

Received August 1, 2019, accepted August 16, 2019, date of publication August 27, 2019, date of current version September 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937914

# A WPCA-Based Method for Detecting Fatigue Driving From EEG-Based Internet of Vehicles System

NA DONG<sup>1</sup>, YINGJIE LI<sup>1</sup>, ZHONGKE GAO<sup>1</sup>, (Senior Member, IEEE),  
WAI HUNG IP<sup>2</sup>, (Senior Member, IEEE), AND KAI LEUNG YUNG<sup>2</sup>

<sup>1</sup>School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

<sup>2</sup>Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong

Corresponding author: Zhongke Gao (zhongkegao@tju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773282, Grant 61873181, and Grant 61922062, and in part by the Department of Industrial and Systems Engineering of The Hong Kong Polytechnic University under Grant H-ZG3K.

**ABSTRACT** Fatigue driving is the main cause of traffic accidents. Analysis of electroencephalogram (EEG) signals has attracted wide attention for identifying fatigue driving. With the development of the Internet of Vehicles (IoV), we hope to establish an EEG-based IoV traffic management system to improve traffic safety. In the proposed system, real-time diagnosis is a significant factor, and improvement of the detection speed is our main concern. EEG signals generate a large amount of spatially oriented data over a relatively short duration; hence, their dimension needs to be reduced effectively before being analysed. We propose a feature reduction method, based on a novel weighted principal component analysis (WPCA) algorithm for EEG signals. First, the EEG features are extracted by an autoregressive (AR) model. Second, we calculate the influence of different features on the classified performance of fatigue state. The accuracy reduction values of different features are normalised as the weights of the features. Finally, these weights are assigned to the WPCA to reduce the EEG features. To verify the effectiveness of the algorithm, we carried out a simulated driving experiment involving eight participants. For comparison, power spectral density and differential entropy models were also introduced to extract EEG features. Support Vector Machine was adopted as a classifier to establish a fatigue driving classification experiment. The experimental results show that the WPCA method can effectively reduce the feature dimension of different EEG feature extraction methods, speed up calculations, and achieve a much higher classification accuracy of fatigue driving.

**INDEX TERMS** AR model, driving fatigue detection, feature reduction, Internet of Vehicles, WPCA algorithm.

## I. INTRODUCTION

The Internet of Things (IoT) is a methodology that can connect objects through the Internet to work together to achieve new goals [1]. The Internet of Vehicles (IoV) is an application of the IoT, which can collect information about vehicles and drivers [2]. Among this information, fatigue driving has received much attention.

In the past ten years, the number of cars in China has increased dramatically, and the number of traffic accidents has increased [3]. According to relevant reports, China has

become one of the countries defined as having frequent traffic accidents. Many factors can result in traffic accidents, the main one being fatigue driving. In a fatigued state, drivers tend to be distracted, think less actively, and have slower reactions, all of which increase the possibility of traffic accidents [4]. Therefore, it is particularly important to detect driver fatigue state accurately and quickly. Driving fatigue detection, based on physiological information, is an objective means of detecting and identifying driver fatigue state through changes in physiological indicators [5]. Studies have shown that the body's fatigue state can be effectively detected and evaluated by electrophysiological information, such as body temperature, blood pressure, electrocardiogram (ECG),

The associate editor coordinating the review of this article and approving it for publication was Mu-Yen Chen.

electroencephalogram (EEG), and electromyography (EMG). Among them, analysis of EEG signals has accepted wide attention.

EEG reflects the electrophysiological signals of the cranial nervous system [6], and can be used to detect and analyse driver fatigue state. For example, Gao *et al.* proposed an EEG-based spatio-temporal convolutional neural network to detect driver fatigue [7], whilst a graph analysis method of functional brain network topology (using minimum spanning tree) was proposed by Chen *et al.* for detecting driver drowsiness [8]. Wang *et al.* developed a novel real-time driving fatigue detection methodology, based on dry EEG signals [9], and Pei *et al.* proposed a method based on EEG signal analysis to study the fatigue characteristics of drivers of different ages [10]. Wang *et al.* proposed a method for analysing feature fatigue EEG signals based on wavelet entropy [11], whilst Hu *et al.* proposed a driving fatigue detection method based on EEG signals using fuzzy entropy [12]. Min *et al.* proposed a method based on multi-entropy fusion of EEG system to detect driver fatigue [13]. Wang *et al.* proposed an EEG-based system for evaluating driver fatigue with only one electrode by ensemble learning [14]. Pathak and Jayanthi designed a portable, low-cost brain-computer interface drowsiness-detection system [15]. Cheng *et al.* proposed an EEG-based prediction system that transforms the measured EEG record into image-like data for estimating the drowsiness level of drivers [16]. Gao *et al.* proposed a novel relative wavelet entropy complex network for improving EEG-based fatigue driving classification [17]. Foong *et al.* proposed an iterative negative-unlabeled learning algorithm for detecting cross-subject of passive fatigue from labelled alert and unlabeled driving EEG data [18]. Luo *et al.* proposed an adaptive multi-scale entropy feature extraction algorithm for fatigue driving detection [19]. Han *et al.* introduced complex network theory to study the evolution of brain dynamics under different rhythms of EEG signals during several periods of the simulated driving [20].

EEG signals generate a large amount of spatially oriented data over relatively short durations, which leads to a big data problem. Big data requires secure storage and high computing resources for real-time processing [21]–[23]. To increase the processing efficiency, and satisfy the real-time requirement more effectively, the dimension of EEG signals need to be reduced properly. Principal component analysis (PCA) is a commonly used dimension reduction method. It can analyse the main influencing factors from multiple contexts, reveal the essence of entities, and simplify complex problems. Liu *et al.* proposed a hybrid dimension feature-reduction scheme using 14 different features extracted from EEG recordings [24]. To reorder the combined features into max-relevance with the labels and min-redundancy of each feature, maximum relevance minimum redundancy (mRMR) was applied. PCA was used to further reduce the generated features for extracting the principal components. Bousseta *et al.* extracted EEG features by continuous wavelet transform (CWT) and empirical mode decomposition

(EMD) [25]. PCA was introduced for feature dimension reduction, and the left and right hand motion imaging classification was performed by using a linear and radial basis function (RBF) kernel function with a Support Vector Machine (SVM) classifier. Sun *et al.* proposed a fusion algorithm based on PCA for nonlinear global features and power spectral entropy [26]. Combining the power spectral entropy of EEG and the nonlinear attribute features (such as Hurst index), PCA was introduced for dimension reduction and feature fusion, and SVM was used as a classifier for emotion recognition. Neshov *et al.* proposed an algorithm, which could identify five psychological tasks using 6-channel EEG data [27]. The main aim was to divide the original EEG signals into several frames and calculate their spectrum, apply the Gaussian second derivative to extract features, and use PCA to reduce feature dimension. Li *et al.* extracted 8 positive and negative emotions from a dataset, representing the data of 14 channels from the different regions of the brain [28]. Based on wavelet transform,  $\delta$ ,  $\theta$ ,  $\alpha$ , and  $\beta$  rhythms were extracted. On this basis, PCA was used to fuse EEG features of wavelet features, approximate entropy and Hurst exponents, and to reduce feature dimension. Zarei *et al.* proposed a feature extraction method [29], which was combined with PCA and the cross-covariance technique (CCOV). The algorithm extracted discriminant information from the mental state, based on EEG signals in a brain-computer interface technology application, and applied correlation-based variable selection.

In the application of PCA, predecessors are based on data variance reduction and treat each dimension feature equally. However, different features play different roles in the recognition process; therefore, it is necessary to assign different weights to the feature values [30]. Weighted PCA (WPCA) weights the original feature data, and finds the linear combination with the largest variance according to the idea of PCA [31].

With the development of the IoV, we propose an EEG-based IoV traffic management system to improve traffic safety [32]–[37]. The proposed system is composed of a fatigue detection system, the IoV, and a traffic management platform, as shown in Fig. 1.

The fatigue detection system includes an electric source imaging (ESI) neuroscan system and a fatigue diagnosis system. The ESI neuroscan system is equipped with 40 electrodes, which are arranged according to the International 10–20 system, with a sampling frequency of 200 Hz. Among the 40 electrodes, in addition to the 4 electrodes as the internal structure, 2 are defined as reference electrodes and 4 (placed across the horizontal and vertical directions) are used to monitor eye movement. Therefore, the remaining 30 electrodes are used to collect EEG signals. The fatigue diagnostic system is a supercomputer equipped with WPCA and SVM algorithms. WPCA algorithm is a dimension reduction method, and SVM is an efficient supervised two-category classifier.

Working principle of the EEG based IoV traffic management system: The system is able to detect the driver's

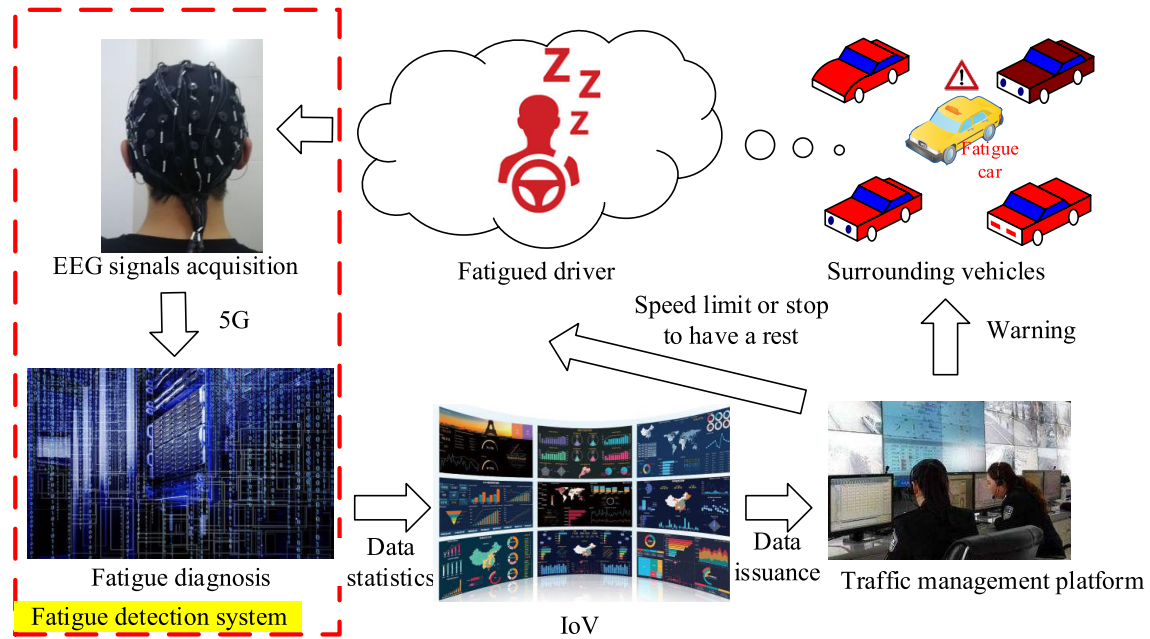


FIGURE 1. EEG-based IoV traffic management system.

fatigue status in real time. It works online and guarantees processing time. The driver's EEG signals are collected by the ESI neuroscan system and transmitted to the fatigue diagnosis system through a 5G network in real-time. Prior to acquisition, the skin impedance of the EEG electrodes is adjusted to below 5 k $\Omega$  by injecting conductive gel. The fatigue diagnosis system reduces the EEG signals using the WPCA algorithm, diagnoses fatigue status using SVM, and uploads the detection result to the IoV. If fatigue is detected, the traffic management platform will remind the driver to either limit their speed or stop and have a rest, whilst warning the surrounding vehicles. Of course, the platform also collects the date, time, location, and frequency statistics of fatigue for each driver. Accordingly, the system can improve traffic safety.

Real-time diagnosis is of significant importance in an EEG-based IoV traffic management system; hence, we need to improve the detection speed. Because EEG is a high-frequency signal, a large amount of data is generated in a short period, meaning that methods to reduce the dimension of EEG signals need to be studied intensively.

In this paper, we propose a feature reduction method based on a novel WPCA algorithm for EEG signals. The autoregressive (AR) model is introduced to extract the features of the EEG signals recorded by drivers during a simulated driving experiment. We calculate the influence of different features on the fatigue state classification performance. Then, the accuracy reduction values of different features are normalised as the weights of the features. SVM is selected as the classifier, and accuracy, sensitivity, and specificity are used as classification evaluation indicators. Simultaneously, three sets of control experiments (SVM, PCA-SVM, and WPCA-SVM) are designed.

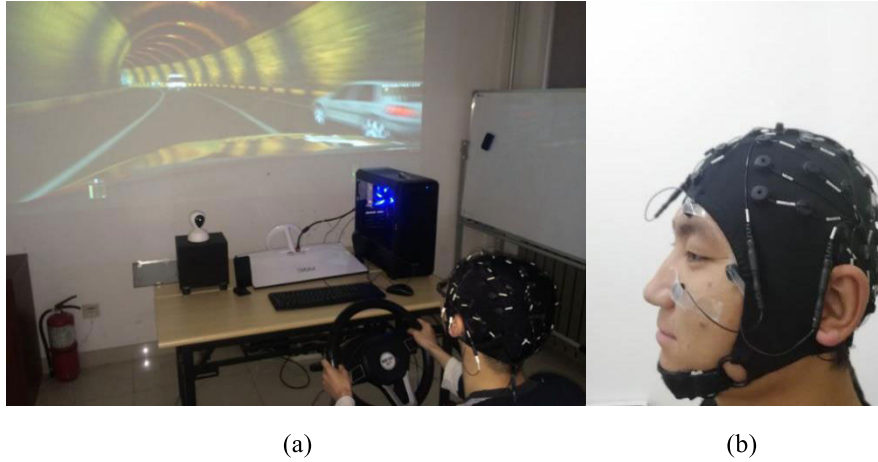
The rest of this paper is organised as follows: Section 2 introduces the simulation driving experiment and data acquisition. Section 3 describes and explains the component of the WPCA. Experimental results and discussions are explained in Section 4. Finally, the summary is presented in Section 5.

## II. SIMULATION DRIVING EXPERIMENT AND DATA ACQUISITION

In the simulated driving experiment, 8 right-handed college students aged 19 to 26 (4 males and 4 females; mean age: 22.73) volunteered to participate in the experiment, and no one has mental illness. In the simulation of EEG, eight subjects are able to reveal the effectiveness of the proposed method [7], [38], [39]. Two days before the experiment, subjects were asked to avoid ingesting any anti-fatigue related products. Further, they were also required to maintain reasonable rest and sleep of more than 7 h a night. Because none of the subjects had access to the driving simulator, they had to practice driving before the experiment until they became proficient.

We conducted the experiment at the Intelligent Systems Laboratory in the Tianjin University Complex Network. We used the PGFD001 driving simulator, which was equipped with a pedal, steering wheel, and clutch. In the virtual driving software 3DInstructor2, we used the ordinary car (Phaeton2.0L), which incorporated automatic shifting by default. In addition, a webcam 360D618, projector, and stereo speakers were added to enhance perception. The experimental setup is shown in Fig.2.

Full-scalp EEG signals were collected in an isolated and quiet room. In addition, we also monitored the subjects' facial status through a front-facing camera to verify fatigue.



**FIGURE 2.** Experimental setup: (a) The experimental scene from the perspective of the researchers. (b) Brain caps of the neuroscan system.

Before the experiment, it took 10 min to set up the scene, and 20 min to practice driving. After the experiment started, the subjects continued driving until they were reported to have mild fatigue, which was usually after 30 minutes, prior to this the drivers were considered to be in an alert state.

Following 10 min of continuous driving (as a transition), the subjects were subjected to another 30 min driving in a fatigued state. The recording time for each subject was close to 90 minutes, which varied slightly due to individual differences.

To record the drivers' fatigued state, the EEG recording device is equipped with an ESI neuroscan system with 40 electrodes, which are arranged according to the International 10-20 system, with a sampling frequency of 200 Hz. Prior to acquisition, the skin impedance of the EEG electrodes was adjusted to below 5 k $\Omega$  by injecting conductive gel. During the experiment, all subjects were required to minimise unnecessary body movements and maintain a constant driving speed to avoid collisions. Among the 40 electrodes, in addition to the 4 electrodes as the internal structure, 2 are defined as reference electrodes and 4 (placed across the horizontal and vertical directions) are used to monitor eye movement. The original EEG signals were subjected to interference based on high frequency and low frequency noise based on eye electricity; therefore, they needed to be pre-processed by the EEGLAB toolbox. We obtained 30-channel EEG signals after eliminating the interference of noise and electro-optical artefacts.

From the collected data we selected the first 10 min of the alert state as the non-fatigue data, and the last 10 min of the fatigue state as the fatigue data. Data were split by a sliding window with a fixed length of 1 s and no overlap. After the data segmentation was completed, we had acquired 1200 sets of samples from each subject.

### III. WEIGHTED PRINCIPAL COMPONENT ANALYSIS

To satisfy the requirement of real-time data acquisition, we need to reduce the dimension of the EEG signals. The commonly used dimension reduction method is PCA,

which can analyse the main influencing factors from multiple contexts, reveal the essence of entities, and simplify complex problems. The purpose of calculating the principal component is to transform the data into a new coordinate system, and project the high-dimensional data into a lower dimensional space. The main idea of PCA is to find a linear combination that can account for the largest change in the value of the initial variable, which also means finding a linear combination with the largest variance. PCA is based on data variance reduction, and treats each dimension feature equally. However, different features play different roles in the recognition process. It is conceivable to strengthen some key features of recognition, while weakening certain non-critical features (such as little correlation information), to improve the recognition accuracy. Based on this, we propose a WPCA method. The WPCA algorithm flow chart is shown in Fig.3.

AR model has been widely introduced into EEG researches [40]–[44]. The advantage of AR model lies in its inherent ability to simulate the peak spectrum of EEG signals. It is an all-pole model that can effectively solve the problem of sharp changes in the spectrum, and requires the selection of the model order number. If the AR model order is too low, the signal cannot be captured successfully. However, if it is too high, more noise is captured.

The principle of AR model is as follows:

$$x(t) = \sum_{k=i}^p a(k)x(t-k) + e(t), \quad (1)$$

where  $x(t)$  represents the EEG data at time  $t$ ,  $p$  represents the AR order number,  $e(t)$  is the white noise sequence, and  $a(k)$  represents the AR model coefficients.

We tested the 3<sup>rd</sup>, 4<sup>th</sup>, ..., 7<sup>th</sup>, and 8<sup>th</sup> order AR models. The experimental results showed that the 3<sup>rd</sup>, 4<sup>th</sup>, and 5<sup>th</sup> order AR models were optimum. Hence the 3<sup>rd</sup>, 4<sup>th</sup>, and 5<sup>th</sup> order AR models were used as feature extractors. The size of the AR features is equal to the AR order number multiplied by 30 EEG channel units. Therefore, the 3<sup>rd</sup>, 4<sup>th</sup>, and 5<sup>th</sup>



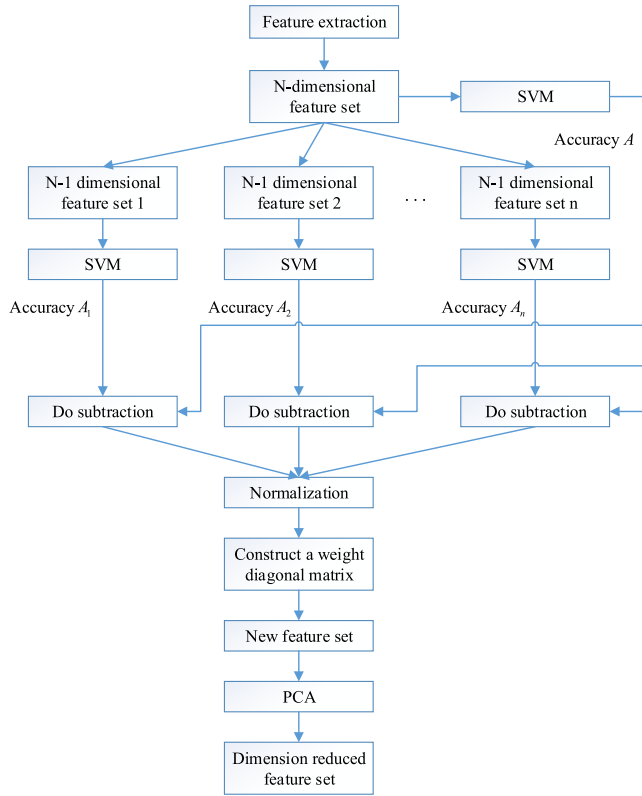


FIGURE 3. The WPCA algorithm flow chart.

order AR models obtained 90, 120, and 150 feature units, respectively.

For comparison, power spectral density (PSD) and differential entropy (DE) were also used as feature extractors. The PSD of each of the frequency bands were calculated by discrete Fourier transform (DFT). The DE is equal to the logarithm of energy spectrum (ES) in a certain frequency band [45]. EEG includes five frequency bands:  $\delta(1 - 3\text{Hz})$ ,  $\theta(4 - 7\text{Hz})$ ,  $\alpha(8 - 13\text{Hz})$ ,  $\beta(14 - 30\text{Hz})$  and  $\gamma(31 - 50\text{Hz})$ . Each of frequency bands can obtain 5 feature units; hence, 30 channels can obtain 150 feature units.

To improve recognition accuracy, reduce the training time of the recognition model, and increase learning speed, we need to focus on dimension reduction of the high-dimensional feature parameters, by selecting some principal elements with high contribution rates to constitute the indicators of fatigue recognition EEG signals [46], [47].

The dimension reduction algorithm is described as follows:

1) Calculate the weights:

Assuming the proposed feature is  $n$ -dimensional, SVM is introduced to train the  $n$ -dimensional data, and obtain the classification accuracy  $A$ . The first dimensional data is removed, and the remaining  $n - 1$  dimensional data is trained by SVM. Hence, we can obtain the classification accuracy  $A_1$ . Then, the second dimensional data is removed, and we can obtain the classification accuracy  $A_2$  by SVM, and so on. Accordingly, the accuracy  $A_1, A_2, \dots, A_n$  of the  $n$

classifications can be obtained. If each accuracy is different from accuracy  $A$ , then  $n$  differences ( $D_1, D_2, \dots, D_n$ ) can be obtained.

$$\begin{cases} D_1 = A - A_1 \\ D_2 = A - A_2 \\ \dots \\ D_n = A - A_n \end{cases} \quad (i = 1, 2, \dots, n). \quad (2)$$

If the difference is positive, this indicates that the dimension feature has a positive influence on the classification, otherwise, there is a negative impact. The  $n$  differences are normalised as weights for each dimension feature.

Normalisation function:

$$w_i = \frac{D_i - D_{\min}}{D_{\max} - D_{\min}} \quad (i = 1, 2, \dots, n), \quad (3)$$

where  $D_i$  is the data before normalisation,  $D_{\max}$  is the maximum value of the sample,  $D_{\min}$  is the minimum value of the sample, and  $w_i$  is the data after normalisation.

2) For the weights  $w_i (i = 1, 2, n)$ , we construct a weight diagonal matrix  $W_{n \times n}$ :

$$W_{n \times n} = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix}. \quad (4)$$

3) Write the extracted data sample set as an  $m \times n$  dimensional matrix:

$$X_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}. \quad (5)$$

4) Introduce a weight diagonal matrix  $W_{n \times n}$  to construct weighted new data:

$$\begin{aligned} Z_{m \times n} &= X_{m \times n} * W_{n \times n} \\ &= \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} * \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix} \\ &= \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1n} \\ z_{21} & z_{22} & \dots & z_{2n} \\ \dots & \dots & \dots & \dots \\ z_{m1} & z_{m2} & \dots & z_{mn} \end{bmatrix}. \end{aligned} \quad (6)$$

5) Calculate the covariance matrix  $C'$  of the input data matrix  $Z_{m \times n}$ :

$$C' = \frac{1}{n-1} Z^T Z. \quad (7)$$

6) Decomposition of the covariance matrix:

$$\lambda_i u_i = C' u_i, \quad i = 1, 2, \dots, n. \quad (8)$$

The eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  of the covariance matrix (arranged in descending order) and the corresponding unitized eigenvectors  $u_1, u_2, \dots, u_n$  are calculated.

7) Select the principal component according to the cumulative contribution rate and construct the mapping matrix  $P'$ :

When taking the first  $n$  principal components, the cumulative contribution rate of the  $i^{th}$  ( $i \leq n$ ) principal components is calculated as follows:

$$\alpha_i = \frac{\sum_{j=1}^i \lambda_j}{\sum_{j=1}^n \lambda_j}. \quad (9)$$

According to the cumulative contribution rate, the first  $k$  feature vectors are selected and combined to form a mapping matrix  $P'$ :

$$P' = [u_1, u_2, \dots, u_k]. \quad (10)$$

8) Use the mapping matrix to obtain the reconstructed EEG data features:

$$Y' = ZP'. \quad (11)$$

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

SVM was adopted as a classifier to establish a fatigue driving classification experiment. SVM is an efficient supervised two-category classifier, which plays an important role in data classification for small samples, nonlinear, and high dimensional modes. SVM has been widely utilized in EEG classification researches [48]–[53]. We established three sets of experimental methods: (1) SVM; (2) PCA-SVM (with a cumulative contribution rate of 0.95); and (3) WPCA-SVM (with a cumulative contribution rate of 0.95).

To verify the reliability of the experimental results, this paper used a 10-fold cross-validation method. Here, the dataset is divided into 10 parts; 1 part as the test set, and the remaining 9 as the training set. The cross-validation is repeated 10 times so that each copy can be tested once as a test set, and the average of the 10 test data is taken as the result.

To assess the classification results, the following three indicators are used [4]:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} * 100\%, \quad (12)$$

$$Sensitivity = \frac{TP}{TP + FN} * 100\%, \quad (13)$$

$$Specificity = \frac{TN}{TN + FP} * 100\%. \quad (14)$$

Here,  $TP$  is the number of positive samples that are correctly identified (the number of samples correctly recognised as fatigue driving);  $TN$  is the number of negative samples that are correctly identified (the number of samples correctly recognised as normal driving);  $FN$  is the number of positive samples that are not recognised (the number of samples for fatigue driving as normal driving); and  $FP$  is the number of negative samples that are not recognised (the number of samples for normal driving as fatigue driving). The accuracy

rate reflects the proportion of the samples with the correct classification to the overall samples. The sensitivity reflects the classification accuracy of the positive samples, and the specificity reflects the classification accuracy of the negative samples.

For the EEG data of eight subjects, the experimental results are shown in Fig.4.

Based on the experimental results of eight subjects, the following was established: Comparing the five feature extraction methods, the 4<sup>th</sup> order AR model achieved the best classification results. When using the WPCA-SVM, all three indicators reached the highest value. Individual differences had a significant impact on the experiment. For example, the accuracy of subject 2 was clearly lower than the other seven subjects. We averaged the test results for eight subjects, as shown in Table 1.

Comparing the three experimental methods, the three performance indicators of WPCA-SVM were better than both PCA-SVM and single SVM without dimension reduction.

Comparing the five feature extraction methods:

- i) Extract features by the 3<sup>rd</sup> order AR model: When using WPCA-SVM, the accuracy, sensitivity, and specificity increased by 5.71%, 5.85%, and 6.12%, respectively, compared with the classification results using SVM alone; the accuracy, sensitivity, and specificity increased by 3.50%, 3.88%, and 3.14%, respectively, compared with the classification results of PCA-SVM.
- ii) Extract features by the 4<sup>th</sup> order AR model: When using WPCA-SVM, the accuracy, sensitivity, and specificity increased by 4.39%, 4.33%, and 4.51%, respectively, compared with the classification results using SVM alone; the accuracy, sensitivity, and specificity increased by 3.11%, 2.97%, and 3.28%, respectively, compared with the classification results of PCA-SVM.
- iii) Extract features by the 5<sup>th</sup> order AR model: When using WPCA-SVM, the accuracy, sensitivity, and specificity increased by 6.78%, 7.24%, and 6.41%, respectively, compared with the classification results using SVM alone; the accuracy, sensitivity, and specificity increased by 5.28%, 5.57%, and 4.98%, respectively, compared with the classification results of PCA-SVM.
- iv) Extract features by PSD: When using WPCA-SVM, the accuracy, sensitivity, and specificity increased by 6.75%, 7.06%, and 6.56%, respectively, compared with the classification results using SVM alone; the accuracy, sensitivity, and specificity increased by 5.12%, 5.05%, and 5.20%, respectively, compared with the classification results of PCA-SVM.
- v) Extract features by DE: When using WPCA-SVM, the accuracy, sensitivity, and specificity increased by 5.82%, 6.01%, and 5.67%, respectively, compared with the classification results using SVM alone; the accuracy, sensitivity, and specificity increased by 3.99%, 4.06%, and 3.89%, respectively, compared with the classification results of PCA-SVM.



**FIGURE 4.** Eight subjects' experimental results.

We could find that the classification results were optimum when the 4<sup>th</sup> order AR model was used. When using the

SVM classifier alone, all three indicators were greater than 93%; when using PCA-SVM, all three indicators were greater

**TABLE 1. Average results for eight subjects.**

Classifier	Feature extractor	Accuracy (%)	Sensitivity (%)	Specificity (%)
SVM	AR3	90.54	90.38	90.56
	AR4	93.23	93.27	93.13
	AR5	86.45	86.04	86.71
	PSD	87.27	86.85	87.52
	DE	89.99	89.79	90.13
PCA-SVM	AR3	92.75	92.35	93.12
	AR4	94.51	94.63	94.36
	AR5	87.95	87.71	88.14
	PSD	88.90	88.86	88.88
	DE	91.82	91.74	91.91
WPCA-SVM	AR3	96.25	96.23	96.26
	AR4	97.62	97.60	97.64
	AR5	93.23	93.28	93.12
	PSD	94.02	93.91	94.08
	DE	95.81	95.80	95.80

**TABLE 2. Comparison of features and the amount of data.**

		SVM	PCA-SVM	WPCA-SVM
Features	AR3	90	41.125	7.375
	AR4	120	33.75	13.75
	AR5	150	26.875	16.375
	PSD	150	28.375	14.875
	DE	150	27.625	15.625
Amount of data	AR3	3240000	1480500	278460
	AR4	4320000	1215000	495000
	AR5	5400000	967500	589500
	PSD	5400000	1021500	535500
	DE	5400000	994500	562500

than 95%. Therefore, when using WPCA-SVM, the three indicators achieved the highest, but only by a small margin.

We compared the number of features and data volume before and after dimension reduction, where the amount of data was calculated through

$$\text{the amount of data} = \text{samples} \times \text{channels} \times \text{features}. \quad (15)$$

We averaged the results for 8 subjects, as shown in Table 2.

It was found that by using PCA for dimension reduction, the feature dimension significantly declined. Using WPCA, the feature dimension (and the amount of data) decreased more obviously, and this made the proposed WPCA an effective solution for the EEG big data processing problem.

The following overall observations can be attained:

- 1) Comparison of the three experimental methods showed that when the SVM classifier was used alone, the data feature dimension was higher, which made the classification task more difficult. When using PCA to reduce the dimension, each dimension feature was treated equally, but the different features played different roles in the identification process. When using WPCA to reduce the dimension, the weights were determined in relation to the importance of different features in the recognition. This enhanced certain features that identified key features and attenuated redundant feature information. Therefore, the experimental results of WPCA-SVM were superior to SVM and PCA-SVM, and all indicators were improved.

- 2) The classification results obtained by extracting features with the 4<sup>th</sup> order AR model were the best, and the three indicators could reach the highest value compared to the other five feature extraction methods. This was because the 4<sup>th</sup> order AR model could track EEG signals accurately, and did not capture too much noise. When the 3<sup>rd</sup> AR model was used as a feature extractor, because only three features were extracted per channel, the EEG signals could not be well characterised. However, more noise was captured due to the higher order number when the 5<sup>th</sup> AR model was used as a feature extractor. When using PSD or DE to extract features, the performance indicators were lower than those of the 4<sup>th</sup> order AR model. Thus, it can be illustrated that better classification results can be obtained when using the 4<sup>th</sup> order AR model.
- 3) Due to individual physiological differences, the accuracy of the results of Subject 2 was slightly lower than for the other subjects. The experimental results showed that the 4<sup>th</sup> order AR model could extract the features of EEG signals excellently, which was the basis for accurate classification in real time. Compared with SVM and PCA-SVM, the proposed WPCA-SVM had a higher accuracy of fatigue detection, although the extent of improvement in accuracy was affected by individual differences. There were differences in the accuracy of subjects' fatigue state detection, but on the whole, the accuracy was improved after using WPCA-SVM. Therefore, these individual differences hardly affected the effectiveness of the proposed WPCA method.
- 4) The high frequency EEG signals in a short time interval contained a large amount of raw sample data, which led to a big data analysis problem. Large-scale data presented challenges for real-time analysis and storage. Accordingly, we proposed a WPCA algorithm to reduce the dimension of the original data, according to the impact of different attributes on the classification results. The algorithm reduced the feature dimension, and improved the accuracy of the classification, which facilitated big data processing.

## V. SUMMARY

In this paper, an EEG-based IoV traffic management system has been proposed to improve traffic safety. We proposed a feature reduction method, based on a novel WPCA algorithm, to reduce the dimension of EEG signals for the purpose of real-time requirement. To verify the algorithm, we carried out a simulated driving experiment involving eight subjects. For comparison, the 3<sup>rd</sup>, 4<sup>th</sup>, and 5<sup>th</sup> order AR models, PSD, and DE were used as feature extractors. We determined the weights according to the importance of different features in the recognition, and SVM was selected as the classifier. The accuracy, sensitivity, and specificity were introduced as the classification evaluation indicators. At the same time, three experimental methods (SVM, PCA-SVM, and WPCA-SVM)



were designed for each feature extractor. The experimental results indicated that when using the 4<sup>th</sup> order AR model to extract features, the classification results of the 8 subjects could attain the highest accuracy. Compared to SVM and PCA-SVM, the proposed WPCA-SVM achieved the best performance with various feature extraction methods. Overall, the algorithm greatly reduced the amount of data and improved the accuracy of the classification, which was more suitable for big data processing.

This method can be effectively applied to EEG-based IoV traffic management systems, meeting real-time requirements. Future research could focus on optimising the above technologies and studying real-time driving fatigue detection systems.

## REFERENCES

- [1] M. A. Saleem, Z. Shijie, and A. Sharif, "Data transmission using IoT in vehicular ad-hoc networks in smart city congestion," *Mobile Netw. Appl.*, vol. 24, no. 1, pp. 248–258, Feb. 2019.
- [2] A. Tolba, "Content accessibility preference approach for improving service optimality in Internet of vehicles," *Comput. Netw.*, vol. 152, pp. 78–86, Apr. 2019.
- [3] F. W. Wang and H. Wang, "EEG characteristic analysis of coach bus drivers in fatigue state," *Chin. J. Sci. Instrum.*, vol. 34, no. 5, pp. 1146–1152, 2013.
- [4] X. Yu and X. Li, "A real-time driving fatigue detection system based on portable EEG," *Mod. Inf. Technol.*, vol. 2, no. 4, pp. 45–47, 2018.
- [5] J. Ye, Q. Liu, and X. Li, "Research on optimization of driving fatigue prediction model based on random forest," *Automobile Appl. Technol.*, vol. 268, no. 13, pp. 46–50, 2018.
- [6] Z. Gao, K. Zhang, W. Dang, Y. Yang, Z. Wang, H. Duan, and G. Chen, "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.
- [7] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dand, and S. Zuo, "EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2755–2763, Sep. 2019. doi: [10.1109/TNNLS.2018.2886414](https://doi.org/10.1109/TNNLS.2018.2886414).
- [8] J. Chen, H. Wang, C. Hua, Q. Wang, and C. Liu, "Graph analysis of functional brain network topology using minimum spanning tree in driver drowsiness," *Cogn. Neurodyn.*, vol. 12, no. 6, pp. 569–581, Dec. 2018.
- [9] H. Wang, A. Dragomir, N. I. Abbasi, J. Li, N. V. Thakor, and A. Bezerianos, "A novel real-time driving fatigue detection system based on wireless dry EEG," *Cogn. Neurodyn.*, vol. 12, no. 4, pp. 365–376, Feb. 2018.
- [10] Y. Pei, Y. Jin, and H. Chen, "Fatigue characteristics in drivers of different ages based on analysis of EEG," *China J. Highway Transport*, vol. 31, no. 4, pp. 59–65 and 77, 2018.
- [11] Q. Wang, Y. Li, and X. Liu, "Analysis of feature fatigue EEG signals based on wavelet entropy," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 8, Aug. 2018, Art. no. 1854023.
- [12] J. Hu and T. Wang, "Analysis of driving fatigue detection based on fuzzy entropy of EEG signals," *China Saf. Sci. J.*, vol. 28, no. 4, pp. 13–18, 2018.
- [13] 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, Dec. 2017, Art. no. e0188756.
- [14] P. Wang, J. Min, and J. Hu, "Ensemble classifier for driver's fatigue detection based on a single EEG channel," *IET Intell. Transp. Syst.*, vol. 12, no. 10, pp. 1322–1328, 2018.
- [15] M. Pathak and A. K. Jayanthi, "Development of a real-time single channel brain-computer interface system for detection of drowsiness," *Biomed. Eng., Appl., Basis Commun.*, vol. 29, no. 3, Jun. 2017, Art. no. 1750019.
- [16] E. J. Cheng, K. Y. Young, and C. T. Lin, "Image-based EEG signal processing for driving fatigue prediction," in *Proc. Int. Autom. Control Conf. (CACCS)*, Taoyuan, Taiwan, Nov. 2018, pp. 1–5.
- [17] Z. Gao, S. Li, Q. Cai, W. Dang, Y. Yang, C. Mu, and P. Hui, "Relative wavelet entropy complex network for improving EEG-based fatigue driving classification," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 7, pp. 2491–2497, Jul. 2019. doi: [10.1109/TIM.2018.2865842](https://doi.org/10.1109/TIM.2018.2865842).
- [18] R. Foong, K. K. Ang, Z. Zhang, and C. Quek, "An iterative cross-subject negative-unlabeled learning algorithm for quantifying passive fatigue," *J. Neural Eng.*, vol. 16, May 2019, Art. no. 056013.
- [19] H. Luo, T. Qiu, C. Liu, and P. Huang, "Research on fatigue driving detection using forehead EEG based on adaptive multi-scale entropy," *Biomed. Signal Process. Control*, vol. 51, pp. 50–58, May 2019.
- [20] C. Han, X. Sun, Y. Yang, Y. Che, and Y. Qin, "Brain complex network characteristic analysis of fatigue during simulated driving based on electroencephalogram signals," *Entropy*, vol. 21, no. 4, p. 353, Apr. 2019.
- [21] M.-P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Optimized deep learning for EEG big data and seizure prediction BCI via Internet of Things," *IEEE Trans. Big Data*, vol. 3, no. 4, pp. 392–404, Dec. 2017. doi: [10.1109/TBDDATA.2017.2769670](https://doi.org/10.1109/TBDDATA.2017.2769670).
- [22] M. P. Hosseini, H. Soltanian-Zadeh, K. Elisevich, and D. Pompili, "Cloud-based deep learning of big EEG data for epileptic seizure prediction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Washington, DC, USA, Dec. 2017, pp. 1151–1155. doi: [10.1109/GlobalSIP.2016.7906022](https://doi.org/10.1109/GlobalSIP.2016.7906022).
- [23] M. P. Hosseini, T. X. Tran, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Deep learning with edge computing for localization of epileptogenicity using multimodal rs-fMRI and EEG big data," in *Proc. IEEE Int. Conf. Auton. Comput. (ICAC)*, Columbus, OH, USA, Jul. 2017, pp. 83–92. doi: [10.1109/ICAC.2017.41](https://doi.org/10.1109/ICAC.2017.41).
- [24] J. Liu, H. Meng, M. Li, F. Zhang, R. Qin, and A. K. Nandi, "Emotion detection from EEG recordings based on supervised and unsupervised dimension reduction," *Concurrency Comput., Pract. Exper.*, vol. 30, no. 23, p. e4446, Dec. 2018.
- [25] R. Bousseta, S. Tayeb, I. E. Ouakouak, M. Gharbi, F. Regragui, and M. M. Himmi, "EEG efficient classification of imagined right and left hand movement using RBF kernel SVM and the joint CWT\_PCA," *AI Soc.*, vol. 33, no. 4, pp. 621–629, Nov. 2018.
- [26] Y. Sun, J. Ma, and X. Zhang, "EEG emotion recognition based on non-linear global features and spectral feature," *Comput. Eng. Appl.*, vol. 54, no. 17, pp. 116–121, 2018.
- [27] N. N. Neshov, A. H. Manolova, I. R. Draganov, K. T. Tonschev, and O. L. Boumbarov, "Classification of mental tasks from EEG signals using spectral analysis, PCA and SVM," *Cybern. Inf. Technol.*, vol. 18, no. 1, pp. 81–92, 2018.
- [28] X. Li, E. Cai, Y. Tian, X. Sun, and M. Fan, "An improved electroencephalogram feature extraction algorithm and its application in emotion recognition," *J. Biomed. Eng.*, vol. 34, no. 4, pp. 510–517, Aug. 2018.
- [29] R. Zarei, J. He, S. Siuly, and Y. Zhang, "A PCA aided cross-covariance scheme for discriminative feature extraction from EEG signals," *Comput. Methods Programs Biomed.*, vol. 146, pp. 47–57, Jul. 2017.
- [30] Y. Wang, H. Zhang, and S. Wang, "An attribute-weighted principal-component analysis algorithm," *J. Univ. Jinan*, vol. 29, no. 6, pp. 438–443, 2015.
- [31] X. Huang, L. Li, and Y. Xin, "WPCA-LDA: New method of data classification," *Appl. Res. Comput.*, vol. 34, no. 6, pp. 1650–1653, 2017.
- [32] W.-J. Chang, L.-B. Chen, and Y.-Z. Chiou, "Design and implementation of a drowsiness-fatigue-detection system based on wearable smart glasses to increase road safety," *IEEE Trans. Consum. Electron.*, vol. 64, no. 4, pp. 461–469, Nov. 2018. doi: [10.1109/TCE.2018.2872162](https://doi.org/10.1109/TCE.2018.2872162).
- [33] M. I. Chacon-Murguia and C. Prieto-Resendiz, "Detecting driver drowsiness: A survey of system designs and technology," *IEEE Consum. Electron. Mag.*, vol. 4, no. 4, pp. 107–119, Oct. 2015. doi: [10.1109/MCE.2015.2463373](https://doi.org/10.1109/MCE.2015.2463373).
- [34] A. Koesdwiady, R. Soua, F. Karray, and M. S. Kamel, "Recent trends in driver safety monitoring systems: State of the art and challenges," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4550–4563, Jun. 2017. doi: [10.1109/TVT.2016.2631604](https://doi.org/10.1109/TVT.2016.2631604).
- [35] X. Zhang, J. Li, Y. Liu, Z. Zhang, Z. Wang, D. Luo, X. Zhou, M. Zhu, W. Salman, G. Hu, and C. Wang, "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.
- [36] G. Li, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based wearable EEG system for driver drowsiness detection," *IEEE Sensors J.*, vol. 15, no. 12, pp. 7169–7180, Dec. 2015. doi: [10.1109/JSEN.2015.2473679](https://doi.org/10.1109/JSEN.2015.2473679).
- [37] L.-B. Chen, W.-J. Chang, J.-P. Su, J.-Y. Ciou, Y.-J. Ciou, C.-C. Kuo, and K. S.-M. Li, "A wearable-glasses-based drowsiness-fatigue-detection system for improving road safety," in *Proc. IEEE 5th Global Conf. Consum. Electron.*, Kyoto, Japan, Oct. 2016, pp. 1–2. doi: [10.1109/GCCE.2016.7800456](https://doi.org/10.1109/GCCE.2016.7800456).

- [38] S. Valenzi, T. Islam, P. Jurica, and A. Cichocki, "Individual classification of emotions using EEG," *J. Biomed. Sci. Eng.*, vol. 7, no. 8, p. 604, Jun. 2014.
- [39] M. Othman, A. Wahab, I. Karim, M. A. Dzulkifli, and I. F. T. Alshaikli, "EEG emotion recognition based on the dimensional models of emotions," *Procedia-Social Behav. Sci.*, vol. 97, pp. 30–37, Nov. 2013.
- [40] R. Chai, G. R. Naik, T. N. Nguyen, S. H. Ling, Y. Tran, A. Craig, and H. T. Nguyen, "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. doi: [10.1109/JBHI.2016.2532354](https://doi.org/10.1109/JBHI.2016.2532354).
- [41] R. Chai, S. H. Ling, P. P. San, G. R. Naik, T. N. Nguyen, Y. Tran, A. Craig, and H. T. Nguyen, "Improving EEG-based driver fatigue classification using sparse-deep belief networks," *Frontiers Neurosci.*, vol. 11, p. 103, Mar. 2017.
- [42] R. Chai, Y. Tran, A. Craig, S. H. Ling, and H. T. Nguyen, "Enhancing accuracy of mental fatigue classification using advanced computational intelligence in an electroencephalography system," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Chicago, IL, USA, Aug. 2014, pp. 1318–1341. doi: [10.1109/EMBC.2014.6943846](https://doi.org/10.1109/EMBC.2014.6943846).
- [43] Y. Zhang, X. Ji, B. Liu, D. Huang, F. Xie, and Y. Zhang, "Combined feature extraction method for classification of EEG signals," *Neural Comput. Appl.*, vol. 28, no. 11, pp. 3153–3161, Nov. 2017.
- [44] J. Nan, L. Ai, and L. Shi, "Bispectrum analysis on EEG for driving fatigue," *J. Comput. Appl.*, vol. 30, no. 7, pp. 1967–1969, 1973.
- [45] J. Pan, Q. Xie, H. Huang, Y. He, Y. Sun, R. Yu, and Y. Li, "Emotion-related consciousness detection in patients with disorders of consciousness through an EEG-based BCI system," *Frontiers Hum. Neurosci.*, vol. 12, p. 198, May 2018.
- [46] Z. Guo, L. Niu, Z. Wu, Q. Xiao, and L. Shi, "Driver's fatigue recognition algorithm based on EEG and its validity verification," *J. Beijing Univ. Technol.*, vol. 43, no. 6, pp. 929–934, 2017.
- [47] C. Zhao, C. Zheng, and M. Zhao, "Driving mental fatigue staging based on kernel learning algorithm," *J. Data Acquisition Process.*, vol. 24, no. 3, pp. 335–339, 2009.
- [48] R. S. Selvakumari, M. Mahalakshmi, and P. Prashalee, "Patient-specific seizure detection method using hybrid classifier with optimized electrodes," *J. Med. Syst.*, vol. 43, no. 5, p. 121, May 2019.
- [49] F. Yan, M. Liu, C. Ding, Y. Wang, and L. Yan, "Driving style recognition based on electroencephalography data from a simulated driving experiment," *Frontiers Psychol.*, vol. 10, p. 1254, May 2019.
- [50] J. Chen, H. Wang, Q. Wang, and C. Hua, "Exploring the fatigue affecting electroencephalography based functional brain networks during real driving in young males," *Neuropsychologia*, vol. 129, pp. 200–211, Jun. 2019.
- [51] D. Zhao, Y. Wang, Q. Wang, and X. Wang, "Comparative analysis of different characteristics of automatic sleep stages," *Comput. Methods Programs Biomed.*, vol. 175, pp. 53–72, Jul. 2019.
- [52] S. Raghu, N. Sriram, A. S. Hegde, and P. L. Kubben, "A novel approach for classification of epileptic seizures using matrix determinant," *Expert Syst. Appl.*, vol. 127, pp. 323–341, Aug. 2019.
- [53] M. V. M. Yeo, 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, pp. 115–124, Jan. 2009.



**NA DONG** received the Ph.D. degree in control theory and control application from Nankai University, in 2011.

She is currently an Associate Professor with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. Her current research interests include intelligent control algorithms, heuristic optimization algorithms, neural networks, data-driven control, deep learning, and image processing.



**YINGJIE LI** received the B.S. degree from the College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China, in 2017.

He is currently pursuing the M.S. degree with the School of Electrical and Information Engineering, Tianjin University, Tianjin. His research interests include EEG analysis and machine learning.



**ZHONGKE GAO** received the M.Sc. and Ph.D. degrees from Tianjin University, Tianjin, in 2007 and 2010, respectively, where he has been a Full Professor with the School of Electrical and Information Engineering, since 2016.

He is currently the Director of the Laboratory of Complex Networks and Intelligent Systems, Tianjin University. His research interests include deep learning, EEG analysis, complex networks, brain-computer interface, and wearable intelligent devices.

Prof. Gao has been serving as an Editorial Board Member for the *Scientific Reports* and an Associate Editor for *IEEE Access*, *Neural Processing Letters*, and *Royal Society Open Science*.



**WAI HUNG IP** received the LL.B. degree (Hons.) from the University of Wolverhampton, the M.Sc. degree in industrial engineering from Cranfield University, the MBA degree from Brunel University London, and the Ph.D. degree from Loughborough University, U.K.

He is currently a Principal Research Fellow with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, and also a Professor Emeritus and an Adjunct

Professor of Mechanical Engineering with the University of Saskatchewan, Canada.

Prof. Ip is a member of The Hong Kong Institutions of Engineers.



**KAI LEUNG YUNG** received the B.Sc. degree in electronic engineering from Brighton University, in 1975, the M.Sc. and DIC degrees in automatic control systems from the Imperial College of Science, Technology and Medicine, University of London, in 1976, and the Ph.D. degree in microprocessor applications in process control from Plymouth University, U.K., in 1985.

He became a Chartered Engineer (C.Eng., MIEE), in 1982. He is currently an Associate

Head and the Chair Professor of the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University. His research interests include space systems, mechatronics, robotics and automatons, AI, and big data.

...