

## Research Article

# The Impact of Truck Proportion on Traffic Safety Using Surrogate Safety Measures in China

Shengdi Chen,<sup>1</sup> Shiwen Zhang,<sup>2</sup> Yingying Xing ,<sup>2</sup> Jian Lu,<sup>2</sup> Yichuan Peng,<sup>2</sup> and H. Michael Zhang<sup>2,3</sup>

<sup>1</sup>School of Transport & Communications, Shanghai Maritime University, 1550 Haigang Street, Shanghai 201306, China

<sup>2</sup>The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China

<sup>3</sup>Civil and Environmental Engineering University of California, Davis 3145 Engr III, Davis, CA 95616, USA

Correspondence should be addressed to Yingying Xing; [yingying199004@tongji.edu.cn](mailto:yingying199004@tongji.edu.cn)

Received 7 September 2019; Revised 20 February 2020; Accepted 18 May 2020; Published 12 August 2020

Academic Editor: Md. Mazharul Haque

Copyright © 2020 Shengdi Chen 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.

The purpose of this study is to investigate the impact of the truck proportion on surrogate safety measures to explore the relationship between truck proportion and traffic safety. The relationship between truck proportion and traffic flow parameters was analyzed by correlation and partial correlation analysis, and the value of the 85th percentile speed minus the 15th percentile speed (85%V–15%V) and the speed variation coefficient were selected as surrogate safety measures to explore the impact of truck proportion on traffic status. The *k*-means algorithm and the support vector machine were employed to evaluate traffic status on a freeway under different truck proportions in different periods. The major results are that the relationship between truck proportion and the value of 85%V–15%V and the speed variation coefficient is consistent in different aggregation periods. With increasing truck proportion, the value of 85%V–15%V, as well as the speed variation coefficient, increases initially and then decreases. In addition, the traffic flow status tends to be dangerous when the truck proportion ranges from 0.4 to 0.6 and when the value of 85%V–15%V and the speed variation coefficient are above 42 km/h and 0.223, respectively. While the truck proportion is from 0.1 to 0.3 and from 0.7 to 0.9, the traffic flow is relatively safe on the condition that the value of 85%V–15%V and the speed variation coefficient were under 42 km/h and 0.223, respectively. Therefore, the relationship between truck proportion and traffic safety could be well revealed by two surrogate safety measures, that is, the value of 85%V–15%V and the speed variation coefficient. In addition, the *k*-means algorithm and the support vector machine can well reveal the impact of truck proportion on traffic safety in different periods. The findings of this study indicate a need for decreasing the disturbance of mixed traffic and the impact of the truck proportion on traffic safety status.

## 1. Introduction and Background

In China, a total of 83,300 traffic crashes occurred (approximately 30% involved trucks) on freeways, the death toll was 5,500, and the number of injuries was 11,500 in 2016 [1]. In 2014, a total of approximately 0.2 million traffic crashes occurred in Shanxi Province in China, and more than 50% of the crashes involved trucks [1]. In addition, the percentage of large crashes involving trucks is high. Therefore, there is a relatively strong relationship between truck proportion and traffic safety.

The truck is the main research object in this paper. The volume of heavy trucks is relatively large in Shanxi Province. The attendant problem of traffic crashes involving trucks is increasing. A high percentage of trucks is an important cause of this phenomenon. On the one hand, when high-speed cars are mixed with low-speed large trucks while driving on the freeway, the running speed of the small car will decrease considerably due to the blocking role of large trucks; thus, the overall traffic flow efficiency is affected. On the other hand, due to the obvious difference between the speed of cars and that of large trucks, the behaviors of lane-changing and

overtaking of vehicles are too frequent; thus, traffic conflict will increase greatly, which could increase the probability of traffic accidents. There is an obvious negative impact on the safe operation of trucks on freeways. Therefore, how to alleviate and eliminate the excessive influence of trucks and truck proportion on freeway operational safety is an urgent problem.

There are few studies on the relationship between truck proportion and traffic safety in literature. Many research efforts have been allocated to the analysis of truck-involved crashes based on historical traffic crash data, such as exploring the performance of trucks on crashes. Dong et al. [2] explored the influence of intersection features on traffic safety, including large truck size, which showed that the truck percentage significantly affects traffic crashes and that the frequencies of truck-involved crashes increase when the truck percentage increases. Based on weather and traffic data in Korea, Choi et al. [3] found that risk factors such as speed were a significant factor affecting truck crash severity related to truck volume. Dong et al. [4] analyzed the influences of risk factors on the frequency and severity of large truck-involved crashes, and the results showed that the severity and frequency of crashes increase significantly when large trucks are involved. Xie et al. [5] explored the different proportions of trucks on truck crash counts, and the results predicted that total truck crashes would decrease by 0.98% with a slight change in the percentage of trucks. Hou et al. [6] examined the factors contributing to traffic crashes in freeway tunnels and found that more crashes occurred on sharply curved segments with a high percentage of trucks. Considering the above reviews, using crash data has proven that truck proportion has a negative effect on other factors in traffic crashes. However, in China, it is difficult to obtain crash data. In addition, there is a certain lag of historical crash data in evaluating traffic safety, and traffic safety status cannot be determined in real time.

Based on the previous literatures, traffic flow safety status could be revealed by some surrogate safety indicators, such as speed standard deviation and speed variation coefficient [7–9]. Thus, through analyzing the relationship between truck proportion and surrogate safety indicators, the relationship between truck proportion and the traffic safety can also be inferred. There is some literature focusing on the relationship between surrogate indicators and truck proportion, which can provide some revelation and reference for the analysis of the impact of truck proportion on traffic safety. Some literature analyzed main truck impacts [10] and found that traffic volume and speed decreased with increasing truck percentage and that large truck size and poor operation performance resulted in the impact of trucks and psychological pressure. Wu et al. [11] analyzed the influence of trucks on traffic flow and velocity differences in Shanghai and found that the traffic flow speed difference is greatest when the truck mix rate is 30%. Li et al. [12] investigated the impact of the heavy truck proportion in heterogeneous traffic condition. The results showed that, with the increase of the truck proportion, the lane-changing frequency increases. Although these studies

described the influence of truck proportion on some parameters of traffic flow, most studies only mentioned the influence of truck proportion on traffic parameters in a certain fixed value or in a small range. In addition, most of these studies also use simulation algorithm to discuss the influence of truck proportion on traffic flow parameters, which still have some limitations in exploring the relationship between them in the real world.

Moreover, the classification of traffic flow parameters was used for surrogate indicators of traffic status in few studies. Statistical techniques have been used to classify the status of traffic parameters and the ability to learn from traffic parameters without being explicitly programmed, which has provided powerful assistance to the classification of traffic flow parameters and has been gradually applied to the classification of traffic safety system [13, 14]. Additionally, *k*-means and support vector machines have been used in some literature to explore the status of traffic safety. Some researchers have studied the increasing amount of road traffic using the *k*-means algorithm [15], and it has been proven that clustering is best for road traffic accident classification, which lays a foundation for further safety classification research [16, 17]. The classification of accident datasets can also help to better understand the complex relationship between injury severity outcomes and other factors [18], and the traffic accident dataset comprises five attributes that can also be segmented [19]. Kumar et al. [20] also verified that *k*-means performs well in reducing the heterogeneity of road accident data. Additionally, SVM is a novel small sample learning method with a solid theoretical foundation, which has been used in many studies [21]. SVM is different from traditional algorithms [22, 23]. This algorithm can not only help us to identify key samples and eliminate a large number of redundant samples but also ensure that the samples are simple and robust [24]. Therefore, SVM is properly suited to the classification of traffic safety status in this paper [25].

Overall, few related studies have focused on the relationship between truck proportion and traffic safety by classifying traffic status. Among them, historical traffic crash data and truck-involved crashes are usually used to evaluate the impact of truck performance or parameters on traffic safety, which delays real-time determination of truck proportion on traffic safety. Therefore, the aim of this study is to explore the influence of truck proportion on traffic safety by collecting data from surveillance videos. According to previous studies, several traffic flow parameters can be used to evaluate the traffic flow safety status, such as the speed difference of 85% vehicle speed minus 15% vehicle speed ( $85\%V - 15\%V$ ), the variation coefficient, and the number of lane changes. Thus, traffic flow parameters are considered as surrogate safety measures for investigating the relationship between truck proportion and traffic safety. The *k*-means clustering algorithm and the support vector machine were used to divide traffic flow safety status into two statuses: dangerous and safe. Therefore, the impact of the truck proportion on freeway traffic safety could be reflected.

## 2. Data Description and Methodology

**2.1. Data Collection and Description.** The traffic parameters on the Taichang, Datong, and Xinbao Freeways in Shanxi Province were collected in this paper. These roads are all two-way four-lane freeways. The designated speed of the inner lane is 100 km/h and the outer lane is 80 km/h, while the upper speed limit is 100 km/h and the lower speed limit is 60 km/h. These sites are all straight roadway segments, and the weather was sunny. The width of each lane is 3.75 m, and the hard shoulder width is 3 m with the right side of the curb bandwidth being 0.5 m.

All the data were collected during daylight hours by video from freeway section monitoring video recording. Frames of the videos were used to collect data, and one second represented 25 frames. The collection times of the three sites of the Datong freeway were from 8:00–9:00, 11:00–13:00, and 16:00–17:00. The collection times of the two sites of the Taichang freeway were from 7:00–17:00, and the collection times of the two sites of the Xinbao freeway were from 8:00–9:00, 11:00–13:00, and 16:30–17:30. Seven sites on 3 freeways were collected, as shown in Table 1, during three periods of 5 minutes, 10 minutes, and 15 minutes. Every two centerlines (9 m) were separated with a gap (6 m) on the freeway. From the monitoring video recording, three centerlines and three gaps were selected as the study travel distances. The end of the last centerline and the start of the first centerline were chosen as the two points of the start and end positions, which could also be collected from the period frame of every vehicle. Therefore, the running speed could be obtained. Thus, traffic parameters such as traffic flow, TP, average vehicle speed, speed difference, and variation coefficient (the standard speed difference divided by the average speed) could be obtained. Types of vehicles were divided into two categories: trucks and other vehicles, such as coach buses with lengths greater than 6 m, and cars, such as private cars and other vehicles less than 6 m in length. According to the Highway Capacity Manual 2000, trucks were converted into 3 passenger car equivalents. As a result, the samples of 969 traffic parameters in 5 minutes, 501 samples in 10 minutes, and 324 samples in 15 minutes were derived from the videos, and the maximum, minimum, mean, and standard deviation of values are shown in Table 1. In general, the mean speed is relatively large, while the truck proportion is low. The number of lane changes is relatively large, while the truck proportion is low.

The trend in traffic flow parameters and truck proportion were used to evaluate the relationship between them. In particular, the speed dispersion parameters, such as speed standard deviation and speed variation coefficient, are important indicators that could represent the relationship between traffic efficiency and traffic safety [8]. The results demonstrated the steady status of traffic flow to reflect traffic safety. Therefore, speed-related variables were selected as surrogate safety indicators to measure traffic flow safety status. In addition, due to the performance and shape of trucks, some small car drivers tended to overtake to obtain a wider field of view while driving, which also increased the probability of conflicts. Thus, driving behavior such as lane-

changing could also be considered an index of traffic safety. The level of service at a design speed of 100 km/h according to Highway Capacity Manual (HCM) was divided into four levels: the first level was free flow of 650 maximum capacity volume passage cars per hour per lane, the second was the upper steady flow of 1,400 passage cars/h/lane, the third was steady flow of 1,800 passage cars/h/lane, and the last was saturated/forced flow of 2,100 passage cars/h/lane.

After calculating the preliminary mathematical statistics of the raw data, traffic volumes above 108 pcu/5 min, 217 pcu/10 min, and 325 pcu/15 minutes were selected. Therefore, 732 samples in 5 minutes, 365 samples in 10 minutes, and 251 samples in 15 minutes were studied. The description of these parameters is shown in Table 2. The mean, maximum, minimum, and standard deviation of every traffic parameter are shown. The data collected in different periods show that the trends of the mean values of these parameters were similar in different time periods. The standard deviation of 15 minutes was relatively low, which reflected that the degree of speed dispersion among individuals in a group was stable.

Figure 1 shows the relationship between the four parameters with the truck proportion. With increasing truck proportion, the trend of speed dispersion increased initially and then fell. When the truck proportion reached 0.5, the degree of dispersion was the greatest in Figure 1. When the truck proportion was less than 0.5 or more than 0.5, the degree of speed dispersion decreased. When the truck proportion was in the high range from 0.6 to 1, the speed standard deviation decreased steadily with increasing truck proportion. In the large truck proportion above 0.6, the small cars accelerated and decelerated frequently, which could result in frequently blocked status. Thus, the speed of the small car was close to the truck speed, which contributed to the aggregate speed difference when the truck proportion was less than 0.5. Therefore, the same trend was observed for the different speeds of the 85<sup>th</sup> percentile minus the 15<sup>th</sup> percentile. The speed variation coefficient is a better index to embody the truck proportion and surrogate measure of safety, which eliminates the impact of measurement scales and dimensions such as traffic volume. Different time periods have the same trend of different truck proportions. The data for the 15-minute time period were steadier and more aggregated than those of the others. The speed of small cars was almost random as the truck proportion was small, as was the interaction between these two kinds of vehicle parameters. Regardless of the traffic parameters related to the truck proportion, the selected time period should ensure that the traffic flow is steady to analyze other parameter relationships.

In Figure 1, unlike the relationship among the other three parameters and the truck proportion, the relationship between the speed difference and truck proportion was relatively dispersed, especially in 5 minutes, which inferred that there was a large difference in speed between these two types of vehicles. However, there was an obvious relationship among the other three parameters and the truck proportion, which shows that the parameters increase first and then decrease with increased truck proportion in different periods, which show the same trends.

TABLE 1: The range of basic data collected from different time periods and sections.

Freeway section	Time period (min)	Truck proportion				Traffic volume			
		Mean	Std.	Max.	Min.	Mean	Std.	Max.	Min.
Datong	5	0.740	0.168	1.000	0.200	75.850	43.065	195	4
	10	0.728	0.153	0.955	0.222	147.486	80.898	292	7
	15	0.727	0.152	0.968	0.250	221.229	122.259	444	26
Taichang	5	0.286	0.101	0.680	0.044	160.808	44.846	321	7
	10	0.287	0.084	0.571	0.128	309.375	92.765	546	7
	15	0.285	0.078	0.569	0.107	483.082	108.818	750	190
Xinbao	5	0.750	0.128	1.000	0.333	110.250	52.981	258	27
	10	0.754	0.107	0.983	0.533	220.500	97.490	492	65
	15	0.756	0.101	0.974	0.580	330.750	141.523	696	124

Freeway section	Time period (min)	Average speed (km/h)				The number of lane changes			
		Mean	Std.	Max.	Min.	Mean	Std.	Max.	Min.
Datong	5	64.110	13.341	98.800	24.545	0.857	1.261	5	0
	10	64.427	12.494	96.419	41.878	1.667	2.214	9	0
	15	64.361	12.352	93.852	39.304	2.500	3.294	17	0
Taichang	5	91.289	10.108	126.313	64.414	1.623	1.802	10	0
	10	90.975	9.670	120.473	69.919	3.123	3.064	17	0
	15	91.257	9.467	121.737	70.945	4.877	4.312	18	0
Xinbao	5	75.693	14.094	98.529	48.503	0.573	0.903	5	0
	10	75.643	13.761	93.847	49.852	1.146	1.502	6	0
	15	75.641	13.906	93.504	50.426	1.719	1.836	8	0

TABLE 2: The data of collected parameters after preliminary statistics in different periods.

Variables	Period (min)	Mean	Std.	Max.	Min.
Truck proportion	5	0.335	0.184	1	0.044
	10	0.326286	0.165	0.983	0.128
	15	0.324583	0.162	0.968	0.107
Traffic volume	5	167.186	38.628	321	108
	10	330.077	70.590	546	217
	15	491.052	99.200	750	329
Average speed (km/h)	5	89.173	12.572	124.223	43.318
	10	89.597	11.957	120.473	49.555
	15	89.358	12.137	121.737	44.407
Speed difference (km/h)	5	24.461	8.976	49.971	0.153
	10	24.547	8.152	45.738	0.426
	15	24.533	7.804	43.796	0.127
Speed standard deviation	5	18.508	4.269	45.924	7.271
	10	18.799	3.885	35.892	7.867
	15	18.875	3.676	31.652	8.145
85%V-15%V	5	38.784	10.711	68.957	11.923
	10	39.963	10.139	64.489	12.457
	15	40.377	9.810	64.067	14.971
Variation coefficient	5	0.212	0.058	0.394	0.090
	10	0.214	0.053	0.348	0.095
	15	0.216	0.053	0.371	0.136
Number of lane changes	5	1.708	1.804	10	0
	10	3.386	3.047	17	0
	15	5.072	4.283	18	0

2.2. *Data Partial Correlation Study.* Correlation analysis can be used to test the degree of correlation of two variables and determine the correlated direction by the sign of the correlation coefficient. However, when deeper factors are involved, the correlation coefficient cannot truly reflect the degree of variables due to the influence or effect of the third variable. Partial correlation analysis can be used to explore

the relationship between two variables under the control of other variables that may be affected. The calculation formula is as follows (equation (1)):

$$r_{12(3)} = \frac{r_{12} - r_{13} \times r_{23}}{\sqrt{1 - r_{13}^2} \times \sqrt{1 - r_{23}^2}} \quad (1)$$

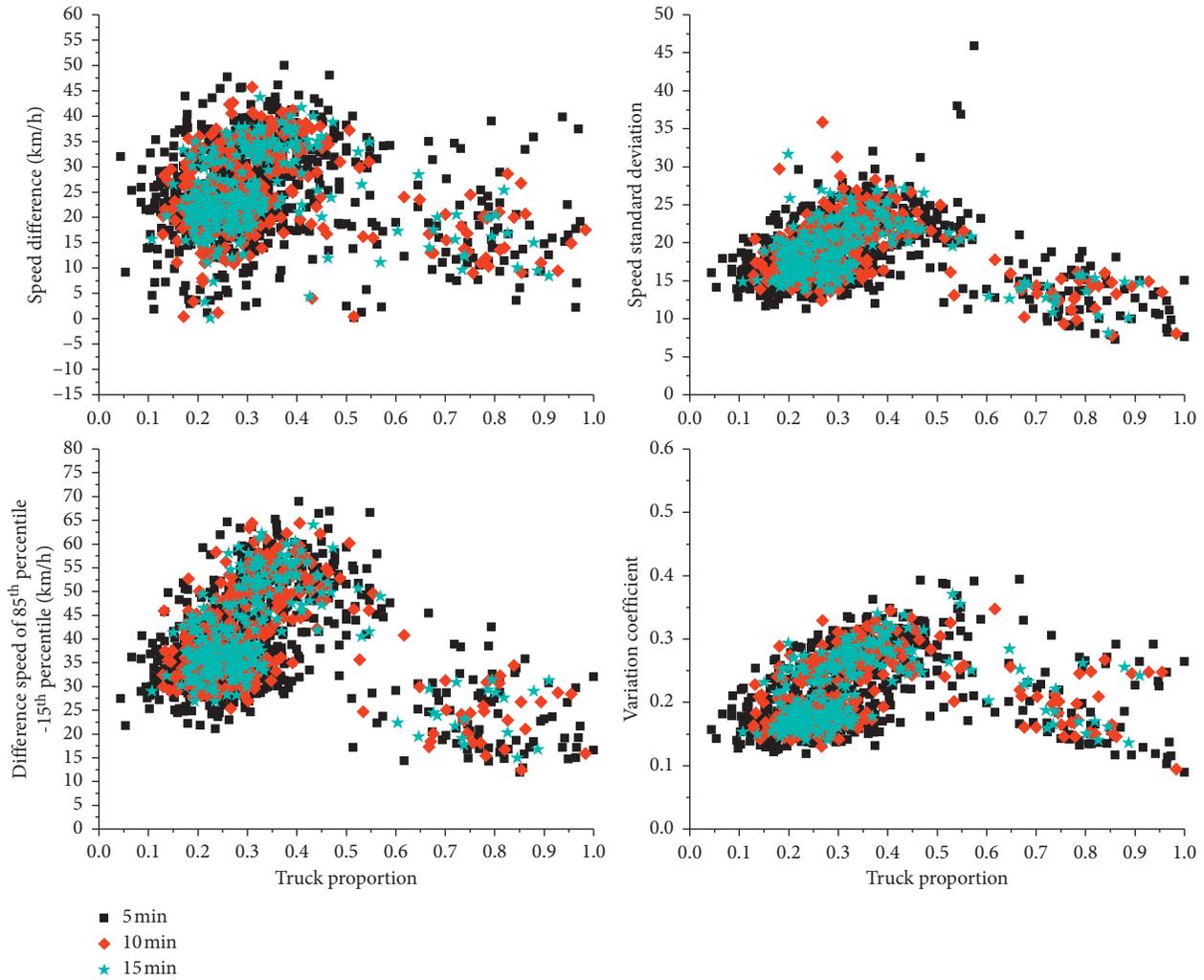


FIGURE 1: The relationship between truck proportion and traffic flow parameters in different periods.

This formula is the coefficient of partial correlation between the first and second variables calculated after controlling for the influence of the third variable. When more than one control variable is considered, the formula follows. Therefore, the correlation of these variables was examined using partial relationship tests. The results are shown in Tables 3 and 4 for different periods.

In Table 3, it can be seen that, in periods of 5 minutes and 10 minutes, there was a significant correlation in truck proportion among mean speed, traffic volume, speed standard deviation, 85%V–15%V, and variation coefficient except for the number of lane changes. The relationship between truck proportion and these five variables was significant at the 0.05 (bilateral) level when there were no control variables in the 15-minute period except for the relationship between speed difference, number of lane changes, and truck proportion without control variables. There was a positive correlation between truck proportion and variation coefficient, while the relationship among other evaluated indexes with truck proportion was negative. The

variable of the mean speed of parameters related to truck proportion was more than 0.5, while other variables were less than 0.3, which indicates that the relationship was relatively weak.

In Table 4, it can be seen that the correlation coefficients of the relationship between truck proportion and 85%V–15%V are  $-0.511$ ,  $-0.590$ , and  $-0.551$  in different periods when the variation coefficient and number of lane changes are controlled, which is higher than the case without control. The correlation coefficients of the speed variation coefficient are  $0.523$ ,  $0.595$ , and  $0.582$ . However, the number of lane changes is not significant in these periods, regardless of being controlled or not. When other variables are controlled, the correlation coefficients of 85%V–15%V and the variation coefficient are above 0.5, which indicates that there is a strong correlation between these variables so that these two indexes can be used as evaluation indexes. Therefore, 85%V–15%V and the variation coefficient were selected as surrogate safety measures to explore the relationship between the truck proportion and traffic safety.

TABLE 3: The relationship between basic traffic flows in the correlation analysis.

Variables (truck proportion) (min)	Basic parameters (-No <sup>a</sup> )	Mean speed	Traffic volume	Speed difference	Speed standard deviation	85% V-15%V	Variation coefficient	Number of lane changes
5	The correlation	-0.626	-0.216	-0.077	-0.151	-0.156	0.191	-0.039
	Significance (bilateral)	0.000	0.000	0.039	0.000	0.000	0.000	0.295
	.df	732	732	732	732	732	732	732
10	The correlation	-0.647	-0.246	-0.110	-0.181	-0.173	0.193	-0.019
	Significance (bilateral)	.000	.000	.036	0.001	0.001	0.000	0.724
	.df	365	365	365	365	365	365	365
15	The correlation	-0.656	-0.245	-0.085	-0.144	-0.126	0.252	-0.001
	Significance (bilateral)	0.000	0.000	0.181	0.023	0.045	0.000	0.993
	.df	251	251	251	251	251	251	251

a: the cell contains a zero-order correlation (Pearson).

TABLE 4: The relationship among indicators of traffic flow parameters in partial correlation analysis.

Variables (truck proportion) (min)	Evaluation indexes (other variables controlled)	85%V-15%V	Variation coefficient	Number of lane changes
5	The correlation	-0.511*	0.523*	0.006
	Significance (bilateral)	0.000	0.000	0.862
	.df	728	728	728
10	The correlation	-0.590*	0.595*	0.033
	Significance (bilateral)	0.000	0.000	0.537
	.df	361	361	361
15	The correlation	-0.551*	0.582*	0.044
	Significance (bilateral)	0.000	0.000	0.492
	.df	247	247	247

a: the cell contains a zero-order correlation (Pearson). \*Significant at the 0.05 level.

### 3. Methodology

As crash data are difficult to obtain in China, traffic flow parameters were utilized as surrogate safety measures to determine the relationship between truck proportion and traffic flow safety status via mathematical-statistical methods of correlation analysis from the conclusions in previous literature, such as specific indicators.

Regarding  $k$ -means clustering algorithm, data mining has been used for data exploration and analysis in many fields, such as traffic safety, by using the  $k$ -means method [18, 26–28], which can help us better understand the complex relationship between different risk factors that contribute to unsafe traffic status. The study conducted by Sohn and Lee [15] also demonstrated that clustering based on a classification algorithm is best for road traffic status classification compared with the Bayesian neural networks and decision trees methods.  $k$ -means clustering is easy to implement, even in large data sets, particularly when using heuristics.  $k$ -means is often used as a preprocessing step to address the relationship between truck proportion and indicators. Thus, the traffic flow parameters were classified as different risk levels to identify the status of safety or unsafety via the  $k$ -means clustering algorithm. Then, the relationship between traffic parameters and truck proportion was

explored. Traffic flow parameters were divided into two kinds of traffic status [29]. Thus, the classification criteria were used in reference to previous literature [30–32].

In this paper, the given sample was divided into two classes:

$$d_{ij} = \min(\|x_i - z_j\|), \quad x_i \in S, z_j \in Z. \quad (2)$$

The clustering criterion function of the  $k$ -means algorithm is

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - z_j^2\|, \quad (3)$$

where  $\|x_i^{(j)} - z_j\|^2$  is a chosen distance measure between a data point  $x_i^{(j)}$  and the cluster center  $z_j$  of the  $n$  data points from their respective clusters. According to variables such as 85%V-15%V and the variation coefficient, the traffic flow status is classified as dangerous and safe, where number 1 represents safety status, meaning there is a lower probability of accidents, and number 2 represents a dangerous status.

The minimum value of the clustering criterion function  $J$  is the goal of algorithm optimization. When  $J$  is the smallest, the partial derivative of the function for each cluster center is 0:

$$\frac{\partial J}{\partial z_j} = -2 \sum_{i=1}^n (x_i^{(j)} - z_j) = -2 \sum_{i=1}^n x_i^{(j)} + 2|S_j| z_j = 0, \quad (4)$$

So

$$z_j = \frac{1}{|S_j|} \sum_{i=1}^n x_i^{(j)}, \quad (5)$$

Equations (4) and (5) show that the clustering center can minimize the clustering criterion function.

A support vector machine (SVM) is also used as a method for analyzing data for classification and regression to explore the relationship of these parameters. It was first proposed by Corinna Cortes and Vapnik [33] and demonstrated many unique advantages in solving small-sample, nonlinear, and high-dimensional pattern recognition. The sample size was small and separable in our research, so SVM is suitable for identifying the relationship between truck proportion and traffic safety. According to previous analyses and studies [34–36], the value of 85%V–15%V and the variation coefficient (CV) can be regarded as surrogate indicators of the probability of accidents.

A support vector machine was employed to evaluate the impacts of the truck proportion on parameters such as 85%V–15%V and the variation coefficient, which represented safety or dangerous status. The 85%V–15%V variable and variation coefficient were divided, so linear discrimination could be used.

The linear discriminant function of the binary classification can be written as in the following equation:

$$G(x) = W^T X + W_0, \quad (6)$$

where  $x$  is the corresponding feature vector,  $w$  is the weight vector, and  $W_0$  is the threshold.

If  $G(x) > 0$ , it can be determined that  $x \in W_1$ ; if  $G(x) < 0$ , then it can be determined that  $x \in W_2$ ; if  $G(x) = 0$ , it cannot be determined which category of  $x$  it belongs to.

At this point, the decision boundary equation is

$$G(x) = W^T X + W_0 = 0. \quad (7)$$

Thus, the linear classification model is as follows:

$$W^T (X_2 - X_1) = 0. \quad (8)$$

## 4. Results and Discussion

**4.1. Results of Data Classification.** First, the selected surrogate safety measures were divided into two types of traffic status via the  $k$ -means clustering algorithm. The traffic flow safety status is evaluated from the relationship between the selected traffic flow parameters above. The parameters represented as surrogate safety indexes were set as the value of 85%V–15%V in every period and the value of the speed variation coefficient in every period of time. The relationship between these indexes and the traffic safety were explored according to many previous studies and the China Statistical Yearbook [1]. For example, the higher the value of 85%V–15%V is, the higher the probability of an unsafe traffic

status is [37]. The higher the variation coefficient is, the higher the probability of traffic crashes occurring is [38]. Thus, the relationship between truck proportion and traffic safety can be inferred from the relationship between the truck proportion and surrogate safety measures such as 85%V–15%V and the variation coefficient. Then, based on the  $k$ -means clustering analysis, the classification of the traffic flow status is given as follows.

SPSS software was used to analyze the selected parameters via  $k$ -means clustering. The values of 85%V–15%V and the speed variation coefficient were divided into two categories. Number 1 represents a normal traffic flow status, while number 2 represents a dangerous traffic flow status. The classification results are shown in Table 5 for different periods.

The classification standard of 85%V–15%V was 41.503 km/h in a period of 5 minutes, 42.127 km/h in a period of 10 minutes, and 42.672 km/h in a period of 15 minutes. The standard variation coefficient was 0.224 in a period of 5 minutes, 0.224 in a period of 10 minutes, and 0.223 in a period of 15 minutes. From Table 5, the results are statistically significant at the level of 0.05 ( $p$  value < 0.05). This is consistent with the classification of traffic flow parameters. Then, the traffic flow status was finally divided into two categories named “safe” and “dangerous” through the classification of 85%V–15%V and the speed variation coefficient. The initial clustering center was consistent in all periods, while the period of 15 minutes was more reasonably compact.

**4.2. Support Vector Machine Results.** Based on the results of the  $k$ -means clustering analysis above, the value of 85%V–15%V and the speed variation coefficient were selected as two indicators according to the partial relationship tests. The greater the value is, the more dangerous the traffic flow status is. The relationship between the selected traffic flow parameters and truck proportion was analyzed as follows. Truck proportion, the value of 85%V–15%V, and the speed variation coefficient were considered as prediction indicators, while the classification of these parameters was regarded as response indicator. For the training data, 70% of the samples were randomly selected, while the rest were used as the testing data.

The classification of 85%V–15%V by the support vector machine in different periods is shown in Figure 2. A red  $x$  indicates the training data for number 1, which means a lower probability of traffic accidents. The purple  $x$  represents the testing data for number 1, the same meaning as the training data of number 1, accounting for 30% of all data. A green dot is the training data for number 2, which indicates a higher probability of traffic accidents. Then, a blue dot is the testing data for number 2. The circles are support vectors, and then the linear decision function is drawn. Thus, two traffic statuses could be classified by the linear classification function.

In the period of 5 minutes in the first picture in Figure 2, when the truck proportion was from 0 to 0.4 and the value of 85%V–15%V was under 41.503 km/h, the traffic status was

TABLE 5: *k*-means clustering analysis in different periods.

Different periods (min)	Clustering & testing	Classification	85%V–15%V	Variation coefficient
5	Initial clustering center	1 (dangerous)	68.957	0.394
		2 (safe)	11.923	0.090
	Final cluster center	1 (dangerous)	50.715	0.276
		2 (safe)	32.290	0.172
	The value of split point		41.503	0.224
	ANOVA	F	1525.788	2388.247
		Sig.	0.000	0.000
10	Initial clustering center	1 (dangerous)	64.489	0.348
		2 (safe)	12.457	0.095
	Final clustering center	1 (dangerous)	51.001	0.273
		2 (safe)	33.252	0.175
	The value of split point		42.127	0.224
	ANOVA	F	945.529	1733.678
		Sig.	0.000	0.000
15	Initial clustering center	1 (dangerous)	64.067	0.371
		2 (safe)	14.971	0.136
	Final clustering center	1 (dangerous)	51.266	0.271
		2 (safe)	34.077	0.174
	The value of split point		42.672	0.223
	ANOVA	F	626.887	1129.481
		Sig.	0.000	0.000

relatively safe, meaning that the probability of traffic accidents was lower. The range from 0.5 to 1.0 of truck proportion was relatively dispersed, while most were under 41.503 km/h, which shows that there was a relatively low probability of traffic accidents. When the truck proportion ranged from 0.2 to 0.6 and the value of 85%V–15%V was above 41.503 km/h, the probability of traffic accidents was relatively high. In this status, there were small cars mixed with trucks, and the performance of trucks was lower than that of small cars, so the speeds of these two types of vehicles were different from each other. Thus, the value 85%V–15%V was also dispersed, especially at truck proportions of 0.4 to 0.6.

In the periods of 10 minutes and 15 minutes in the second and last picture in Figure 2, we have the same conclusion as the period of 5 minutes. When the value of 85%V–15%V was above 42.127 km/h and 42.672 km/h, respectively, it was in dangerous traffic status when the truck proportion was from 0.2 to 0.6. It was in a safe traffic status when the truck proportion was in the ranges of 0.1 to 0.3 and 0.6 to 0.9 and the value of 85%V–15%V was under 42.127 km/h and 42.672 km/h otherwise. Taking a period of 15 minutes as an example, when the value of 85%V–15%V was above 42.672 km/h, the range of truck proportions from 0.3 to 0.5 was more likely to be dangerous than the ranges from 0.2 to 0.3 and from 0.5 to 0.6. When the value of 85%V–15%V was under 42.672 km/h, the range of truck proportions from 0.6 to 0.9 was more likely in a safe traffic status than the range from 0.2 to 0.3. The same trend was in these three periods of time, and the period of 15 minutes was clearer.

Figure 3 is the confusion matrix of testing data of 85%V–15%V in different periods, which is also known as the possibility of form or error matrix and is a specific table layout that allows visualization of the performance of a supervised learning algorithm. It exhibits the true

classification of testing data. Each column represents the predicted value, and each row represents the actual class. In a period of 5 minutes, the correct rate score of the model of training data is 0.996. The testing data of 85%V–15%V that belong to number 1 is classified into the number 2 is just one sample, while the prediction of the value of 85%V–15%V is completely true when classified into number 2. All the correct predicted results are on the diagonal, so it is easy to tell the error from the confusion matrix. The testing data classification verification accuracy is  $99.5\% = (145 + 74) / (145 + 1 + 74) * 100\%$ . In a period of 10 minutes, the correct rate score of the training data was 0.996. The testing data of 85%V–15%V which belongs to number 1 was completely truly classified into number 1. At the same time, the testing data belonging to number 2 was completely truly classified into number 2, so the testing data classification verification accuracy was 100%. In a period of 15 minutes, the correct rate score of the training data was 0.994. The testing data that belong to number 1 were corrected and classified into number 1. There is one sample that the value of 85%V–15%V belonged to number 2 but was wrongly classified into number 1. Therefore, the testing data classification verification accuracy was  $98.7\% = (46 + 29) / (46 + 1 + 29) * 100\%$ .

Then, the model of the index of the speed variation coefficient was also built to explore the relationship between truck proportion and traffic safety. In Figure 4, the data were relatively dispersed in a period of 5 minutes due to a smaller number of trucks and small cars under some situations, so some data were aggregated. When the speed variation coefficient was above 0.224, the truck proportion ranged from 0.2 to 0.6, and the range of 0.8 to 1.0 was in a dangerous traffic status, especially from 0.43 to 0.6, due to the higher variation coefficient. When the truck proportion was from 0.1 to 0.4 and from 0.7 to 1.0 and the variation coefficient was under 0.224, it was in a relatively safe traffic status that had a

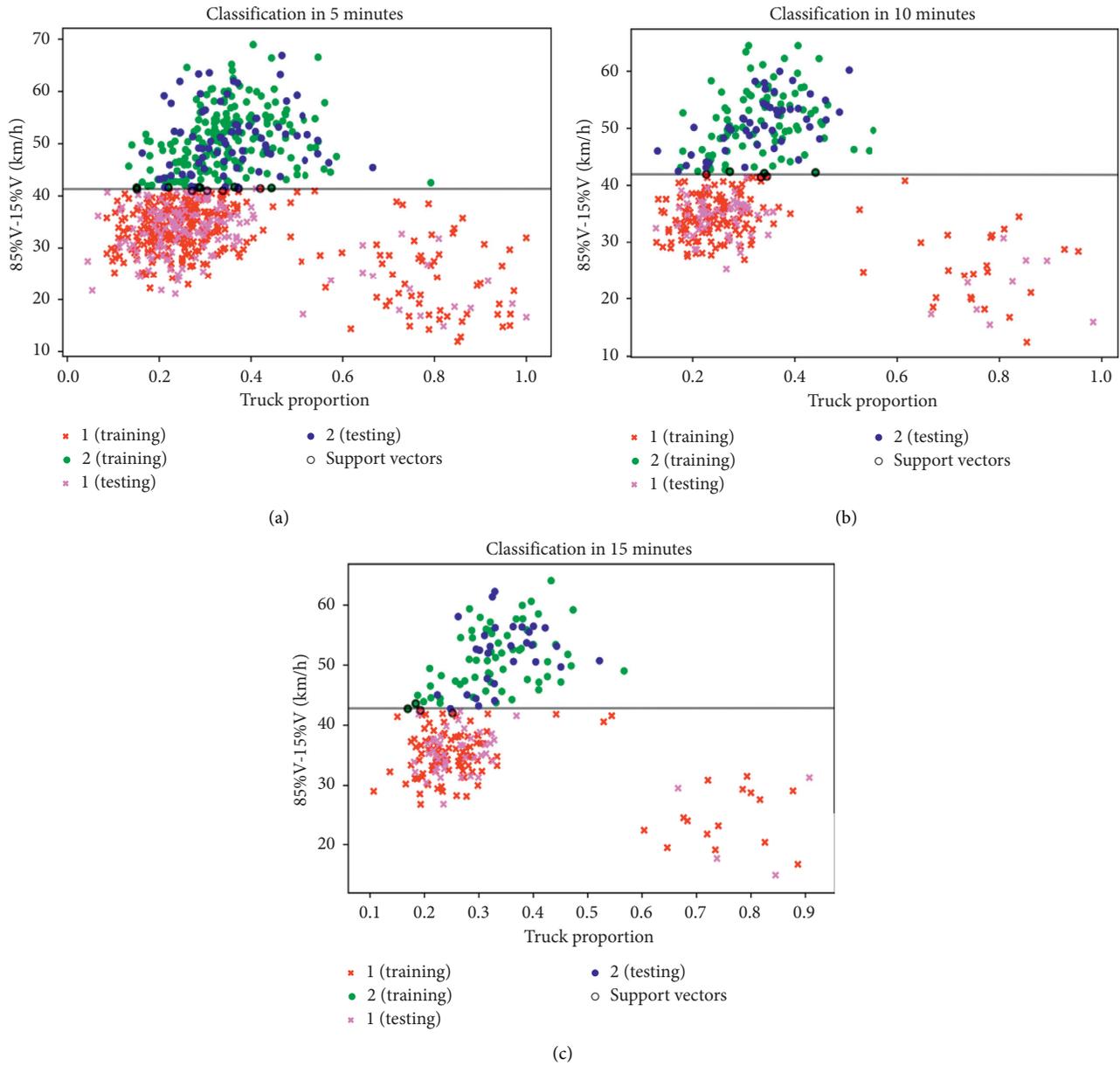


FIGURE 2: Classification diagram of the support vector machine of 85%V-15%V.

lower probability of accidents. In particular, when the truck proportion was from 0.8 to 1.0, it was rather safer than other proportions due to the lower variation coefficient.

In the periods of 10 minutes and 15 minutes in Figure 4, the two kinds of classifications in truck proportion and variation coefficient were clearer. In a period of 10 minutes, when the variation coefficient was above 0.224, the truck proportion ranged from 0.2 to 0.6, and the range of 0.8 to 1.0 was in a dangerous traffic status, especially the proportion from 0.4 to 0.6, due to the higher variation coefficient. The truck proportion ranged from 0.8 to 1.0 and was in a lower dangerous status compared with the proportion from 0.4 to 0.6. When the variation coefficient was under 0.224 and the truck proportion aggregated from 0.1 to 0.4 and from 0.7 to 0.9, there was a relatively safe traffic status. Then, the lower

probability of accidents would be in the situation of truck proportions from 0.1 to 0.4 and from 0.7 to 0.9 and the variation coefficient under 0.223. Truck proportions from 0.2 to 0.3 were more likely to have lower safety than truck proportions from 0.8 to 0.9. When the variation coefficient was above 0.223, there was a higher probability of accidents when the truck proportion was from 0.4 to 0.6, while the other was relatively low.

Figure 5 is the confusion matrix of the testing data of the speed variation coefficient in different periods. In the same situation, it could determine the correction rate of the classification of prediction. The correct rate score of the model of training data of the variation coefficient was 0.986 in a period of 5 minutes. The fact that the testing data of the variation coefficient was classified as 1 is completely true in a

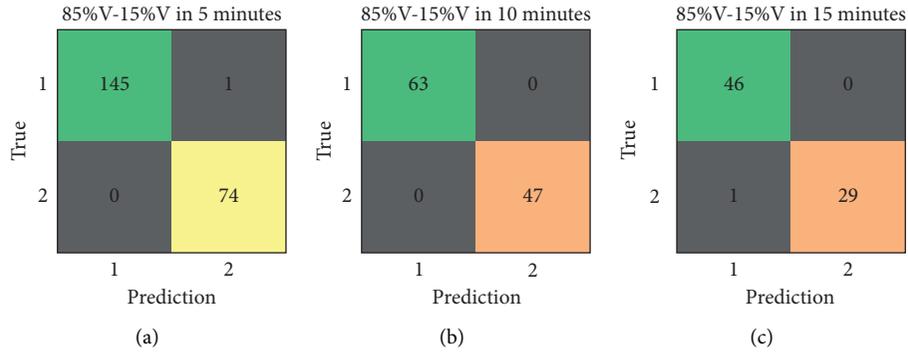


FIGURE 3: Illustration of the confusion matrix of testing data of 85%V-15%V in different periods.

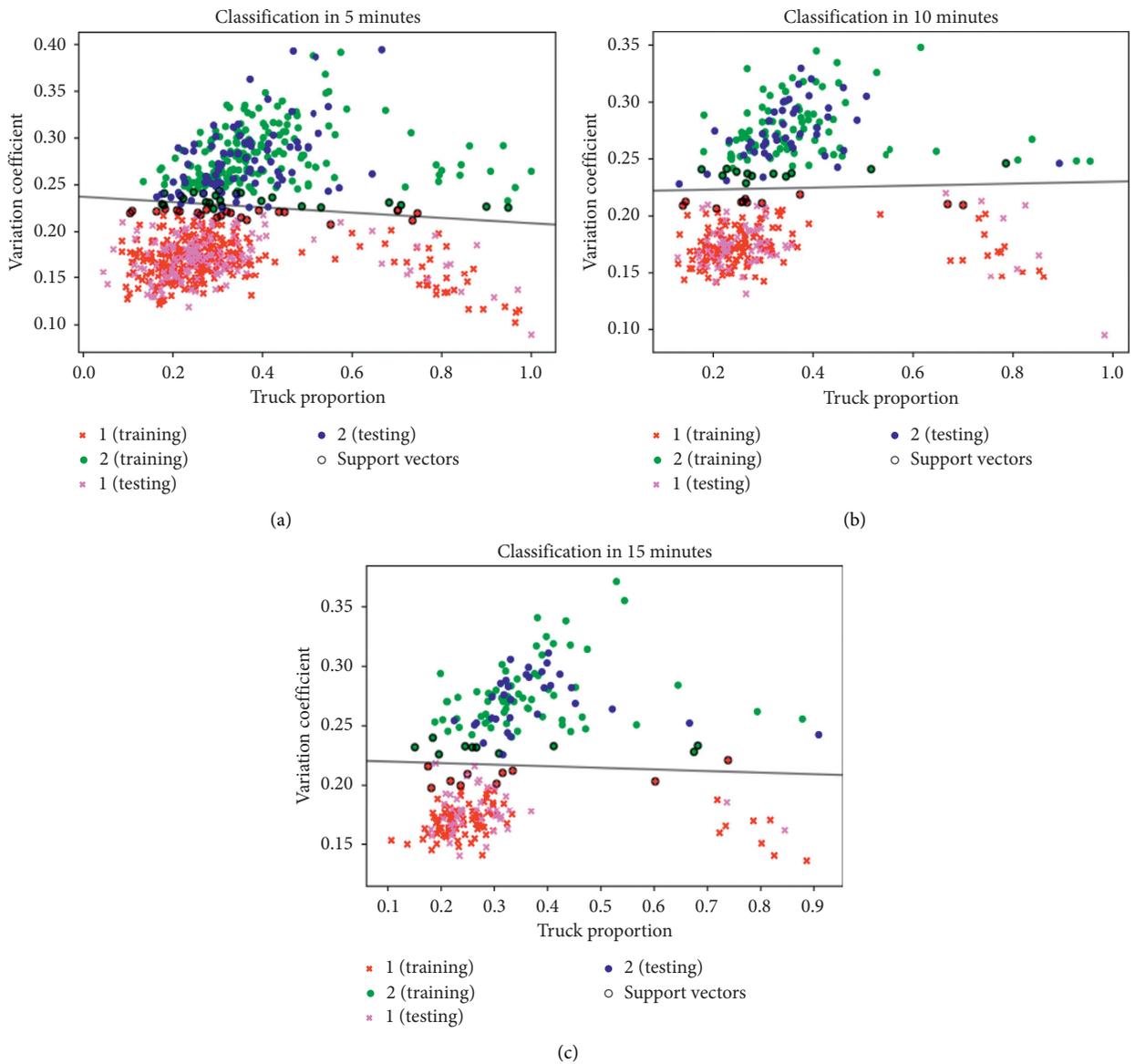


FIGURE 4: Classification diagram of the support vector machine of the speed variation coefficient.

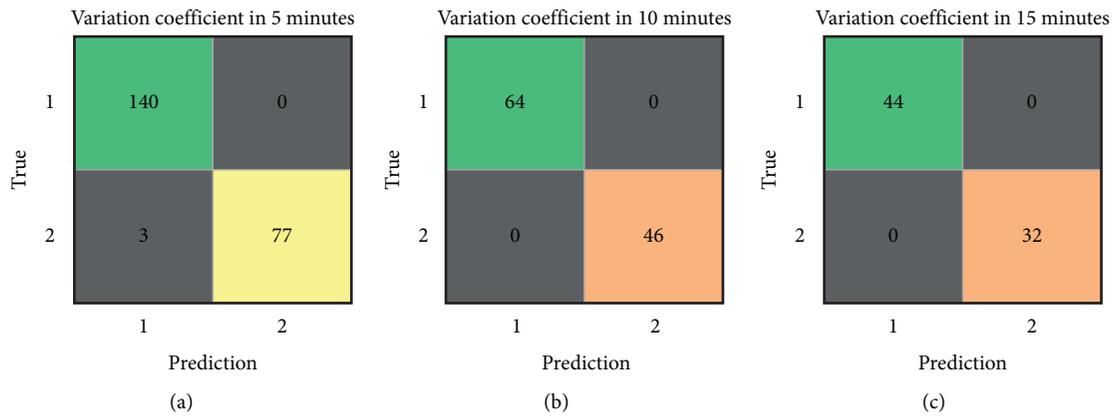


FIGURE 5: Illustration of the confusion matrix of the variation coefficient in different periods.

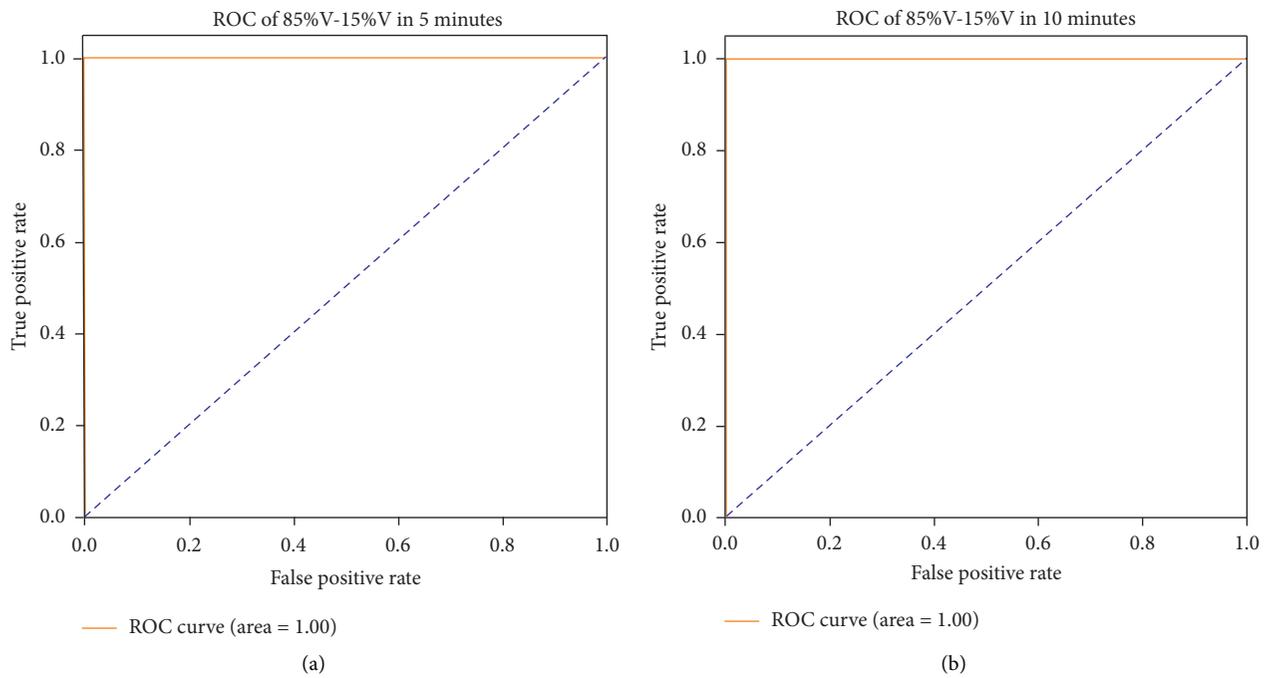


FIGURE 6: Continued.

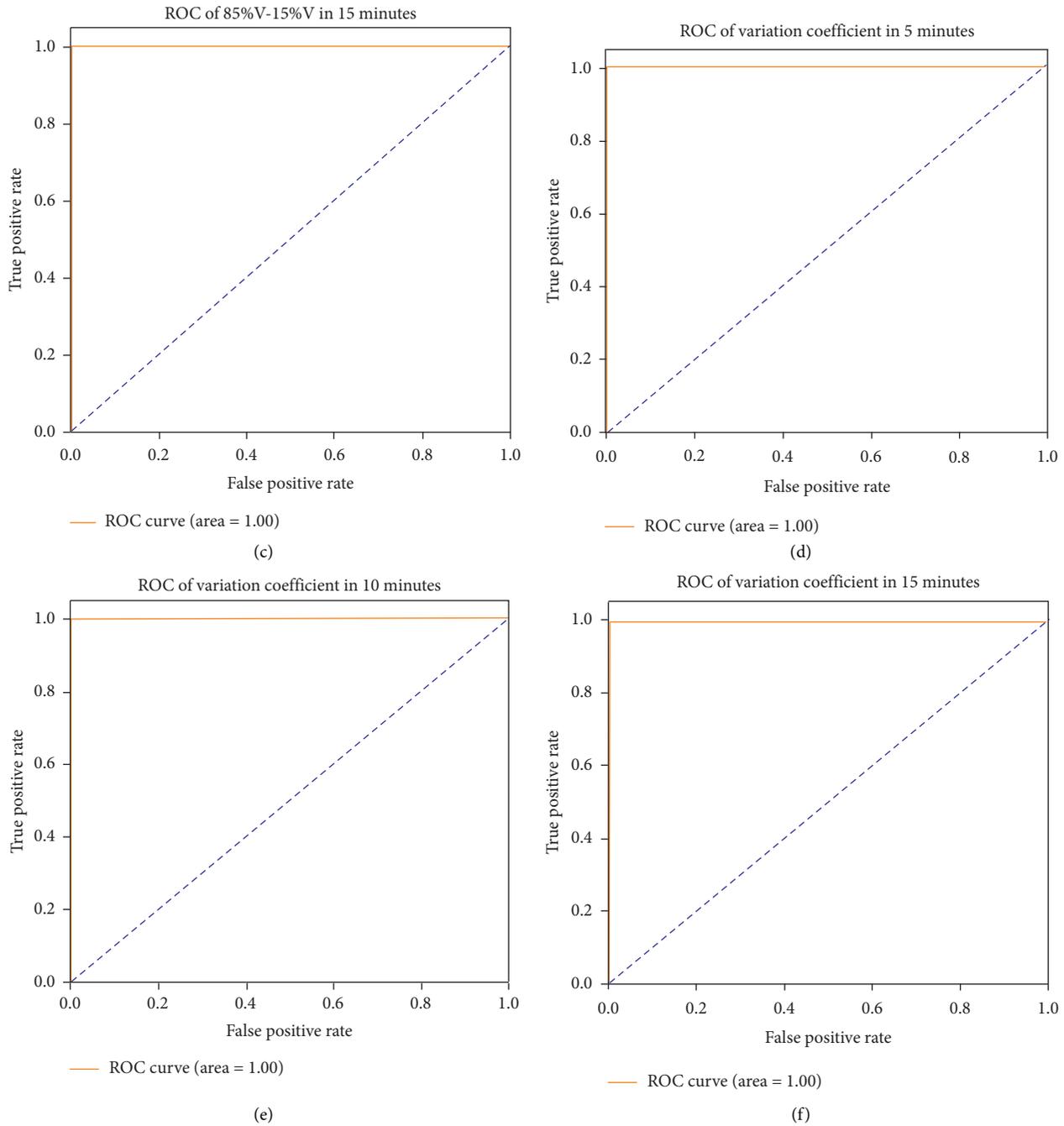


FIGURE 6: The ROC curve for SVM of 85%V-15%V and variation coefficient in different periods.

period of 5 minutes, while there were three samples of number 2 meaning dangerous status classified into false status. Therefore, the accuracy of verification of the testing data of the variation coefficient was  $98.6\% = (140 + 77) / (140 + 3 + 77) * 100\%$ . The correct rate scores of the training data of the variation coefficient were 1.0 and 0.994 in periods of 10 minutes and 15 minutes, respectively. The accuracy of verification of the testing data variation coefficient was 100%, which indicates good classification of traffic status.

In addition, the receiver operating characteristic curve (the ROC curve), also known as the susceptibility curve

(sensitivity curve), can also be used to evaluate the performance of SVM because it can determine the accuracy of the classification of the parameter. It is based on a series of different methods of binary classification (boundary value or decision threshold). The value of 85%V-15%V and the variation coefficient can represent the traffic flow safety status to some extent. Therefore, the area under the ROC curve (AUC) of each test could also be calculated separately for comparison. The larger AUC was, the better the diagnostic result was. The value of AUC was from 0 to 1, and the classification algorithm of SVM was an uninformative

TABLE 6: Traffic status level of the classification of indicators and truck proportion.

Indicators of evaluation	Period of time	Threshold of the classification	Truck proportion	Traffic flow status
85%V-15%V	5 minutes	$\geq 41.503$ km/h	0.2-0.6	Dangerous
		$< 41.503$ km/h	0-0.4, 0.5-1.0	Safe
	10 minutes	$\geq 42.127$ km/h	0.2-0.6	Dangerous
		$< 42.127$ km/h	0.1-0.3, 0.6-0.9	Safe
	15 minutes	$\geq 42.672$ km/h	0.2-0.6	Dangerous
		$< 42.672$ km/h	0.1-0.3, 0.6-0.9	Safe
Speed variation coefficient	5 minutes	$\geq 0.224$	0.2-0.6, 0.8-1.0	Dangerous
		$< 0.224$	0.1-0.4, 0.7-1.0	Safe
	10 minutes	$\geq 0.224$	0.2-0.6, 0.8-1.0	Dangerous
		$< 0.224$	0.1-0.4, 0.7-0.9	Safe
	15 minutes	$\geq 0.223$	0.4-0.6	Dangerous
		$< 0.223$	0.1-0.4, 0.7-0.9	Safe

classifier if it was less than 0.5; the classification algorithm had a certain accuracy if the AUC ranged from 0.7 to 0.9, and it was the best classifier when the high accuracy of AUC was greater than 0.9. Nevertheless, if the AUC varied from 0.5 to 0.7, the algorithm had low accuracy. When the AUC was less than 0.5, the diagnosis method was completely ineffective and had no diagnostic value.

In Figure 6, the first three pictures are the ROC curve and the area under the curve (AUC) of 85%V-15%V in three periods, while the rest of the pictures are the ROC curve and the area under the curve (AUC) of the variation coefficient in three periods. The AUCs of 85%V-15%V in these periods were both 1, which indicates the high accuracy of the SVM model. Therefore, it was the same situation for the variation coefficient. In summary, the confusion matrix, ROC curve, and area under the curve (AUC) were used to evaluate the accuracy of the SVM model.

Above all, the value of 85%V-15%V and speed variation coefficient could reveal the impact of the truck proportion on the traffic safety. Therefore, the selected parameters served as surrogate safety measures to indirectly express the impact of truck proportion on freeway traffic safety. Table 6 is the result of the classification of 85%V-15%V and speed variation coefficient. From Table 6, it can be seen that, regardless of the indicator of 85%V-15%V and the indicator of speed variation coefficient, the truck proportion in the range from 0.4 to 0.6 was in a safety traffic status when the threshold of 85%V-15%V was above approximately 42 km/h and the threshold of speed variation coefficient was above approximately 0.223. However, truck proportions ranging from 0.1 to 0.3 and from 0.7 to 0.9 were in a dangerous traffic status otherwise. In particular, in the threshold of the classification of 85%V-15%V above approximately 42 km/h and the variation coefficient above 0.223, the safety status was at truck proportions of 0.2 and 0.8. Otherwise, it was in a dangerous status at a truck proportion of 0.5, specifically.

In addition, it is known that the traffic flow safety status is related to speed parameters, such as the speed of the vehicle, speed difference, speed standard deviation, 85% V-15%V, and variation coefficient. It depends not only on the absolute speed but also on the speed variation coefficient, which is related to the degree of dispersion of the vehicle speed. Thus, the control of the speed [39] will be helpful for

improving the safety of trucks on the freeway. The average speed, speed dispersion, variation coefficient of speed, traffic regulations, and the possibility of implementation of speed control standards and other factors should be considered in the development of speed control standards. According to the relationship between the truck proportion and the average speed of the truck and the relationship between the truck proportion and the average speed of the cars, the value of 85%V-15%V should be used to determine the speed limit; for example, 85%V could be used for the highest speed and 15%V could be used for the lowest speed [40, 41]. Then, the value of 85%V-15%V is less than the threshold of the classification, which could reduce the probability of traffic accidents to some extent.

## 5. Conclusion

This study explored the impact of truck proportion on traffic safety by using surrogate safety measures such as traffic flow parameters. Through *k*-means clustering, the traffic surrogate parameters were classified into two statuses that could determine the traffic status in different periods by different truck proportions. Then, the support vector machine algorithm was used for training, testing, and verifying the classification and finding the threshold of the classification of different indicators. In this process, the characteristics of traffic flow parameters, including the average speed, speed difference, number of lane changes, speed variation coefficient, and value of 85%V-15%V with truck proportion, were analyzed by correlation analysis and partial correlation analysis. The correlation of the number of lane changes is less than 0.3, and the *p* value is more than 0.05, which is not significant bilaterally; the partial relationship is still weaker. Thus, 85%V-15%V and the speed variation coefficient were selected as surrogate safety measures to evaluate traffic flow safety status. Finally, the conclusion of the impact of the truck proportion on freeway traffic safety can be drawn.

The results show that the relationship between truck proportion and the value of 85%V-15%V and the speed variation coefficient is generally consistent in different aggregation periods, showing the impact of truck proportion on traffic safety. When the value of 85%V-15%V is above 42 km/h, the traffic flow status tends to be dangerous with

truck proportions ranging from 0.2 to 0.6, especially from 0.4 to 0.6. When the value of  $85\%V-15\%V$  is under 42 km/h, the traffic flow status is relatively safe with truck proportions ranging from 0.1 to 0.3 and from 0.6 to 0.9, especially from 0.6 to 0.9. When the speed variation coefficient is above 0.223, the traffic flow status tends to be dangerous, with the truck proportion ranging from 0.4 to 0.6. When the speed variation coefficient is under 0.223, traffic flow status is relatively safe, with truck proportions ranging from 0.1 to 0.4 and from 0.7 to 0.9, especially from 0.7 to 0.9. In conclusion, the traffic flow status tends to be dangerous when the truck proportion ranges from 0.4 to 0.6 and when the value of  $85\%V-15\%V$  and the speed variation coefficient are above 42 km/h and 0.223, respectively. While the truck proportion is from 0.1 to 0.3 and from 0.7 to 0.9, the traffic flow is relatively safe on the condition that the value of  $85\%V-15\%V$  and the speed variation coefficient were under 42 km/h and 0.223, respectively.

In addition, the  $k$ -means clustering and SVM are both appropriate algorithms for exploring traffic status classification. The results indicate that surrogate safety measures could be effectively classified as different risk levels to identify the traffic flow status of safety or unsafety by using the  $k$ -means clustering algorithm, while the support vector machine could establish the relationship among parameters quite well and indicate the influence of the truck proportion on traffic safety by surrogate safety measures. Compared with existing statistical methods, it does not involve the probability measure and the law of large numbers and greatly simplifies the usual classification and regression problems. Therefore, it is quite likely to fit well with two-classification problems in the traffic field, such as classifying driving patterns and traffic flow status. However, it should be noted that the classical SVM is only limited to two-classification problems. The multiclassification problems should resort to the multiclass SVM or other methods.

The obtained conclusions could be meaningful in making an effective countermeasure on freeways of different truck proportions in different situations. To improve freeway traffic safety, the type of vehicles, especially trucks, and traffic speed on freeways should be controlled by traffic management departments. The improvement measures could be developed based on the above-mentioned analysis and real-time control of truck proportion from freeway entrances. For example, traffic flow restrictions, including the speed limit and truck traffic flow control, could be utilized in toll station entrances. Therefore, the truck proportion should be considered in setting reasonable speed limits if possible, which may improve freeway safety. In addition, effective freeway freight routes could be developed to control the truck proportion on the freeway, and the safe operation and management of trucks on the freeway should also be promoted.

### Data Availability

The data used in this paper to support the findings of our study were supplied by Yingying Xing under license and so cannot be made freely available. Requests for access to these

data should be made to Yingying Xing, and contact e-mail is [yingying199004@tongji.edu.cn](mailto:yingying199004@tongji.edu.cn).

### Conflicts of Interest

The authors declare no conflicts of interest.

### Acknowledgments

This research was supported by projects of the National Key R&D Program of China (no. 2018YFB201403); the Science and Technology Project of Shanxi Provincial Department of Transportation (Project no. 16002380370); research in the impact of truck proportion on freeway traffic safety, China; National Natural Science Foundation's youth science fund project (Project no.71601143); Shanghai Pujiang talent planning project (Project no. 16PJC088); and the China Scholarship Council.

### References

- [1] National Bureau of Statistics of China: China Statistical Yearbook 2016: Road Safety, National Bureau of Statistics of China Beijing, China 2016, <http://data.stats.gov.cn/easyquery.htm?cn=C01>.
- [2] C. Dong, D. B. Clarke, S. H. Richards, and B. Huang, "Differences in passenger car and large truck involved crash frequencies at urban signalized intersections: an exploratory analysis," *Accident Analysis & Prevention*, vol. 62, pp. 87–94, 2014.
- [3] S. Choi, C. Oh, and M. Kim, "Risk factors related to fatal truck crashes on Korean freeways," *Traffic Injury Prevention*, vol. 15, no. 1, pp. 73–80, 2014.
- [4] C. Dong, Q. Dong, B. Huang, W. Hu, and S. S. Nambisan, "Estimating factors contributing to frequency and severity of large truck-involved crashes," *Journal of Transportation Engineering, Part A: Systems*, vol. 143, no. 8, p. 04017032, 2017.
- [5] K. Li, K. Ozbay, H. Yang, J. Holguín-Veras, and E. F. Morgul, "Modeling safety impacts of off-hour delivery programs in urban areas. Transportation Research Record," *Journal of the Transportation Research Board*, vol. 2478, pp. 19–27, 2015.
- [6] Q. Hou, A. P. Tarko, and X. Meng, "Analyzing crash frequency in freeway tunnels: a correlated random parameters approach," *Accident Analysis & Prevention*, vol. 111, pp. 94–100, 2018.
- [7] T. F. Golob, W. Recker, and Y. Pavlis, "Probabilistic models of freeway safety performance using traffic flow data as predictors," *Safety Science*, vol. 46, no. 9, pp. 1306–1333, 2008.
- [8] C. Xu, P. Liu, and W. Wang, "Evaluation of the impacts of traffic states on crash risks on freeways," *Accident Analysis & Prevention*, vol. 47, pp. 162–171, 2012.
- [9] R. Yu and M. Abdel-Aty, "Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data," *Safety Science*, vol. 63, no. 4, pp. 50–56, 2014.
- [10] D. Kong, X. Guo, B. Yang et al., "Analyzing the impact of trucks on traffic flow based on an improved cellular automaton model," *Discrete Dynamics in Nature and Society*, vol. 2016, Article ID 1236846, 14 pages, 2016.
- [11] W. Zhizhou, L. Hao, and Y. Hongsu, "Variable speed limit control on highway work zone considering large vehicle mix rates," *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 1784–1791, 2013.

- [12] X. Li, X. Li, Y. Xiao, and B. Jia, "Modeling mechanical restriction differences between car and heavy truck in two-lane cellular automata traffic flow model," *Physica A: Statistical Mechanics and Its Applications*, vol. 451, pp. 49–62, 2016.
- [13] H. El-Sayed and Y.-A. Daraghmi, "Accurate traffic flow prediction in heterogeneous vehicular networks in an intelligent transport system using a supervised non-parametric classifier," *Sensors*, vol. 18, no. 6, p. 1696, 2018.
- [14] J. Liu and E. Zio, "A scalable fuzzy support vector machine for fault detection in transportation systems," *Expert Systems with Applications*, vol. 102, pp. 36–43, 2018.
- [15] S. Y. Sohn and S. H. Lee, "Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea," *Safety Science*, vol. 41, no. 1, pp. 1–14, 2003.
- [16] S. Feng, Z. Li, Y. Ci, and G. Zhang, "Risk factors affecting fatal bus accident severity: their impact on different types of bus drivers," *Accident Analysis & Prevention*, vol. 86, pp. 29–39, 2016.
- [17] L. Li, S. Sharad, and G. Hu, "Analysis of road traffic fatal accidents using data mining techniques," in *Proceedings of the 2017 IEEE 15th International Conference on Management and Applications (SERA)*, London, UK, July 2017.
- [18] M. G. Mohamed, N. Saunier, and L. F. Miranda-Moreno, "A clustering regression approach: a comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US and Montreal, Canada," *Safety Science*, vol. 54, pp. 27–37, 2013.
- [19] K. Polat and S. S. Durduran, "Automatic determination of traffic accidents based on KMC-based attribute weighting," *Neural Computing and Applications*, vol. 21, no. 6, pp. 1271–1279, 2012.
- [20] S. Kumar, D. Toshniwal, and M. Parida, "A comparative analysis of heterogeneity in road accident data using data mining techniques," *Evolving Systems*, vol. 8, no. 2, pp. 147–155, 2017.
- [21] N. Dong, H. Huang, and L. Zheng, "Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects," *Accident Analysis & Prevention*, vol. 82, pp. 192–198, 2015.
- [22] M. Effati, "Prediction of crash severity on two-lane, two-way roads based on fuzzy classification and regression tree using geospatial analysis," *Journal of Computing in Civil Engineering*, vol. 29, no. 6, Article ID 04014099, 2014.
- [23] A. Theofilatos, C. Antoniou, A. Chaziris, and D. Sermpis, "Time series and support vector machines to predict powered-two-wheeler accident risk and accident type propensity: a combined approach," *Journal of Transportation Safety & Security*, vol. 10, no. 5, pp. 471–490, 2018.
- [24] Z. Li, "Comparison of drunk driving recognizing methods based on KNN and SVM," *Journal of Transportation Systems Engineering and Information Technology*, vol. 5, p. 035, 2015.
- [25] W. Zou, X. Wang, and D. Zhang, "Truck crash severity in New York city: an investigation of the spatial and the time of day effects," *Accident Analysis & Prevention*, vol. 99, pp. 249–261, 2017.
- [26] K. Moses and E. Y. Yamashita, "Using a k-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii," *Journal of Advanced Transportation*, vol. 41, no. 1, pp. 69–89, 2007.
- [27] C. G. Prato, S. Bekhor, A. Galtzur, D. Mahalel, and J. N. Prashker, "Exploring the potential of data mining techniques for the analysis of accident patterns," in *Proceedings of the 12th World Conference on Transportation Research*, Lisbon, Portugal, July 2010.
- [28] S. Saharan and R. Baragona, "A cluster analysis on road traffic accidents using genetic algorithms," in *Proceedings of the American Institute of Physics Conference Series (AIP) Conference Proceedings*, College Park, MA, USA, April 2017.
- [29] A. Kocatepe, E. E. Ozguven, A. Vanli, and R. Moses, "Analysis of speed patterns on inter-urban parallel highways: a case study in the southeast Florida," *Transportation Research Procedia*, vol. 22, pp. 479–488, 2017.
- [30] D. Vapnik, Z. Deng, Y. He, C. Wu, C. Sun, and Z. Lu, "Curve speed model for driver assistance based on driving style classification," *Intelligent Transport Systems*, vol. 11, no. 8, pp. 501–510, 2017.
- [31] S. Chen, M. Li, and N. Zhao, "A study on traffic risk assessment of road section with high rate of bridge and tunnel of freeway," in *Proceedings of the International Conference on Intelligent Computation Technology and Automation*, IEEE, Changsha, China, pp. 102–105, 2016.
- [32] H. U. Li, Y. Lin, and J. P. Gao, "Real-time safety probability evaluation of the expressway traffic flow," *Journal of Safety & Environment*, vol. 131, 2015.
- [33] C. Cortes, "Support vector network," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [34] H. Chen and L. Chen, "Support vector machine classification of drunk driving behaviour," *International Journal of Environmental Research and Public Health*, vol. 14, no. 1, p. 108, 2017.
- [35] Y. L. Xin, H. Zhen, W. C. Zhong, Q. L. Qiao, Z. D. Yao, and B. Ran, "Driving mode selection of intelligent vehicles based on risky situation identification," *Journal of South China University of Technology*, vol. 44, pp. 139–146, 2016.
- [36] R. Yu and M. Abdel-Aty, "An optimal variable speed limits system to ameliorate traffic safety risk," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 235–246, 2014.
- [37] L. Aarts and I. Van Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.
- [38] M. Shirazi, D. Lord, and S. R. Geedipally, "Sample-size guidelines for recalibrating crash prediction models: recommendations for the Highway Safety Manual," *Accident Analysis & Prevention*, vol. 93, pp. 160–168, 2016.
- [39] F. K. Afukaar, "Speed control in developing countries: issues, challenges and opportunities in reducing road traffic injuries," *Injury Control and Safety Promotion*, vol. 10, no. 1–2, pp. 77–81, 2003.
- [40] Y. Hou, C. Sun, and P. Edara, "Statistical test for 85th and 15th percentile speeds with asymptotic distribution of sample quantiles. Transportation Research Record," *Journal of the Transportation Research Board*, vol. 2279, pp. 47–53, 2012.
- [41] B. J. Russo, P. T. Savolainen, T. J. Gates, J. J. Kay, and S. Frazier, "Driver speed selection on high-speed two-lane highways: comparing speed profiles between uniform and differential speed limits," *Traffic Injury Prevention*, vol. 18, no. 5, pp. 521–527, 2016.