythm waves: $E = E_\delta + E_\theta + E_\alpha + E_\beta + E_\gamma$

8) Frontal Asymmetry Ratio: The asymmetry of frontal EEG refers to the difference of brain activity between left and right frontal lobe. According to some neuropsychology researches, the frontal EEG asymmetry is an important indicator of emotion and cognition [33]–[35]. In this study, we used the features ratio of Fp1 and Fp2 to calculate the frontal EEG asymmetry. The proposed frontal EEG asymmetry features is a new feature extraction method to fill the requirement of cognitive distraction detection.

B. CatBoost Feature Selection (CatB-FS)

Feature selection can reduce the comparison time and overfitting in machine learning. The CatB-FS algorithm returns a ranking of the features by training a CatBoost model, which is a new gradient boosting algorithm based on the decision tree [36]. According to the feature importance rank, we excluded the useless features from the dataset. We used the prediction values change (PVC) method to evaluate the feature importance. For each feature, PVC shows the average changes in the prediction when the feature value changes. If this feature is changed, then the larger the average change to the prediction value, the greater the feature importance of this feature. The detailed calculation principle is as follows:

Leaf pairs that are compared have different split values in the node on the path to these leaves. If the split condition is met (this condition depends on the feature F), then the object goes to the left subtree; otherwise, it goes to the right one. The calculation equation for the feature importance of feature F is as follows:

$$FI_F = \sum_{trees, leafs_F} (v_1 - avr)^2 \cdot c1 + (v_2 - avr)^2 \cdot c2 \tag{14}$$

$$avr = \frac{v_1 \cdot c_1 + v_2 \cdot c_2}{c_1 + c_2} \tag{15}$$

where c_1 and c_2 represent the total weight of objects in the left and right leaves, respectively. This weight is equal to the number of objects in each leaf. v_1 and v_2 represent the formula values in the left and right leaves, respectively.

C. Ensemble Method

Due to the complexity of EEG signals, it is difficult to build an EEG-based driver fatigue and distraction detection model through conventional statistical methods. Machine learning classifiers extract information from noisy data and are trained without prior knowledge. Based on the advantages, ensemble learning as a part of machine learning has attracted increasing attention due to powerful classification performance in different areas. All ensemble methods used in this study can be briefly explained as follows:

1) Bagging Algorithm: Bagging [37] is an integrated technique which trains multiple base learners, each from a different bootstrap sample. As a variant of bagging, random forest (RF) is one of the most potent ensemble learning methods in the world [37]. RF chooses the CART decision tree as the base learner: unlike conventional bagging, it uses randomly selected features to train base learners.

2) Boosting Algorithms: Boosting algorithms can iteratively adjust the weights of the training set instances and the weights of the base learning algorithms. Because of its efficiency, accuracy, and interpretability, the gradient boosting decision tree (GBDT) has become the widest-used boosting algorithm. In this study, two novel GBDT algorithms were chosen.

- XGBoost: XGBoost is a novel sparsity-aware algorithm for sparse data and weighted quantile sketching for approximate tree learning [38].
- LightGBM: LightGBM is a distributed and efficient boosting algorithm which uses two novel techniques, gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB), to solve the problem of time consumption [39].

3) Stacking Algorithm: Stacking is an ensemble learning method to combine multiple classification models via a meta-classifier [40]. For reducing overfitting, stackingCV, which uses cross-validation to prepare the input data for the level-2 classifier, was chosen in this study. The second layer classifier is logistic regression.

4) Time-Series Voting Algorithm: In this study, the features were extracted from the train driver EEG data with 0.5s sliding windows. However, EEG has been considered as typical time-series data with the time-relativity of the neighboring data. And the driving states were not instantaneous states, but long-term performances. The 0.5 s predicted results were usually not sufficient for detecting the driving states with high accuracy. In this study, a time-series voting algorithm was presented. After input feature vector, the RLX model output the class-probability values of different classes per sample. We combined the class-probability values of continuous samples in temporal dimension by voting algorithm. The class labels are predicted by averaging the class-probabilities. The voting formula is as follow:

$$\hat{y_j} = \arg\max_i \sum_{j}^{j+t} P_{ij} \tag{16}$$

where t is the number of combined samples, P_{ij} is the probability of j_{th} sample belong to i class.

IV. Results

A. Driving Performance and SSS Score Results

The LD was calculated to reflect the driving performance of a train driver. Longer LD indicates better driving performance. In this study, subjects exhibited different lookout distances with respect to the different tasks. In the alert task, the mean of lookout distance was 427.75 m; the mean values were 330.82 m and 272.56 m in the distraction task and fatigue task respectively. There was a significant difference between the alert task and distraction task, which also existed between the alert task and fatigue task. The SSS scores have a high validity for measuring fatigue. In the alert task, the mean of lookout distance was 1.55; the mean values were 1.55 and 4.7 for the distraction task and fatigue task respectively. The SSS scores of the alert task and distraction task were significantly lower than those of the fatigue task. All of the results are shown in Table I.

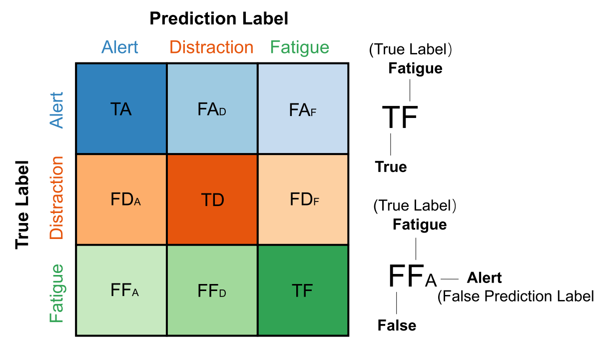


Fig. 3. Multiclass classification confusion matrix and symbol definition.

B. Feature Importance

The feature importance was obtained by the CatB-FS algorithm. As Table II. shows, the total feature importance values of the four feature types were 56.68, 12.32, 16.46, and 14.54 respectively. The features whose feature importance less than 0.3 were excluded in our datasets because of their negligible contributions. The new features set were only 41 in number, which is smaller than before. The comparison of detection performance is presented in Table II.

C. Performance of RLX and Time-Series Voting

This study is a multiclass classification problem. We referred to the evaluation criteria of dual-classification, and we proposed some evaluation criteria. Fig. 3 and Equations (17)-(20) show the definition of the evaluation criteria used in this study.

$$Accuracy = \frac{TA + TD + TF}{Total} \tag{17}$$

$$\begin{aligned} Total = & TA + TD + TF \\ & + FA_D + FA_F + FD_A + FD_F \\ & + FF_A + FF_D \end{aligned} \tag{18}$$

$$Recall_{Alert} = \frac{TA}{TA + FA_D + FA_F} \tag{19}$$

$$Prediction_{Alert} = \frac{TA}{TA + FD_A + FF_A} \tag{20}$$

Support vector machines (SVM), K-nearest neighbor (KNN), and decision tree (DT) were selected as the baseline models due to their stable and excellent performances in EEG

TABLE I
LOOKOUT DISTANCE AND SSS SCORES

	Alert	Distraction	Fatigue	P-value (Alert -Distraction)	P-value (Alert -Fatigue))
LD (m) Mean (Std)	427.75 (9.92)	330.82 (6.25)	272.56 (5.59)	<0.01[a]	<0.01[a]
SSS Mean (Std)	1.43 (0.53)	1.29 (0.49)	4.57 (0.79)	0.59[a]	<0.01[a]

[a] The Kruskal-Wallis test was used to obtain P-values.

TABLE II
FEATURE IMPORTANCE RESULTS

Feature name	Energy	Entropy	Rhythmic Energy Ratio	Frontal Asymmetry Ratio	Total
Original Num_F[a]	10	40	18	29	97
Selected Num_F[a]	10	14	12	5	41
Origin Imp_F[b]	56.68	12.32	16.46	14.54	100
Selected Imp_F[b]	56.68	8.86	15.67	11.00	92.22

[a] Num_F is the abbreviation for feature number.
[b] Imp_F is the abbreviation for feature importance.

signal classification [8],[40],[41]. In this study, due to the lack of the proper pre-trained models, we did not do any fine-tuning before training the detection models. All models used the same labeled data set and the grid-search and 5-fold cross-validation were used to find optimal hyper-parameter of them. We also applied 5-fold cross-validation to test the detection model performance.

Support vector machines (SVM), K-nearest neighbor (KNN), and decision tree (DT) were selected as the baseline models due to their stable and excellent performances in EEG signal classification. [8], [41], [42].In this study, due to the lack of the proper pre-trained models, we did not do any fine-tuning before training the detection models. All models used the same labeled data set and the grid-search and 5-fold cross-validation were used to find optimal hyper-parameter of them. We also applied 5-fold cross-validation to test the detection model performance.The performance comparison results are listed in Table III. Compared with the baseline models, the RLX model presented the best average detection accuracy achieved: 79.35%. The RLX model with CatB-FS presented slightly better detection performance than RLX without CatB-FS. The detection accuracies between different cross-validations ranged from 78.13% to 81.91%, which showed that our proposed method exhibited effective detection stability with respect to different datasets.

Fig. 4 additionally shows the accuracy of different base model which used time-series voting algorithm and Long short-term memory (LSTM) network. LSTM is a widely used deep network in the field of time series data analysis, which is an artificial recurrent neural network (RNN) architecture. Unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only single data points, but also entire sequences of data. LSTM networks are well-suited to classify, process and make predictions based on

TABLE III
DETECTION PERFORMANCE COMPARISON OF BASE MODEL

Model	Cross-Validation Accuracy (%)						
	CV1	CV2	CV3	CV4	CV5	Mean	Std
SVM	58.06	58.61	58.49	61.33	53.51	58.60	1.74
KNN	58.99	63.05	60.50	65.56	61.21	61.86	2.53
DT	70.39	70.87	71.87	74.70	72.18	72.00	1.67
RLX without CatB-FS	77.81	77.91	78.80	**82.05**	79.33	79.18	1.72
RLX with CatB-FS	**78.13**	**78.30**	**78.89**	81.91	**79.51**	**79.35**	**1.53**

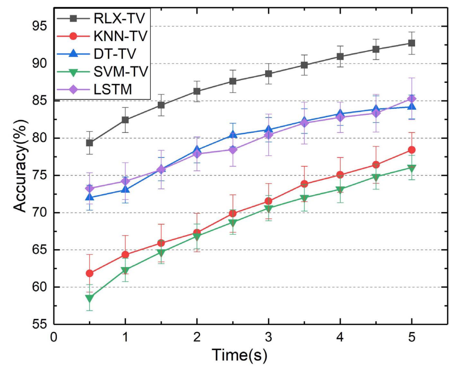


Fig. 4. Detection accuracy of different base models in different time.

time series data, since there can be lags of unknown duration between important events in a time series [43], [44]. And we can find that the RLX-TV model have the best performance between all the comparative models in all length of time. The RLX-TV detection accuracy achieved was 92.7% for 5s fusion time.

D. Performances of Different Subjects and Different Status

Table III presents the detailed detection performance of our proposed model for a 5s fusion time. The mean detection accuracy achieved was 92.73%. The recall and precision of alert detection were 92.92% and 91.00% respectively. For distraction detection, recall and precision were 89.44% and 93.48% respectively. Fatigue detection performance was the best, of which the mean precision and recall reached 95.89% and 95.96%. Each subject exhibited different detection performances, and the detection accuracy of subject 6 reached 99.67%. In contrast, the detection accuracy of subject 2 was only 81.90%.

V. DISCUSSIONS

A. Simulated Experiment Results

The primary requirement for train drivers is keeping sharp lookout. LD directly evaluates the attention and reaction of the driver under emergencies. Furthermore, the LD influenced not only the safety of the GoA-0, GoA-1 train recently, but also the GoA-2 semi-automatic train operation train in the future. This study proposed the LD as the driving performance of the train drivers and proved that train driver fatigue and distraction can lead to less LD, which might lead to higher safety risk.

Detecting train driver fatigue and distraction in time can avoid such dangerous situation.

The results of the SSS scores showed that all of the subjects were fatigued before the fatigue task was initiated. In contrast, subjects were in the alert state before both the alert task and the distraction task. Successful research need to be closely related to effective experimental design. Different LD and SSS scores between the three tasks proved the rationality of our fatigue and distraction task division.

B. Feature Analysis

In this study, four types of feature extraction methods were used to explore the information of forehead EEG: energy, entropy, rhythmic energy ratio, and frontal asymmetry ratio. Energy is the most effective and intuitive way to acquire EEG information. This is consistent with many previous studies [28]. The energy features directly contribute more than half of the feature importance. It is noted that the frontal EEG asymmetry, a newly proposed feature, which contributes 14.54 feature importance. Moreover, beside the energy features, there were some entropy, rhythmic energy ratio, and frontal asymmetry ratio with high rank in the feature importance list. All types of extracted features in this study are helpful and meaningful. The CatB-FS method reduced the feature dimension from 98 to 41, which not only improved the detection performance but significantly reduced the computational complexity

C. RLX-TV Method Performance

Ensemble methods combine the predictions of several classifiers to obtain better predictive performance classification models [40]. The variance and bias of the classification model can be reduced, and the dependency of results on characteristics of a single training set is eliminated by combining classifiers. It has been verified that ensemble methods can help weak classifications make very accurate predictions according to their ensembles in many fields. Comparing with base classifiers, ensemble models have much stronger generalization ability. In this study, as the baseline models, SVM, KNN, and DT had detection accuracies achieved 58.60%, 61.86%, and 72.00% respectively. The proposed RLX model presented better average detection accuracy achieved 79.35%. What's more, a practical driver fatigue and distraction detection system should be regard as a dynamic system that consider both current and past information. Therefore, it is difficult for a person's driving state to be accurately detected by only 0.5s EEG data. In most studies, when longer sliding windows were used to extract features of EEG data, higher values of detection accuracy were achieved. Researchers must extract features and train models repeatedly to meet different detection time demands [8]. To reduce the time consumption of extracting features and training models repeatedly, as well as to satisfy the requirements of different detection accuracies and detection times, a novel time-series voting algorithm was proposed. Comparing with other time-series model like RNN, LSTM, the proposed method needs to train the base model only once, then by combing output of baseline model in temporal dimension, we can get higher detection accuracy. whichever the base model we selected, the detection accuracies increased by 12.20% to 18.7% by using proposed time-series voting algorithm with 5s fusion time. Among all the models, the RLX-TV method proposed by us achieved highest accuracy: 92.7%.

TABLE IV

DETECTION PERFORMANCE COMPARISON OF RELATED WORK

Reference	Fatigue (%)	Distraction (%)	Detection Time	Data
[45]	66	None	>55s	Pressure, Force Strain, Temperature
[46]	90.9	None	30s	Eye move
[47]	91.3	None	Only driver speak	Voice signal
[8]	90.7	None	4s	8 channels EEG
Proposed method	**96.0**	**93.5**	**5s**	**2 channels EEG**

D. Different Detection Performance Between Subjects

The mean detection accuracy of this study was 92.31%. However, the detection performances of different subjects were different, ranging from 81.90% to 99.67%. It is normal that different subjects exhibit different detection performance. This situation always occurred in previous researches [45], [46]. In addition, we found that subject 2 whose detection performance was the worst exhibited low precision (73.53%) and high recall (92.43%) for alert state detection, and low recall (65.43%) and high prediction (91.08%) for distraction state detection. This represent that many distraction samples were classified as alert. In contrast, fatigue detection performance was stable. The main reasons for detection errors of subject 2 were shown as follows. The distraction task of this paper attempts to induce distraction by presenting auditory mathematics tasks to the subjects continuously. Although the frequency of the secondary task was high, if some subjects had powerful computing abilities, then the math task cannot sufficiently increase mental load to cause cognitive distraction. Therefore, there is no guarantee that the subject will always be in the cognitive distraction state during the half-hour distraction task. The subjects may return to the alert state quickly, and thus are classified as being in the alert state by classifiers. In short, the individual difference between subjects lead to difference of detection performance.

E. Comparison and Application

There Has Been Some Research Focusing on Train Driver Driving State Detection. Mechanical Measurements Such as Pressure, Force, Strain, and Temperature Which Were Installed on the Train Seat Were Used in Train Driver Alertness Detection. Noted That It Was the Only Research That Didn't Need Human Information. However, the Detection Performance Was Not Good (66%) [47]. Yan *et al.* [48] Used Eye Movement Data Collected by Non-Contact Means as an Objective Measurement, to Detect the Fatigue in High-Speed Railway Driving. The Average Detection Accuracy of This Method Was 90.9% With 30s Detect Time. In Addition, There

TABLE V
DETECTION PERFORMANCE OF PROPOSED METHOD WHEN FUSION TIME WAS 5 S

Subject	Alert		Distraction		Fatigue		
	Recall (%) Mean (Std)	Precision (%) Mean (Std)	Recall (%) Mean (Std)	Precision (%) Mean (Std)	Recall (%) Mean (Std)	Precision (%) Mean (Std)	Accuracy (%) Mean (Std)
1[a]	94.67 (5.46)	94.95 (7.65)	94.90 (4.82)	95.20 (4.42)	94.74 (8.03)	94.72 (3.97)	94.77 (4.80)
2[a]	92.43 (6.27)	73.53 (15.48)	65.43 (19.69)	91.08 (8.76)	87.99 (10.09)	88.35 (6.88)	81.90 (10.24)
3[a]	78.21 (25.5)	78.90 (15.75)	76.76 (27.6)	80.00 (15.73)	96.13 (2.24)	99.06 (0.72)	83.65 (10.72)
4[a]	95.22 (3.09)	99.34 (0.65)	98.93 (1.26)	97.72 (2.73)	99.55 (0.50)	96.79 (2.81)	97.90 (1.39)
5[a]	90.77 (6.48)	91.34 (9.39)	90.43 (10.14)	90.1 (5.62)	93.71 (4.23)	93.51 (5.50)	91.63 (6.03)
6[a]	99.97 (0.08)	99.25 (1.04)	99.76 (0.36)	99.90 (0.09)	99.37 (1.03)	99.97 (0.08)	**99.67 (0.33)**
7[a]	99.14 (1.64)	99.65 (0.53)	99.86 (0.22)	99.72 (0.26)	99.72 (0.53)	99.35 (1.35)	99.57 (0.60)
Mean[b]	**92.92 (7.28)**	**91.00 (10.64)**	**89.44 (13.37)**	**93.48 (7.03)**	**95.89 (4.26)**	**95.96 (4.15)**	**92.73 (7.37)**

[a] Mean and Std here are the calculated results of 5-fold cross-validation.
[b] Mean and Std here are the calculated results for the 7 subjects.

Was Research Finding That the Characteristic of Voice Signal Can Indicate the Fatigue Level of Train Driver. A Fuzzy SVM Model Was Established for Fatigue Recognition of Drivers' Speech samples [49]. Though the Detection Accuracy Can Achieved 91.3%, There Was a Great Weakness That the Method Work Only When Train Driver Speak In truth, the train driver can't speak continuously for a long time.

EEG has been used in train driver fatigue detection before. Zhang *et al.* [8] proposed a novel high speed train driver fatigue detection system using a wireless wearable electroencephalograph (EEG). In this study, different time windows such as 1s, 2s, 3s, and 4s were used in this study, and the features and train model need to be extracted four times. Compared with them, our proposed method can reduce the time consumption of extracting features and training models. Because our proposed voting method directly uses classification results of base model, which only needs to be trained once. What's more, 8 channels EEG was more difficult to be used in practice than 2 channels forehead EEG.

The comparison results between our study and other related work are listed in Table IV. In the field of fatigue detection of train drivers, the proposed method using forehead EEG has best fatigue detection performance with short time and convenient data acquisition method. Noted that to the authors' knowledge, there has been no study focusing on train driver distraction detection.

Currently, with the development of wearable devices and remote sensor technologies, a headband [50] can stably collect forehead EEG data from people. Compared with the present 32-128 channel EEG acquisition system, it is more comfortable and convenient to collect forehead EEG data. Using the limited forehead EEG to obtain more useful information about humans was called a "future feasible solution" in the field of brain-computer interface (BCI) [50]. However, forehead EEG had not been enough to detect train driver's fatigue and distraction accurately before. This research can fill this research gap. Our proposed method used dual-channel forehead EEG and the RLX-TV method to detect train driver's fatigue and distraction simultaneously. The experimental results show that the forehead EEG with the proposed method can be applied to practice in train driver fatigue and distraction detection.

VI. CONCLUSION

This study proposed a method for the detection of train driver fatigue and distraction using forehead EEG data, which involved a "future feasible solution" for train driver fatigue and distraction detection. Multi-type feature extraction, CatB-FS algorithm, and time-series ensemble learning method were used to improve the model accuracy with respect to the detection of train driver fatigue and distraction. For a 5s fusion time, the average detection accuracy of our proposed method reached 92.7%. The detection precision of fatigue and that of distraction reached 95.96% and 93.48% respectively. Compared with several baselines including SVM, KNN, DT, LSTM model, the detection accuracy of this method is improved by 34.10% to 20.7%. What's more, the accuracy can be improved by longer fusion time, if necessary. Compared with other studies on this topic, the proposed method achieves better performance with more accessible data and fill the gap of train driver distraction detection. Future works will include designing new secondary tasks for train driver cognitive distraction and designing more driving environments. In addition, designing a complete system which includes detection and reactive control systems also will be our future direction.

REFERENCES

[1] A. J. Filtness and A. Naweed, "Causes, consequences and countermeasures to driver fatigue in the rail industry: The train driver perspective," *Appl. Ergonom.*, vol. 60, pp. 12–21, Apr. 2017.

[2] M. Kyriakidis, A. Majumdar, and W. Y. Ochieng, "Data based framework to identify the most significant performance shaping factors in railway operations," *Saf. Sci.*, vol. 78, pp. 60–76, Oct. 2015.

[3] N. Brandenburger, A. Naumann, and M. Jipp, "Task-induced fatigue when implementing high grades of railway automation," *Cognition Technol. Work*, vol. 23, pp. 273–283, Nov. 2019.

[4] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.

[5] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019.

[6] H. Zhang, Y. Peng, L. Hou, D. Wang, G. Tian, and Z. Li, "Multistage impact energy distribution for whole vehicles in high-speed train collisions: Modeling and solution methodology," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2486–2499, Apr. 2020.

[7] T. Ersal, H. J. A. Fuller, O. Tsimhoni, J. L. Stein, and H. K. Fathy, "Model-based analysis and classification of driver distraction under secondary tasks," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 692–701, Sep. 2010.

[8] 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.

[9] A. Němcová *et al.*, "Multimodal features for detection of driver stress and fatigue: Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3214–3233, Jun. 2021.

[10] H. Zhang, Y. Peng, L. Hou, G. Tian, and Z. Li, "A hybrid multi-objective optimization approach for energy-absorbing structures in train collisions," *Inf. Sci.*, vol. 481, pp. 491–506, May 2019.

[11] X. Tan, K. Xu, Y. Cao, Y. Zhang, L. Ma, and R. W. Lau, "Night-time scene parsing with a large real dataset," *IEEE Trans. Image Process.*, vol. 30, pp. 9085–9088, 2021.

[12] A. Yuce, H. Gao, G. L. Cuendet, and J.-P. Thiran, "Action units and their cross-correlations for prediction of cognitive load during driving," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 161–175, Apr. 2017.

[13] J. Engström, E. Johansson, and J. Östlund, "Effects of visual and cognitive load in real and simulated motorway driving," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 8, no. 2, pp. 97–120, Mar. 2005.

[14] E. Q. Wu *et al.*, "Novel nonlinear approach for real-time fatigue EEG data: An infinitely warped model of weighted permutation entropy," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2437–2448, Jun. 2020.

[15] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2352–2359, 2009.

[16] C.-T. Lin, S.-A. Chen, T.-T. Chiu, H.-Z. Lin, and L.-W. Ko, "Spatial and temporal EEG dynamics of dual-task driving performance," *J. Neuroeng. Rehabil.*, vol. 8, no. 1, p. 11, Dec. 2011.

[17] S. di Fronso, P. Fiedler, G. Tamburro, J. Haueisen, M. Bertollo, and S. Comani, "Dry EEG in sports sciences: A fast and reliable tool to assess individual alpha peak frequency changes induced by physical effort," *Frontiers Neurosci.*, vol. 13, p. 982, Sep. 2019.

[18] G. A. M. Vasiljevic and L. C. de Miranda, "Brain–computer interface games based on consumer-grade EEG devices: A systematic literature review," *Int. J. Hum.-Comput. Interact.*, vol. 36, no. 2, pp. 105–142, Jan. 2020.

[19] Y. Peng *et al.*, "Tunnel driving occupational environment and hearing loss in train drivers in China," *Occupational Environ. Med.*, vol. 76, no. 2, pp. 97–104, Feb. 2019.

[20] *Products Service*. [Online]. Available: http://www.yd-tec.com/productinfo/472412273923588096.html

[21] E. Hoddes, V. Zarcone, H. Smythe, R. Phillips, and W. C. Dement, "Quantification of sleepiness: A new approach," *Psychophysiology*, vol. 10, no. 4, pp. 431–436, 1973.

[22] M. A. Regan, C. Hallett, and C. P. Gordon, "Driver distraction and driver inattention: Definition, relationship and taxonomy," *Accident Anal. Prevention*, vol. 43, no. 5, pp. 1771–1781, Sep. 2011.

[23] D. L. Strayer, J. M. Watson, and F. A. Drews, "Cognitive distraction while multitasking in the automobile," in *Psychology of Learning and Motivation*, vol. 54. Amsterdam, The Netherlands: Elsevier, 2011, pp. 29–58.

[24] A. Naweed, J. Rose, S. Singh, and D. Kook, "Risk factors for driver distraction and inattention in tram drivers," in *Proc. AHFE*. Cham, Switzerland: Springer, 2017, pp. 257–268.

[25] Y.-K. Wang, S.-A. Chen, and C.-T. Lin, "An EEG-based brain–computer interface for dual task driving detection," *Neurocomputing*, vol. 129, pp. 85–93, Apr. 2014.

[26] E. Y. Hamid, R. Mardiana, and Z. I. Kawasaki, "Method for RMS and power measurements based on the wavelet packet transform," *IEE Proc.-Sci., Meas. Technol.*, vol. 149, no. 2, pp. 60–66, Mar. 2002.

[27] J.-Z. Xue, H. Zhang, C.-X. Zheng, and X.-G. Yan, "Wavelet packet transform for feature extraction of EEG during mental tasks," in *Proc. Int. Conf. Mach. Learn.*, vol. 1, 2003, pp. 360–363.

[28] F. Wang, J. Lin, W. Wang, and H. Wang, "EEG-based mental fatigue assessment during driving by using sample entropy and rhythm energy," in *Proc. IEEE CYBER*, Jun. 2015, pp. 1906–1911.

[29] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.

[30] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Nat. Acad. Sci. USA*, vol. 88, no. 6, pp. 2297–2301, Mar. 1991.

[31] J. Hu, "Automated detection of driver fatigue based on AdaBoost classifier with EEG signals," *Frontiers Comput. Neurosci.*, vol. 11, p. 72, Aug. 2017.

[32] W. Li, Q.-C. He, X.-M. Fan, and Z.-M. Fei, "Evaluation of driver fatigue on two channels of EEG data," *Neurosci. Lett.*, vol. 506, no. 2, pp. 235–239, 2012.

[33] W. Haehl, A. Mirifar, M. Luan, and J. Beckmann, "Dealing with failure: Prefrontal asymmetry predicts affective recovery and cognitive performance," *Biol. Psychol.*, vol. 155, Sep. 2020, Art. no. 107927.

[34] S. M. M. Farahi, M. J. A. Ebrahimabad, A. Gorji, I. Bigdeli, and S. M. M. M. Farahi, "Neuroticism and frontal EEG asymmetry correlated with dynamic facial emotional processing in adolescents," *Frontiers Psychol.*, vol. 10, p. 175, Feb. 2019.

[35] P. C. Petrantonakis and L. J. Hadjileontiadis, "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 5, pp. 737–746, Sep. 2011.

[36] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6638–6648.

[37] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[38] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2016, pp. 785–794.

[39] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.

[40] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.

[41] B. Hu, X. Li, S. Sun, and M. Ratcliffe, "Attention recognition in EEG-based affective learning research using CFS+KNN algorithm," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 1, pp. 38–45, Jan. 2018.

[42] S. Guan, K. Zhao, and S. Yang, "Motor imagery EEG classification based on decision tree framework and Riemannian geometry," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–13, Jan. 2019.

[43] X. Wang, R. Jiang, L. Li, Y. Lin, X. Zheng, and F.-Y. Wang, "Capturing car-following behaviors by deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 910–920, Mar. 2018.

[44] Y. Chen, Y. Lv, X. Wang, L. Li, and F. Wang, "Detecting traffic information from social media texts with deep learning approaches," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 3049–3058, Aug. 2019.

[45] W.-L. Zheng *et al.*, "Vigilance estimation using a wearable EOG device in real driving environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 170–184, Jan. 2020.

[46] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F. Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5379–5390, Jun. 2019.

[47] D. Scaccabarozzi, L. Mazzola, M. F. Bocciolone, F. Resta, and B. Saggin, "Monitoring of train driver's alertness: A feasibility study," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2017, pp. 1–5.

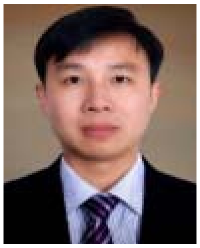
[48] R. Yan, C. Wu, and Y. Wang, "Exploration and evaluation of individual difference to driving fatigue for high-speed railway: A parametric SVM model based on multidimensional visual cue," *IET Intell. Transp. Syst.*, vol. 12, no. 6, pp. 504–512, 2018.

[49] X. Li, N. Tan, T. Wang, and S. Su, "Detecting driver fatigue based on nonlinear speech processing and fuzzy SVM," in *Proc. IEEE ICSP*, Oct. 2014, pp. 510–515.

[50] C. T. Lin *et al.*, "Forehead EEG in support of future feasible personal healthcare solutions: Sleep management, headache prevention, and depression treatment," *IEEE Access*, vol. 5, pp. 10612–10621, 2017.

Chaojie Fan (Student Member, IEEE) received the B.S. degree in communications and transportation engineering from Central South University, Changsha, China, in 2017, where he is currently pursuing the Ph.D. degree with the Key Laboratory of Traffic Safety on Track of Ministry of Education.

He is a joint Ph.D. Student with the Department of Computer Science, City University of Hong Kong. His research interests focuses on brain–computer interaction and its application in intelligence transportation systems.



Yong Peng (Member, IEEE) received the Ph.D. degree from the Strasbourg of University, France, in 2012. He is currently a Professor with the School of Transportation and Traffic Engineering, Central South University, China. He has authored over 60 peer-reviewed publications, including over 50 articles in SCI/SSCI-indexed journals. His research interests include vehicle and traffic safety technology, epidemiology of accidents, and injury biomechanics. He is a Journal Reviewer of SCI/SSCI-indexed international journals, including *Accident Analysis & Prevention*, *Safety Science*, *Traffic Injury Prevention*, *International Journal of Crashworthiness*, *Proceedings of the Institution of Mechanical Engineers Part D: Journal of Automobile Engineering*, and *Transport and Computational & Applied Mathematics*.

Shuangling Peng received the B.S. and M.S. degrees in transportation engineering from Central South University, China, in 2016 and 2019, respectively. She is currently working with Guangzhou Railway Polytechnic. Her main research is focused on behavior of electric locomotive drivers.

Honghao Zhang (Member, IEEE) received the B.S. degree in communications and transportation engineering from the Shandong University of Technology, Zibo, China, in 2014, the M.S. degree in vehicle operation engineering from Northeast Forestry University, Harbin, China, in 2016, and the Ph.D. degree from the School of Transportation Engineering, Central South University, Changsha, China, in 2021. He is currently a Lecturer with the School of Mechanical Engineering, Shandong University, Jinan, China. His research interests include high-speed train collision and remanufacturing and green manufacturing.

Yuankai Wu (Member, IEEE) received the Ph.D. degree from the School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China, in 2019.

He was a Visiting Ph.D. Student with the Department of Civil Environmental Engineering, University of Wisconsin-Madison, from November 2016 to November 2017. Currently, he is a Post-Doctoral Researcher with the Department of Civil Engineering and Applied Mechanics, McGill University. His research interests include intelligent transportation systems and machine learning.

Sam Kwong (Fellow, IEEE) received the B.Sc. degree in electrical engineering from the State University of New York, Buffalo, NY, USA, in 1983, the M.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1985, and the Ph.D. degree from the University of Hagen, North Rhine-Westphalia, Germany, in 1996.

From 1985 to 1987, he was a Diagnostic Engineer with Control Data Canada, Mississauga, ON, Canada. Later, he joined Bell Northern Research Canada, Ottawa, ON, Canada, as a Member of Scientific Staff; and the City University of Hong Kong (CityU), Hong Kong, as a Lecturer with the Department of Electronic Engineering in 1990. He is currently a Chair Professor with the Department of Computer Science, CityU. His research interests include video coding, pattern recognition, and evolutionary algorithms. He is currently the Vice President of conferences and meetings with the IEEE Systems, Man and Cybernetics Society. He also serves as an Associate Editor for the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and the *Journal of Information Science*.