



# Automated Detection of Driver Fatigue Based on AdaBoost Classifier with EEG Signals

Jianfeng Hu\*

The Center of Collaboration and Innovation, Jiangxi University of Technology, Nanchang, China

**Purpose:** Driving fatigue has become one of the important causes of road accidents, there are many researches to analyze driver fatigue. EEG is becoming increasingly useful in the measuring fatigue state. Manual interpretation of EEG signals is impossible, so an effective method for automatic detection of EEG signals is crucial needed.

**Method:** In order to evaluate the complex, unstable, and non-linear characteristics of EEG signals, four feature sets were computed from EEG signals, in which fuzzy entropy (FE), sample entropy (SE), approximate Entropy (AE), spectral entropy (PE), and combined entropies (FE + SE + AE + PE) were included. All these feature sets were used as the input vectors of AdaBoost classifier, a boosting method which is fast and highly accurate. To assess our method, several experiments including parameter setting and classifier comparison were conducted on 28 subjects. For comparison, Decision Trees (DT), Support Vector Machine (SVM) and Naive Bayes (NB) classifiers are used.

**Results:** The proposed method (combination of FE and AdaBoost) yields superior performance than other schemes. Using FE feature extractor, AdaBoost achieves improved area (AUC) under the receiver operating curve of 0.994, error rate (ERR) of 0.024, Precision of 0.969, Recall of 0.984, F1 score of 0.976, and Matthews correlation coefficient (MCC) of 0.952, compared to SVM (ERR at 0.035, Precision of 0.957, Recall of 0.974, F1 score of 0.966, and MCC of 0.930 with AUC of 0.990), DT (ERR at 0.142, Precision of 0.857, Recall of 0.859, F1 score of 0.966, and MCC of 0.716 with AUC of 0.916) and NB (ERR at 0.405, Precision of 0.646, Recall of 0.434, F1 score of 0.519, and MCC of 0.203 with AUC of 0.606). It shows that the FE feature set and combined feature set outperform other feature sets. AdaBoost seems to have better robustness against changes of ratio of test samples for all samples and number of subjects, which might therefore aid in the real-time detection of driver fatigue through the classification of EEG signals.

**Conclusion:** By using combination of FE features and AdaBoost classifier to detect EEG-based driver fatigue, this paper ensured confidence in exploring the inherent physiological mechanisms and wearable application.

**Keywords:** driver fatigue, electroencephalogram (EEG), adaboost, fuzzy entropy, receiver operating characteristic (ROC)

## OPEN ACCESS

### Edited by:

Daniela Iacoviello,  
Sapienza Università di Roma, Italy

### Reviewed by:

Petia D. Koprinkova-Hristova,  
Institute of Information and  
Communication Technologies (BAS),  
Bulgaria  
Jingxin Nie,  
South China Normal University, China

### \*Correspondence:

Jianfeng Hu  
huguess211@hotmail.com

**Received:** 22 May 2017

**Accepted:** 20 July 2017

**Published:** 03 August 2017

### Citation:

Hu J (2017) Automated Detection of  
Driver Fatigue Based on AdaBoost  
Classifier with EEG Signals.  
*Front. Comput. Neurosci.* 11:72.  
doi: 10.3389/fncom.2017.00072

## INTRODUCTION

Electroencephalogram (EEG) is a very important monitoring technique to reflect the instantaneous state of the brain. Various computational ways based on EEG signals have been successfully used to assist the diagnosis of seizure (Amal Feltane et al., 2013), stroke, Alzheimer's, schizophrenia (Boostani et al., 2009), epilepsy (Guo et al., 2010), depression, Attention Deficit Hyperactivity Disorder, and even fatigue. Driver fatigue is very important factor to traffic safety and automated detection is necessary urgently (Lal and Craig, 2001). Many EEG-based studies have been performed to analyze and detect driving fatigue (Kar et al., 2010; Mu et al., 2017a; Yin et al., 2017).

Correa et al. got 83.6% accuracy using a Neural Network classifier (Correa et al., 2014). Mousa Kadhim et al. yielded the highest accuracy of 85% using Discrete Wavelet Transforms method (Mousa Kadhim et al., 2013). Li et al. achieved the highest accuracy of 91.5% based on 12 types of energy parameters (Li et al., 2012). Fu et al. reached a highest accuracy of 92.5% based on Hidden Markov Model (HMM; Fu et al., 2016). Zhao et al. hit a higher accuracy (98.7%) based on a KPCA-SVM classifier (Zhao et al., 2010). Recently, entropy has been broadly applied in the analysis of EEG signals, considering the fact that EEG is a complex, unstable, and non-linear signal (Acharya et al., 2012; Hu et al., 2015; Mu et al., 2016). A diverse varied collection of these methods has been proposed in the last few decades, including spectral entropy (PE), permutation entropy, distribution entropy, fuzzy entropy (FE), Renyi entropy, approximate entropy (AE), sample entropy (SE), and some others. Specially, in the field of EEG processing, four of the most widely used and successful entropy estimators are FE (Chen et al., 2009), SE (Richman and Moorman, 2000), AE (Pincus, 1991), and PE (Reyes-Sanchez et al., 2016). AE has demonstrated its capability to detect complexity changes. SE is a similar statistic, which has not yet been used as extensively as AE. AE and SE are very successful entropy features, but they also have their weaknesses. AE is biased because it includes self-matching in the count, while SE needs to avoid the  $\log(0)$  problem. They are also very sensitive to input parameters. More recently, FE has been proposed to alleviate these problems. FE is based on a continuous function to compute the dissimilarity between two zero-mean subsequences, so it is more stable in noise and parameter initialization.

Liu et al. got 84% accuracy with the combination of kernel principal component analysis and HMM utilizing AE and Kolmogorov complexity to detect the fatigue state (Liu et al., 2010). Mu et al. yielded accuracy of 85% with FE and Support Vector Machine (SVM) classifier (Mu et al., 2017a). Xiong et al. proposed a feature combination of AE and SE with SVM classifier to test driving fatigue, and achieved the best accuracy of 91.3% (Xiong et al., 2016). Khushaba et al. exploited a feature extraction by using fuzzy mutual-information method and achieved 92.8% (Khushaba et al., 2011). Hu hit highest accuracy of 96.6% with FE and Random Forest classifier (Hu, 2017).

**Abbreviations:** FE, Fuzzy entropy; SE, Sample entropy; AE, Approximate entropy; PE, Spectral entropy; EEG, Electroencephalogram; AUC, Areas under ROC curves; SVM, Support Vector Machine; DT, Decision Tree; NB, Naive Bayes; AdaBoost, Adaptive Boosting; ERR, Error rate.

From the literature review, it has been observed that few studies have been conducted for using ensemble classifier based on EEG to study driver fatigue detection. Keeping this in mind, the prime motivation of this work is to develop an automated detection system for driver fatigue based on ensemble classifier. The scheme employs four types of entropy for feature extraction and AdaBoost (Freund and Schapire, 1997; Hastie et al., 2009) for classification of EEG signals into normal and fatigue. Several experiments on 28 subjects indicate that the proposed scheme earns better detection performance and robustness in comparison to other existing schemes.

The rest of this article is described as below. In Materials and Methods, data acquisition, feature extraction, and classification are illustrated. The results are discussed in Section Results presents the evaluation of the method with the obtained results, followed by a general discussion about classifier accuracy in Section Discussion.

## MATERIALS AND METHODS

### Subjects

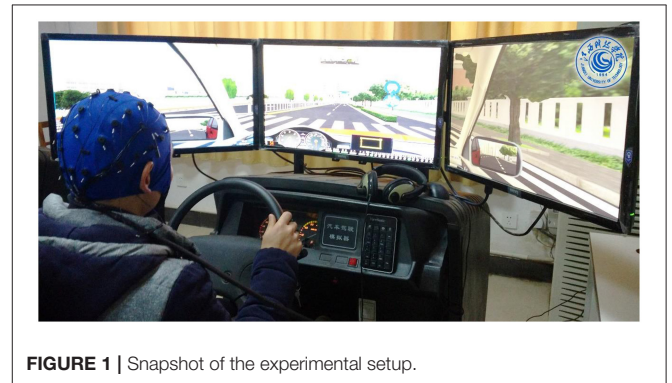
Twenty-eight university students (14 male, 19–24 years) participated in this experiment, which all had a current driver's license. Before the experiment, they practiced driving for several minutes to familiarize themselves with the process and purpose of the experiment. The experiment was approved by the Academic Ethics Committee of the Jiangxi University of Technology according to the standards of the Declaration of Helsinki. Written informed consent was obtained from each subject.

### Experiment

In the static driving simulator (ZY-31D, ZhongYu CO., LTD, China), the driver's fatigue simulation test was performed on each subject, as shown in **Figure 1**. The driving environment selected for this work was a highway with low traffic density so as to induce monotonous driving, which easily leads to driver fatigue state.

### Data Recording

Similar to former experiments (Chai et al., 2017; Hu, 2017; Mu et al., 2017b), when the driving procedure started 20 min, the last 5-min EEG recordings were marked as normal state. When



**FIGURE 1** | Snapshot of the experimental setup.

the continuous driving procedure lasted 60–120 min until the questionnaire results (Lee’s subjective fatigue scale and Borg’s CR-10 scale; Borg, 1990; Lee et al., 1991), participants’ responses and electrooculogram (EOG) signals show the subject was in driving fatigue state, the last 5-min recorded EEG recordings were marked as fatigue state. EOG signals were used to determine fatigue state using the blink rate and eye closure such as, the small and slow blinks.

All channel data were referenced to two reference electrodes A1 and A2, and digitized at 1 kHz from a 32-channel electrode cap (including 2 reference electrodes) based on the international 10–20 system.

After the acquisition of EEG signals, the main procedures of data preprocessing was implemented by Scan 4.3 software of Neuroscan (Compumedics, Australia). The original signals were first filtered and a 0.15–45 Hz band-pass filter was used. Then 5-min EEG signals from 30 channels were sectioned into 1-s epochs, resulting in 300 epochs. With the 28 subjects and 30 channels, a total of 504,000 units were randomly formed for dataset (each state having 252,000 units).

### Feature Extraction

The EEG is assumed to be a non-stationary time series and most feature extraction methods are only applicable to stationary signal. To deal with this problem, the EEG time series were divided into several short windows and its statistics is assumed to be approximately stationary within each window. The following feature extraction methods are applied to each 1-s windowed signal. EEG signals are segmented without overlap, finally feature sets are extracted from all channels in each 1 s window.

The ability to distinguish between normal state and fatigue state depends mainly on the quality of the input vectors of the classifier. In order to capture EEG features, four feature sets are computed, including FE, SE, AE, and PE. In this section, the computational methods of these feature sets in EEG recordings are described in detail.

### Spectral Entropy (PE)

PE is evaluated using the normalized Shannon entropy (Kannathal et al., 2005), which quantifies the spectral complexity of the time series. The power level of the frequency component is indicated by  $Y_i$  and  $y_i$  is normalized as:

$$y_i = \frac{Y_i}{\sum Y_i} \tag{1}$$

The PE of the time series is calculated using the following equation:

$$PE = \sum y_i \log\left(\frac{1}{y_i}\right) \tag{2}$$

### Approximate Entropy (AE)

AE, as proposed by Pincus (1991), is a statistically quantified non-linear dynamic parameter that measures the complexity of a time series. The procedure for the AE-based algorithm is described as follows:

- (1) Considering a time series  $t(i)$ , a set of  $m$ -dimensional vectors are obtained according to the sequence order of  $t(i)$ :

$$T_i^m = [t(i), t(i + 1), \dots, t(i + m - 1)]; i \leq L - m + 1 \tag{3}$$

$d[T_i^m, T_j^m]$  is the distance between two vectors  $T_i^m$  and  $T_j^m$ , defined as the maximum difference values between the corresponding elements of two vectors:

$$d[T_i^m, T_j^m] = \max\{|t(i + k) - t(j + k)|\}, (i, j = 1 \sim L - m + 1, i \neq j) \tag{4}$$

- (2) Define  $S_i$  is the number of vectors  $T_j$  that are similar to  $T_i$ , subject to the criterion of similarity  $d[T_i^m, T_j^m] \leq s$

$$S_i^m(s) = \frac{1}{L - m + 1} S_i \tag{5}$$

- (3) Define the function  $\gamma^m(s)$  as:

$$\gamma^m(s) = \frac{1}{L - m + 1} \sum_{i=1}^{L-m+1} \ln S_i^m(s) \tag{6}$$

- (4) Set  $m = m + 1$ , and repeat steps (3) to (6) to obtain  $S_i^{m+1}(s)$  and  $\gamma^{m+1}(s)$ , then:

$$\gamma^{m+1}(s) = \frac{1}{L - m} \sum_{i=1}^{L-m} \ln S_i^{m+1}(s) \tag{7}$$

- (5) The AE can be expressed as:

$$AE = \gamma^m(s) - \gamma^{m+1}(s) \tag{8}$$

### Sample Entropy (SE)

SE’s algorithm is similar to that of AE (Yentes et al., 2013), which is a new measure of time series complexity proposed by Richman and Moorman (2000). The step (1) can be defined in the same way as the AE-based algorithm; other steps in the SE-based algorithm are described as follows:

- (1) Define  $A_i$  is the number of vectors  $T_j$  that are similar to  $T_i$  subject to the criterion of similarity  $d[T_i^m, T_j^m] \leq s$

$$A_i^m(s) = \frac{1}{L - m - 1} A_i \tag{9}$$

- (2) Define the function  $\gamma^m(s)$  as:

$$\gamma^m(s) = \frac{1}{L - m} \sum_{i=1}^{L-m} A_i^m(s) \tag{10}$$

- (3) Set  $m = m + 1$ , and repeat above steps to obtain  $A_i^{m+1}(s)$  and  $\gamma^{m+1}(s)$ , then

$$\gamma^{m+1}(s) = \frac{1}{L - m - 1} \sum_{i=1}^{L-m-1} A_i^{m+1}(s) \tag{11}$$

- (4) The SE can be expressed as:

$$SE = \log(\gamma^m(s) / \gamma^{m+1}(s)) \tag{12}$$

## Fuzzy Entropy (FE)

To deal with some of the issues with SE, Xiang et al. proposed the use of fuzzy membership function in computing the vector similarity to replace the binary function in SE algorithm (Xiang et al., 2015), so that the entropy value is continuous and smooth. The procedure for the FE-based algorithm is described in detail as follows:

- (1) Set a  $L$ -point sample sequence:  $\{v(i) : 1 \leq i \leq L\}$ ;
- (2) The phase-space reconstruction is performed on  $v(i)$  according to the sequence order. The reconstructed vector can be written as:

$$T_i^m = \{v(i), v(i+1), \dots, v(i+m-1)\} - v_0(i) \quad (13)$$

where  $i = 1, 2, \dots, L - m + 1$ , and  $v_0(i)$  is the average value described as the following equation:

$$v_0(i) = \frac{1}{m} \sum_{j=0}^{m-1} v(i+j) \quad (14)$$

- (3)  $d_{ij}^m$ , the distance between two vectors  $T_i^m$  and  $T_j^m$ , is defined as the maximum difference values between the corresponding elements of two vectors:

$$d_{ij}^m = d[T_i^m, T_j^m] = \max_{k \in (0, m-1)} \{|v(i+k) - v_0(i) - (v(j+k) - v_0(j))|\} \\ (i, j = 1 \sim L - m, i \neq j) \quad (15)$$

- (4) According to the fuzzy membership function  $\sigma(d_{ij}^m, n, s)$ , the similarity degree  $D_{ij}^m$  between two vectors  $T_i^m$  and  $T_j^m$  is defined as:

$$D_{ij}^m = \sigma(d_{ij}^m, n, s) = \exp(-(d_{ij}^m)^n / s) \quad (16)$$

where the fuzzy membership function  $\sigma(d_{ij}^m, n, s)$  is an exponential function, while  $n$  and  $s$  are the gradient and width of the exponential function, respectively.

- (5) Define the function  $\gamma^m(n, s)$ :

$$\gamma^m(n, s) = \frac{1}{L-m} \sum_{i=1}^{L-m} \frac{1}{L-m-1} \sum_{j=1, j \neq i}^{L-m} D_{ij}^m \quad (17)$$

- (6) Repeat the steps from (1) to (4) in the same manner. Define the function:

$$\gamma^{m+1}(n, s) = \frac{1}{L-m} \sum_{i=1}^{L-m} \frac{1}{L-m-1} \sum_{j=1, j \neq i}^{L-m} D_{ij}^{m+1} \quad (18)$$

- (7) The FE can be expressed as:

$$FE(m, s, n) = \ln \gamma^m(n, s) - \ln \gamma^{m+1}(n, s) \quad (19)$$

In these four entropies,  $m$  and  $s$  are the dimensions of phase space and similarity tolerance, respectively. In the present study,  $m = 2$ ,  $n = 4$  while  $s = 0.2 * SD$ , where  $SD$  denotes the standard deviation of the time series.

For optimizing the detection quality, the feature sets were normalized for each subject and each channel by scaling between  $-1$  and  $1$ .

## Classification

To avoid over-fitting problem, the datasets were separated into train sets and test sets in the following pattern. In the train phase, 10-fold cross validation applied on the features such that 10% feature vectors are dedicated as test set and other 90% feature vectors are considered as the train set. In the next iteration, another 10% feature vectors consider as test set and the rest for the train set, till all of feature vectors involved one time in the test process. The final result was obtained by averaging the results of corresponding turns. By this evaluation scheme, the dependencies of the train and test sets were eliminated.

Since there is no uniform classification method suitable for all subjects and all applications, usually it may be useful to test multiple methods (Zhang et al., 2017). In this work, three types of base classifiers namely Decision Trees (DT), Support Vector Machine (SVM), and Naive Bayes (NB) were used. DT is a non-parametric supervised learning method used for classification. DT establishes several binary decision functions on the features. DT1 and DT9 represent DT with the maximum depth of the tree being 1 and 9 in this work, respectively. In the case of non-linear classification, kernels, such as, radial basis functions (RBF), are used to map the data into a higher dimensional feature space in which a linear separating hyper-plane could be found. Naive Bayes method is based on applying Bayes' theorem with the "naive" assumption. The likelihood in NB of the features is assumed to be Gaussian. In this study, grid parameter search was used to achieve better results.

AdaBoost is an eminent ensemble learning based classification model (Amal Feltane et al., 2013; Yang et al., 2016), which was first proposed by Freund and Schapire (1997). AdaBoost produces the final output by weighting the decisions of all these weak classifiers using majority vote method. The AdaBoost algorithm is described as follow:

### Algorithm AdaBoost

*Definition train dataset*  $(X, Y) = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_N, y_N)\}$ ,  $y_i \in \{-1, 1\}$

*iterator*:  $M$ ;

*Initialize each weight*  $W_{1,i} = \frac{1}{N}$ ,  $i = 1, 2, \dots, N$ ,

*Linear combination function of basic classifiers*  $f_0(x) = 0$   
for  $m = 1$  to  $M$  do

*train a base learner*:  $D_m(x)$

*calculate error rate*:  $e_m = \sum_{i=1}^N W_{m,i} I(D_m(x_i) - y_i)$

$\alpha_m = \frac{1}{2} \ln\left(\frac{1-e_m}{e_m}\right)$

*update weight*:  $W_{m+1,i} = \frac{W_{m,i}}{Z_m} \exp(-\alpha_m y_i D_m(x_i))$ ,

*normalization factor*  $Z_m = \sum_{i=1}^N W_{m,i} \exp(-\alpha_m y_i D_m(x_i))$

$f_m(x) = f_{m-1}(x) + \alpha_m D_m(x)$

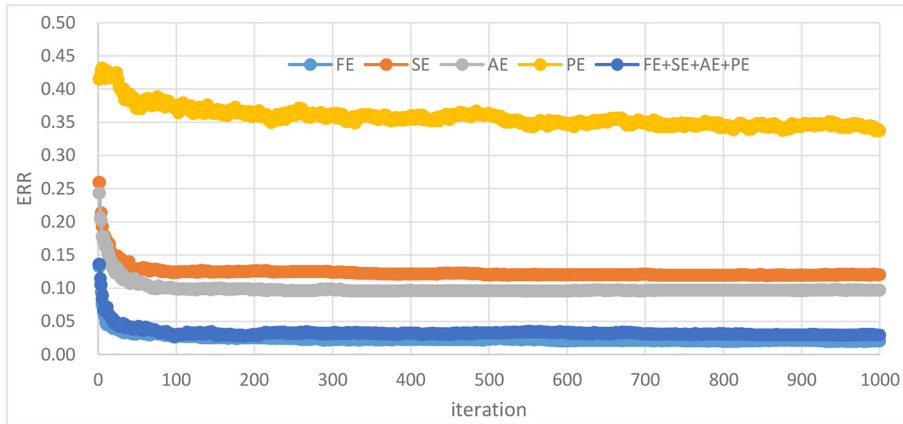
end for

*Output sign*  $(f_M(x)) = \text{sign}(\sum_{m=1}^M \alpha_m D_m(x))$

In this work, the DT9 was used as base classifiers.

## Performance Evaluation

To provide an easier-to-understand method to assess the classification quality, the results of classification and the performance of classifiers are expressed in terms of Error rate, Precision, Recall, F1 score, MCC, and AUC which are defined as follows:



**FIGURE 2** | ERR for AdaBoost based on different feature sets.

Error rate (ERR) calculates the total number of EEG segments which are incorrectly classified

$$\text{ERR} = \frac{(\text{FP} + \text{FN})}{(\text{FP} + \text{TN} + \text{FP} + \text{FN})}$$

The precision intuitively reflects the ability of the classifier to determine the whole sample—which the positive is identified as positive and the negative is identified as negative.

$$\text{Precision} = \frac{TP}{TP + FP}$$

The recall intuitively reflects the proportion of positive samples that are correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

The F1 score can be interpreted as a weighted average of precision and recall, where an F1 score reaches the optimum value at 1 and the worst score at 0.

$$\text{F1} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

The Matthews correlation coefficient (MCC) is used in machine learning as a measure of the quality of two-class classifications. The MCC is in essence a correlation coefficient value between -1 and +1.

$$\text{MCC} = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Where

- TP (True Positive) = correctly identified normal segments
- TN (True Negative) = correctly identified fatigue segments
- FN (False Negative) = incorrectly identified normal segments
- FP (False Positive) = incorrectly identified fatigue segments

AUC illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by drawing true positive rates from positive (true positive rate) and false positive rates (false positive rates) in a variety of threshold settings.

## Statistical Analysis

In order to investigate differences of average accuracy among various classifiers and feature sets, the paired sample *t*-test was used to evaluate effectiveness on each comparison. The results are averages over 10 independently turns of combination of train set and test set in each experiment.

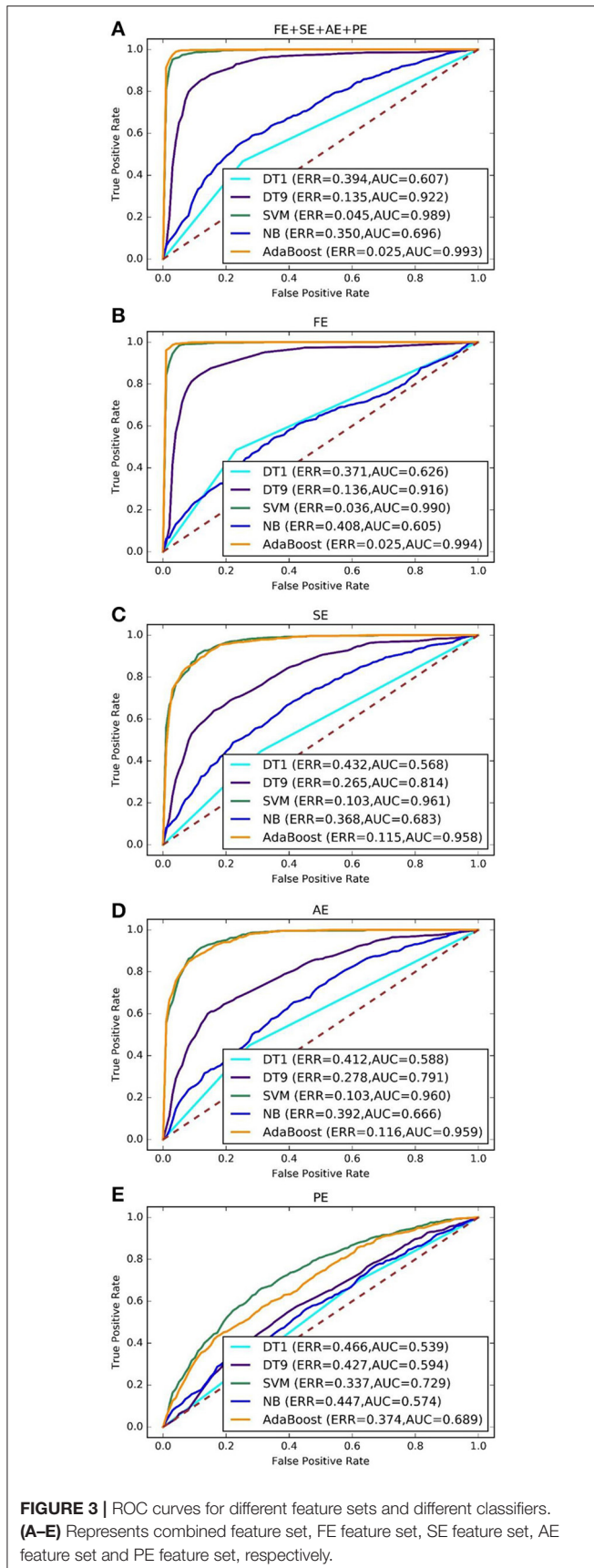
## RESULTS

In order to verify the validity, effectiveness, and robust of proposed method, some experiments were performed on 28 subjects.

### Comparison with Different Feature Sets and Different Classifiers

As shown in **Figure 2**. FE feature set performs slightly better than the combined entropy (FE + SE + AE + PE) feature set (0.020 against 0.029). A paired *t*-test across the 10 independent comparisons indicates a significant difference with *p*-value around 0.003. It can be seen that the FE feature set performs about 0.098 and 0.075 better than the SE and AE feature set at ERR index, respectively. A paired *t*-test over 10 independent comparisons shows a significant difference with *p*-value lower than 0.001. AE feature set performs slightly better than SE feature set. It can also show that the PE feature set performs worst with the lowest ERR being about 0.337.

The results of 10 independently rounds are used to draw mean ROC curves. Different feature sets or classifiers were compared by analyzing their ROC curves and areas under ROC curves (AUC). In **Figures 3A–E**, their performance in ROC curves produced was compared by different classifiers on combined entropy feature set, FE feature set, SE feature set, AE feature set



and PE feature set, respectively. It shows that the FE feature set and combined feature set outperform other feature sets, which similar to **Figure 2**. For example, the best ERR of FE feature set and combined feature set are both 0.025, while the best ERR of SE, AE, and PE are 0.115, 0.116, and 0.374, respectively. The best AUC of FE feature set and combined feature set is 0.993 and 0.994, respectively while the best ERR of SE, AE, and PE are 0.961, 0.960, and 0.729, respectively. Consequently, adding more features makes nothing changes for driver fatigue detection. Therefore, the FE feature set is selected for the next experiments.

As shown in **Figure 4**, it illustrate that AdaBoost outperform other classifiers. For instance, the best ERR and AUC is 0.025 and 0.994 for AdaBoost based on FE feature set, while the best ERR and AUC is 0.036 and 0.990 for SVM based on FE feature set. The  $p$ -value is 0.0062 between AdaBoost and SVM. AdaBoost classifier is significantly better than other classifiers. The  $p$ -values are 0.0032 and 0.0001, by paired  $t$ -test between DT9 and AdaBoost, and between NB and AdaBoost, respectively. It's conjectured that AdaBoost models work best because they may be more robust than other models such as, DT and NB when dealing with scalar data sets that are not too larger.

To evaluate the effectiveness of AdaBoost in the classification of EEG signals, the classification indexes including ERR, Precision, Recall, F1, score and MCC of the four classifiers were compared based on FE feature set. As shown in **Figure 4** and **Table 1**, the overall performance of AdaBoost is the best of the four classifiers in terms of ERR, Precision, Recall, MCC, and F1 score. The ERR of AdaBoost can reach  $0.024 \pm 0.002$ , which is almost 0.011 lower than SVM ( $0.035 \pm 0.005$ ). The ERR of DT1, DT9, and NB is  $0.0369 \pm 0.014$ ,  $0.0142 \pm 0.008$ , and  $0.405 \pm 0.012$ , respectively.

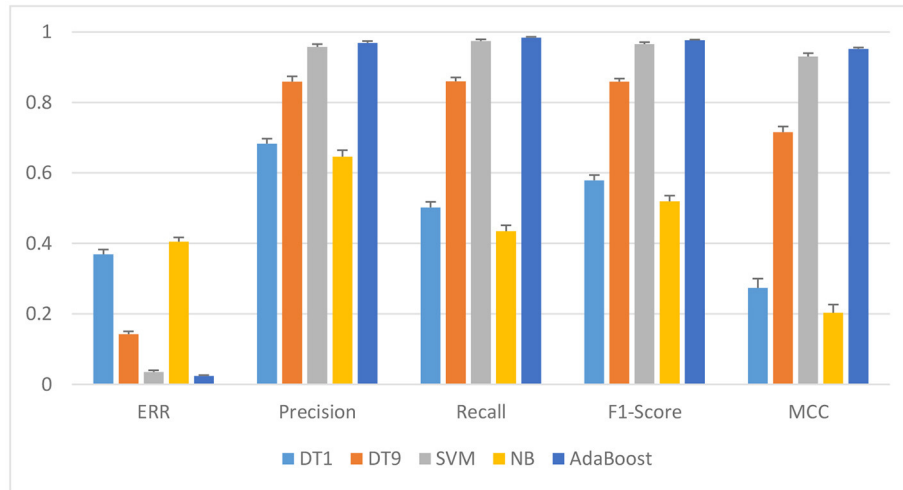
## Parameter Setting

The main parameters to be adjusted in AdaBoost method are parameter  $max\_depth$  of base classifier DT and  $lr$ . Best performance of AdaBoost model can be yielded through carefully choosing the optimal combination of these parameters. The parameter  $max\_depth$  is the most important one in the DT, which controls the maximum depth of the tree. **Figure 5A** shows the error rates under different  $max\_depth$  and fixed iteration ( $=500$ ) based on FE feature set. It is showed that the average error rate attains the minimal point  $0.022 \pm 0.004$  when  $max\_depth$  equals to about 8. From **Figure 5B**, it can be seen that the average error rate starts to even out at  $0.022 \pm 0.003$  when the value of  $lr$  smaller than about 1.5. According to these results, the final AdaBoost classifier in next experiments is set with the parameters  $max\_depth = 9$  (DT9) and  $lr = 1.0$ .

## Comparison with Different Size of Test Samples

The ratio of train samples for test samples is important for the performance of classifier. To determine the robustness of the classifier against size of test sample or train size, the ratio of test samples for all samples is set varying from 0.03 to 0.97. The ERR of AdaBoost against different ratio is shown in **Figure 6**.

It is observed that the average error rate begins to stabilize at about 0.03 when the ratio being about 0.5. When the ratio



**FIGURE 4** | Performance of different classifiers based on FE feature set.

**TABLE 1** |  $p$ -value between AdaBoost and other classifiers with paired  $t$ -test.

Classifiers	Index				
	ERR	Precision	Recall	F1 score	MCC
SVM	1.05e-5	4.81e-4	5.60e-4	1.31e-4	8.92e-5
DT1	2.12e-7	4.82e-7	4.10e-7	4.46e-7	4.41e-7
DT9	2.07e-7	1.32e-6	9.46e-7	6.61e-7	6.10e-7
NB	2.06e-7	5.33e-7	3.98e-7	4.27e-7	3.97e-7

becomes larger, the ERR also becomes larger, but when ratio reaches close to 1.0, the ERR is close to 10% and becomes worse, possibly because of the lack of training samples. On the contrary, when the ratio becomes smaller, ERR is stable at around 0.02, which indicating that ratio is more appropriate in the 0.1.

$P$ -value between AdaBoost and SVM, between AdaBoost and DT9, are  $3e-8$  and  $4e-16$ , respectively. Compared to SVM and DT, AdaBoost seems to have better robustness against changes of ratio of test samples for all samples.

## Comparison with Different Number of Subjects

The number of subjects is also an important parameter in the driving fatigue detection system. More subjects can provide more information that may improve or reduce detection performance. Generally speaking, when average performance is poor, any subject with higher accuracy can improve the overall performance, and vice versa. Sometimes, the classifier model that is suitable for small samples may lose performance when large samples are used. However, when more subjects are involved, the system costs, including hardware and computation time, will also increase. Therefore, a tradeoff between the system performance and system cost should be based on the specificity of the application.

To answer the question of how many subjects are needed to train for a satisfactory detection system, system performance

was evaluated with respect to the number of subjects. For each number  $n$  (from 2 to 28), a random combination ( $n$  out of 28 subjects) was repeated 20 times for calculating classification accuracy using 10-fold cross validation. Three classifiers approaches were calculated for comparison. Furthermore, for each condition ( $n$  from 2 to 28), the paired  $T$ -test was used as a post-hoc test to evaluate if the performance of AdaBoost was significantly better than that of other two classifiers.

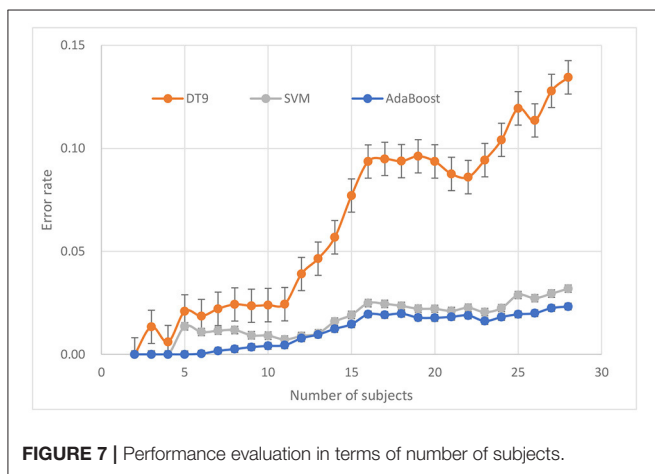
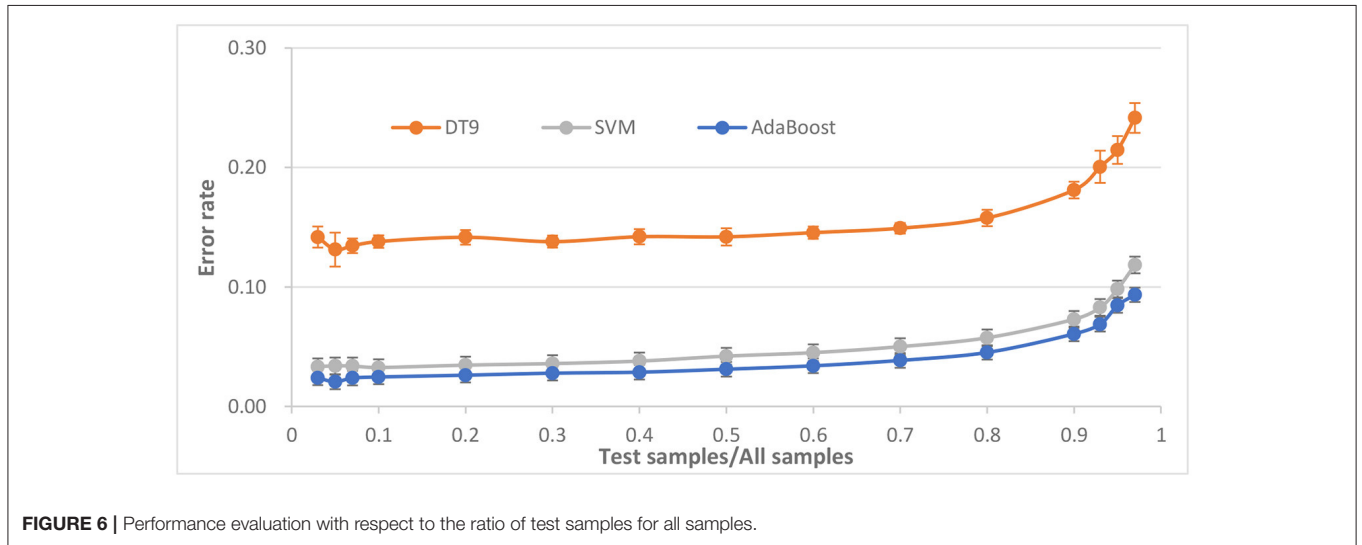
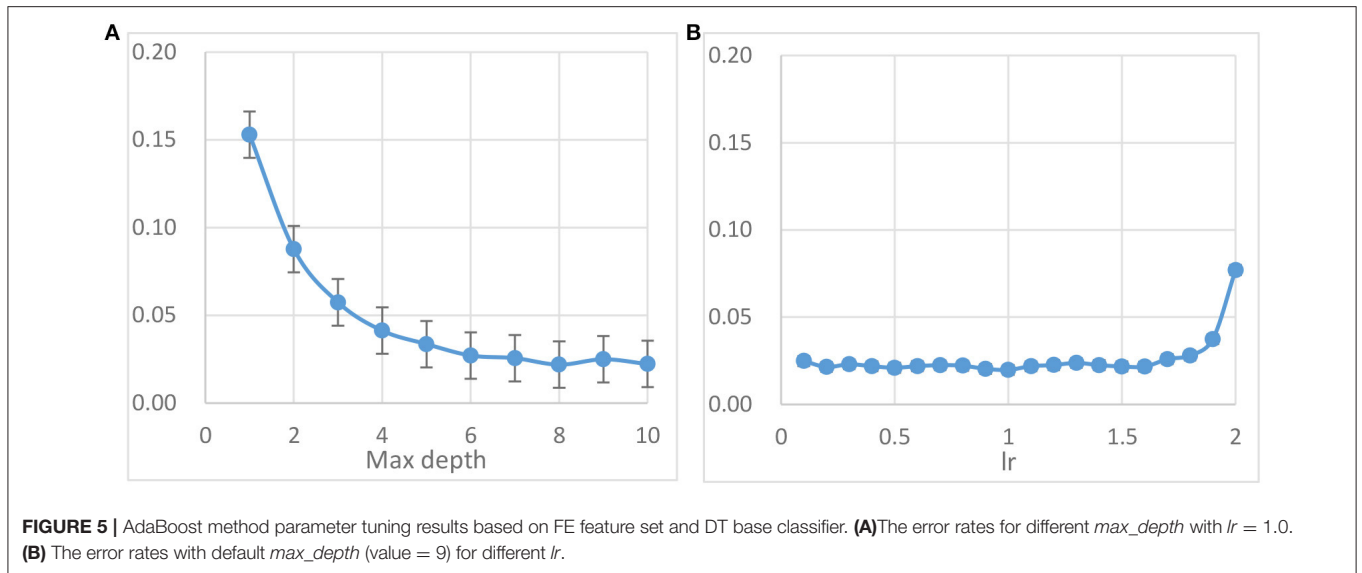
The ERR of AdaBoost against different number of subjects is shown in **Figure 7**. It can be seen that, for AdaBoost classifier, when the number of subjects is  $<13$ , ERR is  $<0.01$ , when the number of subjects continue to increase, ERR also increases, and is stable at about 0.02. ERR is not increasing monotonically with the number of subjects but tending to reach equilibrium.

$P$ -values between AdaBoost and SVM, between AdaBoost and DT9, are  $4.665e-8$  and  $1.409e-8$ , respectively. Compared to SVM and DT, AdaBoost seems to have better robustness against changes of number of subjects.

## DISCUSSION

As see in **Table 2**, it is found that the classification performance of proposed method was better than that in the others research using entropy feature sets. Although, based on the existing EEG data, the optimal performance of detection of driving fatigue by using AdaBoost-based method showed well application on the real-time detection of driving fatigue.

Among the state-of-art classifier schemes, four representative algorithms, DT, NB, SVM, and AdaBoost were experimented for classification tasks of some data sets. These classifiers have been shown very effective in many pattern recognition applications. These classifiers are applied on the extracted features and their results are shown in **Figures 3, 4** in which AdaBoost



**TABLE 2 |** Studies regarding driver fatigue detection using entropy feature sets.

Research group	Feature method	Highest accuracy (%)
Liu et al., 2010	AE and others	84
Mu et al., 2017a	FE	85
Xiong et al., 2016	AE and SE	91.3
Khushaba et al., 2011	FE	92.8
Hu, 2017	FE	96.6
This paper	FE	97.5

showed a better results in comparison with the three other classifiers.

Also to evaluate robustness of the classifiers, different combinations of train set and test sets were employed and the classification results were brought in **Figures 6, 7**. A repeated



progressive method with various sample sizes was applied to find out if there is any relationship between data set size and the performance. It can be seen that AdaBoost is also more robust than the three other classifiers.

The experiment confirmed that, in comparison with the AE, PE, and SE, the FE had a better consistency and better discrimination ability. The results also showed that the differences between the normal state and the fatigue state were relative larger from the FE from the AE, SE, or the PE, confirming that the FE had a better performance in distinguishing fatigue state. The result achieved in this study ensured confidence in probing the theoretical reason for the different discrimination ability and, hence, leads to new ideas for exploring the inherent physiological mechanisms when using the entropy methods. This indicated that the FE could be an effective method for the driver fatigue detection.

However, there are several limitations in this study. First, it is worth noting that the parameter settings for the SE, AE, and PE method are the local similarity and parameters may not be the optimal solution. Second, the number of subject is relatively small. Although according to the existing literature in the Introduction section, the 28 subjects are not too small, but the number still needs to be increased. Third, only three commonly used classifiers and the four feature sets were compared in this study. Last, the different impacts of different channels haven't been took into account.

## REFERENCES

- Acharya, U. R., Molinari, F., Sree, S. V., Chattopadhyay, S., Ng, K. H., and Suri, J. S. (2012). Automated diagnosis of epileptic EEG using entropies. *Biomed. Signal Process. Control* 7, 401–408. doi: 10.1016/j.bspc.2011.07.007
- Amal Feltane, G., Boudreaux-Bartels, F., and Besio, W. (2013). Automatic seizure detection in rats using laplacian EEG and verification with human seizure signals. *Ann. Biomed. Eng.* 41, 645–654. doi: 10.1007/s10439-012-0675-4
- Boostani, R., Sadatnezhad, K., and Sabeti, M. (2009). An efficient classifier to diagnose of schizophrenia based on the EEG signals. *Expert Syst. Appl.* 36, 6492–6499. doi: 10.1016/j.eswa.2008.07.037
- Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scand. J. Work Environ. Health.* 16, 55–58. doi: 10.5271/sjweh.1815
- Chai, R., Ling, S. H., San, P. P., Naik, G. R., Nguyen, T. N., Tran, Y., et al. (2017). Improving EEG-based driver fatigue classification using sparse-deep belief networks. *Front. Neurosci.* 11:103. doi: 10.3389/fnins.2017.00103
- Chen, W., Zhuang, J., Yu, W., and Wang, Z. (2009). Measuring complexity using fuzzyen, apen, and sampen. *Med. Eng. Phys.* 31, 61–68. doi: 10.1016/j.medengphy.2008.04.005
- Correa, A. G., Orosco, L., and Laciari, E. (2014). Automatic detection of drowsiness in EEG records based on multimodal analysis. *Med. Eng. Phys.* 36, 244–249. doi: 10.1016/j.medengphy.2013.07.011
- Freund, F., and Schapire, R. (1997). A decision-theoretic generalization of on-line learning algorithms and an application to boosting. *J. Comput. Syst. Sci.* 55, 119–139. doi: 10.1006/jcss.1997.1504
- Fu, R. R., Wang, H., and Zhao, W. B. (2016). Dynamic driver fatigue detection using hidden Markov model in real driving condition. *Expert Syst. Appl.* 63, 397–411. doi: 10.1016/j.eswa.2016.06.042
- Guo, L., Rivero, D., and Pazos, A. (2010). Epileptic seizure detection using multiwavelet transform based approximate entropy and

## CONCLUSION

In this paper, a method to develop an ensemble classifier for recognizing fatigue was proposed. A new EEG feature vector based on FE, SE, AE, and PE was used as input into four different classifiers: DT, NB, SVM, and AdaBoost. It was concluded that the combination of these feature sets or FE feature set with the AdaBoost provided the best performance on EEG dataset. The proposed method had very high accuracy classifying driver fatigue events. Further, it was showed how the method for detecting fatigue segments was robust.

## AUTHOR CONTRIBUTIONS

JH conceived and designed the experiments; JH analyzed the data and wrote the paper.

## ACKNOWLEDGMENTS

This work was supported by Project of Department of Science and Technology, Jiangxi Province (No 20151BBE50079), Project of Department of Education, Jiangxi Province (No GJJ151146 and No GJJ161143) and Patent transformation Project of Intellectual Property Office of Jiangxi Province [The application and popularization of the digital method to distinguish the direction of rotation photoelectric encoder in identification]. Thanks P. Wang, J. L. Min and Z. D. Mu for collecting and preprocessing EEG data.

- artificial neural networks. *J. Neurosci. Methods* 193, 156–163. doi: 10.1016/j.jneumeth.2010.08.030
- Hastie, T., Rosset, S., Zhu, J., and Zou, H. (2009). Multi-class AdaBoost. *Stat. Interf.* 2, 349–360. doi: 10.4310/sii.2009.v2.n3.a8
- Hu, J. F. (2017). Comparison of different features and classifiers for driver fatigue detection based on a single EEG channel. *Comput. Math. Methods Med.* 2017:9. doi: 10.1155/2017/5109530
- Hu, J. F., Mu, Z. D., and Wang, P. (2015). Multi-feature authentication system based on event evoked electroencephalogram. *J. Med. Imaging Health Inform.* 5, 862–870. doi: 10.1166/jmhi.2015.1471
- Kannathal, N., Choo, M. L., Acharya, U. R., and Sadasivan, P. (2005). Entropies for detection of epilepsy in EEG. *Comput. Methods Programs Biomed.* 80, 187–194. doi: 10.1016/j.cmpb.2005.06.012
- Kar, S., Bhagat, M., and Routray, A. (2010). EEG signal analysis for the assessment and quantification of driver's fatigue. *Transp. Res. F Traffic Psychol. Behav.* 13, 297–306. doi: 10.1016/j.trf.2010.06.006
- Khushaba, R. N., Kodagoda, S., Lal, S., and Dissanayake, G. (2011). Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE Trans. Biomed. Eng.* 58, 121–131. doi: 10.1109/TBME.2010.2077291
- Lal, S. K., and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* 55, 173–194. doi: 10.1016/S0301-0511(00)00085-5
- Lee, K. A., Hicks, G., and Nino-Murcia, G. (1991). Validity and reliability of a scale to assess fatigue. *Psychiatry Res.* 36, 291–298. doi: 10.1016/0165-1781(91)90027-M
- Li, W., He, Q. C., Fan, X. M., and Fei, Z. M. (2012). Evaluation of driver fatigue on two channels of EEG data. *Neurosci. Lett.* 506, 235–239. doi: 10.1016/j.neulet.2011.11.014
- Liu, J. P., Zhang, C., and Zheng, C. X. (2010). EEG-based estimation of mental fatigue by using KPCA-HMM and complexity parameters. *Biomed. Signal Process. Control* 5, 124–130. doi: 10.1016/j.bspc.2010.01.001

- Mousa Kadhim, W., Murugappan, M., and Ahmmad, B. (2013). Wavelet packet transform based driver distraction level classification using EEG. *Mathem. Probl. Eng.* 3, 841–860. doi: 10.1155/2013/297587
- Mu, Z. D., Hu, J. F., and Min, J. L. (2016). EEG-based person authentication using a fuzzy entropy-related approach with two electrodes. *Entropy* 18:432. doi: 10.3390/e18120432
- Mu, Z., Hu, J., and Yin, J. (2017a). Driving fatigue detecting based on EEG signals of forehead area. *Int. J. Pattern Recognit. Artif. Intell.* 31:12. doi: 10.1142/S0218001417500112
- Mu, Z., Hu, J., and Min, J. (2017b). Driver fatigue detection system using electroencephalography signals based on combined entropy features. *Appl. Sci.* 7:150. doi: 10.3390/app7020150
- Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* 88, 2297–2301. doi: 10.1073/pnas.88.6.2297
- Reyes-Sanchez, E., Alba, A., Mendez, M. O., Milioli, G., and Parrino, L. (2016). Spectral entropy analysis of the respiratory signal and its relationship with the cyclic alternating pattern during sleep. *Int. J. Mod. Phys. C* 27:10. doi: 10.1142/S0129183116501400
- Richman, J. S., and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049.
- Xiang, J., Li, C., Li, H., Cao, R., Wang, B., Han, X., et al. (2015). The detection of epileptic seizure signals based on fuzzy entropy. *J. Neurosci. Methods* 243, 18–25. doi: 10.1016/j.jneumeth.2015.01.015
- Xiong, Y., Gao, J., Yang, Y., Yu, X., and Huang, W. (2016). Classifying driving fatigue based on combined entropy measure using EEG signals. *Int. J. Control Autom.* 9, 329–338. doi: 10.14257/ijca.2016.9.3.30
- Yang, T., Chen, W. T., and Cao, G. T. (2016). Automated classification of neonatal amplitude-integrated EEG based on gradient boosting method. *Biomed. Signal Process. Control.* 28, 50–57. doi: 10.1016/j.bspc.2016.04.004
- Yentes, J. M., Hunt, N., Schmid, K. K., Kaipust, J. P., McGrath, D., and Stergiou, N. (2013). The appropriate use of approximate entropy and sample entropy with short data sets. *Ann. Biomed. Eng.* 41, 349–365. doi: 10.1007/s10439-012-0668-3
- Yin, J. H., Hu, J. F., and Mu, Z. D. (2017). Developing and evaluating a mobile driver fatigue detection network based on electroencephalograph signals. *Healthcare Technol. Lett.* 4, 34–38. doi: 10.1049/htl.2016.0053
- Zhang, C. S., Liu, C. C., Zhang, X. L., and Alpanidis, G. (2017). An up-to-date comparison of state-of-the-art classification algorithms. *Expert Syst. Appl.* 82, 128–150. doi: 10.1016/j.eswa.2017.04.003
- Zhao, C., Zheng, C., Zhao, M., and Liu, J. (2010). Physiological assessment of driving mental fatigue using wavelet packet energy and random forests. *Am. J. Biomed. Sci.* 2, 262–274. doi: 10.5099/aj100300262

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.