Random forest–based feature selection and detection method for drunk driving recognition

ZhenLong Li¹, HaoXin Wang¹, YaoWei Zhang¹ and XiaoHua Zhao¹

Abstract
A method for drunk driving detection using Feature Selection based on the Random Forest was proposed. First, driving behavior data were collected using a driving simulator at Beijing University of Technology. Second, the features were selected according to the Feature Importance in the random forest. Third, a dummy variable was introduced to encode the geometric characteristics of different roads so that drunk driving under different road conditions can be detected with the same classifier based on the random forest. Finally, the linear discriminant analysis, support vector machine, and AdaBoost classifiers were used and compared with the random forest. The accuracy, F1 score, receiver operating characteristic curve, and area under the curve value were used to evaluate the performance of the classifiers. The results show that Accelerator Depth, Speed, Distance to the Center of the Lane, Acceleration, Engine Revolution, Brake Depth, and Steering Angle have important influences on identifying the drivers’ states and can be used to detect drunk driving. Specifically, the classifiers with Accelerator Depth outperformed the other classifiers without Accelerator Depth. This means that Accelerator Depth is an important feature. Both the AdaBoost and random forest classifiers have an accuracy of 81.48%, which verified the effectiveness of the proposed method.

Keywords
Drunk driving detection, feature selection, random forest, traffic safety, driving behavior

Date received: 8 August 2019; accepted: 21 December 2019
Handling Editor: Paolo Bellavista

Introduction
With the increasing number of traffic accidents, the casualties and property losses from traffic accidents have caused increasingly more damage to the whole society. According to the World Health Organization,¹ traffic accidents have become one of the top 10 causes of death since 2016. Globally, alcohol consumption is the seventh most important factor for both deaths and disabilities; meanwhile, in the age range of 15–49 years, nearly 12.2% of males died from alcohol-related causes.² According to the National Highway Traffic Safety Administration,³ the number of fatal alcohol-related crashes in the United States was 34,748 in 2016, which means that alcohol-impaired traffic accidents account for 29% of the total number of traffic accidents. In China, the number of dangerous traffic accidents is approximately several hundred thousand every year, of which the accidents that are caused by alcohol account for 34.1%.⁴ Drunk driving frequently results in dangerous accidents. Therefore, increasingly more public attention has been paid to traffic safety research. If we can perceive the states of the drivers as soon as possible and intervene in advance, the accidents

¹College of Metropolitan Transportation, Beijing University of Technology, Beijing, China

Corresponding author:
ZhenLong Li, College of Metropolitan Transportation, Beijing University of Technology, 100 Pingleyuan Street, Chaoyang, Beijing 100124, China. Email: lzl@bjut.edu.cn

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
and subsequent property losses could be avoided to some degree.

Currently, many studies have reported on driver behavior detection systems that aim to decrease the number of accidents. The methods can be classified into the following two types according to whether the methods are obtrusive or not.

**Obtrusive methods**

In the first category, a driver’s physiological state changes are used to detect drunk driving. These parameters include their breath alcohol concentration, blood alcohol concentration (BAC), brain wave changes, electroencephalogram (EEG) signal changes, and eye movements. At present, the breath alcohol concentration index is still the primary method to detect drunk driving. Sakairi⁵ used water cluster detection sensors and alcohol sensors to observe the electrical signals that result from the alcohol use of the participants. Wu et al.⁶ focused on electrocardiogram (ECG) signals and used a support vector machine (SVM henceforth) to detect the states of the drivers. Khardi and Vallet⁷ used EEG signals to study low vigilance periods, which are related to a driver’s fatigue.

However, two problems exist. First, it is difficult to acquire physiological data, such as EEG signals and heart rate changes, and their acquisition depends on the driving environment. Second, due to the surrounding environment and the collection methods being based on intrusive equipment, drivers may be annoyed.

**Unobtrusive methods**

**Image-based features to detect drunk driving.** Carswell and Chandran⁸ used an image processing method to construct a model for the extracted vehicle trajectory and used an artificial neural network to predict the states of drivers who were in different states of intoxication. Chen and Chia-Tseng⁹ developed a method to monitor a driver’s face using an image capturing unit, an image processing unit, and a warning unit. With the development of deep learning, progress has been made in image recognition and target detection. Although the performances of the models have been greatly improved, there are tens of thousands of parameters that need to be learned during the training process, which imposes some requirements on the computing devices. In addition, the methods that were given above require much environmental information, which can be significantly affected by light.

**Infrared sensor-based method to detect drunk driving.** Ljungblad et al.¹⁰ used sensors based on infrared spectroscopy and a video camera to record alcohol, CO₂, and the image of drivers’ body in a vehicle. The feasibility of passive driver breath alcohol detection was verified. It should be noted that the sensor acquisition results depend on the surrounding environment. Further verification of the accuracy needs to be made.

**Vehicle-based features to detect drunk driving.** These methods use the transportation hardware systems that are controlled by drivers to monitor driver behavior. The frequently used features are changes to the steering wheel, accelerator pedal depth, brake pedal depth, speed, and so on. By utilizing these significant driving behavior features, the states of the drivers can be determined by manipulating the vehicles. Some vehicle-based behaviors such as turning the steering wheel, moving the vehicle, and changing the speed can reflect the state of the driver. These indicators are totally based on the transportation hardware systems, which may not influence the driving process. As data-driven machine learning continues to develop, it will play an increasingly important role in data analysis and prediction. Therefore, if we can apply an appropriate model to detect drunk driving and timely predict abnormal driving behaviors early enough, the damages that are related to casualties and property losses may be avoided.

Due to the excessive number of driving behavior features, it is well worth studying methods that can select the important features that can improve the robustness of the classifier and accurately detect the states of drivers. With respect to feature selection, Ding and Peng¹¹ proposed a minimum redundancy–maximum relevance (MRMR) method to select the key features and significantly improved class predictions using extensive experiments. Malhi and Gao¹² introduced the principal component analysis (PCA)-based method to select the features and used both supervised and unsupervised approaches to verify the effectiveness of the feature selection scheme. PCA is a popular method for dimensionality reduction. However, due to the projection transformation, the reduced features that are produced by PCA lack interpretability. The random forest (RF henceforth), which is an ensemble learning algorithm based on decision trees, has been widely used in different fields and provides an ideal prediction ability. In addition, the model is even more robust than other popular models. Feature Selection based on the Random Forest (FSRF henceforth) can evaluate the importance of the features and select a subset of the most significant features with good interpretability. Pan and Shen¹³ extracted the global, local, and evolutionary features from protein data using the FSRF to select the key features that served the two-stage inputs to construct the support vector regression model, which improved the robustness of predicting the B-factor. Lefkovits et al.¹⁴
used the RF and feature importance to process a large data set in order to eliminate the irrelevant features and improve the effects of brain tumor segmentation. For the drunk driving behavior model and feature selection, El Masri et al.\textsuperscript{15} used the logistic regression model based on the time window and PCA to select the features and model drunk driving behaviors. Chen and Chen\textsuperscript{16} conducted feature selection using PCA with the feature weight values, and the SVM was selected for modeling. Li et al.\textsuperscript{17} selected the steering angle and the vehicle’s lateral position as the features and constructed the model of the driving states under different road geometries using the SVM and K-nearest neighbors based on multiple time series. Zhang et al.\textsuperscript{18} analyzed the characteristics of significant driving changes under different conditions and selected the Gauss hybrid hidden Markov model to represent the driving behaviors of drivers under different road conditions. Different classifiers were created for different road conditions in existing studies which led to the lack of certain integrity due to there being too many sub-models. In the field of machine learning, there are some different variable types. For example, there are numerical variables and categorical variables. In actual processing, with respect to categorical variables, the common coding method is dummy variable encoding. Alkharusi\textsuperscript{19} detailed the method and effect of categorical variable coding and found that the coding category variables would not affect the overall result. Zeng\textsuperscript{20} introduced a dummy variable to encode a tree’s origin, which is natural or planted, and used a nonlinear mixed model to develop individual trees above and below the ground’s biomass, which resulted in a low mean prediction error. The paper used dummy variables to encode road conditions (geometric characteristics) so that only one classifier was needed for the different road conditions.

Furthermore, it is worthwhile to verify the performance of the RF and the performance of the FSRF with respect to the final model identification. Hence, we used the driving simulator that was located in the Traffic Research Center of Beijing University of Technology, as shown in Figure 1. The simulator car was remolded from a Toyota car, and it included six computers, various detectors, and hardware interfaces that can detect the driver's behaviors in a timely manner. The sampling frequency of the equipment was 30 Hz. The front of the driving simulator was equipped with screens that present the three-dimensional virtual driving scenario with a 130° field of vision. The simulator also had two-side mirrors beside the car and one rearview mirror. Another equipment was a breath alcohol detector used in this experiment.

Nagoshi et al.\textsuperscript{21} stated that male drivers are more impetuous and sensitive than female drivers under the same BAC level. According to Mayhew et al.,\textsuperscript{22} traffic accidents happen more frequently among younger drivers than among older drivers. Therefore, in this article, we only focus on young male drivers. To collect the data of different driving states, the experiment recruited 25 drivers who have 3–4 years of driving experience and relatively regular sleeping habits. A Chinese Liquor (Erguotou, 46°) and drinking water were selected in this experiment. To make sure that the experimental results are only affected by the inebriation level, the BAC was introduced. Referring to the Chinese traffic regulations, BAC values of 0.09% and 0.00% were selected as the experimental group (drunk group) and the control group (normal group), respectively, to observe the differences in the behaviors between drunk and normal drivers.

According to Zhao et al.,\textsuperscript{23} both fatigued driving and drunk driving can weaken the driving behaviors of a driver. To distinguish the two different states of fatigued driving and drunk driving, a questionnaire that consists of seven different levels was used to determine the fatigue of the participants (fatigue increases as the level rises). At the beginning and end of the experiment, the researchers inquired about the states of the drivers.
The data about the fatigue levels were collected using the questionnaires. In addition, each participant was asked to remain healthy without any substance abuse and have at least 1 h of rest before the experiment.

To create a real driving environment, each driver was asked to practice driving for a while on a given route. Approximately 15 min after drinking, the participants' inebriation levels were measured using a breath detector every 5 min. It was not until the value reached to the target that the experiment was conducted. Each recruited driver drove for approximately 35 min. Three driving routes were designed, as shown in Figure 2. Each route consists of a straight-line segment and a curved road segment with three left and right turns. The specific curves were the following: 200L (a left turn with a radius of 200 m), 200R (a right turn with a radius of 200 m), 500L (a left turn with a radius of 500 m), 500R (a right turn with a radius of 500 m), 800L (a left turn with a radius of 800 m), and 800R (a right turn with a radius of 800 m). More than 50 features can be obtained through the experiment including Speed, Vehicle Acceleration, Engine Speed, Accelerator Pedal Depth, Clutch, Brake Pedal Depth, Steering, Switches, Gear, Wheel Slip, and Distance to the Center of the Lane.

Methods

**Dummy variable**

In machine learning, categorical variables (such as geometric characteristics) can only range over a series of fixed values. Generally, a feature with \( k \) possible values needs to be encoded as a set of \( k \) derived dummy variables so that all the categories within the feature can be represented. In this article, we encode the radius and direction of the curve, which results in six possible values (200L, 200R, 500L, 500R, 800L, and 800R). For each sample, there is only one setting for each of the derived dummy variables that has a value of 1. For the remaining parts of the derived dummy variables, they are equal to 0.

**Feature selection based on the RF**

**Bootstrap sampling.** Bootstrap sampling is a resampling method that was proposed by Efron and Gong.\(^{24}\) The training set contained \( M \) samples. Using the put-back sampling method, \( M \) samples are randomly extracted. After repeating the process \( K \) times, \( K \) bootstrap sampling sets that have the same size as the original training set can be obtained.

Using the bootstrap sampling method, we can implement parallel computing to decrease the training time of the RF. Because of the random sampling, each sample data set is not quite the same as the original data set. It can be known that approximately 36.8% of the samples in the original training set will not appear in the sampling set. To some extent, bootstrap sampling helps to improve the generalizability of the model. The bootstrap sampling method is more suitable for a bagging algorithm with smaller data sets and it can meet the construction needs of different training sets.

**Feature selection.** Using the Bootstrap sampling method \( K \) times, we can obtain \( K \) classification and regression trees (CART henceforth), and the importance of features based on permutation can be defined as the variation in accuracy of classification between the features after permuting and the features before permuting. The specific pipeline can be seen in the work by Yao et al.\(^ {25} \)

The feature selection method can not only save computation costs but also perform at an acceptable level. Different decision trees are constructed by selecting several subsequent feature sets. According to Kalousis et al.,\(^ {26} \) the feature selection result is proportional to the generated decision tree. To ensure the reliability and stability of feature selection, this article conducted the experiment using the RF. The original data set that we obtained came from the driving simulation. After data preprocessing, the experiment was implemented on a Windows 10 machine with a 1.80 GHz CPU and 16.0 GB of RAM. The number of training trees (\( n \)) was set as 1000, and the importance of each feature was calculated accordingly.
Drunk detection based on the RF

CART. The CART is used as the single decision tree in the RF. It is a binary tree that was proposed by Breiman\textsuperscript{27} and can complete the tasks such as classification and regression. In the training process of the RF, the CART completely grows without any pruning. In this article, we just refer to the development of the CART. The pruning process can be found in the work by Esposito et al.\textsuperscript{28} In the process of CART development, the minimum value of the Gini index is recursively used from the root node to select and divide the features. The definition of training set $D$ can be divided into two parts, $D_1$ and $D_2$, according to feature $\alpha$ and its division value $\delta$. To provide a more explicit description, in this article, we just refer to the two-class problem

$$D_1 = \{(x, y) \in D | \alpha(X) = 1\}$$
$$D_2 = D - D_1$$

Definition 1
The Gini index\textsuperscript{27} of the variable $\alpha$ can be defined as follows

$$\text{Imp}(\alpha) = 1 - p_1^2 - p_0^2$$

where $p_k = (n_k/n)$, $k = \{0, 1\}$, $n_k$ is the number of outputs at the node $\tau$; $n$ is the total number of samples at the node $\tau$.

Definition 2
The Gini index of training set $D$ is assigned with the characteristic $\alpha$

$$\text{Imp}(D, \alpha) = \frac{|D_1|}{|D|} \text{Imp}(D_1) + \frac{|D_2|}{|D|} \text{Imp}(D_2)$$

where $|D_k|$, $k = \{0, 1\}$, $|D_k|$ is the number of the samples in $|D_k|$, $|D|$ is the number of the samples in the training set.

The generation process of the CART will stop when the number of samples in the node is less than a preset threshold or when the Gini index is less than a predetermined threshold. Similar to entropy, the Gini index also indicates the uncertainty of the feature. As the Gini index increases, the uncertainty increases accordingly.

Recognition method. Based on bagging, the concept of the RF is introduced. In addition to collecting $K$ sample sets using the bootstrap sampling method, which had the same size as the original training set, $N$ subsets of feature sequences for training each decision tree were randomly selected. The final output was determined by voting or calculating the mean value. Using bootstrap sampling and randomly selecting the feature sequence, we can not only improve the generalizability but also improve the prediction performance.

To evaluate the effectiveness of the FSRF algorithm for drunk driving detection, this article selected three other commonly used classification methods, and we evaluated and compared their effects. Because of its simple operation and no need to tune the hyperparameters, LDA\textsuperscript{29} is one of the most popular methods in actual projects. By optimizing the largest interclass variance and the smallest intraclass variance, LDA enables the model to select the direction with the best classification performance, and the calculation is relatively simple. In this article, we selected the Fisher criterion to obtain the weights. The SVM,\textsuperscript{30} which was proposed by Cortes and Vapnik,\textsuperscript{31} is a statistical learning method that is usually adopted in both linear and nonlinear fields. Based on the principle of minimizing the structural risk function, the model finds the weights with the largest interval between the training samples and the separated hyperplane from the feature space under the premise that the training samples were linearly inseparable. The SVM uses only a small number of support vectors, which makes it able to remove a large number of redundant samples while capturing key samples and obtain low algorithmic complexity and good robustness. In this article, we selected the Gaussian kernel function, which is a widely used function that has wide convergence, regardless of whether the dimension in the feature space is low or high. AdaBoost, which is a boosting algorithm in ensemble learning, is also a statistical learning algorithm.\textsuperscript{32} Through iterative learning and changing the weights of training samples, the samples that are misclassified by the base learner in the previous iteration will have greater weights and draw more attention in the process of the next iteration. After a series of iterations, the learned weak classifiers will be linearly combined to form a strong classifier. Finally, these weak classifiers form the final decision function.

To verify the feature selection performance and obtain a proper model for identifying the states of the drivers, 25 drivers were recruited to collect their driving behaviors under different conditions (drunk and normal). The total number of valid samples in the experiment was 265, which included both driving states. The fivefold cross-validation method was used to assess the performance of the classifiers. The training set (three-fifths of the data set) was used to train the model. The validation set (one-fifth of the data set) was used to select the appropriate hyperparameter, such as the $C$ and $\sigma$ in SVM and the number of decision trees in the RF. The test set (one-fifth of the data set) was used to evaluate the final performance of the classifier.
Results

Considering the different road conditions, we introduced dummy variables in this article for drivers to encode the radius and direction of the curve. The final performance was measured in terms of the accuracy rate, F1, receiver operating characteristic curve (ROC curve), and area under the curve value (AUC).

The results of feature selection

In this article, the FSRF was selected to extract the driving behavior features based on the change of the Gini index. Accelerator Depth, Speed, Distance to the Center of the Lane, Vehicle Acceleration, Engine Revolutions, Brake Depth, and Steering Angle were extracted from the feature set. The other features were filtered out since they had little contribution in terms of the FSRF. The feature selection results are shown in Figure 3.

Classification results with different numbers of features

Considering that the importance levels of the first seven features (Accelerator Depth, Speed, Distance to the Center of the Lane, Vehicle Acceleration, Engine Revolutions, Brake Depth, and Steering) are obviously higher than those of the others, we selected these features and combined them with different numbers of features using the four classifiers that were mentioned above. The best hyperparameter was selected through the validation set. To analyze the performance of the different classifiers and the effectiveness of the FSRF in detail, the number of selected features was reduced gradually from the first seven features to the first two features. In addition, the last six features (the first seven features without Accelerator Depth) were selected as another feature combination set to explore the role of Accelerator Depth. Accordingly, there were seven feature combinations, as shown in Table 1. A checkmark means the combination contains the feature. The classification results are shown in Table 2 and Figure 4.

Classification results without dummy variables

To evaluate the effect of dummy variables, we trained a general classifier without dummy variables. In other words, the general classifiers with the same features

---

**Figure 3.** The importance of features.

**Table 1.** Feature combination.

| Feature                   | Accelerator Depth | Speed | Distance to the Center of the Lane | Vehicle Acceleration | Engine Revolutions | Brake Depth | Steering | Gear | Wheel Slip | Switches |
|---------------------------|-------------------|-------|-----------------------------------|----------------------|-------------------|-------------|----------|-----|------------|----------|
| Seven features            | ✓                 | ✓     | ✓                                 | ✓                    | ✓                 | ✓           | ✓        | ✓   | ✓          | ✓        |
| Six features              | ✓                 | ✓     | ✓                                 | ✓                    | ✓                 | ✓           | ✓        | ✓   | ✓          | ✓        |
| Last six features         | ✓                 | ✓     | ✓                                 | ✓                    |                   |             |          |     |            |          |
| Five features             | ✓                 |       |                                    |                      |                   |             |          |     |            |          |
| Four features             | ✓                 |       |                                    |                      |                   |             |          |     |            |          |
| Three features            | ✓                 |       |                                    |                      |                   |             |          |     |            |          |
| Two features              | ✓                 |       |                                    |                      |                   |             |          |     |            |          |

---

International Journal of Distributed Sensor Networks
were compared between cases with versus without taking the geometric characteristics into account. The classification results of the classifier without dummy variables are shown in Table 3.

Classification results for particular types of roads

To compare the results without the dummy variables with those using the dummy variables, we further subdivided the road geometry and trained the classifiers for different geometric features using the first seven features. In other words, six LDA classifiers, six SVM classifiers, six AdaBoost classifiers, and six RF classifiers were created for six road conditions (200L, 200R, 500L, 500R, 800L, and 800R). The accuracy, AUC, and F1 were selected to measure the results, which can be seen in Table 4. The average accuracy of six LDA classifiers for 200L, 200R, 500L, 500R, 800L, and 800R was calculated and compared with the accuracy of the LDA classifier with the dummy variables. The same calculation applied to six SVM classifiers and six AdaBoost classifiers. The results can be seen in Table 4 and Figure 5.

Significance test

To further verify the role of Accelerator Depth from the other perspective, the significance test was selected to verify the impact of Accelerator Depth on drunk driving identification. The Wilcoxon signed ranks test was used to determine whether there is a difference in Accelerator Depth between normal drivers and drunk drivers. The Accelerator Depth distribution of the two groups is not consistent, as shown in Figure 6. The Accelerator Depths for drunk drivers and normal drivers are statistically significantly different since the mean of the negative ranks is 75.16 and the mean of the positive ranks is 36.76 ($Z = -7.613, p = 0$, two-tailed).

Discussion

The performance of feature selected

From Figure 3, the sum of the feature importance including Accelerator Depth, Speed, Distance to the Center of the Lane, Vehicle Acceleration, Engine Revolution, Brake Depth, and Steering Angle is 0.876, which accounts for 92.81% of the sum of the first 10 features. Furthermore, there is quite a gap between the feature importance of Steering Angle and Gear. Therefore, the first seven features are extracted to identify the states of drunk drivers. When using these seven features to train the model, the accuracies of the RF and AdaBoost are greater than 80%; meanwhile, the accuracies of LDA and SVM are 75.93% and 74.07%, respectively. The classifiers have high recognition abilities. As the number of features decreases, the classifiers can still maintain a certain degree of accuracy. From Table 2, it can be seen that the classifiers are sensitive to the selected features. The results that are reported here confirm the effectiveness of the FSRF. In the light

### Table 2. The performance of the classifiers with different features.

| Features    | Metrics | Classifier |
|-------------|---------|------------|
|             |         | LDA        | SVM        | AdaBoost   | Random forest |
|             | Accuracy| 75.93%     | 74.07%     | 81.48%     | 81.48%       |
|             | AUC     | 0.78       | 0.73       | 0.81       | 0.77         |
|             | F1      | 0.73       | 0.77       | 0.83       | 0.71         |
| Six features| Accuracy| 79.63%     | 62.96%     | 77.78%     | 75.93%       |
|             | AUC     | 0.78       | 0.63       | 0.78       | 0.76         |
|             | F1      | 0.74       | 0.6        | 0.76       | 0.75         |
| Last six features| Accuracy| 63.58%     | 62.96%     | 69.75%     | 66.05%       |
|             | AUC     | 0.64       | 0.64       | 0.7        | 0.66         |
|             | F1      | 0.6        | 0.63       | 0.7        | 0.64         |
| Five features| Accuracy| 77.78%     | 61.11%     | 74.07%     | 79.63%       |
|             | AUC     | 0.76       | 0.6        | 0.73       | 0.8          |
|             | F1      | 0.71       | 0.53       | 0.7        | 0.79         |
| Four features| Accuracy| 72.22%     | 66.67%     | 73.15%     | 74.07%       |
|             | AUC     | 0.73       | 0.67       | 0.76       | 0.75         |
|             | F1      | 0.7        | 0.6        | 0.73       | 0.73         |
| Three features| Accuracy| 71.30%     | 69.44%     | 70.37%     | 74.07%       |
|             | AUC     | 0.72       | 0.71       | 0.71       | 0.75         |
|             | F1      | 0.69       | 0.67       | 0.68       | 0.74         |
| Two features| Accuracy| 68.52%     | 67.59%     | 73.15%     | 71.30%       |
|             | AUC     | 0.65       | 0.64       | 0.72       | 0.72         |
|             | F1      | 0.69       | 0.67       | 0.73       | 0.71         |

LDA: linear discriminant analysis; SVM: support vector machine; AUC: area under the curve.
of the relevant literature, it can be found that most authors focus on the features of Speed, Distance to the Center of the Lane, and Steering Angle. These features are useful for driving pattern identification. However, using the FSRF, the feature importance of Accelerator Depth is 0.087 higher than that of Speed and 0.133 higher than that of Steering Angle when the number of CARTs is 1000. The classifier including Accelerator Depth has better performance in each metric. Accelerator Depth, as one of the significant features, does have an important influence on drunk driving detection. The results of the significance test also verify this conclusion. Igarashi et al. proved that the degree of accelerator pressure by drivers significantly

Figure 4. ROC curves of the four classifiers with different features.
varies, and thus, it can be used as the key feature to assess the performance of drivers. Wahab et al.\textsuperscript{34} used the accelerator and brake pedal as the original features and conducted feature extraction based on the Gaussian mixture model to form a fuzzy neural network to observe the driving patterns of drivers. Hence, Accelerator Depth is a significant feature for drunk driving detection.

### Table 3. The performance of the classifiers without dummy variables.

| Features      | Metrics | LDA          | SVM          | AdaBoost     | Random forest |
|---------------|---------|--------------|--------------|--------------|---------------|
| Seven features| Accuracy| 75.93%       | 64.81%       | 74.07%       | 77.78%        |
|               | AUC     | 0.76         | 0.65         | 0.74         | 0.78          |
|               | F1      | 0.72         | 0.7          | 0.72         | 0.76          |
| Six features  | Accuracy| 77.78%       | 59.26%       | 72.22%       | 75.93%        |
|               | AUC     | 0.78         | 0.59         | 0.72         | 0.76          |
|               | F1      | 0.74         | 0.61         | 0.69         | 0.73          |
| Last six features | Accuracy | 66.30% | 63.70% | 70.00% | 66.30% |
|               | AUC     | 0.67         | 0.65         | 0.7          | 0.67          |
|               | F1      | 0.65         | 0.62         | 0.65         | 0.65          |
| Five features | Accuracy| 76.92%       | 61.54%       | 71.15%       | 73.08%        |
|               | AUC     | 0.77         | 0.62         | 0.71         | 0.73          |
|               | F1      | 0.75         | 0.66         | 0.72         | 0.74          |
| Four features | Accuracy| 70.37%       | 62.96%       | 75.93%       | 72.22%        |
|               | AUC     | 0.7          | 0.63         | 0.76         | 0.72          |
|               | F1      | 0.65         | 0.58         | 0.76         | 0.74          |
| Three features| Accuracy| 69.81%       | 64.15%       | 67.92%       | 75.47%        |
|               | AUC     | 0.7          | 0.64         | 0.68         | 0.76          |
|               | F1      | 0.64         | 0.56         | 0.65         | 0.76          |
| Two features  | Accuracy| 69.23%       | 53.85%       | 73.08%       | 67.31%        |
|               | AUC     | 0.69         | 0.54         | 0.73         | 0.67          |
|               | F1      | 0.64         | 0.57         | 0.68         | 0.67          |

LDA: linear discriminant analysis; SVM: support vector machine; AUC: area under the curve.

### Table 4. The performance of the classifiers with different geometric features.

| Features | Metrics | LDA          | SVM          | AdaBoost     | Random forest |
|----------|---------|--------------|--------------|--------------|---------------|
| 200L     | Accuracy| 88.89%       | 88.89%       | 88.89%       | 77.78%        |
|          | AUC     | 0.9          | 0.88         | 0.88         | 0.78          |
|          | F1      | 0.89         | 0.91         | 0.91         | 0.8           |
| 200R     | Accuracy| 77.78%       | 77.78%       | 88.89%       | 88.89%        |
|          | AUC     | 0.75         | 0.65         | 0.78         | 0.88          |
|          | F1      | 0.8          | 0.8          | 0.89         | 0.86          |
| 500L     | Accuracy| 72.73%       | 81.82%       | 72.73%       | 81.82%        |
|          | AUC     | 0.73         | 0.83         | 0.73         | 0.83          |
|          | F1      | 0.73         | 0.8          | 0.73         | 0.8           |
| 500R     | Accuracy| 75%          | 87.5%        | 87.5%        | 87.5%         |
|          | AUC     | 0.83         | 0.75         | 0.92         | 0.92          |
|          | F1      | 0.67         | 0.67         | 0.8          | 0.8           |
| 800L     | Accuracy| 75%          | 75%          | 87.5%        | 87.5%         |
|          | AUC     | 0.67         | 0.67         | 0.8          | 0.8           |
|          | F1      | 0.73         | 0.73         | 0.83         | 0.83          |
| 800R     | Accuracy| 81.82%       | 72.73%       | 81.82%       | 81.82%        |
|          | AUC     | 0.82         | 0.73         | 0.82         | 0.8           |
|          | F1      | 0.8          | 0.73         | 0.8          | 0.75          |
| Average  | Accuracy| 78.54%       | 80.62%       | 84.55%       | 84.22%        |
|          | AUC     | 0.78         | 0.75         | 0.82         | 0.83          |
|          | F1      | 0.77         | 0.77         | 0.83         | 0.81          |

LDA: linear discriminant analysis; SVM: support vector machine; AUC: area under the curve.
The classifiers without dummy variables were used to compare with our approach and evaluate the value of taking the road geometry into account. From Tables 2 and 3, it can be seen that the performances of classifiers without dummy variables are generally lower than that with dummy variables. From the results, we can find that the road geometry is beneficial to identify drunk driving. From Tables 2 and 4, and Figure 5, we can find that the results when using dummy variables are a little worse than the results without using them for particular types of roads. It is obvious that building the models separately can allow the model to more easily learn the data under specific cases. However, the complexity of the classifier with dummy variables is reduced because only one classifier is used for the different road conditions. Therefore, from a comprehensive perspective, the classifiers with the dummy variables and different classifiers for different road conditions seem to have their own advantages and disadvantages. In short, there is a trade-off between the complexity of the model and the accuracy. The classifier with dummy variables significantly reduces the computational complexity with minor loss in the accuracy. Encoding road conditions using dummy variables is a feasible method to detect drunk driving on different road conditions.

Limitations

The method for drunk driving detection was studied using the driving simulator. As we all know, driving simulators are able to provide a safe environment without harming drivers. Meanwhile, events can be identically repeated for participants which provide adequate guarantees for examining drunk driving behaviors. However, driving simulators do not always provide an accurate description of on-road driving behavior. Limited physical and perceptual fidelity may produce invalid results. In addition, simulator sickness may affect research outcomes. It is noted that speed perception and lateral control are worse in simulators. From the viewpoint of kinematics, we can find that only accelerations can generate powers which drivers will directly experience. Drivers in on-road vehicles can
easily feel the forces. Accordingly, the driver can adjust the current driving operation through the brake pedal and the accelerator pedal. However, drivers in simulators are unlikely to feel the forces. Pedal operation characteristics might be different from that of the on-road vehicles. Therefore, whether the performance of the accelerator pedal shown in the simulator is consistent with the real-world situation needs further study. In addition, the existing study reported that simulators can indicate relative validity for lateral position, but they cannot show absolute validity. Drivers probably drive further from the centerline in the simulator than in the real road. Therefore, lateral position and steering wheel angle are likely to be greater in simulators than real-world environments.

The features selected by FSRF, Accelerator Depth, Speed, Distance to the Center of the Lane were also used for other driving impairments detection such as distraction and fatigue. These features not only can help to identify the drunk driving behavior, but also can be used to detect other driving impairments. According to the literature, drivers generally tend to drive slowly when they were talking on hand-held or hands-free phones (distraction driving), while drunk drivers tend to drive faster under the influence of alcohol than the normal. In addition, driving smoothly (standard deviation of lane position) when drunk is worse than the distraction and normal. Accordingly, the different impairments might be detected in terms of the driving characteristics. Of course, the detection accuracy needs further study. Some equipment such as cameras can be used to capture facial and eye information to improve the accuracy.

Conclusion

This article collects the features under drunk and normal conditions using a driving simulator and it proposes a feature selection method based on the RF. By applying the dummy variables, different radii and directions are encoded. To evaluate the performance of the FSRF, the main classification models with different feature combinations are analyzed using metrics including the accuracy, AUC value, F1 score, and ROC curve. The conclusions are as follows:

1. To detect the drunk driving behaviors, a feature selection method based on the RF is proposed and its effectiveness is tested by applying the selected features into different models. Accelerator Depth, Speed, Distance to the Center of the Lane, Vehicle Acceleration, Engine Revolution, Brake Depth, and Steering Angle all have significant impacts on drunk driving detection.
2. Comparing the classifier using Accelerator Depth with the classifier without Accelerator Depth shows that Accelerator Depth is a crucial feature that is effective for identifying the states of the participants.
3. RF and AdaBoost achieve the highest accuracy when using seven features.
4. Encoding road conditions (geometric characteristics) using dummy variables is a feasible method to detect drunk driving under different road conditions.

In the future, the convenience of the above feature collection method will be considered and tested. We will try to extend the present studies in detail to evaluate the influences of the selected features, especially Accelerator Depth, and expand the adopted input features to obtain improved performance. Our experiment was implemented based on young male drivers. Therefore, whether the drunk driving behaviors will be affected by age and gender remains to be investigated. In addition, whether the selected features are suitable for real-road driving environments still needs further study.

Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Traffic Research Center, Beijing University of Technology, China.

ORCID iD
HaoXin Wang https://orcid.org/0000-0002-5438-2847

References
1. Organization WH. The top ten causes of death, 2018, https://www.who.int
2. Griswold MG, Fullman N, Hawley C, et al. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018; 392: 1015–1035.
3. Administration NHTS. Fatal crashes and percent alcohol-impaired driving, 2016, https://crashstats.nhtsa.dot.gov
4. Li Y, Xie D, Nie G, et al. The drink driving situation in China. Traffic Inj Prev 2012; 13(2): 101–108.
5. Sakairi M. Water-cluster-detecting breath sensor and applications in cars for detecting drunk or drowsy driving. IEEE Sens J 2012; 12(5): 1078–1083.
6. Wu C, Tsang K, Chi H, et al. A precise drunk driving detection using weighted kernel based on electrocardiogram. Sensors 2016; 16(5): 659.
7. Khardi S and Vallet M. Drowsiness of the driver: EEG (electroencephalogram) and vehicle parameters interaction. In: Proceedings: International technical conference on the enhanced safety of vehicles, Munich, Germany, 23–26 May 1995, pp.443–461. NHTSA.
8. Carswell B and Chandran V. Automated recognition of drunk driving on highways from video sequences. In: Proceedings of the 1st international conference on image processing, Austin, TX, 13–16 November 1994. New York: IEEE.
9. Chen Y-S and Chia-Tseng C. Facial image recognition system for a driver of a vehicle. Google patents 8300891, USA, 2012.
10. Ljungblad J, Hök B, Allalou A, et al. Passive in-vehicle driver breath alcohol detection using advanced sensor signal acquisition and fusion. Traffic Inj Prev 2017; 18: S31–S36.
11. Ding C and Peng H. Minimum redundancy feature selection from microarray gene expression data. J Bioinf Comput Biol 2008; 3(2): 185–205.
12. Malhi A and Gao RX. PCA-based feature selection scheme for machine defect classification. IEEE T Instrum Meas 2004; 53(6): 1517–1525.
13. Pan XY and Shen HB. Prediction of B-factor profile from sequence using two-stage SVR based on random forest feature selection. Protein Pept Lett 2009; 16(12): 1447–1454.
14. Lefkovits L, Lefkovits S, Emerich S, et al. Random forest feature selection approach for image segmentation. In: 9th international conference on machine vision (ICMV 2016), Bellingham, 6 August 2017, paper no. 1034117. SPIE.
15. El Masri AEB, Artail H and Akkary H. Toward self-policing: detecting drunk driving behaviors through sampling CAN bus data. In: 2017 international conference on electrical and computing technologies and applications (ICECTA), Ras Al Khaimah, United Arab Emirates, 21–23 November 2017, pp.1–5. New York: IEEE.
16. Chen H and Chen L. Support vector machine classification of drunk driving behavior. Int J Environ Res Public Health 2017; 14(1): 108.
17. Li Z, Jin X and Zhao X. Drunk driving detection based on classification of multivariate time series. J Safety Res 2015; 54: 61–64.
18. Zhang X, Zhao X and Rong J. A study of individual characteristics of driving behavior based on hidden Markov model. Sens Transducers 2012; 167: 194–202.
19. Alkharusi H. Categorical variables in regression analysis: a comparison of dummy and effect coding, 2012, https://www.researchgate.net/
20. Zeng W-S. Using nonlinear mixed model and dummy variable model approaches to develop origin-based individual tree biomass equations. Trees 2015; 29(1): 275–283.
21. Nagoshi CT, Wilson JR and Rodriguez LA. Impulsivity, sensation seeking, and behavioral and emotional responses to alcohol. Alcohol Clin Exp Res 1991; 15(4): 661–667.
22. Mayhew DR, Donelson AC, Beirness DJ, et al. Youth, alcohol and relative risk of crash involvement. Accident Anal Prev 1986; 18(4): 273–287.
23. Zhao X, Zhang X and Rong J. Study of the effects of alcohol on drivers and driving performance on straight road. Math Probl Eng 2014; 2014: 1–9.
24. Efren B and Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am Stat 1983; 37(1): 36–48.
25. Yao D, Yang J and Zhan X. An improved random forest scheme for machine defect classification. In: Proceedings of the 2005 IEEE international conference on robotics and automation, Barcelona, Spain, 18–22 April 2005, pp.1730–1735. IEEE.
33. Igarashi K, Miyajima C, Itou K, et al. Biometric identification using driving behavioral signals. In: 2004 IEEE international conference on multimedia and expo (ICME) (IEEE Cat. No.04TH8763), Taipei, Taiwan, 27–30 June 2004, pp.65–68. New York: IEEE.
34. Wahab A, Quek C, Tan CK, et al. Driving profile modeling and recognition based on soft computing approach. IEEE Trans Neural Netw 2009; 20(4): 563–582.
35. Greenberg JA and Blommer M. Physical fidelity of driving simulators, 2011, https://www.researchgate.net/publication/330310893_Physical_fidelity_of_driving_simulators
36. Kaptein NA, Theeuwes J and Van Der Horst R. Driving simulator validity: some considerations. Transp Res Record 1996; 1550(1): 30–36.
37. Blaauw GJ. Driving experience and task demands in simulator and instrumented car: a validation study. Hum Factors 1982; 24(4): 473–486.
38. Wade MG and Hammond C. Simulation validation: evaluating driver performance in simulation and the real world. St. Paul, MN: Minnesota Department of Transportation, 1998.
39. Fisher DL, Rizzo M, Caird J, et al. Handbook of driving simulation for engineering, medicine, and psychology. Boca Raton, FL: CRC Press, 2011.
40. Burns PC, Parkes A, Burton S, et al. How dangerous is driving with a mobile phone? Benchmarking the impairment to alcohol, vol. 547. Crowthorne: TRL, 2002.