

## Research Article

# A Case Study in China to Determine Whether GPS Data and Derivative Indicator Can Be Used to Identify Risky Drivers

Rui Fu <sup>1</sup>, Tong Liu <sup>1</sup>, Yuxi Guo,<sup>1</sup> Shiwei Zhang,<sup>1</sup> and Wendong Cheng<sup>2</sup>

<sup>1</sup>Chang'an University, School of Automobile, Middle Section of Nan'er Huan Road, Xi'an, 710064, China

<sup>2</sup>Xi'an Technological University, School of Mechatronic Engineering, Xi'an, 710021, China

Correspondence should be addressed to Tong Liu; [liutong@chd.edu.cn](mailto:liutong@chd.edu.cn)

Received 30 November 2018; Revised 9 April 2019; Accepted 2 July 2019; Published 24 November 2019

Academic Editor: Filomena Mauriello

Copyright © 2019 Rui Fu 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.

This paper presents an investigation of the relationship between driver risk and factors indicating vehicle's speed and driver's acceleration behavior. The main objective is to examine whether GPS data and derivative indicator can be used to identify risky drivers by means of factor analysis. In doing so, a real road driving experiment is conducted to collect data. Fifty drivers are asked to drive along a route which includes both rural highways and urban roads. The trajectories are recorded by GPS devices to calculate speed and derive acceleration measures. Driver's behavior is also recorded by cameras and analyzed by another group of volunteers to determine whether the driver is risky or not. The drivers are then classified into five groups with different levels of risk based on the scores obtained through factor analysis. The results are verified by the volunteer's categorization and further evaluated by symbolic aggregate approximation. A binary logistic regression model is established ultimately for predicting high-risk drivers. The potential applications of this study include developing quantitative measures to identify risky drivers, especially for auto-insurance companies with usage-based insurance (UBI) applications, bus companies, and transport enterprises.

## 1. Introduction

With an ever-increasing number of cars on the roads, traffic safety has become a progressively more important research issue worldwide [1]. According to statistical data of road accidents in China [2], nearly 95% of road accidents are attributed to human factors and driver maneuvers. Among these factors, inappropriate speed control or speeding behavior is one of the most risky factors contributing to road accidents [3]. Furthermore, factors like inappropriate speed control or speeding account for 3.62% of accidents, with death and injury rates of 5.59% and 3.34%, respectively [4]. Research has shown that there is a significant correlation between speed-related risk indices and historical accident data [5]. To this end, measures should be taken to monitor driver performance, particularly typical risky driving behaviors such as hard braking or abrupt acceleration, and speed-limit violation.

Usage-based insurance (UBI) or insurance telematics appears to have become increasingly popular in the insurance industry, especially among innovative auto-insurance

companies worldwide [6]. Various UBI applications for personalized auto-insurance, including smartphone-based products and other on-board devices, are emerging to evaluate driving performance and determine the premiums or benefits according to the driving data collected during trips. Generally, the data used for UBI analysis include running period (time of day), location of vehicle, speed condition (e.g., average speed and speeding percentage), usage of gas pedal or brake pedal (e.g., the frequency of abrupt acceleration or hard braking), steering behavior, and total trip length [7, 8]. Furthermore, UBI schemes also provide an insight into other possible uses for the information gathered during the identification of risky driving behaviors. With the help of UBI schemes, drivers could receive performance feedback and understand factors that may stimulate or motivate them to behave aggressively. Some UBI schemes that have been implemented around the world are summarized in the work by Husnjak et al. [8] and Tselentis et al. [9].

The main contribution of this paper is to explore the relationship between driver risk and factors indicating vehicle's speed and driver's acceleration behavior from a real road

driving study. The purpose is to examine whether GPS data and derivative indicator can be used to identify risky drivers through factor analysis. Categorization by volunteers and symbolic representation are both used to verify and evaluate the results of the case study. A binary logistic regression model is developed for predicting high-risk drivers. Moreover, for the purpose of identifying individual differences and comparing driving behavior among drivers, nearly the same trip length and driving period are selected for each participant.

The remainder of this paper is organized as follows. Section 2 discusses some previous work on the identification of risky driving behaviors through various monitoring systems or devices. Section 3 presents the variables extracted from GPS data and determines the threshold for abrupt acceleration and hard braking in the test. Section 4 describes the details of the real road driving experiment and introduces the rules for extracting risky events. Section 5 compares the risk scores obtained through factor analysis with the scores evaluated by volunteers and develops a binary logistic regression model for predicting high-risk drivers. In Section 6, symbolic aggregate approximation is used to evaluate the case study. Finally, Section 7 compares the methods and results described here with previous research and offers conclusions and some suggestions for future work.

## 2. Literature Review

Some previous research has been done on the identification of risky driving behaviors in normal driving situations, and some systems have been proposed that utilize smartphones, on-board data recorders, multiple sensors such as the Controller Area Network (CAN) bus, camera systems, 3-axis accelerometers, on-board diagnostics (OBD) readers, and GPS (Global Positioning System) receivers. Wang et al. [10] proposed a system with multi-sensors in a bid to detect risky driving events on the basis of vehicle dynamic parameters and driver physiological data. Imkamon et al. [11] developed a fuzzy logic system to infer the risk level of drivers with information collected from an OBD reader, accelerometer, and camera during a real road driving test. This method performed well when compared to passenger responses questionnaires, which were collected during the trips. A self-diagnosis system equipped with multi-sensors was proposed by Takeda et al. [12] to detect hazardous situations and potential risky events in terms of the rule-based and thresholding method. The number of risky driving behaviors was reduced by half following the use of this system.

The high cost and complexity of multi-sensor-based systems mean that it is not practical to deploy within a vehicle for daily users, and an efficient approach utilizing available sensors in smartphones or data recorders (including a GPS receiver or 3-axis accelerometer) is required. Unlike some costly methods, smartphones with some applications installed could also provide reliable information for identifying some typical risky driving behaviors, such as the percentage in excess of the legal speed limit, or the frequency of abrupt acceleration and hard braking [13]. Researchers Johnson and Trivedi [14] were able to determine driving style with

smartphones by using the dynamic time warping algorithm. Zeeman and Booyesen [15] used GPS data (speed and location) and acceleration information collected from minibus taxis to detect reckless driving and presented a reckless driving detection system to visualize the tracked vehicle and the driving behavior. Similarly, Fazeen et al. [16] utilized GPS data and acceleration information derived from smartphones to detect risky driving behaviors in different road conditions. The data used in this approach included vehicle position, maximum speed, maximum acceleration, and minimum deceleration. Results show that it is feasible to detect aberrant driving behaviors and identify risky situations automatically with smartphones, even in different road conditions.

Furthermore, Eboli et al. [17] investigated the relationship between drivers' characteristics and driving behaviors by analyzing speed information derived from smartphones with embedded GPS chips through a real road driving test, and classified driving behaviors in terms of the travelling speed adopted by drivers. Also, Eboli et al. [18] utilized the information of longitudinal/lateral acceleration and speed derived from smartphones equipped with accelerometer and GPS receiver to identify the risk level of a driver, and made an overall judgment on the driving style based on objective measure (the percentage of external points at borderline of a safety domain) and subjective judgment (a questionnaire provided by the driver at the end of each survey). Considering that both speed and acceleration information can be derived from GPS data, devices with only a GPS receiver can be used for risky behavior identification. Therefore, this research will discuss the application of GPS-measured data for identifying some typical risky situations during a trip and determine the risk level of each driver with the help of video verification.

## 3. Variables and Data Analysis

In this test, GPS data and some derivative indicators are used to identify the risky events and evaluate the risk level of drivers. Therefore, a simple method that can detect typical risky situations efficiently is proposed here, and a data analysis is conducted in the following sections.

**3.1. Average Speed.** The instantaneous speed [19] is calculated between two consecutive GPS coordinates with the sampling frequency of 1 HZ, which is shown in Eq. (1). The average speed  $v_a$  is described in Eq. (2).

$$v_i = \frac{d}{t} * 3.6, \quad (1)$$

$$v_a = \frac{1}{m} \sum_{i=1}^m v_i, \quad (2)$$

where  $v_i$  or  $v_a$  is in the unit of km/h,  $d$  is in the unit of meters, and  $t$  is recorded in seconds.

**3.2. Average Acceleration and Acceleration Standard Deviation.** Acceleration is the representation of the driver's

speed control operating the gas pedal or brake pedal and is closely associated with characteristic driving behavior. If a driver applies the gas pedal or brake pedal heavily and frequently, they are more likely to be considered an aggressive driver [20]. Furthermore, although the second derivative of the vehicle's speed, or 'jerk', has been recognized as a risk indicator [21], it is not included in this paper. The average and standard deviation of the positive acceleration and negative acceleration (deceleration) are calculated based on the following equations:

$$a_i = \frac{v_{i+1} - v_i}{3.6 * \Delta t}, \quad (3)$$

$$a^+ = \frac{1}{p} \sum_{i=1}^p a_i^+, \quad (4)$$

$$a_s^+ = \sqrt{\frac{1}{p} \sum_{i=1}^p (a_i^+ - a^+)^2}, \quad (5)$$

$$a^- = \frac{1}{q} \sum_{i=1}^q a_i^-, \quad (6)$$

$$a_s^- = \sqrt{\frac{1}{q} \sum_{i=1}^q (a_i^- - a^-)^2}, \quad (7)$$

where  $a_i$  ( $i=1, 2, \dots, n$ ) is the derivative of the vehicle's speed in the unit of  $m/s^2$  and  $\Delta t$  is the time interval of sampling (set to 1 s). In addition,  $a_i$  is classified into positive sequence ( $a_1^+, a_2^+, \dots, a_p^+$ ) and negative sequence ( $a_1^-, a_2^-, \dots, a_q^-$ ), with the absolute value above  $0.01 m/s^2$  ( $p + q \leq n$ ).

**3.3. Event Rate and Duration of Acceleration or Deceleration Exceeding a Certain Value.** Generally, the more abruptly acceleration or hard braking occurs, the more risky the driver will be [22]. The available threshold reported in the literature for abrupt acceleration or hard braking is varied [23]. The threshold of abrupt acceleration or hard braking is acknowledged to be  $\pm 0.3g$  by Mortimer et al. [24] and Anderson & Baldock [25], while it is set at  $\pm 0.5g$  by Fazeen et al. [16]. Deceleration (in absolute value) exceeding  $0.49g$  [26] or greater than  $0.4g$  in the euroFOT project [27] is regarded as hard braking, while deceleration above  $0.15g$  is considered a moderate risk level [28]. Bergasa et al. [29] used the threshold of  $0.4g$  in both longitudinal and lateral directions to identify risky driving behavior and utilized the frequency of risky events of a different level to determine the risk score of drivers. In contrast, the threshold for abrupt acceleration or hard braking used in some studies is  $0.1g$  [30],  $0.0875g$  [31], and  $0.15g$  [32]. In addition, according to the findings in Zeeman and Booyens [33], Eboli et al. [34], and Eboli et al. [35], speed or road condition should be considered for characterizing drivers' behavior and the acceleration threshold is found to be reduced with an increase in speed. Also, the threshold range for positive acceleration on city roads in Brussels, as defined in Vlieger et al. [36], is  $0.85\text{--}1.10 m/s^2$

for aggressive driving. Based on this, in this case, the threshold of  $0.12g$ ,  $0.1g$ , and  $0.0875g$  is used for abrupt acceleration or hard braking (in absolute value) in three different kinds of routes: urban road, city ring road, and expressway, respectively.

Correspondingly, the average duration of the acceleration or deceleration threshold violated is considered along with the event rate in this paper. The event of abrupt acceleration (or deceleration) exceeding a certain value is defined as follows: only if the acceleration is above (or below) the identified threshold in different kinds of routes and lasts for a continuous period of time  $T$  ( $T \geq 2$  s), it will be counted as one abrupt acceleration (or hard braking incidence). The cumulative number is denoted as  $n_i^+$  (or  $n_i^-$ ). Likewise, the average duration of the acceleration or deceleration exceeding a certain value in different kinds of routes is described as  $T^+$  or  $T^-$ . Furthermore, the event rate is calculated as average number of acceleration or deceleration exceeding a certain value in different kinds of routes per 10 km traveled:

$$R^+ = \frac{\text{Number of Events (positive)}}{\text{Distance Traveled (10km)}} = \frac{1}{3} \left( \frac{n_1^+}{d_1} + \frac{n_2^+}{d_2} + \frac{n_3^+}{d_3} \right), \quad (8)$$

$$R^- = \frac{\text{Number of Events (negative)}}{\text{Distance Traveled (10km)}} = \frac{1}{3} \left( \frac{n_1^-}{d_1} + \frac{n_2^-}{d_2} + \frac{n_3^-}{d_3} \right), \quad (9)$$

$$T^+ = \frac{1}{3} \left( \frac{T_{p_1}}{n_1^+} + \frac{T_{p_2}}{n_2^+} + \frac{T_{p_3}}{n_3^+} \right), \quad (10)$$

$$T^- = \frac{1}{3} \left( \frac{T_{q_1}}{n_1^-} + \frac{T_{q_2}}{n_2^-} + \frac{T_{q_3}}{n_3^-} \right), \quad (11)$$

where  $T_{p_i}$  and  $T_{q_i}$  represent the cumulative time of acceleration and deceleration exceeding a certain value in different kinds of routes.

**3.4. Maximum Speed, Acceleration, and Deceleration.** The maximum speed  $v_{\max}$ , acceleration  $a_{\max}^+$ , and deceleration  $a_{\max}^-$  (in absolute value) are selected from durations lasting for three seconds at least to avoid a sudden extreme value.

The values of GPS data and derivative indicator mentioned above are provided in Table 1.

## 4. Methodology

**4.1. Participants and Routes.** A total of 50 drivers (three females among them) of an age group from 27 to 50 years old were recruited to participate in the experiment. The average driving experience of the participants was 14.0 year with a standard deviation of 8.3 years. All participants were asked to drive as they usually do in everyday driving, and the test session lasted for 90–110 minutes for each participant. Additionally, the dataset used here was collected using an instrumented vehicle in real traffic between Huzhou and Changxing and

TABLE 1: GPS data and derivative indicator.

Driver	$v_a$	$a^+$	$a_s^+$	$a^-$	$a_s^-$	$R^+$	$T^+$	$R^-$	$T^-$	$v_{\max}$	$a_{\max}^+$	$a_{\max}^-$
1	64.7	0.23	0.23	0.26	0.31	3.04	1.99	4.35	2.19	109.7	1.62	2.36
2	63.8	0.27	0.25	0.31	0.33	3.22	2.36	4	2.69	117.6	1.88	2.40
3	63.9	0.25	0.24	0.31	0.34	3.16	2.18	5.62	2.35	129.9	1.71	2.15
4	58.2	0.19	0.18	0.22	0.27	1.06	2.03	3.03	2.62	104.6	1.12	2.88
5	61.9	0.28	0.23	0.31	0.29	2.82	1.95	4.6	2.41	129.3	1.55	2.50
6	64	0.23	0.21	0.27	0.28	1.24	1.88	3.69	1.86	106.2	1.38	2.33
7	71.4	0.28	0.29	0.32	0.43	3.53	3.06	3.91	3.12	114.7	1.80	3.37
8	63.7	0.26	0.27	0.29	0.29	2.98	2.55	3.24	2.37	119.7	2.15	1.92
9	64.3	0.28	0.23	0.33	0.33	1.95	1.87	4.16	1.98	131.6	1.82	2.76
10	64.2	0.26	0.21	0.32	0.36	0.93	2.18	3.79	2.37	122.9	1.20	2.73
$\vdots$	$\vdots$	$\vdots$										
41	62.4	0.34	0.29	0.42	0.47	3.38	2.85	5.28	3.29	117.8	1.57	2.83
43	84.1	0.20	0.23	0.20	0.22	2.15	1.87	3.62	2.39	135.7	1.78	1.90
44	76.6	0.14	0.16	0.17	0.26	1.6	2.2	2.36	1.42	132.5	0.89	2.06
45	60.1	0.36	0.34	0.40	0.42	0.18	1.5	1.67	1.96	109.7	1.99	2.66
46	59.9	0.30	0.24	0.35	0.33	3.25	3.77	4.12	3.07	123.1	1.37	2.13
47	61.9	0.33	0.27	0.45	0.45	2.03	2.67	4.5	2.86	114.4	1.72	2.46
48	65.4	0.30	0.26	0.35	0.39	2.67	2.25	4.08	2.69	113.7	1.32	3.19
49	64.1	0.31	0.32	0.38	0.41	1.94	2.1	3.56	3.24	121.7	1.96	2.37
50	60.5	0.35	0.28	0.44	0.51	3.03	3.2	4.3	1.96	114.2	1.69	3.40

between Yanta and Lantian in China, with similar road conditions at both test sites. The participants mainly drove three kinds of routes at both test sites: urban road, city ring road, and expressway, with the speed limit ranging from 40 to 50 km/h, 70 to 90 km/h, and 100 to 110 km/h, respectively. The test routes between Yanta and Lantian are provided as an example in Figure 1 using map-matching, where urban road, city ring road, and expressway are marked in green, yellow, and red, respectively. In addition, traffic jam and congested roads rarely exist in the expressway session, but several traffic lights and crossroads are present on the city ring road.



FIGURE 1: Test routes between Yanta and Lantian.

**4.2. Apparatus.** Tests were conducted to evaluate the accuracy of the speed and acceleration using GPS devices as outlined by Zito et al. [37]. The position and speed of a vehicle can be obtained by modern GPS devices with a much higher accuracy however. In this test, the speed of the subject vehicle was extracted using a Differential Global Positioning System (DGPS) with an accuracy of 0.1 km/h (resolution of 0.01 km/h), and the data used were recorded with a sampling rate of 1 HZ. Also, this test provides an opportunity to explore the validity of GPS measurements. Along with GPS, acceleration data from inertia sensor are also available. Figure 2 shows GPS and inertia measurements from a single trip by time. Visually, these two data streams seem to correlate well ( $r = .68$ ). In addition, six cameras installed on the instrumented vehicle (Figure 3) were used to record the front view, rear view, two side views, the driver's face, and the pedal area. The videos recorded by these cameras were mainly used to pick out typical risky events (the relevant extracting rules will be introduced in Section 4.3). Note that the purpose of this research is to detect characteristic risky events and identify risky drivers with the data extracted from

GPS, and in addition to GPS data and derivative indicator, inertia measurements (for the comparison of accuracy), and radar data (for extracting risky situations), together with video verification, were also used as an assistance to GPS identification.

**4.3. Risky Event Extraction.** According to the findings of a risky driving report [38] on the 100-Car Naturalistic Driving Study (lasting for 12 to 13 months with 109 drivers participating), the number of crashes and near crashes was 69 and 761, respectively. In comparison, crash and near crash events rarely exist in the dataset of this paper. Also in many cases, crashes or near crashes may not actually happen due to the prompt reaction of traffic participants from both sides to avoid them (e.g., pedestrian stopping first before entering the roadway or another vehicle taking evasive actions). Therefore, it is also of great importance to detect and record those potential risky situations or events to improve the overall traffic safety, as frequent risky maneuvers or events may lead to crashes or near crashes at any time.

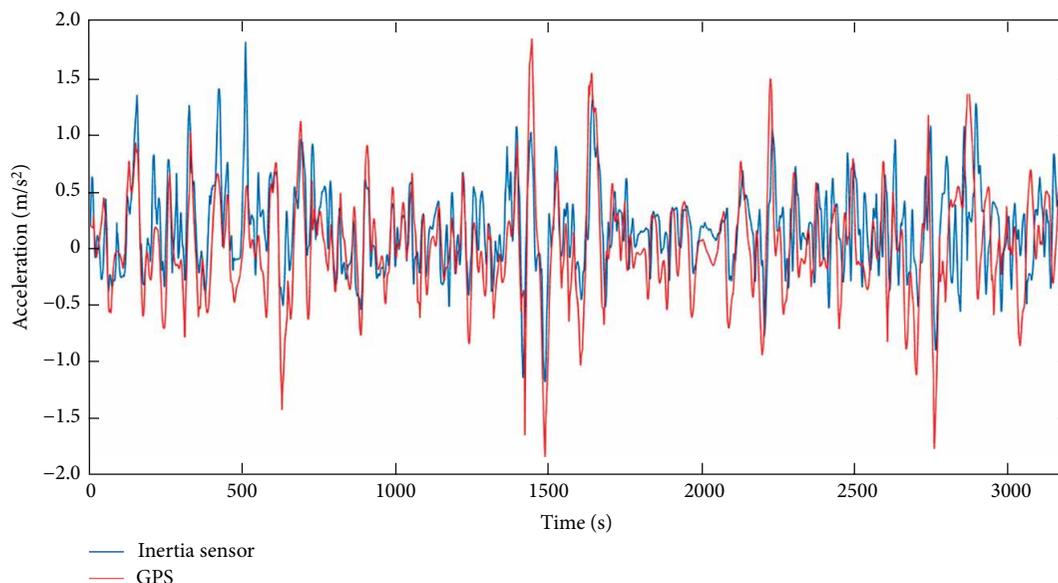


FIGURE 2: Acceleration data from GPS device and inertia sensor.



FIGURE 3: Instrumented vehicle.

For the purpose of extracting potential risky situations, one data reductionist watched videos from the front-view camera and manually extracted all risky events from the 50 participants according to basic rules. Risky events include: if a vehicle cuts in suddenly with a longitudinal gap lower than 10 meters, the subject vehicle’s speed is simultaneously greater than 20 km/h, following too closely to the lead vehicle frequently in urban areas, running on the line, no evasive action or reacting too late, other vehicles bursting into view at poor-visibility intersections, driving through an intersection without traffic lights, failing to notice someone at the pedestrian crossings, failing to give way to vehicles with priorities,

TABLE 2: KMO and Bartlett’s test.

Kaiser–Meyer–Olkin Measure of Sampling Adequacy.		.774
Bartlett’s Test of Sphericity	Approx. Chi-square	706.76
	df	66
	Sig.	.000

zigzag driving, and abrupt turning or hard braking [21, 38–40].

After extracting the risky event video segments from 50 drivers, ten volunteers (five professors, five PhD students) mainly engaged in traffic safety research were recruited to evaluate the severity level of those risky events. They were asked to complete a questionnaire after viewing the video segments extracted from 50 drivers, in a process that took nearly four hours (two sessions were carried out on consecutive days) for each volunteer to complete the evaluation. The risk level ranged from 1 to 3, with 1 indicating the least risky driving behavior and 3 indicating a most risky driving behavior in the test. Note that although some near crashes and one minor crash occurred in this test, the authors believe that most of the risky events in level 3 corresponded to the severity of incidents described in the risky driving report [38]. Following this, the output from the questionnaires was compared with the risk scores calculated through factor analysis. The factor analysis was conducted with SPSS 19.0 software.

## 5. Results

5.1. *Factor Extraction and Analysis.* Before conducting a factor analysis, it is necessary to check whether this method is appropriate for the dataset used here. Note that the measure of Kaiser–Meyer–Olkin (KMO) could verify the sampling adequacy for the factor analysis. In this analysis, KMO = .774 (Table 2), which is greater than .70 and well

TABLE 3: Correlation coefficients among GPS data and derivative indicators.

	$v_a$	$a^+$	$a_s^+$	$a^-$	$a_s^-$	$R^+$	$T^+$	$R^-$	$T^-$	$v_{\max}$	$a_{\max}^+$	$a_{\max}^-$
$v_a$	1.0	-	-	-	-	-	-	-	-	-	-	-
$a^+$	-.29	1.0	-	-	-	-	-	-	-	-	-	-
$a_s^+$	-.20	.92	1.0	-	-	-	-	-	-	-	-	-
$a^-$	-.34	.97	.86	1.0	-	-	-	-	-	-	-	-
$a_s^-$	-.26	.83	.80	.91	1.0	-	-	-	-	-	-	-
$R^+$	-.17	.82	.89	.73	.65	1.0	-	-	-	-	-	-
$T^+$	-.08	.72	.83	.69	.71	.67	1.0	-	-	-	-	-
$R^-$	-.31	.80	.75	.78	.71	.80	.57	1.0	-	-	-	-
$T^-$	-.20	.42	.38	.49	.60	.28	.44	.31	1.0	-	-	-
$v_{\max}$	.59	.21	.18	.15	.04	.25	.12	.29	-.2	1.0	-	-
$a_{\max}^+$	-.20	.66	.75	.59	.50	.82	.58	.71	.14	.26	1.0	-
$a_{\max}^-$	-.10	.54	.56	.55	.70	.54	.52	.65	.41	.15	.46	1.0

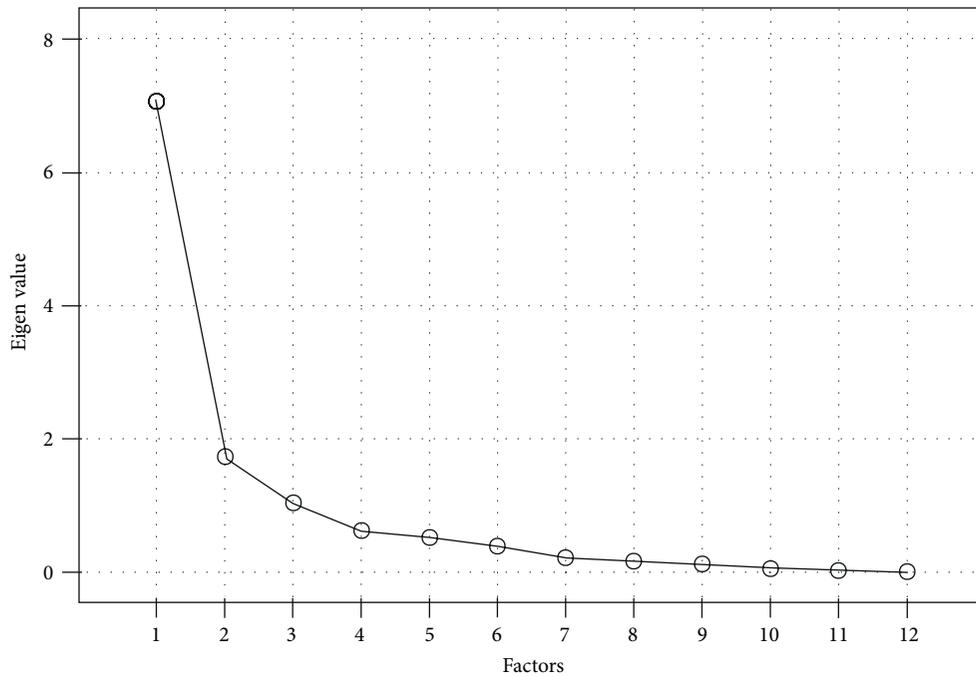


FIGURE 4: Scree plot.

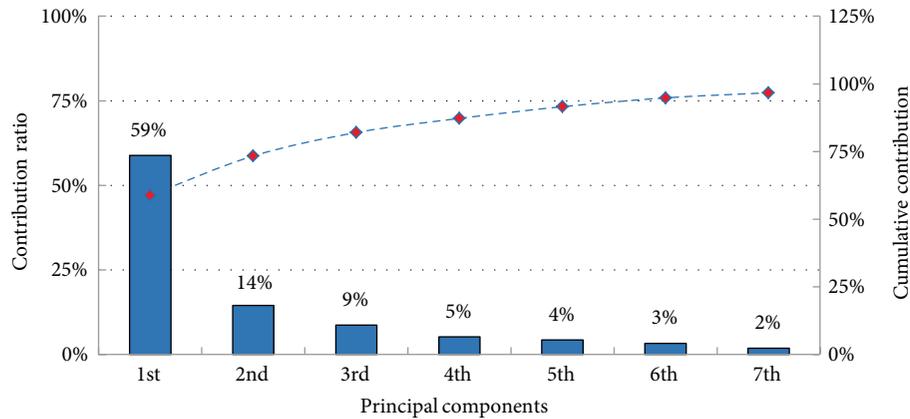


FIGURE 5: Contribution ratio of principal components.

TABLE 4: Contribution ratio before and after rotation.

Component	Principal Component Analysis			Factor Analysis			
	Total	Contribution ratio (%)	Cumulative contribution (%)	Component	Total	Contribution ratio (%)	Cumulative contribution (%)
PC1	7.067	58.892	58.892	RC1	5.589	46.577	46.577
PC2	1.735	14.456	73.348	RC2	2.621	21.844	68.421
PC3	1.040	8.668	82.015	RC3	1.631	13.594	82.015

TABLE 5: Factor loading matrix.

Variables	Principal Component Analysis			Factor Analysis		
	PC1	PC2	PC3	RC1	RC2	RC3
$v_a$		<b>0.776</b>	0.477			<b>0.908</b>
$a^+$	<b>0.943</b>			<b>0.839</b>	0.431	
$a_s^+$	<b>0.946</b>			<b>0.865</b>		
$a^-$	<b>0.927</b>			0.523	<b>0.766</b>	
$a_s^-$	<b>0.89</b>				<b>0.715</b>	
$R^+$	<b>0.884</b>			<b>0.909</b>		
$T^+$	<b>0.808</b>			<b>0.63</b>	0.529	
$R^-$	<b>0.87</b>			0.580		
$T^-$	0.499	-.432	0.61		<b>0.880</b>	
$v_{\max}$		<b>.899</b>				<b>0.858</b>
$a_{\max}^+$	0.765		-.401	<b>0.891</b>		
$a_{\max}^-$	0.70			0.485	<b>0.567</b>	
Interpretation	Acceleration & deceleration	Speed	None	Acceleration	Deceleration	Speed

Note: Factor loading below .4 appears in blank for brevity, and factor loading in bold indicates strong correlation.

above the acceptable threshold of .5 [41]. In addition, the results of the sphericity test (Table 2) illustrate that the significance level is close to zero, showing that it is suitable to conduct a factor analysis. The Pearson correlation coefficients among all pairs of variables are provided in Table 3. One can see that these variables are not completely independent of each other and some variables are even strongly correlated with other features. Therefore, it is appropriate to reduce the dimension of variables through factor analysis. The 12 original variables were thus reduced into a subset of principal components (PCs) for better interpretation by using their combinations.

Communality indicates the proportion of common variance within each variable that could be explained by the extracted factors. According to the factor analysis results, the communalities of the 12 variables are all greater than .7, so the factors extracted through factor analysis could reflect most of the information about the original variables, and Kaiser's criterion, the SPSS software default for extracting factors, is suitable [41].

In addition, considering the eigenvalues represent the amount of variation that could be explained by a factor, in theory a larger eigenvalue should be retained and Kaiser [42] recommends retaining factors with eigenvalues greater than 1. Kaiser's criterion is based on the idea that an eigenvalue of 1 represents a substantial amount of variation, and the same criterion is used in this analysis. Figure 4 shows the eigenvalues in the descending order through a scree plot, and the contribution ratio of several principal components (or factors)

together with a cumulative curve is shown in Figure 5. Considering the eigenvalues should be greater than 1 and the cumulative contribution was above 80 percent within three factors, the three factors were retained. Furthermore, each contribution ratio of the first three factors before and after rotation is depicted in Table 4.

One can observe that although the same cumulative contribution is achieved for the principal component analysis and factor analysis, the contribution ratio of each component became more balanced after rotation. Note that rotation components (RCs) obtained by orthogonal rotation were more explanatory than PCs and could explain most of the original variables from different perspectives, which is shown in Table 5. A factor loading matrix obtained from the principal component analysis and factor analysis is also included in Table 5. According to the factor loading matrix, the variables that clustered on the same factor suggest that factor RC1 mainly represents acceleration information, factor RC2 represents deceleration information, and factor RC3 denotes speed-related information.

*5.2 Comparison of Driving Risk Scores and Outputs from Questionnaires.* In this section, the factor scores are calculated using the formula shown in Eq. (12).

$$F_j = \sum_{i=1}^n a_{ij}x_{ij}, \quad (j = 1, 2, \dots, m), \quad (12)$$

TABLE 6: Risk score obtained through factor analysis.

Driver	F1	F2	F3	Score	Driver	F1	F2	F3	Score
1	.25	-.63	-.35	-0.069	26	1.80	-1.90	-1.06	0.280
2	0.40	0.01	-0.07	0.179	27	2.29	0.12	1.12	1.245
3	0.72	-0.62	0.40	0.256	28	0.01	-0.90	-0.36	-0.241
4	-1.09	0.48	-0.85	-0.519	29	0.84	-0.03	0.13	0.406
5	0.41	-0.46	0.29	0.127	30	1.59	-0.84	0.42	0.615
6	-0.21	-0.72	-0.56	-0.333	31	-0.95	0.11	1.26	-0.247
7	0.12	1.83	0.64	0.541	32	-1.69	-1.40	0.80	-0.987
8	0.70	-0.86	-0.06	0.127	33	-0.17	0.92	1.47	0.323
9	0.59	-0.67	0.57	0.207	34	-0.31	1.26	-0.09	0.117
10	-0.45	0.56	0.28	-0.05	35	-2.00	-1.02	1.10	-1.004
11	-0.35	-0.47	-1.92	-0.53	36	-1.90	0.53	0.19	-0.745
12	0.10	1.51	0.93	0.505	37	1.08	1.76	-0.02	0.884
13	-1.24	0.31	-2.13	-0.799	38	-0.17	1.06	-1.09	0.004
14	-0.61	-0.66	-0.43	-0.485	39	-2.09	0.21	1.00	-0.792
15	0.09	-0.87	-0.82	-0.260	40	-1.74	-0.92	1.47	-0.813
16	0.53	-0.30	-0.30	0.139	41	0.43	1.86	-0.09	0.594
17	-0.84	-0.19	-1.20	-0.597	42	0.30	-0.40	-0.66	-0.037
18	0.06	0.27	0.33	0.131	43	0.03	-1.64	2.18	-0.050
19	1.74	-0.03	1.37	0.989	44	-1.50	-0.38	1.56	-0.573
20	0.33	-0.92	0.19	-0.020	45	0.84	1.37	-0.54	0.617
21	0.27	-0.46	-0.28	-0.011	46	0.04	0.56	-0.13	0.124
22	0.24	-0.98	-1.12	-0.26	47	0.49	0.83	-0.43	0.349
23	0.37	0.28	0.79	0.342	48	-0.57	1.90	0.03	0.153
24	-1.16	-0.27	-3.15	-1.027	49	1.25	-0.3	0.17	0.536
25	0.82	-1.24	-0.71	0.012	50	0.34	2.37	-0.26	0.641

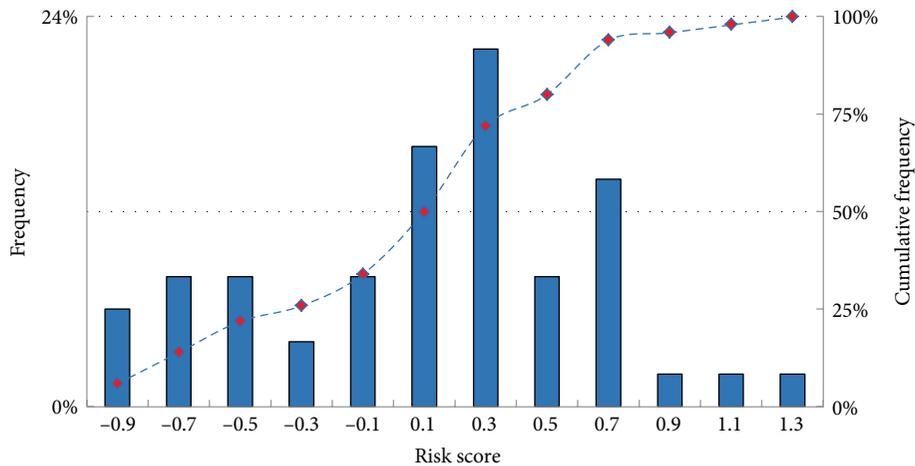


FIGURE 6: Distribution of risk score from 50 drivers through factor analysis.

where  $n$  is the number of initial variables, which is 12 in this test. The number of principal factors retained is  $m$ , which is 3 in the study. In addition,  $a_{ij}$  stands for the value of GPS data or derivative indicator (listed in Table 1 through standardization and  $x_{ij}$  represents the component score coefficient matrix calculated by SPSS software automatically, which is useful in understanding how the factor score has been computed. The factor score was recorded in three new columns of data (one for each rotation factor RC) labelled  $F1$ ,  $F2$ , and  $F3$ ,

respectively. The risk score was calculated (Table 6) based on the following equation:

$$Z = \sum_{j=1}^m \alpha_j F_j = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3, \quad (m = 3), \quad (13)$$

where  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  represent the contribution ratio of variance after rotation with the value of .4658, .2184, and .1359, respectively (Table 4). The higher the sum score is, the more

TABLE 7: One-sample Kolmogorov–Smirnov test.

		Risk score
N		50
Normal parameters <sup>a,b</sup>	Mean	.00000
	Std. Deviation	.532106
Most Extreme Differences	Absolute	.109
	Positive	.059
	Negative	-.109
Kolmogorov–Smirnov Z		.769
Asymp. Sig. (two-tailed)		.595

<sup>a</sup>Test distribution is normal. <sup>b</sup>Calculated from data.

TABLE 8: The scores of risky events evaluated by volunteers.

Event	Volunteer									
	1	2	3	4	5	6	7	8	9	10
1	2	2	2	3	2	2	2	2	3	2
2	2	3	3	2	2	2	2	2	2	2
3	1	2	1	1	1	1	1	1	1	1
4	1	1	2	1	1	1	1	1	1	1
5	2	2	1	1	2	2	2	1	2	2
6	1	2	1	1	1	1	2	1	2	1
7	1	1	1	1	1	1	1	1	1	1
8	1	2	1	2	1	1	1	1	2	1
9	1	1	1	1	1	1	1	1	1	1
10	3	3	2	3	3	3	3	2	3	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
420	1	2	1	1	1	1	1	1	2	1
421	2	2	2	2	2	2	2	2	3	2
422	3	3	3	3	3	3	3	3	3	3
423	1	1	1	1	1	1	1	1	2	1
424	1	2	1	1	1	1	1	1	2	1
425	1	2	1	1	1	1	1	1	2	1
426	1	2	2	1	1	1	1	1	2	1
427	1	1	1	1	1	1	1	1	1	1
428	1	2	1	1	1	1	1	1	1	1
429	1	2	2	1	2	1	1	1	1	1

risky the driver will be. Figure 6 shows the distribution of the risk score obtained through factor analysis. Nonparametric estimation reveals that the risk score obeys the law of normal distribution (Table 7). One can see from Table 7 that the  $P$  value (.595) is greater than 0.05. Therefore, the null hypothesis that the risk score conforms to the normal distribution is accepted.

In addition, as mentioned in Section 4.3, ten volunteers (five professors and five PhD students) mainly engaged in traffic safety research were recruited to evaluate the severity level of risky events from 50 participants by viewing video segments. A risky event level was assigned to each event with a score ranging from 1 to 3 accordingly. The scores of risky events evaluated by volunteers are shown in Table 8. Given that Kendall’s coefficient of concordance (also known as Kendall’s  $W$ ) is a measure of agreement among the judges, when running the Kendall’s  $W$  test, the value of Kendall’s  $W$

was 0.615 and the  $P$  value came out to be significant (.000), which means that there is a good agreement among the scores evaluated by volunteers [43]. In order to get a much lower variance, for each risky event, an average score was taken from the ten volunteers after removing the maximum and minimum scores and then rounded to the nearest whole number. The number of risky events with a different level (score) and the sum score are shown in Table 9.

To compare whether the score acquired through factor analysis and evaluated by volunteers is highly correlated, the risk score obtained through factor analysis (Figure 7) is normalized and evaluated by volunteers in a range from 1 to 5 (the bigger the value is, the more risky a driver will be), respectively, without changing the original distribution of both scores. The correlation between the volunteers’ score and the score obtained through factor analysis is illustrated in Figure 8 with a Pearson’s correlation coefficient of 0.52. This result indicates that the risk score obtained through factor analysis has a relatively high correlation to the score evaluated by volunteers. Furthermore, when the score evaluated by volunteers is taken as a benchmark, it can be observed that the score obtained through factor analysis is relatively consistent with the volunteer’s score at the high-risk level ( $\geq 4$ ), rather than the low-risk level ( $\leq 3$ ). Considering the aim and focus of similar research is to identify high-risk drivers rather than low-risk ones, it is feasible to determine risky drivers by using the approach described here.

**5.3. Driving Risk Model Establishment.** After the risk level of each driver was identified through factor analysis and then verified by the volunteer’s categorization, a binary logistic regression model was developed to evaluate the probability of being a high-risk driver based on the RC factors.

Research has shown that there is a significant correlation between speed-related risk indices and historical accident data [5]. To this end, the accident record of each driver during last three years was considered as the dependent variable. The model setup is as follows. Define

$$Y_i = \begin{cases} 1 & \text{If driver } i \text{ has a history of accidents } (\geq 1) \text{ in the last 3 years} \\ 0 & \text{Otherwise.} \end{cases} \quad (14)$$

The observed  $Y_i$  has two categories: driver with and without a record of car accident during last 3 year, which is assumed to follow a Bernoulli distribution.

$$Y_i \sim \text{Bernoulli}(p_i). \quad (15)$$

Let  $p_i$  be the probability of being a high-risk driver for driver  $i$ . This probability is associated with a set of continuous covariates by a logit link function,

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_i X_i, \quad (16)$$

where  $\beta_i$  is the vector of regression parameters and  $X_i$  is the matrix of RC factors (predictors) for driver  $i$ . A driver will be predicted as a high-risk one if  $p_i$  is above a predefined threshold  $p_0$  (0.5 by default).

TABLE 9: Number of risky events in different levels (score) and the sum score.

Driver ID	Level 1 (1)	Level 2 (2)	Level 3 (3)	Risk score	Driver ID	Level 1 (1)	Level 2 (2)	Level 3 (3)	Risk score
1	1	6	0	13	26	7	4	1	18
2	3	4	0	11	27	10	5	0	20
3	4	6	1	19	28	9	2	0	13
4	1	4	0	9	29	6	7	0	20
5	2	5	2	18	30	4	4	1	15
6	3	6	0	15	31	4	4	0	12
7	7	8	1	26	32	4	0	1	7
8	6	5	0	16	33	5	4	0	13
9	5	4	0	13	34	2	3	0	8
10	3	6	2	21	35	2	4	0	10
11	3	5	0	13	36	3	1	0	5
12	7	7	3	30	37	1	3	1	10
13	4	3	0	10	38	9	1	1	14
14	5	2	0	9	39	4	2	0	8
15	5	3	0	11	40	1	4	0	9
16	8	3	0	14	41	3	6	1	18
17	2	5	1	15	42	2	4	1	13
18	9	6	1	24	43	2	0	0	2
19	11	9	0	29	44	0	2	0	4
20	4	2	0	8	45	2	2	0	6
21	4	5	1	17	46	2	6	1	17
22	2	4	0	10	47	1	6	0	13
23	4	4	0	12	48	2	0	0	2
24	4	1	1	9	49	5	5	0	15
25	1	5	2	17	50	10	0	1	13

TABLE 10: Binary logistic regression model output.

Effect	Parameter estimate	Sig.	Odds ratio	95% Odds ratio confidence	
				Lower limit	Upper limit
RC_1	2.298	.012**	9.957	1.653	59.959
RC_2	.923	.032**	2.516	1.084	5.838
RC_3	1.366	.056*	3.921	.968	15.886
Intercept	-2.502	.001	.082	-	-

Note: \* $p < .1$ , \*\* $p < .05$ .

The null hypothesis of the model is that the RC factors have no effect on model outputs, and model outputs are listed in Table 10. One can see from Table 10 that all three RC factors can improve the model significantly when using a significance level of 0.1 and variables 'RC\_1' and 'RC\_2' are significant predictors when using 0.05 as a significance level. Note that the best effect size to use in the context of logistic regression is the odds ratio. One can also see from Table 10, that the odds ratios of three RC factors are all greater than 1, which indicates that as the predictor increases, the odds ratio of the outcome occurring increases. Finally, the predictive model is developed as follows:

$$\begin{aligned}
 & \text{Probability}(\text{high} - \text{risk driver}) \\
 &= \frac{\exp(-2.502 + 2.298 \times RC\_1 + 0.923 \times RC\_2 + 1.366 \times RC\_3)}{1 + \exp(-2.502 + 2.298 \times RC\_1 + 0.923 \times RC\_2 + 1.366 \times RC\_3)}, \tag{17}
 \end{aligned}$$

where RC\_1 represents acceleration information, RC\_2 represents deceleration information, and RC\_3 denotes speed-related information.

### 6. Experimental Evaluation

In this section, the method of Symbolic Aggregate Approximation (SAX) proposed by Lin et al. [44] is used to evaluate and validate the results of the case study. As a time-series analysis method, SAX is commonly used to reduce dimensions by converting the continuous values to Piecewise Aggregate Approximation (PAA) representations. Furthermore, SAX also supports data visualization and allows distance measurement for discrete characters that lower the bound rather than the distance measurement on the original time series [45].

To this end, speed and acceleration data used in this paper were analyzed again with the SAX algorithm and the major notation concerning this algorithm is summarized in Table 11. Following the steps of the SAX algorithm, the time series  $L = l_1, \dots, l_n$  must be transformed into a string of length  $w$  ( $w < n$ , typically  $w \ll n$ ). Time-series data with the length of  $n$  were first normalized with a mean of zero and a standard variance of one before being converted into symbolic representation, denoted as  $L' = l'_1, \dots, l'_n$ . Next, the standardized time series was further transformed into a  $w$ -dimensional space by a vector  $\bar{L} = \bar{l}_1, \dots, \bar{l}_w$  on the basis of PAA method. The  $k$ th element of  $\bar{L}$  was calculated by the following equation:

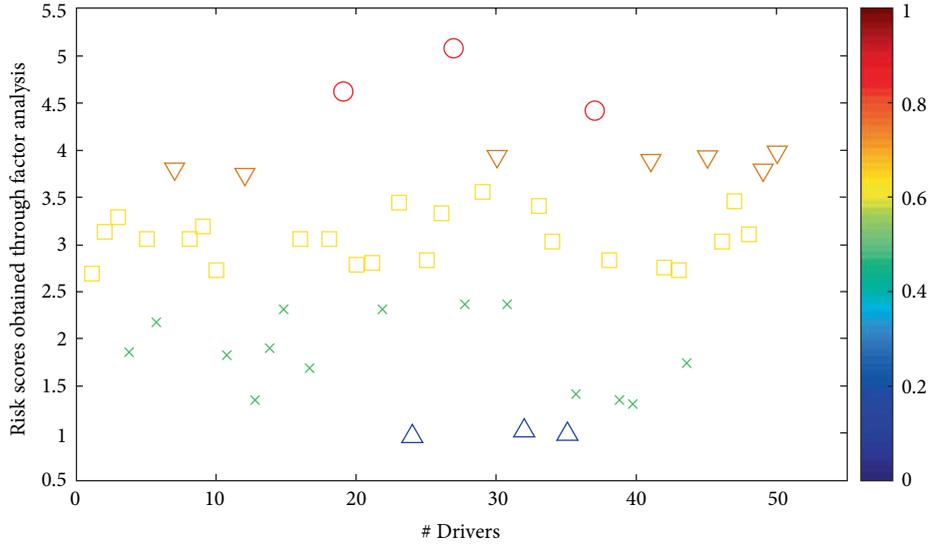


FIGURE 7: Risk score of 50 drivers obtained through factor analysis in the new range from 1 to 5.

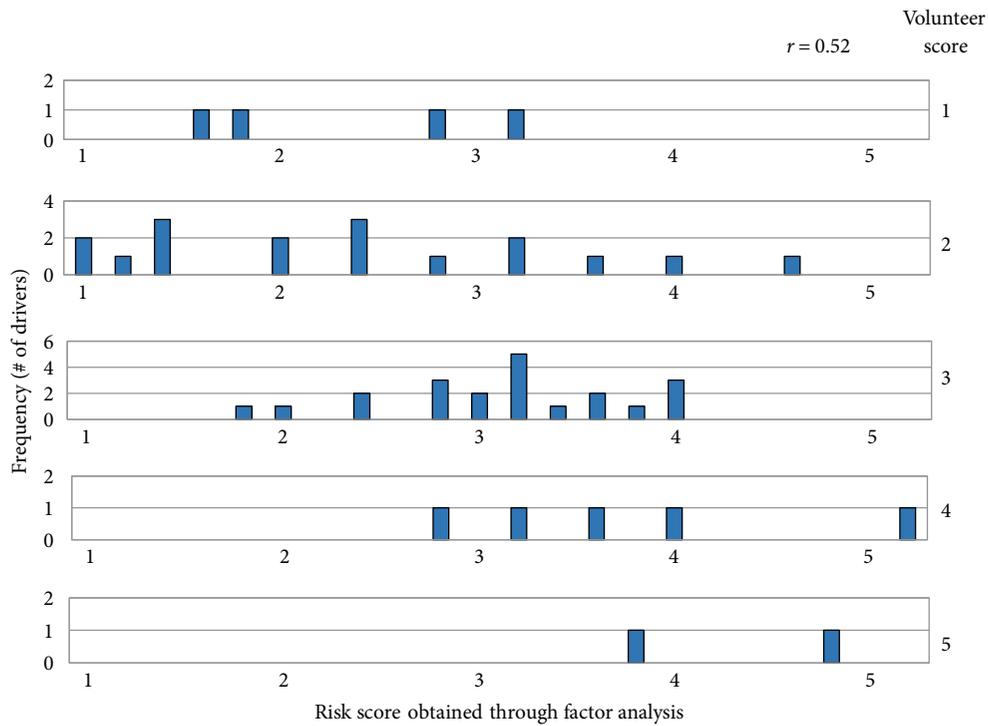


FIGURE 8: Correlation between risk score obtained through factor analysis and volunteer's score.

$$\bar{l}_k = \frac{w}{n} \sum_{t=1+n(k-1)/w}^{nk/w} l_t. \quad (18)$$

In other words, to transform an original time series  $L$  into  $w$  dimensions, the data are converted to  $w$ -equal sized sections under the Gaussian curve yielding breakpoints. The PAA representation is then mapped into symbols (a discrete string) with reference to the breakpoint lookup table (see Table 12).

Further analysis could be conducted on the symbolic representation instead of the original series.

In this approach, both speed and acceleration data with the length of 1000 s were taken as an example to briefly illustrate the SAX algorithm. After normalization, the time series of speed and acceleration were converted from  $[0, 70]$  km/h to  $[-2, 3]$  and from  $[-2, 1]$  m/s<sup>2</sup> to  $[-6, 4]$ , respectively. The normalized speed and acceleration data were then transformed into 10 and 20 segments based on the PAA method.

TABLE 11: Summarization of the notation used in this case.

Parameter	Notation	Example	
		Speed (km/h)	Acceleration (m/s <sup>2</sup> )
$L$	Time series $L=l_1, \dots, l_n$	$L_v$	$L_a$
$n$	Length of a time series, $n$	$n_v=1000$ s	$n_a=1000$ s
$L'$	Standardized time series $L'=l'_1, \dots, l'_n$	$L'_v$	$L'_a$
$\bar{L}$	PAA of a time series $\bar{L}=\bar{l}_1, \dots, \bar{l}_w$	$\bar{L}_v$	$\bar{L}_a$
$w$	Number of PAA segments representing time series $L$ ( $w \ll n$ )	$w_v=10$	$w_a=20$
$\hat{L}$	Symbol representation of a time series $\hat{L}=\hat{l}_1, \dots, \hat{l}_w$	$\hat{L}_v$	$\hat{L}_a$
$\alpha$	Alphabet size (e.g., for the alphabet = { $a, b, c, d$ }, $\alpha=4$ )	$\alpha_v=4$	$\alpha_a=4$

TABLE 12: Breakpoint of dividing character  $\alpha$  (from 2 to 10).

$\beta$	$\alpha$								
	2	3	4	5	6	7	8	9	10
$\beta_1$	0	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
$\beta_2$	—	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
$\beta_3$	—	—	0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
$\beta_4$	—	—	—	0.84	0.43	0.18	0	-0.14	-0.25
$\beta_5$	—	—	—	—	0.97	0.57	0.43	0.14	0
$\beta_6$	—	—	—	—	—	1.07	0.67	0.43	0.25
$\beta_7$	—	—	—	—	—	—	1.15	0.76	0.52
$\beta_8$	—	—	—	—	—	—	—	1.22	0.84
$\beta_9$	—	—	—	—	—	—	—	—	1.28

Finally, these segments were replaced with four letters ( $a, b, c$ , and  $d$ ), and the original speed and acceleration data were reduced to symbols with the representation of  $aaccbacddd$  and  $cbbcbcbcbcbcbcbcbcc$  (Figure 9).

As such, the speed and acceleration data from 50 drivers were sorted out and encoded as a whole with a sequence from 1 to 50. The length  $n$  of each original time series is nearly 1.5 h to 2 h. Given that the dataset from 50 drivers is a high number when using the representation of the original time series as depicted in Figure 9, the distribution of the speed and acceleration data from the drivers was displayed instead (Figure 10). In addition, a discrete string with the alphabet size of nine was selected for each distribution with the SAX algorithm.

To the authors' knowledge, a driver is more prone to be considered as a high-risk driver when applying the gas pedal or brake pedal heavily and frequently, especially at a higher running speed. This study evaluates whether a driver is risky or not in this regard, though it does not mean the scenarios in which a driver is operating in lower speed conditions (i.e., urban roads) are considered safe, however.

According to the symbolic representation of the speed and acceleration data from 50 drivers depicted in Figure 10, a vehicle's speed greater than 88 km/h, with the representation of  $h$  or  $i$ , together with acceleration or deceleration (in absolute value) exceeding  $0.5 \text{ m/s}^2$  with representation of  $a$  or  $i$ , was taken into consideration; that is, the symbolic representation of  $ha, hi, ia$ , or  $ii$  was considered as a risky driving event. Analysis was then conducted for each driver in the same approach, and the number of risky events created by the 50 drivers was calculated. The results show that a total of 2,585 risky driving events were generated from the driving data of 50 drivers, with an average number of 51.7 for each participant. Note that the number of risky driving events from drivers (no. 3, 5, 9, 12, 19, 27, and 33) accounted for 46.1% (namely 1191), and these seven drivers were from the high- or medium-risk group. The number of risky driving events from drivers (no. 4, 6, 11, 13, 14, 17, and 24) only accounted for 1.2% (namely 31), and these seven drivers were all from the low-risk group. The results obtained on the basis of the SAX algorithm can also verify the case study by using the number of risky driving events described here, which provides a technical means to identify risky drivers from a different perspective.

## 7. Discussion and Conclusion

The methods and results generated in this study are compared with previous research in this section. In work by Wu et al. [46], GPS data of the connected vehicles were collected and factor analysis and cluster analysis were conducted to aggregate commercial vehicle drivers into different groups with four speed-related clusters. However, except for two drivers, the risk level of drivers in each cluster was not same, and no prediction model for risky drivers was presented or validation research conducted. The SAX algorithm is commonly applied in motif discovery and anomaly detection, such as in areas including medical scenarios and power consumption [47], but it is rarely used in the field of driving safety analysis. Additionally, according to the findings in Zeeman and Booyens [33], Eboli et al. [34], and Eboli et al. [35], speed or road condition should be considered for characterizing drivers' behavior and the acceleration threshold is found to be reduced with an increase in speed. The threshold of rapid acceleration or hard braking is set at  $0.12 \text{ g}/0.1 \text{ g}/0.0875 \text{ g}$  here for abrupt acceleration or hard braking (in absolute value) in three different kinds of routes, which is in line with the threshold described in Vlieger et al. [36], Boonmee and Tangamchit [31], and Paefgen et al. [30]. However, the threshold is acknowledged to be  $0.3 \text{ g}$  by Mortimer et al. [24] and Anderson & Baldock [25],  $0.4 \text{ g}$  in the euroFOT project [27], and  $0.5 \text{ g}$  by Fazeen et al. [16]. This could be attributed to the limited dataset (only 1.5 h to 2 h driving data) for each participant used in this test when compared with other larger datasets (i.e., the euroFOT project).

This paper predominantly explored the relationship between driver risk and factors indicating vehicle's speed and driver's acceleration behavior and examined whether a driver is risky or not by using GPS data and derivative indicator collected through a real road driving experiment. Fifty drivers

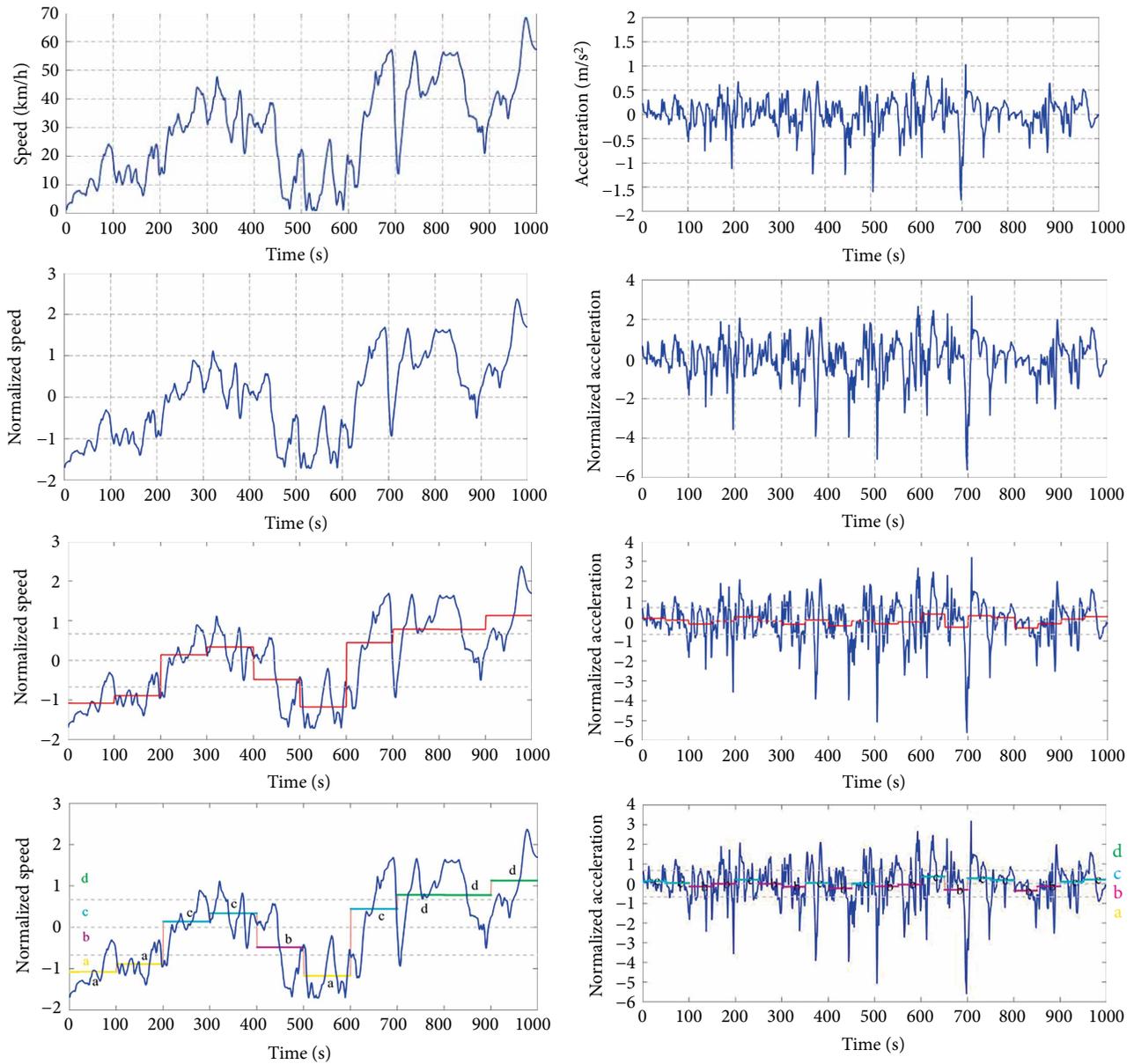


FIGURE 9: Discretization and symbolization of speed (left) and acceleration (right) data with the SAX algorithm.

were asked to drive a route that included both rural highways and urban roads, and their behaviors were recorded by cameras. All potential risky situations from the 50 drivers were extracted by one data reductionist using some basic rules, and the footage was then analyzed by another group of volunteers to evaluate the risk level of each event and determine whether the driver was risky or not through viewing the video segments. The risk score obtained through factor analysis and evaluated by volunteers was compared afterwards, and SAX algorithm was used to validate the case study through normalization, discretization, and symbolization. The results demonstrated that three factors closely associated with speed and acceleration measures were extracted by reducing the dimensions, and the drivers were classified into five groups with different levels of risk based on the risk score obtained through factor analysis. In addition, the risk score obtained through

factor analysis had a relatively high correlation ( $r = 0.52$ ) with the score evaluated by volunteers. Drivers (no. 3, 5, 9, 12, 19, 27, and 33) from the high- and medium-risk groups accounted for 46.1% of the total risky driving events, while the number of risky driving events for drivers (no. 4, 6, 11, 13, 14, 17, and 24) from the low-risk group only accounted for 1.2%. This kind of research could provide a technical means to identify risky drivers by using signals easily obtained through conventional sensors, especially for auto-insurance companies with UBI applications, bus companies, and transport enterprises.

It is acknowledged that the low dataset of 50 drivers with only 1.5 h to 2 h driving data for each participant created limitations in this study. Furthermore, although the same threshold of rapid acceleration or hard braking was used in research by Vlieger et al. [41], Boonmee and Tangamchit [8] and Paefgen et al. [37], the threshold value used here was still

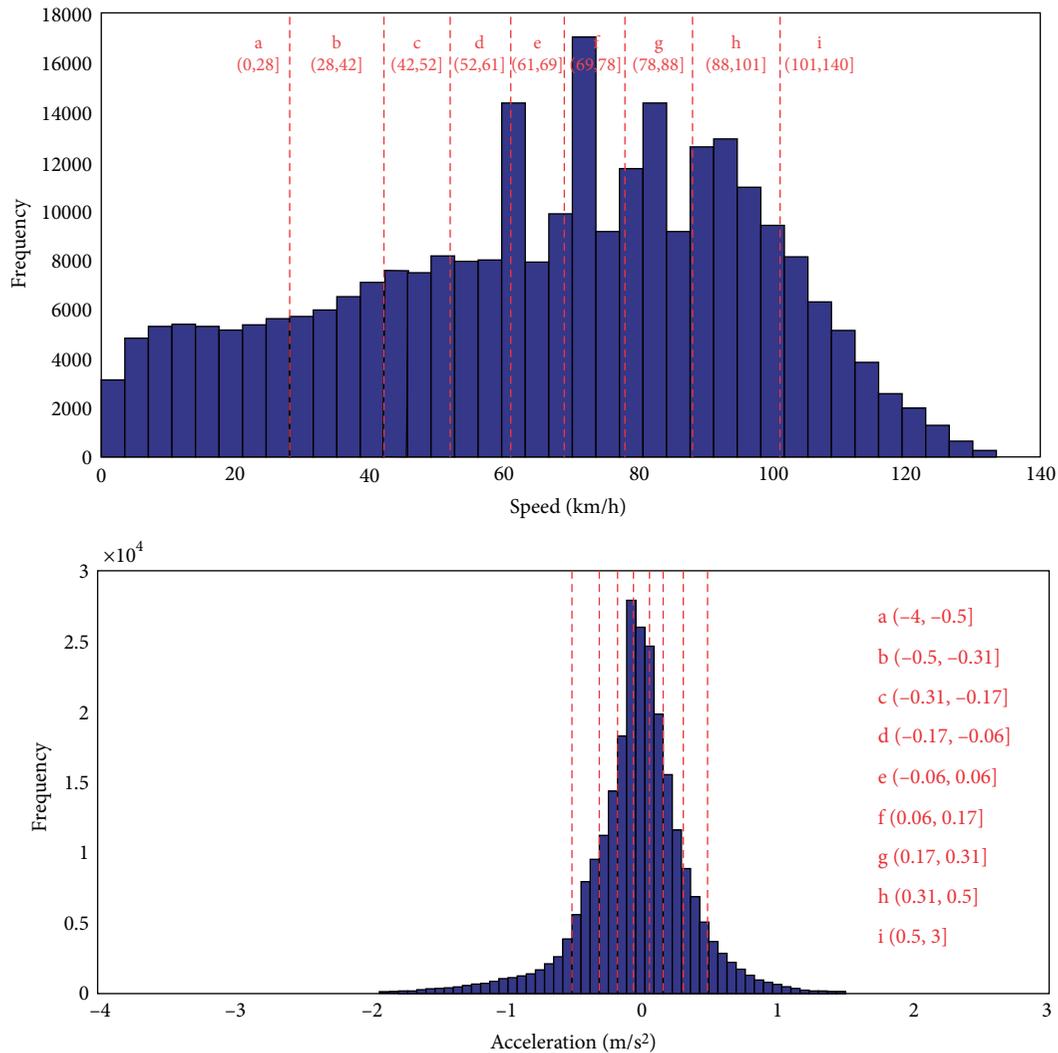


FIGURE 10: Distribution of the speed (above) and acceleration (below) from 50 drivers.

relatively lower. Note that threshold also depends on the dataset size to some extent, so an even lower threshold was used here. However, it was still possible to identify risky drivers as nearly the same trip length and running time was selected for each participant. Almost all participants recruited in this test are male drivers, and this means that the findings may only represent this specific driver population.

To expand on this study in the future, exploration into how the identification of risky driving is affected by the driver's experience, gender, or type of instrumented vehicle (e.g., car vs bus) requires investigation. Examination into the variability within individual driver data using a larger dataset is also necessary. Furthermore, identifying individual differences in more specific areas such as car-following and lane-changing situations will improve the development of the advanced driver assistance system in relation to driving styles.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they no conflicts of interest.

### Acknowledgments

This work was supported by National Key R&D Program of China [2018YFB1600501], Changjiang scholars and an innovative project team development plan [IRT\_17R95], the National Natural Science Foundation of China [51775053], and the Natural Science Foundation of Shaanxi Province of China [2018JM5158].

### References

- [1] World Health Organization, "Fact Sheet: The Top Ten Causes of Death," World Health Organisation, 2010.
- [2] CRTAS, *China Road Traffic Accidents Statistics*, Traffic Administration Bureau of China State Security Ministry, Beijing, China, 2010.

- [3] L. Aarts and I. V. Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.
- [4] CRTAS, *China Road Traffic Accidents Statistics*, Traffic Administration Bureau of China State Security Ministry, Beijing, China, 2014.
- [5] T. Lotan and T. Toledo, "In-vehicle data recorder for evaluation of driving behavior and safety," *Transportation Research Record Journal of the Transportation Research Board*, pp. 112–119, 2006.
- [6] J. Paefgen, T. Staake, and F. Thiesse, "Resolving the misalignment between consumer privacy concerns and ubiquitous IS design: the case of usage-based insurance," in *International Conference on Information Systems*, 2012.
- [7] J. Cen, Z. Wang, C. Wang, and F. Liu, "A system design for driving behavior analysis and assessment," in *IEEE International Conference on Internet of Things (iThings)*, IEEE, Chengdu, China, 2017.
- [8] S. Husnjak, D. Peraković, I. Forenbacher, and M. Mumdzic, "Telematics system in usage based motor insurance," *Procedia Engineering*, vol. 100, pp. 816–825, 2015.
- [9] D. I. Tselentis, G. Yannis, and E. I. Vlahogianni, "Innovative motor insurance schemes: a review of current practices and emerging challenges," *Accident Analysis & Prevention*, vol. 98, pp. 139–148, 2017.
- [10] J. Wang, W. Xu, and Y. Gong, "Real time driving danger level prediction," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 8, pp. 1247–1254, 2010.
- [11] T. Imkamon, P. Saensom, P. Tangamchit, and P. Pongpaibool, "Detection of hazardous driving behavior using fuzzy logic," in *IEEE International Conference on Electrical Engineering/electronics, Computer, Telecommunications and Information Technology*, IEEE, Krabi, Thailand, 2008.
- [12] K. Takeda, C. Miyajima, T. Suzuki et al., "Self-coaching system based on recorded driving data: learning from one's experiences," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1821–1831, 2012.
- [13] S. V. Hosseinioun, H. Al-Osman, and A. E. Saddik, "Employing sensors and services fusion to detect and assess driving events," in *IEEE International Symposium on Multimedia*, IEEE, Miami, FL, USA, 2016.
- [14] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *IEEE International Conference on Intelligent Transportation Systems*, IEEE, Washington, DC, USA, 2011.
- [15] A. S. Zeeman and M. J. Booyesen, "Public transport sector driver behaviour: measuring recklessness using speed and acceleration," in *South African Transport conference*, 2014.
- [16] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya, and M. C. González, "Safe driving using mobile phones," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1462–1468, 2012.
- [17] L. Eboli, G. Guido, G. Mazzulla, G. Pungillo, and R. Pungillo, "Investigating on car users' driving behaviour through the speed analysis," *PROMET, Traffic & Transportation*, vol. 29, no. 2, pp. 187–196, 2017.
- [18] L. Eboli, G. Mazzulla, and G. Pungillo, "How to define the accident risk level of car drivers by combining objective and subjective measures of driving style," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 49, pp. 29–38, 2017.
- [19] F. L. Hall, "Chapter 2. Traffic Stream Characteristics," *Traffic Flow Theory*, U.S. Department of Transportation, 38 pages, 2011.
- [20] F. Feng, S. Bao, J. R. Sayer, C. Flannagan, M. Manser, and R. Wunderlich, "Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data," *Accident Analysis & Prevention*, vol. 104, pp. 125–136, 2017.
- [21] O. Bagdadi and A. Várhelyi, "Development of a method for detecting jerks in safety critical events," *Accident Analysis & Prevention*, vol. 50, pp. 83–91, 2013.
- [22] E. Gilman, A. Keskinarkaus, S. Tamminen, S. Pirttikangas, J. Rönning, and J. Riekkki, "Personalised assistance for fuel-efficient driving," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 681–705, 2015.
- [23] J. W. Joubert, D. de Beer, and N. de Koker, "Combining accelerometer data and contextual variables to evaluate the risk of driver behaviour," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 41, pp. 80–96, 2016.
- [24] R. G. Mortimer, "Brake force requirement study: driver-vehicle braking performance as a function of brake system design variables: final report—Deceleration," 1970.
- [25] R. Anderson and M. Baldock, *Vehicle Improvements to Reduce the Number and Severity of Rear End Crashes. CASR Report Series: CASR052*, University of Adelaide, 2008.
- [26] GEOTAB INC, "Geotab management by measurement," 2015, <https://www.geotab.com>
- [27] C. Kessler, A. Etemad, G. Alessandretti et al., "EuroFOT-European large-scale field operational test on in-vehicle systems: EuroFOT DL11.3," 2012.
- [28] V. L. Neale, T. A. Dingus, S. G. Klauer, J. D. Sudweeks, and M. J. Goodman, "An overview of the 100-car naturalistic study and findings," in *19th International Technical Conference on the Enhanced Safety of Vehicles*, Washington DC, United States, 2005.
- [29] L. M. Bergasa, D. Almeria, J. Almazan, J. J. Yebe, and R. Arroyo, "Drive Safe: an app for alerting inattentive drivers and scoring driving behaviors," in *IEEE Intelligent Vehicles Symposium Proceedings*, IEEE, Dearborn, MI, USA, 2014.
- [30] J. Paefgen, F. Kehr, Y. Zhai, and F. Michahelles, *Driving Behavior Analysis with Smartphones: Insights From a Controlled Field Study*, ACM Press, New York, pp. 36–38, 2012.
- [31] S. Boonmee and P. Tangamchit, "Portable reckless driving detection system," in *IEEE Conference on Computer, Telecommunications and Information Technology*, IEEE, Pattaya, Chonburi, Thailand, 2009.
- [32] K. C. Baldwin, D. D. Duncan, and S. K. West, "The driver monitor system: a means of assessing driver performance," *Johns Hopkins APL Technical Digest*, vol. 25, pp. 269–277, 2004.
- [33] A. S. Zeeman and M. J. Booyesen, "Combining speed and acceleration to detect reckless driving in the informal public transport industry," in *16th International IEEE Conference on Intelligent Transportation Systems*, IEEE The Hague, Netherlands, 2013.
- [34] L. Eboli, G. Mazzulla, and G. Pungillo, "Combining speed and acceleration to define car users' safe or unsafe driving behavior," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 113–125, 2016.
- [35] L. Eboli, G. Mazzulla, G. Pungillo, and R. Pungillo, "Analysing car users' driving behaviour: safety domains for different types of roads," *Advances in Transportation Studies*, vol. 46, pp. 111–120, 2018.

- [36] I. De Vlieger, D. De Keukeleere, and J. G. Kretzschmar, "Environmental effects of driving behaviour and congestion related to passenger cars," *Atmospheric Environment*, vol. 34, no. 27, pp. 4649–4655, 2000.
- [37] R. Zito, G. D'Este, and M. A. P. Taylor, "Global positioning systems in the time domain: How useful a tool for intelligent vehicle-highway systems?" *Transportation Research Part C: Emerging Technologies*, vol. 3, no. 4, pp. 193–209, 1995.
- [38] J. S. Hickman and V. L. Neale, "How risky is it? An assessment of the relative risk of engaging in potentially unsafe driving behaviors," *Behavior*, pp. 38–40, 2006.
- [39] O. Bagdadi, "Assessing safety critical braking events in naturalistic driving studies," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 16, pp. 117–126, 2013.
- [40] T. A. Dingus, S. G. Klauer, V. L. Neale et al., *The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment DOT HS 810 593*, National Highway Traffic Safety Administration, Washington, DC, 2006.
- [41] A. P. Field, *Discovering Statistics Using IBM SPSS Statistics: and Sex and Drugs and Rock 'n' Roll*, Sage, London, 4th edition, 2013.
- [42] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 141–151, 1960.
- [43] D. G. Altman, *Practical Statistics for Medical Research*, Chapman and Hall, London, 1991.
- [44] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, ACM, New York, NY, USA, 2003.
- [45] Y. Sun, J. Li, J. Liu, B. Sun, and C. Chow, "An improvement of symbolic aggregate approximation distance measure for time series," *Neurocomputing*, vol. 138, no. 11, pp. 189–198, 2014.
- [46] C. Wu, C. Sun, D. Chu, Z. Huang, J. Ma, and H. Li, "Clustering of several typical behavioral characteristics of commercial vehicle drivers based on GPS data mining," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2581, no. 2581, pp. 154–163, 2016.
- [47] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time series," *Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp. 107–144, 2007.



Hindawi

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

