

Research Article

Rapid Driving Style Recognition in Car-Following Using Machine Learning and Vehicle Trajectory Data

Qingwen Xue , **Ke Wang** , **Jian John Lu** , and **Yujie Liu**

College of Transportation Engineering, Tongji University, 4800 Cao'an Road, Shanghai 201804, China

Correspondence should be addressed to Ke Wang; kew@tongji.edu.cn

Received 4 October 2018; Revised 27 December 2018; Accepted 6 January 2019; Published 23 January 2019

Guest Editor: Mohammad H. Y. Moghaddam

Copyright © 2019 Qingwen Xue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rear-end collision crash is one of the most common accidents on the road. Accurate driving style recognition considering rear-end collision risk is crucial to design useful driver assistance systems and vehicle control systems. The purpose of this study is to develop a driving style recognition method based on vehicle trajectory data extracted from the surveillance video. First, three rear-end collision surrogates, Inversed Time to Collision (ITTC), Time-Headway (THW), and Modified Margin to Collision (MMTC), are selected to evaluate the collision risk level of vehicle trajectory for each driver. The driving style of each driver in training data is labelled based on their collision risk level using K-mean algorithm. Then, the driving style recognition model's inputs are extracted from vehicle trajectory features, including acceleration, relative speed, and relative distance, using Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), and statistical method to facilitate the driving style recognition. Finally, Supporting Vector Machine (SVM) is applied to recognize driving style based on the labelled data. The performance of Random Forest (RF), K-Nearest Neighbor (KNN), and Multi-Layer Perceptron (MLP) is also compared with SVM. The results show that SVM overperforms others with 91.7% accuracy with DWT feature extraction method.

1. Introduction

Driving style refers to the ways that drivers choose to habitually drive and the driver states that represent the common parts of varied driving behavior [1]. Driving style of drivers plays an important role in driving safety as well as vehicle energy consumption. Different driving styles may lead to different possibilities for traffic incidents. Recognition of a driver's driving style based on rear-end collision risk is of great significance to improve the safety of driving. With the development of connected autonomous vehicles and Advanced Driver Assistance System (ADAS), there is an urgent demand for enhancing recognition of driving style. It is not only important to guarantee the safety and adequate performance of drivers, but also essential to meet drivers' need, adjust to the drivers' preference, and ultimately improve the safety of the driving environment. Driving style recognition also has potential value to help traffic agencies design control strategies effectively [2, 3].

The availability of high-definition surveillance camera makes it possible to collect numerous vehicle motions from

real world traffic flow. The advanced video extraction software can extract vehicle trajectory data accurately and efficiently from the surveillance video. The technologies provide a good opportunity to recognize driving style using the video-extracted vehicle trajectory data. Moreover, the machine learning technique is playing a crucial role in driving behavior recognition. A growing amount of studies on machine learning algorithms have been conducted in recent years [4–7]. This paper builds a driving style recognition model based on vehicle trajectory data. Four supervised machine learning algorithms, including Supporting Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), and Multi-Layer Perceptron (MLP), are used in model training. A new method based on rear-end collision risk is proposed to label the driving style of each driver in the sample data. Three feature extraction methods, including Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), and statistical method, are also adopted to extract the most effective features of driving style recognition.

To the best knowledge of the authors, there are three main contributions in this paper: (1) This paper proposes

TABLE 1: Comparison of different driving behavior data collection approaches.

Collection Approaches	Advantages	Disadvantages
In-vehicle camera [29–33], sensors, hardware [10, 24]	Real-world driving data; high-accuracy data; Access to driver's personal data and vehicle control data	Expensive and time consuming; Lack of data in extreme and dangerous driving condition
Driving simulator [24–26]	Collect drivers' behavior in designed and controlled driving scenarios	Driving behavior observed in the simulator may not always correspond to real-world driving
Traffic video [35–37]	Low expense; Easy to collect enormous vehicle data in a short time; Observe vehicle interaction in real traffic flow	Video extraction is challenging; No access to driver's personal data and vehicle control data
Smartphone-based method [52–55]	Real-world driving data; Low expense in the smartphone	Data accuracy is critical

a new method based on rear-end collision risk to evaluate driving style. The trajectory of each driver is divided into segments with different risk level by the threshold of rear-end collision surrogates. (2) The DFT, DWT, and statistical feature extraction methods are all applied on vehicle trajectory data, and their performance is compared. (3) This paper builds a driving style recognition model based on vehicle trajectory data with 92.7% accuracy rate. The recognition results of SVM and other popular classification algorithms including RF, MLP, and KNN are compared.

This paper is organized as follows. Section 2 presents the related work on driving behavior data analysis and machine learning algorithms. Section 3 introduces the data analyzed in this paper. Section 4 details the driving style recognition method implemented in this paper. Section 5 shows the results and discussion. Section 6 concludes this paper and raises the possible future work.

2. Literature Review

In recent years, the machine learning algorithms applied to the driving behavior recognition have been studied in many previous works. Different types of neural network (NN) algorithms have been used. Molchanov et al. [8] proposed a convolutional deep neural network (CDNN) to recognize the risky driving. Other types such an artificial neural network (ANN) [9] and pulse coupled neural network (PCNN) [10] were adopted to classify the driving behaviors. In the study by Srinivasan [11], the effectiveness of three types of NN methods was compared. The results show that the Multi-Layer Perceptron (MLP) model can achieve excellent classification results. However, the learning rate of NN is difficult to be determined, resulting in higher possibility to be trapped in local minima. A larger size of the network could lead to a long training time [12]. The tree-like structures including decision tree algorithm [13] and Random Forest algorithm [14] are also adopted to detect the driving behaviors according to the extracted features. Some researchers proposed Hidden Markov Model (HMM) to effectively detect dangerous driving behaviors. Berndt et al. [15] established the HMM to identify lane change, steering, and follow-up intention. The recognition accuracy of left-change and right-change is, respectively, 76% and 74%. Meng et al. [16] trained the HMM by collecting driver's operation

data on acceleration pedal, brake pedal, and steering wheel to recognize the driver's profiles online. Some researchers also combined the HMM with dynamic Bayesian networks or ANN to predict the driving behavior by learning the driving data [17, 18]. While HMM requires long training time, especially for a high number of states, the recognition time also increases with the number of states [19]. Therefore, a more suitable and effective method should be found to identify the driving style. SVM has been widely applied to various kinds of pattern recognition problems, including voice identification, text categorization, and face detection [6, 20, 21]. In addition, SVM performs well with a limited number of training samples, and SVM has fewer parameters to be determined [22, 23]. Therefore, many studies employed SVM to build driving style recognition models [24–28].

Along with machine learning algorithms, driving behavior data collection is crucial to the success of driving style recognition. Table 1 summarizes the advantage and the disadvantages of different driving data collection approaches. Researchers used instrumented vehicles to conduct naturalist driving experiments to identify behaviors [29–31]. Some instrumented vehicles were equipped with in-vehicle mounted cameras to capture video images of drivers [32, 33], while others got help from specialized hardware and sensors to acquire throttle opening, pedal brake, wheel steering, vehicle speed, acceleration rate, and yaw rate [10, 24]. Although the driver controlling data and vehicle kinematic data can be collected on the instrumented vehicles, the requirement of expensive devices and sensors is a major obstacle to large scale naturalist driving experiments. In addition, extreme driving conditions, like extreme weather and driving under the influence, could be unobservable in naturalist driving studies. Some research adopted driving simulators to collect driving behavior data [24–26] in the designed and controlled driving environment. However, the results heavily relied on the fidelity and validity of the driving simulator used in research, because the driving behavior observed in the simulator may not always correspond to real-world driving.

Besides Naturalist Driving Studies (NDS) and driving simulator, another important data source is traffic video, because surveillance cameras deployed on the roadside can provide a large amount of traffic environment data and vehicle trajectory data [34]. Traffic video contains all vehicle

trajectory data on the road and can offer a full view of vehicle's interactions with other during car-following and lane-change, etc. However, extracting vehicle trajectory from video could be challenging, which depends on video quality and algorithms used [35–37].

Except for unsupervised machine learning algorithms, for example, clustering, other machine learning algorithms require labelled or partially labelled driving behavior data. In the field of driving style recognition, the method of driving style labelling for each driver in the sample is of great importance to the reliability of the recognition model. There are several methods to label driving style. One is the behavior-based or accident-based method. The driver's driving style depends on risky behavior or accident happened during observation. Chen et al. [20] defined the dangerous driving behaviors according to criteria as frequent lane changes, abrupt double lane change, and illegal lane occupation. The accidents data are also adopted to determine the risk level of driving behavior [38]. However, risky behavior or accident is hardly observable in daily traffic. Therefore, driver self-reported questionnaire [39] and expert scoring [13] are also adopted to evaluate driving style. However, these two methods rely on subjective judgments of drivers or experts and can be very time-consuming when the number of drivers in the sample is hundreds or even thousands. Some research used the facial movement or driving duration to label driver's drowsiness or fatigue driving [9, 10]. The unsupervised clustering methods including the K-means [40] and fuzzy clustering [41] are also used to label drivers in each clustering group.

This paper proposes a new driving data labelling method based on collision surrogates. There are many effective surrogates to evaluate the collision risk [42, 43]. Mahmud et al. [44] compared the advantages and disadvantages between temporal proximity indicators, i.e., Time to Collision (TTC), Time to Accident (TA), Time-Headway (THW), and distance based proximal indicators, i.e., Margin to Collision (MTC), Proportion of Stopping Distance (PSD). Many automobile collision avoidance systems or driver assistance systems used TTC as an important warning criterion for its theoretical and reliable reasons [45–47]. Since TTC can not handle zero relative speed in car-following, the Inversed TTC (ITTC) was adopted to measure the collision risk [41]. THW is another surrogate used to estimate the criticality of a follow-up situation, which is applicable in all traffic environments [44]. MTC provides the possibility of conflict when the preceding and following vehicle at the same time decelerate abruptly [48]. Modified MTC (MMTC) considers the reaction time for drivers when preceding vehicle abruptly decelerates. These three surrogates can be adopted to label the driving style corresponding to different rear-end collision effectively.

In this paper, the vehicle trajectory data extracted from traffic video is analyzed to study the driving style. Three surrogates, i.e., ITTC, THW, and MMTC, are used to effectively measure the rear-end collision risk and label the driving style. This labeling method is more efficient and objective compared with questionnaires [10] and expert scoring [20]. Then the SVM is applied to build a driving style recognition model. The vehicle trajectory features are extracted using

the Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), and statistical methods. The performance of SVM is also compared with RF, KNN, and MLP. This paper provides an efficient method to identify driving style based on the trajectory data.

3. Data

A high-fidelity vehicle trajectory dataset, Next Generation Simulation (NGSIM), was collected by U.S. Federal Highway Administration (FHWA) in 2005. This dataset is still widely used in transportation research, especially in traffic flow analysis and modelling, traffic-related estimation and prediction, and vehicular ad hoc network-related studies [49]. It has rarely been applied to driving style recognition. Since this dataset was collected more than a decade ago, the accuracy of NGSIM dataset was questioned in recent years [50]. The measurement errors in NGSIM dataset were found to be far beyond negligible, partially due to low-resolution cameras and mis-tracking of vehicles from video images. Montanino et al. [51] removed outliers and noise and reconstructed the I-80 dataset 1 (from 4:00 p.m. to 4:15 p.m.), which showed significant improvement over the original NGSIM dataset.

In this paper, the I-80 trajectory dataset is adopted to study driving style. The trajectory data was collected on a segment of I-80 freeway in Emeryville, California. The segment contains 6 lanes, where lane 1 is a high occupancy vehicle (HOV) lane. The frequency of data collection is 10 Hz, and each leader-follower pair of dataset contains detailed information including the vehicle ID, position, length, and width of the vehicle, velocity, acceleration, lane ID, and following and preceding vehicles. About 206,000 records of vehicle trajectory for 370 Leader-follower Vehicle Pairs (LVP) on HOV lane are chosen to study the driving style in this paper since there are fewer interrupting vehicles from other lanes.

4. Methodology

The flow of driving style recognition in this paper is depicted in Figure 1. Three collision risk surrogates are used to determine the risk level of every moment in the car-following process for each LVP. K-means algorithm is applied to group the drivers as normal or aggressive driving style based on their trajectory risk levels. Given the labeled driving data, driving style recognition model is built using machine learning algorithms. The input features of machine learning algorithms are extracted by DFT, DWT, and statistical methods from trajectory features, without using surrogates and risk levels. The recognition results recognized by SVM are compared with other machine learning algorithms.

4.1. Collision Risk Surrogates. For each driver, it is essential to find the most effective surrogates to describe the collision risk when driving on the road. Vehicle trajectory data such as velocity and acceleration of the vehicle usually are not good enough to estimate the rear-end collision risk. Three collision surrogates are considered to measure the collision risk, including Time to Collision (TTC), Time-Headway (THW),

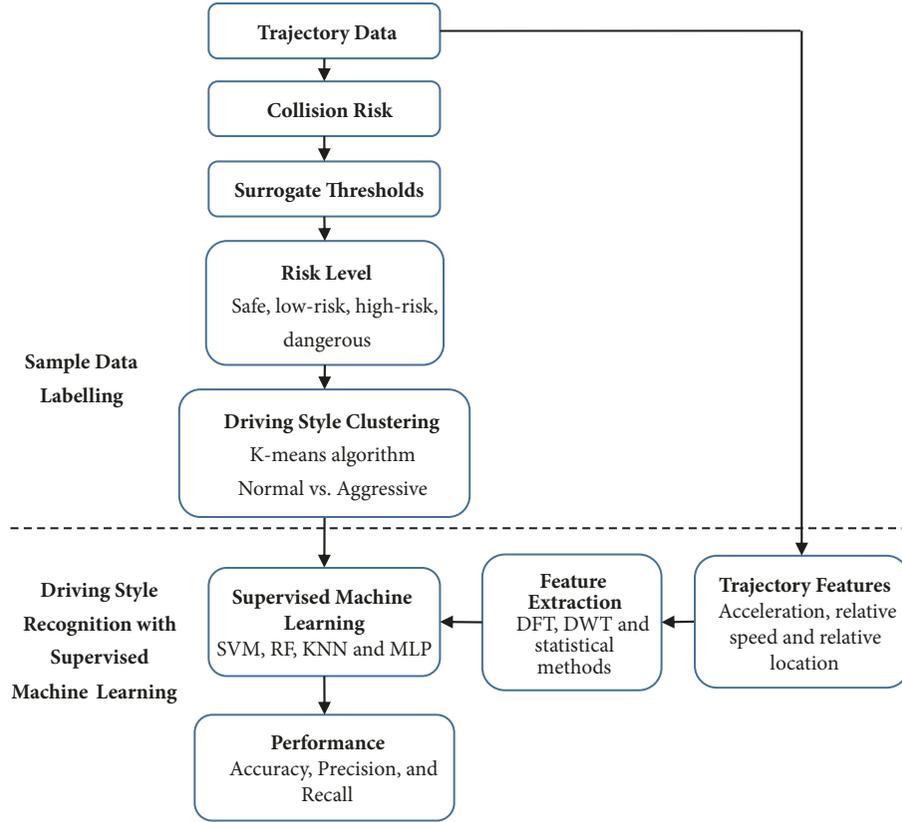


FIGURE 1: Research methodology framework.

and Margin to Collision (MTC). These three collision risk surrogates are defined and modified as follows.

Inversed Time to Collision (ITTC). TTC is the predicted time to collision between the preceding vehicle (PV) and following vehicle (FV) when two vehicles remain the current relative velocity.

$$TTC = -\frac{x_r}{v_r} = -\frac{(x_f - x_p)}{(v_f - v_p)} \quad (1)$$

where x_r and v_r denote relative distance and velocity between two following vehicles, respectively. x_f and x_p denote the front position of FV and rear position of PV, respectively. v_f and v_p , respectively, denote the velocity of FV and PV, respectively. However, TTC can be very large with lower relative velocity for two following vehicles, which happened a lot in the real driving environment. To reduce the scope of TTC, the ITTC is adopted to measure the collision risk in the paper. The risk of rear-end collision is higher with larger ITTC value.

$$ITTC = \frac{1}{TTC} \quad (2)$$

Time-Headway (THW). THW indicates the time for FV to reach the present position of PV with the current velocity. The potential collision risk of drivers is determined by THW in the steady vehicle following situation.

$$THW = -\frac{x_r}{v_f} = -\frac{(x_f - x_p)}{v_f} \quad (3)$$

The potential collision risk can be evaluated by THW when FV approaches PV with constant v_f . Lower THW indicates a higher potential collision risk.

The Modified Margin to Collision (MMTC). MTC indicates the final relative position of PV and FV if two vehicles decelerate abruptly.

$$MTC = \frac{(-x_r + d_p)}{d_f} = \frac{(-x_r - v_p^2/2a_p)}{(-v_f^2/2a_f)} \quad (4)$$

where a_f and a_p denote the deceleration for FV and PV, respectively. Usually, both are defined as $0.7G$. A modified MTC (MMTC) is used in the paper to include the reaction time of the following vehicle when the PV abruptly decelerates. The equation is modified as follows.

$$\begin{aligned} MMTC &= \frac{(-x_r + d_p - d_f)}{v_f} \\ &= \frac{(-x_r - v_p^2/2a_p + v_f^2/2a_f)}{v_f} \end{aligned} \quad (5)$$

MMTC evaluates the minimum reaction time needed for FV to avoid a collision when PV abruptly decelerates at $0.7G$.

The collision risk is higher with lower MMTC value since there is little time for drivers to react. MMTC can evaluate potential collision risk with abrupt deceleration of PV.

4.2. Driving Style Clustering. The threshold values of surrogates are adopted to divide the trajectory of each driver into several collision risk levels. Then the K-means method is used to group the drivers into normal or aggressive driving style based on their components of collision risk levels. The purpose of the method is to provide an objective and stable label of driving style for each driver in the sample data and then make it ready to use in supervised machine learning.

Assume that there are n sets of driving data, and each set consists of v dimensional features denoting λ_i , which belongs to a class y_i . Therefore, the driving data of each driver can be described as $\{\lambda_i, y_i\}$. The K-means method finds the best class y_i for each driving data. The objective function of the K-means algorithm is to minimize the total in-class error squares shown as follows.

$$\min \sum_i^k \sum_{y \in y_i} \left\| (\lambda, y_i) - (\hat{\lambda}_i, y_i) \right\|^2 \quad (6)$$

where k is the number of classes. $(\hat{\lambda}_i, y_i)$ is the mean vector of all points in class y_i .

4.3. Trajectory Feature Extraction. In this paper, the vehicle acceleration a_f , relative distance x_r , and relative velocity v_r are adopted to recognize the driving style. The Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), and statistical method are used, respectively, to extract the effective features from the vehicle acceleration a_f , relative distance x_r , and relative velocity v_r .

4.3.1. Discrete Fourier Transform. DFT has been applied to convert time series of trajectory data to signal amplitude in the frequency domain [7]. The DFT of a given time series (x_1, x_2, \dots, x_N) is defined as a sequence of N complex numbers $(DFT_0, DFT_1, \dots, DFT_{N-1})$:

$$DFT_k = \sum_{n=0}^{N-1} x_n e^{-(2\pi i/N)kn} \quad (7)$$

where i is the imaginary unit. The first 10 DFT coefficients of trajectory data are used to recognize the driving style.

4.3.2. Discrete Wavelet Transform. DWT is shown to be more suitable to analyze and decompose a given signal in some studies [54]. This paper follows the DWT method described in [54] and uses the energy of approximation sub-time series and detail sub-time series, which are decomposed from vehicle acceleration a_f , relative distance x_r , and relative velocity v_r , to recognize the driving style.

4.3.3. Statistical Method. The key statistical parameters that can capture most of the distribution information of vehicle acceleration a_f , relative distance x_r , and relative velocity v_r ,

TABLE 2: Pearson correlation analysis of ITTC, THW, and MMTC.

Surrogate	THW	MMTC	ITTC
THW	1	.980**	-.240**
MMTC	.980**	1	-.385**
ITTC	-.240**	-.385**	1

** : significant correlation at 0.01 level (bilateral).

are also selected for recognition. The statistical parameters are the maximum, minimum, mean, standard deviation, and 85% percentiles, which were proved useful in previous driving behavior study [20].

4.3.4. Feature Combinations. For each driver, during car-following process, there are three time series: acceleration a_f , relative distance x_r , and relative velocity v_r . This paper tries 7 different feature combinations as the input of driving style recognition model:

Single-source features: use only one time series out of acceleration a_f , relative distance x_r , and relative velocity v_r , and extract features from this time series.

Two-source features: use two time series out of acceleration a_f , relative distance x_r , and relative velocity v_r . Therefore, there are three combinations: $a_f + x_r$, $x_r + v_r$, and $v_r + a_f$. Features are extracted from two time series separately.

Three-source features: use all three time series and extract features from three time series separately.

5. Results and Discussion

5.1. The Sample Data Labelling

5.1.1. Threshold Value of Collision Risk Surrogates. The correlation analysis among three surrogates is shown in Table 2.

Table 2 shows that the Pearson coefficient between THW and MMTC is 0.980, indicating a strong positive correlation. ITTC and THW have a weak negative correlation. Therefore, ITTC and THW are selected to measure driving behavior risk. The classification result will not be influenced by the adopting of THW instead of MMTC because of the strong correlation between the two surrogates.

To make a reasonable adjustment on collision risk along the car-following process, each surrogate has a risk threshold that can be obtained through the probability density distribution and fitting results of ITTC, THW shown in Figure 2.

Figure 2(a) shows the fitting results of ITTC, THW by adopting three distributions, i.e., normal distribution, logistic distribution, and t distribution. The t distribution achieves a better fitting performance than other two distributions on probability density distribution of ITTC and THW. Therefore, the t distribution is adopted to determine the threshold value of features. The percentile values of ITTC are shown in Figure 2(b). The 25%, 45%, 65%, 85%, and 95% percentile values of ITTC are 0.02, 0.08, 0.12, 0.19, and 0.28 s^{-1} , respectively. The 25%, 45%, 65%, and 85% percentile values of THW are 1.26, 1.71, 2.13, and 2.73 s, respectively.

ITTC. The upper threshold of ITTC is 0.28 s^{-1} , which is equivalent to 3.5 s for TTC. Previous studies show that

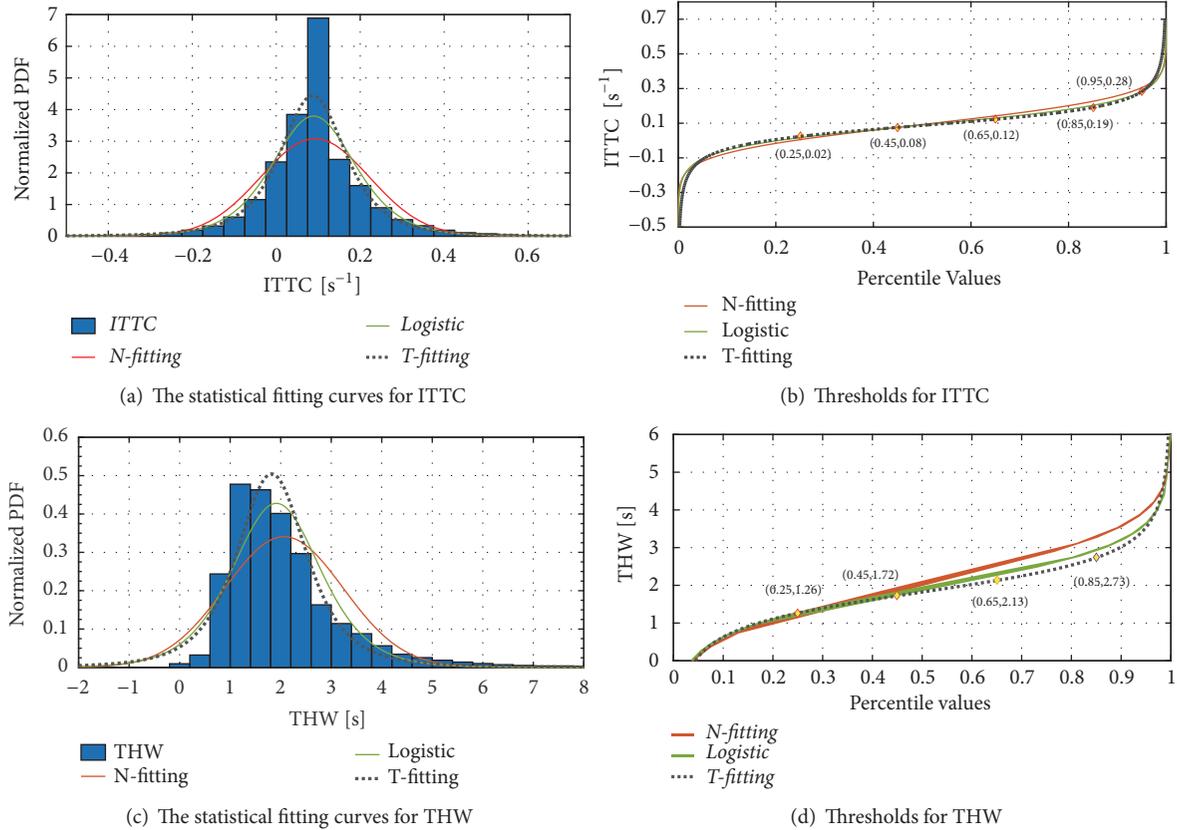


FIGURE 2: Fitting results and thresholds of surrogates.

the desirable TTC is 4 s for urban road [46] and 3.5 s for nonsupported drivers [45]. The desirable TTC for signalized intersection and two-lane rural roads is 3 s [47]. Therefore, 3.5 s is adopted in this paper as the rear-end collision risk threshold. When TTC is lower than 3.5 s, the FV is labeled as having a higher collision risk.

THW. Since a lower THW indicates a higher collision risk, the author first chose the 25% percentile, which is 1.26 s. However, many road administrations in European countries recommend a safe THW of 2 s [48]. The THW below 2 s may cause uncomfortable driving feelings and potential risk for drivers. Finally, 2 s is used as the threshold value for THW in this study.

5.1.2. Trajectory Risk Level. The threshold values of ITTC and THW, i.e., 0.28 s^{-1} and 2 s, are used to divide the driving trajectory into different risk levels. To be more specific, the different values of ITTC and THW are corresponding to different driving risk level. The driving trajectory for each driver can be divided into four risk levels: safe, low-risky, high-risky, and dangerous driving behavior, shown in Figure 3.

Safe Driving Behavior. The FV has THW above 2 s and ITTC below 0.28 s^{-1} , which indicates that the FV keeps low velocity and a large gap with the PV at car-following state.

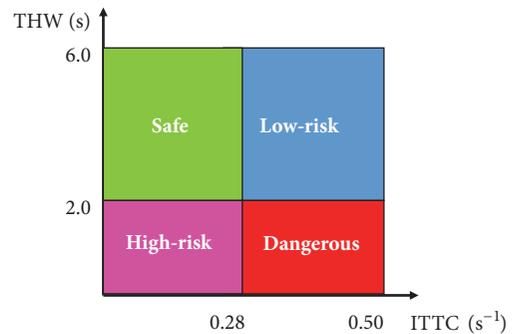


FIGURE 3: The threshold values of surrogates indicating different driving risk.

Low-Risky Driving Behavior. The FV has THW above 2 s and ITTC above 0.28 s^{-1} , which indicates that the FV keeps low velocity and a small gap with the PV at car-following state.

High-Risky Driving Behavior. The FV has THW below 2 s and ITTC below 0.28 s^{-1} , which indicates that the FV remains high velocity and a large gap with the PV at car-following state.

Dangerous Driving Behavior. The FV has THW below 2 s and ITTC above 0.28 s^{-1} , which indicates that the FV remains

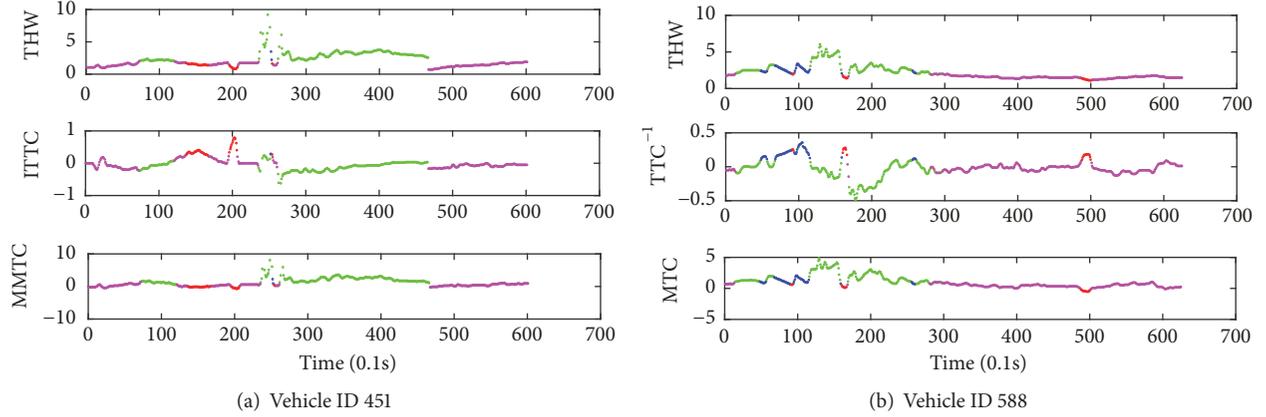


FIGURE 4: Trajectory segments for four drivers based on threshold values of ITTC and THW.

high velocity and a small gap with the PV at car-following state.

The driving trajectory of each driver can be divided into several segments, which belongs to different driving risk levels. Two drivers are selected to show the trajectory segments according to the threshold values of ITTC and THW, shown in Figure 4.

As Figure 4 shows, for most drivers, the safe and high-risk driving behaviors account for over 80% of driving trajectory. The proportion of dangerous driving and low-risk driving behaviors is limited to 10% and 5%, respectively. The driving style of each driver can be determined by the proportions of trajectory segments with different risk levels. The 370 drivers are clustered into two groups in Section 5.1.3.

5.1.3. Driving Style Clustering. Based on the proportions of trajectory segments determined by the threshold values of ITTC and THW, the drivers can be grouped into two classes using the K-means algorithm. The results show one class has 246 drivers and the other has 124 drivers. On average, drivers in the first class have 45.5% safe driving behavior, 37.5% high-risk driving behavior, and 11.4% dangerous driving behavior, and drivers in the second class have 7.4% safe driving behavior, 77.8% high-risk driving behavior, and 13.5% dangerous driving behavior. Therefore, drivers in the first class are labelled as normal drivers, while drivers in the second class are labelled as aggressive drivers. The driving style labels provided by K-means are used to train SVM in Section 5.2.

5.2. Driving Style Recognition. The SVM method is adopted to recognize the driving style for 370 drivers. In this paper, the trajectory data including the vehicle acceleration a_f , relative distance x_r , and relative velocity v_r are adopted to recognize the driving style, respectively. The DFT, DWT, and statistical methods are both applied to extract effective features from trajectory data. Every single feature can also be combined with other features as multisource features to recognize the driving style. The recognition accuracy rates are compared to find the best feature extraction method and the most important trajectory features. The z-score method is adopted to standardize features before model training.

In the study, the accuracy, precision, and recall rates are assessed to evaluate the model's ability to recognize aggressive drivers among all vehicles on the road. The performance of the recognition model is evaluated using the "leave-one-out" cross-validation method. Driving style recognition results based on different feature extraction methods and SVM are shown in Tables 3–7. Except mentioned, the SVM algorithm uses linear kernel function.

5.2.1. Discrete Fourier Transform. Shown in Table 3, the recognition accuracy rate is 83.2% based on v_r and 88.9% based on x_r . The recognition accuracy rate is 88.9% based on x_r and a_f , and 87.8% based on x_r and v_r . In general, the features x_r and v_r are better than a_f in recognizing the driving style. A possible reason is that the driving style label is determined by the rear-end collision risk, the feature a_f can not accurately describe the relative motivation between two following vehicles. The accuracy rate based on all three features can achieve 87.6%. Surprisingly, using DFT coefficients of x_r along has the highest accuracy rate.

5.2.2. Discrete Wavelet Transform. For DWT, there are two parameters to be determined, which could affect the performance of the recognition model. One is an appropriate wavelet mother function; the other is the number of decomposition levels. This paper tried 15 different wavelet mother functions (listed in Table 4) and 5 decomposition levels (listed in Table 5). The results show that Daubechies 4 mother function can generate the highest accuracy rate: 91.7%. The best decomposition level is 1, while decomposing time series further does not help to improve the accuracy rate.

With Daubechies 4 mother function and 1 decomposition level, SVM performance is assessed with different combinations of features. Shown in Table 6, the recognition accuracy rate is 83.8% based on v_r and 86.8% based on x_r . Therefore, when using x_r along in SVM, DFT extraction method works better than DWT. The recognition accuracy rate is 88.7% based on x_r and a_f and 90.2% based on x_r and v_r . The accuracy rate based on all three features can achieve 91.7%. Compared with DFT coefficients, DWT methods also get higher precision rate 92.8% and higher recall rate 81.8%.

TABLE 3: The evaluation results of driving style based on SVM using DFT.

Features	Accuracy	Precision	Recall
a_f	67.0%	55.0%	17.5%
v_r	83.2%	80.2%	67.5%
x_r	88.9%	86.3%	80.2%
$a_f + v_r$	83.2%	80.8%	66.7%
$v_r + x_r$	87.8%	86.5%	76.2%
$a_f + x_r$	88.9%	84.6%	82.5%
$a_f + v_r + x_r$	87.6%	85.1%	77.0%

TABLE 4: The accuracy of driving style recognition using DWT-SVM with different wavelet mother functions.

Wavelet Mother	Accuracy	Wavelet mother	Accuracy	Wavelet mother	Accuracy
Daubechies 1	90.3%	Symlet 1	90.3%	Coiflet 1	90.8%
Daubechies 2	90.0%	Symlet 2	90.0%	Coiflet 2	90.3%
Daubechies 3	91.1%	Symlet 3	91.1%	Coiflet 3	89.7%
Daubechies 4	91.7%	Symlet 4	89.7%	Coiflet 4	90.0%
Daubechies 5	91.1%	Symlet 5	90.5%	Coiflet 5	88.4%

TABLE 5: The performance of DWT-SVM with different decomposition levels.

Decomposition Level	Accuracy	Precision	Recall
1	91.7%	92.8%	81.8%
2	90.5%	91.0%	80.2%
3	86.5%	83.3%	75.4%
4	87.0%	83.1%	77.8%
5	85.1%	80.8%	73.8%

5.2.3. *Statistical Method.* Driving style recognition results based on the features extracted by statistical method and SVM are shown in Table 7. With any combinations of features, the accuracy rate of the statistical method is lower than that based on DFT and DWT. The highest accuracy rate in Table 7 is 85.7% when adopting three features.

5.2.4. *Machine Learning Algorithms.* This section tests the performance of four machine learning algorithms: RF, MLP, KNN, and SVM using all three features and DWT method. The accuracy, precision, and recall rates are listed in Table 8. SVM outperforms other machine learning algorithms. Random Forest is the second best algorithm. MLP gives the highest recall rate among all candidates. KNN, as the simplest classification method, unsurprisingly obtains the worst performance.

6. Conclusion

In this study, a novel driving style labelling method is proposed to assign normal and aggressive labels based on collision risk, which is critical to sample data needed in supervised machine learning. The method is based on the vehicle trajectory extracted from traffic video. The rear-end collision risk surrogates are adopted to evaluate the

risk during the car-following process. The study also applies the SVM algorithm to recognize the driving style based on the trajectory features. Three feature extraction methods are tested. Other machine learning algorithms including RF, MLP, and KNN are also adopted to compare with the SVM. Several conclusions can be obtained from this study.

(1) Three effective rear-end collision risk surrogates, namely, ITTC, THW, and MMTC, are selected to evaluate the collision risk in the car-following process. Since THW and MMTC show a strong positive correlation, only ITTC and THW are kept to evaluate driving risk level. This paper gives threshold values of ITTC and THW based on their distribution and previous studies. Each driver's trajectory can be divided into four risk levels, and all drivers can be grouped into two classes using the K-means algorithm. Using NGSIM dataset, this method labels 246 normal drivers and 124 aggressive drivers. On average, normal drivers have 45.5% safe driving behavior, 37.5% high-risk driving behavior, and 11.4% dangerous driving behavior, and aggressive drivers have 7.4% safe driving behavior, 77.8% high-risk driving behavior, and 13.5% dangerous driving behavior.

(2) DFT, DWT, and statistical methods are adopted to extract the effective features from trajectory data to facilitate the driving style recognition. Using relative distance along DFT method can convert relative distance time series into coefficients in the frequency domain and help SVM reach the accuracy rate of 88.9%, the precision rate of 86.3%, and the recall rate of 80.2%. However, when using multiple features, including acceleration, relative distance, and relative speed, DWT method can improve the accuracy rate to 91.7%, precision rate to 92.8%, and recall rate to 81.8%. Among 15 wavelet mother functions tested, Daubechies 4 mother function provides the best results.

(3) The driving style can be accurately recognized by the proposed SVM model based on the trajectory features with 91.7% accuracy rate. The recognition accuracy is superior to other famous and frequently used classifiers: RF, MLP, and

TABLE 6: The evaluation results of driving style based on SVM using DWT.

Features	Accuracy	Precision	Recall
a_f^*	69.5%	74.1%	15.9%
v_r	83.8%	83.0%	65.9%
x_r	86.8%	88.1%	70.6%
$a_f + v_r$	82.4%	81.4%	62.7%
$v_r + x_r$	90.2%	90.2%	80.2%
$a_f + x_r$	88.7%	89.6%	75.4%
$a_f + v_r + x_r$	91.7%	92.8%	81.8%

*: using polynomial kernel function in SVM to produce better results.

TABLE 7: The evaluation results of driving style based on SVM using the statistical method.

Features	Accuracy	Precision	Recall
a_f^*	66.2%	51.6%	12.7%
v_r	79.7%	72.6%	65.1%
x_r	83.5%	80.4%	68.3%
$a_f + v_r$	78.7%	71.2%	62.7%
$v_r + x_r$	83.2%	79.6%	68.3%
$a_f + x_r$	84.6%	79.5%	73.8%
$a_f + v_r + x_r$	85.7%	79.2%	78.6%

*: using polynomial kernel function, in SVM to produce better results.

TABLE 8: The performance comparison of multiple algorithms.

Features	Algorithm	Accuracy	Precision	Recall
$a_f + v_r + x_r$	RF	91.6%	89.2%	82.0%
	MLP	88.1%	73.0%	87.3%
	KNN	87.6%	85.7%	76.2%
	SVM	91.7%	92.8%	81.8%

KNN. This result indicates that the SVM method is a more appropriate method for driving style recognition based on the trajectory features.

(4) The proposed method can be effectively used to label and recognize the driving style based on the traffic video surveillance systems. The development of network connected vehicles can help to collect the data more precisely. The model with machine learning algorithm can be trained to better recognize driving style. It can help to evaluate the collision risk on the road network and also provide real-time decision support to drivers.

This study offers the possibility of developing more sophisticated driving style recognition methods. For further work, the proposed method can be extended by selecting other features that can reflect the driving style more accurately. As we know, the driving style is also influenced by the road conditions and traffic flow level. Such results can also be used to improve the driving style recognition. It is possible to use some semi-supervised and unsupervised methods to save the label time in the future.

Data Availability

The reconstructed NGSIM dataset can be accessed at <http://www.multitude-project.eu/reconstructed-ngsim.html>. The

original NGSIM data is open to download at <https://data.transportation.gov/>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study has been funded by the National Key Research and Development Program of China (No. 2017YFC0803902).

References

- [1] J. Elander, R. West, and D. French, "Behavioral correlates of individual differences in road-traffic crash risk: an examination method and findings," *Psychological Bulletin*, vol. 113, no. 2, pp. 279–294, 1993.
- [2] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014.
- [3] Y. Zheng, S. E. Li, J. Wang, D. Cao, and K. Li, "Stability and scalability of homogeneous vehicular platoon: Study on the

- influence of information flow topologies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 14–26, 2015.
- [4] B. Yu, H. Wang, W. Shan et al., "Prediction of bus travel time using random forests based on near neighbors," *Computer-aided Civil and Infrastructure Engineering*, vol. 33, no. 2, 2017.
 - [5] B. Yu, X. Song, F. Guan et al., "k-nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *Journal of Transportation Engineering*, vol. 142, no. 6, Article ID 04016018, 2016.
 - [6] J. Yao, H. Xia, B. Yu et al., "Prediction on building vibration induced by moving train based on support vector machine and wavelet analysis," *Journal of Mechanical Science and Technology*, vol. 28, no. 6, pp. 2065–2074, 2014.
 - [7] Y. Zhang, W. C. Lin, and Y.-K. S. Chin, "A pattern-recognition approach for driving skill characterization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 905–916, 2010.
 - [8] P. Molchanov, S. Gupta, K. Kim et al., "Multi-sensor system for driver's hand-gesture recognition," in *Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–8, IEEE, 2015.
 - [9] A. Hashemi, V. Saba, and S. N. Resalat, "Real time driver's drowsiness detection by processing the EEG signals stimulated with external flickering light," *Basic and Clinical Neuroscience*, vol. 5, no. 1, pp. 22–27, 2014.
 - [10] H. Wang, C. Zhang, T. Shi et al., "Real-time EEG-based detection of fatigue driving danger for accident prediction," *International Journal of Neural Systems*, vol. 25, no. 2, pp. 498–369, 2015.
 - [11] D. Srinivasan, X. Jin, and R. L. Cheu, "Adaptive neural network models for automatic incident detection on freeways," *Neurocomputing*, vol. 64, no. 1-4, pp. 473–496, 2005.
 - [12] N. K. Singh, A. K. Singh, and M. Tripathy, "A comparative study of BPNN, RBFNN and ELMAN neural network for short-term electric load forecasting: a case study of Delhi region," in *Proceedings of the International Conference on Industrial and Information Systems*, IEEE, 2015.
 - [13] M. M. Bejani and M. Ghatee, "A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 303–320, 2018.
 - [14] G. Li, S. E. Li, B. Cheng, and P. Green, "Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 113–125, 2017.
 - [15] H. Berndt and K. Dietmayer, "Driver intention inference with vehicle onboard sensors," in *Proceedings of the International Conference on Vehicular Electronics and Safety, ICVES '09*, pp. 102–107, IEEE, 2009.
 - [16] X. Meng, K. K. Lee, and Y. Xu, "Human driving behavior recognition based on hidden markov models," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics*, 2007.
 - [17] T. Kumagai, Y. Sakaguchi, M. Okuwa et al., "Prediction of driving behavior through probabilistic inference," in *Proceedings of the 8th International Conference on Engineering Applications of Neural Networks*, 2003.
 - [18] C. F. Zong, X. Yang, C. Wang et al., "Driver's driving intention identification and behavior prediction during vehicle steering," *Journal of Jilin University*, vol. s1, pp. 27–32, 2009.
 - [19] G. S. Aoude, V. R. Desaraju, L. H. Stephens et al., "Driver behavior classification at intersections and validation on large naturalistic data set," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 724–736, 2012.
 - [20] Z. Chen, C. Wu, Z. Huang et al., "Dangerous driving behavior detection using video-extracted vehicle trajectory histograms," *Journal of Intelligent Transportation Systems*, vol. 21, no. 11, 2017.
 - [21] W. Sun, X. Zhang, S. Peeta et al., "A real-time fatigue driving recognition method incorporating contextual features and two fusion levels," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–13, 2017.
 - [22] L. Feng, Y. Yao, and B. Jin, "Research on credit scoring model with svm for network management," *Journal of Computational Information Systems*, vol. 6, no. 11, pp. 1032–1040, 2010.
 - [23] C. M. Martinez, M. Heucke, F. Y. Wang et al., "Driving style recognition for intelligent vehicle control and advanced driver assistance: a survey," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–11, 2017.
 - [24] M. Wu, S. Zhang, and Y. Dong, "A novel model-based driving behavior recognition system using motion sensors," *Sensors*, vol. 16, no. 10, p. 1746, 2016.
 - [25] W. Wang and J. Xi, "A rapid style-recognition method for driving styles using clustering-based support vector machines," in *Proceedings of the American Control Conference*, IEEE, 2016.
 - [26] W. Wang, J. Xi, A. Chong et al., "Driving style classification using a semi-supervised support vector machine," *IEEE Transactions on Human-Machine Systems*, 2017.
 - [27] C. Zhang, M. Patel, S. Buthpitiya et al., "Driver classification based on driving behaviors," in *Proceedings of the International Conference on Intelligent User Interfaces*, pp. 80–84, ACM, 2016.
 - [28] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 300–311, 2010.
 - [29] B. Coifman, "Using LIDAR to validate the performance of vehicle classification stations," *Journal of Intelligent Transportation Systems*, vol. 19, no. 4, pp. 355–369, 2015.
 - [30] J. Schorr, S. H. Hamdar, and C. Silverstein, "Measuring the safety impact of road infrastructure systems on driver behavior: vehicle instrumentation and real world driving experiment," *Journal of Intelligent Transportation Systems*, 2017.
 - [31] R. Sun, K. Han, J. Hu, Y. Wang, M. Hu, and W. Y. Ochieng, "Integrated solution for anomalous driving detection based on BeiDou/GPS/IMU measurements," *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 193–207, 2016.
 - [32] T. Liu, Y. Yang, G.-B. Huang et al., "Driver distraction detection using semi-supervised machine learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1108–1120, 2016.
 - [33] N. P. Chandrasiri, K. Nawa, and A. Ishii, "Driving skill classification in curve driving scenes using machine learning," *Journal of Modern Transportation*, vol. 24, no. 3, pp. 1–11, 2016.
 - [34] C. Yan, F. Coenen, Y. Yue, X. Yang, and B. Zhang, "Video-based classification of driving behavior using a hierarchical classification system with multiple features," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 30, no. 5, 2016.
 - [35] M. H. Zaki, T. Sayed, and K. Shaaban, "Use of drivers' jerk profiles in computer vision-based traffic safety evaluations," *Transportation Research Record*, 2014.

- [36] C. Oh, E. Jung, H. Rim et al., "Intervehicle safety warning information system for unsafe driving events," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2324, pp. 1–10, 2012.
- [37] R. Chhabra, S. Verma, and C. R. Krishna, "A survey on driver behavior detection techniques for intelligent transportation systems," in *Proceedings of the International Conference on Cloud Computing, Data Science and Engineering - Confluence*, pp. 36–41, IEEE, 2017.
- [38] F. Teimouri and M. Ghatee, "A real-time warning system for rear-end collision based on random forest classifier," 2018, <https://arxiv.org/abs/1803.10988>.
- [39] M. Ishibashi, M. Okuwa, S. Doi et al., "Indices for characterizing driving style and their relevance to car following behavior," in *Proceedings of the SICE*, 2007.
- [40] N. Li, T. Misu, and A. Miranda, "Driver behavior event detection for manual annotation by clustering of the driver physiological signals," in *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*, 2016.
- [41] B. Zhu, W. Li, N. Bian et al., "Identification of driver individualities using random forest model," *SAE Technical Papers*, 2017.
- [42] J. Ambros, R. Turek, and J. Paukert, "Road safety evaluation using traffic conflicts: Pilot comparison of micro-simulation and observation," in *Proceedings of the International Conference on Traffic and Transport Engineering*, 2014.
- [43] S. Kitajima, Y. Marumo, T. Hiraoka et al., "Comparison of evaluation features concerning estimation of driver's risk perception," *Transactions of the Society of Automotive Engineers of Japan*, vol. 40, no. 2, pp. 191–198, 2009.
- [44] S. M. S. Mahmud, L. Ferreira, M. S. Hoque et al., "Application of proximal surrogate indicators for safety evaluation: a review of recent developments and research needs," *IATSS Research*, vol. 41, no. 4, 2017.
- [45] H. J. Hogema and W. H. Janssen, "EFFECTs of intelligent cruise control on driving behavior: a simulator study," in *Proceedings of the Intelligent Transportation: Realizing the Future Abstracts of the 3rd World Congress on Intelligent Transport Systems*, 1996.
- [46] Q. Meng and X. Qu, "Estimation of vehicle crash frequencies in road tunnels," *Accident Analysis & Prevention*, vol. 48, no. 5, pp. 254–263, 2012.
- [47] T. Sayed, M. H. Zaki, and J. Autey, "Automated safety diagnosis of vehicle-bicycle interactions using computer vision analysis," *Safety Science*, vol. 59, pp. 163–172, 2013.
- [48] L. Yang, C. F. Xing, and H. B. Zhao, "Study on driver's reaction time (DRT) during car following," *Computing Technology and Automation*, vol. 34, no. 3, pp. 33–37, 2015.
- [49] V. G. Kovvali, V. Alexoadis, and P. E. Zhang, "Video-based vehicle trajectory data collection," in *Proceedings of the Transportation Research Board 86th Annual Meeting*, 2007.
- [50] Z. He, "Research based on high-fidelity NGSIM vehicle trajectory datasets: a review," *Research Gate*, pp. 1–33, 2017.
- [51] M. Montanino and V. Punzo, "Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns," *Transportation Research Part B: Methodological*, vol. 80, pp. 82–106, 2015.
- [52] J. Wahlström, I. Skog, and P. Handel, "Smartphone-based vehicle telematics: a ten-year anniversary," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2802–2825, 2017.
- [53] R. B. Zadeh, M. Ghatee, and H. R. Eftekhari, "Three-phases smartphone-based warning system to protect vulnerable road users under fuzzy conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 7, pp. 2086–2098, 2018.
- [54] H. R. Eftekhari and M. Ghatee, "Hybrid of discrete wavelet transform and adaptive neuro fuzzy inference system for overall driving behavior recognition," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 58, pp. 782–796, 2018.
- [55] H. R. Eftekhari and M. Ghatee, "An inference engine for smartphones to preprocess data and detect stationary and transportation modes," *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 313–327, 2016.

