

Research Article

Exploring Relationships between Months and Different Crash Types on Mountainous Freeways Using a Combined Modeling Approach

Changjian Zhang,¹ Jie He,¹ Chunguang Bai ,¹ Xintong Yan,¹ Chenwei Wang,¹ and Yazhong Guo²

¹School of Transportation, Southeast University, Nanjing, China

²Jiangsu Communications Planning and Design Institute Limited By Share Ltd, Nanjing, China

Correspondence should be addressed to Chunguang Bai; baichg@seu.edu.cn

Received 21 March 2022; Revised 31 May 2022; Accepted 30 June 2022; Published 19 July 2022

Academic Editor: Jaeyoung Lee

Copyright © 2022 Changjian Zhang 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.

Investigating the relationship between the months and traffic crashes is a foremost task for the safety improvement of mountainous freeways. Taking a mountainous freeway located in China as an example, this paper proposed a combined modeling framework to identify the relationships between months and different crash types. *K*-means and Apriori were initially used to extract the monthly distribution patterns of different types of crashes. A graphical approach and a risk calculation equation were developed to assess the output of *K*-means and Apriori. Then, using the assessment results as the input, a logistic regression model was constructed to quantify the effects of each month on crashes. The results indicate that the monthly distribution patterns of different crash types are inconsistent, i.e., for a specific month, the high risk of a certain crash type may be covered up if experts only focus on the total number of crashes. Moreover, when identified as high-risk months by *K*-means and Apriori, the crash-proneness will significantly increase several times than months identified as high-risk by only one of *K*-means and Apriori, thereby illustrating the superior performance of the mix-method. The conclusions can assist local relevant organizations in formulating strategies for preventing different types of traffic crashes in different months (e.g., the risk of rear-end crashes in August, the risk of fixed-object hitting crashes in February, and the risk of overturning crashes in October) and provide a methodological reference for relevant studies in other regions.

1. Introduction

According to the World Health Organization (WHO), road traffic injury is the eighth leading cause of death in all age groups and the leading cause of death in children and young adults, aged 5 to 29 years [1]. To reduce the damage of crashes, recent studies have focused on exploring the relationship between traffic crashes and other factors [2–5]. These microlevel studies aim to investigate the impact of contributing factors on the probability of crashes. Compared with microlevel research, macrolevel research can provide more direct insight into guide practice. For instance, examining the spatial distribution patterns of crashes and determining the black spots can assist the road agencies in developing the

targeted countermeasures to reduce the frequency of crashes in crash-prone locations. Despite these achievements in revealing the spatial distribution of crashes based on the black-spot theory, many subtleties have not been explicitly understood relating to the temporal distribution patterns of crashes. Furthermore, while the recent technological advances have made traffic conditions increasingly convenient, the crash rate of mountainous roads remains significantly higher than that of ordinary roads because of many detrimental factors, such as the complicated driving environment and ineffective safety system [6]. Hence, to help enhance traffic safety and mitigate crashes in mountainous areas, there is a great necessity to understand the temporal distribution characteristics of crashes on mountainous roads.

Another issue worthy of attention is that there are also discrepancies in the temporal distribution patterns of crashes in different countries and regions because of the inconsistency of travel habits, national policies, traffic control regulations, road design, etc., [7, 8]. Thus, caution should be exercised in projecting some of the conclusions in one study to another, especially when they are conducted based on different contexts. Because of owning the largest population and the most motor vehicles in the world, China is always confronted with formidable challenges regarding the appropriate planning and operation to ameliorate road safety. Toward this end, it is necessary to explore the temporal distribution patterns of crashes in mountainous areas in China. The rest of this paper is structured as follows: sections 1.1 and 1.2 provide a review of the relevant research. Section 2 introduces the data sources and data processing procedures. Then, the combined modeling framework is described in Section 3, and the estimation results are presented in Section 4. Finally, Section 5 summarizes the findings and limitations of the study.

1.1. Time Distribution of Traffic Crashes. There is a lack of relevant literature on exploring the temporal distribution patterns of crashes. Most related studies regard “time” as a variable in the causality model to investigate its effect on the crash frequencies/severities. For such types of studies, the contribution of different periods to crashes can be revealed based on the estimation results of the model. Some researchers point out that relative to off-peak hours in the daytime, crashes during peak hours are found to be more severe (both serious injury and fatality). This result might be attributed to stress, frustration, aggression, and fatigue when driving on congested roads during peak hours [9–11]. It is consistent with findings from the study [12], which proposed that 17:00 to 20:00 is the period with the highest incidence of traffic crashes. Besides, the crash severity of nighttime is found to be more serious than that of daytime for multi-vehicle crashes [13, 14]. Ackaah et al. (2020) speculated that the poor night visibility coupled with poor visual guidance on roads is the critical factor leading to more serious crashes [15]. Because of the divergencies of different periods in influencing the crashes, the effects should be separately investigated with disaggregate models to ensure the accuracy of the analysis [16]. Note that some researchers have investigated the weekly/monthly/quarterly crash distribution characteristics. For instance, Li et al. (2013) simply analyzed the temporal distribution characteristics of roads in mountainous areas using a broken line diagram and found that the first and second quarters experience more crashes than the remaining two quarters. For the monthly distribution patterns, they pointed out that the frequency of crashes in January and June is significantly higher than that in other months [17]. Yadollahi (2019) documented a retrospective cohort study using traffic crashes in Shiraz city [7], with the results suggesting that most crashes tended to occur on Thursdays (15.09%) instead of Mondays (13.85%) or Saturdays (13.95%). In all months, August and September have witnessed the most crashes, while December and April

have witnessed the least crashes. In other studies, April, October, and weekends are found to result in serious crashes [18, 19]. Obviously, there exist differences in the conclusions of these studies, which may be caused by the different climate, socioeconomic characteristics, and topography [20].

Several limitations can be summarized from the existing research. Firstly, the past work pays more attention to the hourly distribution pattern of crashes. However, the monthly distribution patterns of crashes can provide more effective insights into formulating the macrolevel safety strategies. Secondly, most studies only regard temporal factors as the explanatory variables in the model rather than devoted to the analysis of crash time distribution patterns. Finally, the methods used in the previous studies are usually very simple (e.g., histogram and broken line diagram), which may lead to inaccurate and unsystematic conclusions. Consequently, to systematically explore the inherent relationship between traffic crashes and months, this study aims to develop a combined modeling approach. A mountainous freeway located in China is taken as the research object. Besides, the monthly distribution of different crash types (overall crashes, rear-end, fixed-object hitting, and overturning) are investigated in this study. This study may contribute to improving analysis accuracy and providing practical insights for road crash interventions.

1.2. Approach Analysis Review. Regarding models in traffic crash analysis, the nonparametric models represented by machine learning and parametric models represented by the regression model are the most common. Previous studies prefer to compare different models and then select the one with a better performance. It is certainly a valid approach. However, different types of models often have different advantages. This study integrates the results of various types of modeling approaches and tries to improve the reliability of the study from the perspective of the combined application. Parametric techniques are good at quantitatively analyzing the influencing factors and model results, while nonparametric models have good performance at qualitative fitting and classification but could have a weak interpretation of results [21, 22]. These two types of models can just complement each other’s characteristics, e.g., using machine learning algorithms to identify the relationship between crashes and months and then employing the parametric model to quantitatively analyze the patterns extracted by nonparametric models. It not only improves the accuracy but also ensures the interpretability of the results.

1.2.1. Nonparametric Models. The nonparametric model is used to investigate the corresponding relationship between different months and different types of crashes. Since this paper only involves crash frequency and the occurrence time of each crash, the basic machine learning model is sufficient for the current research.

Apriori, originally introduced by Agrawal et al. (1993), is a representative algorithm that states the association rules between attributes and objectives [23]. Because of the satisfactory data-processing capability and low requirements

for the data types, it has become the foundation of many other data mining techniques [24]. Usually, it is utilized as a desirable instrument to untangle the repeated patterns from the dataset [25, 26]. John and Shaiba (2019) applied the Apriori to mine frequent item sets and identify the major causes and trends associated with road crashes [27]. Mohammed et al. (2018) implemented the Apriori algorithm and a clustering method for traffic datasets to discover the factors associated with crashes [28]. Deng et al. (2018) presented a causation analysis model for traffic crashes based on a hybrid Apriori-Genetic algorithm [29].

To enhance the reliability of the output obtained from the subsequent parametric model, *K*-means clustering will also be employed to identify the monthly distribution patterns of crashes. The results of these two machine learning algorithms will be combined as the input to the parametric model. *K*-means is a typical clustering method that has been extensively used in identifying the road black spots [30]. For instance, Dadashova et al. (2016) applied *K*-means to identify the crash patterns and trends for off-road truck-related crashes in the U.S. [31]. Almjewail et al. (2018) employed *K*-means to understand traffic crash characteristics and identify black spots [32]. Qu et al. (2013) attempted to unveil the spatial similarities in traffic crashes, with the finding indicating that the clustering method could explore and visually symbolize the crash attributes [12]. Zhang et al. (2019) improved the *K*-means algorithm for filtering traffic crashes in black spots [33]. Similarly, if the “month” is compared to the “road segment”, *K*-means can also be employed to determine the “black spots” in the temporal dimension.

1.2.2. Parametric Model. Logistic regression (LR) has been widely applied in modeling the relationship between crash risks (dependent variable) and various contributors (independent variables) [34–36]. Since there is little literature on the temporal distribution patterns of traffic crashes, the following review mainly focuses on the application of the LR model in the field of crash analysis. Specifically, Abdel-Aty and Pemmanaboina (2006) developed a linear LR model to estimate the impact of real-time traffic flow and weather on crashes [37]. Li et al. (2019) combined the mixed LR model and the latent class LR models to investigate driver injury severities in rural single-vehicle crashes under rain conditions and provided beneficial references for severe injury prevention [38]. Dong et al. (2018) explored the differences between single-vehicle (SV) and multivehicle (MV) crash probability using a mixed LR model [39]. The results indicated that the length of segments and wet road surfaces are significant for SV and MV crashes, while most of the other variables are significant only for MV crashes. Ahmed et al. (2021) applied random parameters binary LR models to explain the heterogeneity of unobservable variables associated with deer-vehicle collisions and their resulting injury severities [40]. Zhang et al. (2021) used the random parameters LR model to identify the relationship between tire forces and road characteristics [41].

The outcomes of the LR model can predict the probability of crashes and estimate the marginal effects of each explanatory variable. However, it should be noted that this method is highly dependent on the assumptions for data distribution. Referring to other research studies, the pre-defined basic relationships between variables might be restricted by the model itself, which is another important limitation [42, 43]. To fill this gap, this study will propose a joint method by combining the *K*-means, Apriori, and LR model.

1.3. Research Goals and Contributions. The innovations and contributions of this paper can be summarized in the following four points:

- (i) This paper explores the temporal distribution characteristics of the overall crashes, rear-end crashes, fixed-object hitting crashes, and overturning crashes, respectively. Besides, a comprehensive comparison among them is also conducted to understand their differences in temporal distribution.
- (ii) In terms of methodology, a combined method integrating three basic nonparametric and parametric models is proposed and applied to extract the temporal distribution patterns of traffic crashes. The proposed framework of integrated application can provide a reference for other studies, and the results also verify the necessity and effectiveness of this approach.
- (iii) This study calculates how much higher the risk of crashes in the months that experience more crashes is than in months that experience fewer crashes. Such results provide a deeper insight into the degree of danger in different months to warn drivers to drive more cautiously in these months by establishing preventive measures.
- (iv) The research conclusions can not only guide relevant departments to prevent different types of traffic crashes in a targeted manner but also lay the foundation for the in-depth cause analysis in the future.

2. Data

This study utilizes an eight-year dataset, from 2006 to 2013. The target freeway is located in Zhejiang Province, China, characterized by mountainous terrain. As the length of tunnels and bridges accounts for up to 47%, the complex road environment increases the risks of crashes and near-crash events, resulting in a high crash frequency every year.

Traffic crash data provided by the Zhejiang Provincial Department of Transportation include the following three types of statistics: (1) time of each crash, (2) the location of each crash, and (3) different types of crashes, such as rear-ends, fixed-object hitting, and overturning.

The total number of crashes is 3998, including 975 rear-end crashes, 2276 fixed-object hitting crashes, 422

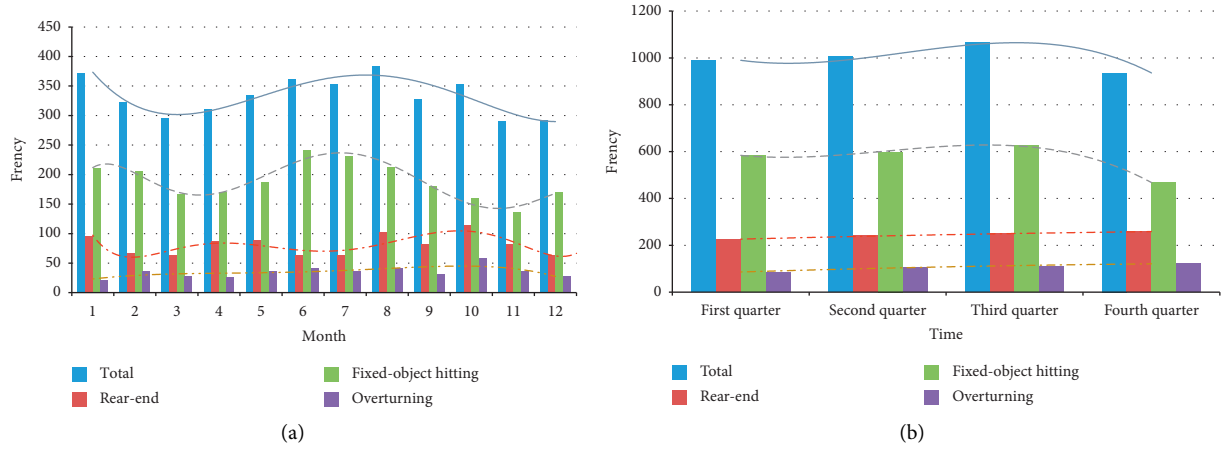


FIGURE 1: Temporal distribution of crashes: (a) crash type distribution by month and (b) crash type distribution by quarter.

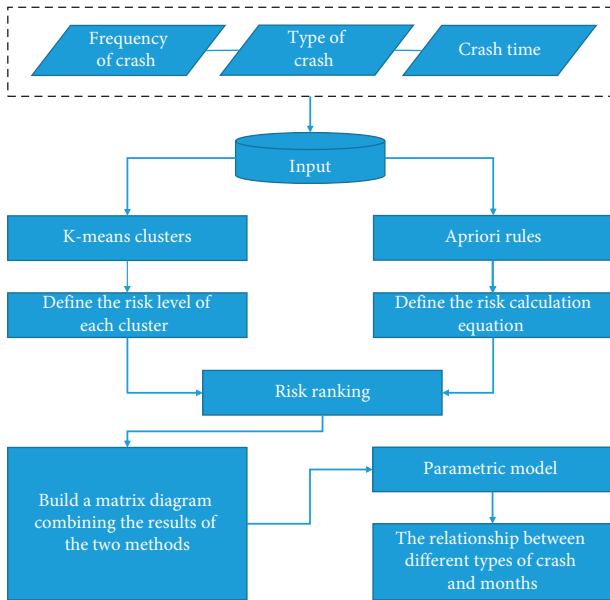


FIGURE 2: The framework of the combined method.

overturning crashes, and 325 other types of crashes. The possible reason for the low proportion of rear-end crashes may be that the speed limit of the target freeway is low (80 km/h–100 km/h) and drivers are more cautious when driving on roads with too many bridge-tunnel groups. Considering that the main objective of this study is to identify the monthly distribution patterns of different types of traffic crashes, each crash is classified into a monthly dataset according to the crash time. The given rear-end, fixed-object hitting, and overturning crashes are the most typical types, and this study will focus on an in-depth analysis of these three crash types. The statistics of crashes are shown in Figure 1.

3. Methodology

The framework of the proposed combined method is shown in Figure 2.

3.1. K-Means Clusters. The distance between data is taken as the criterion of similarity measurement (cosine distance, European distance, Manhattan distance, etc.) of data objects by the K -means [33]. According to Yan et al. (2020) [44], this paper took Euclidean distance as a calculating standard because of its reliability and generality.

The European distance formula is as follows:

$$\text{dist}(X_i, X_j) = \sqrt{\sum_{d=1}^D (x_{i,d} - x_{j,d})^2}. \quad (1)$$

The data objects can be clustered into K categories based on Equation (1). For the dataset, the mean value of all data in the relative class is initially selected as the class center, which needs iterating until the class center changes slowly or stops changing (the squared error between the empirical mean of a cluster and the points in the cluster is minimized [45]). The class center can be defined as follows:

$$\text{center}_k = \frac{1}{|C_k|} \sum_{X_i \in C_k} x_i, \quad (2)$$

where C_k represents class k and $|C_k|$ represents the number of data objects in class k .

The iterating process can be calculated through the following:

$$J = \sum_{k=1}^K \sum_{X_i \in C_k} \text{dist}(X_i, \text{center}_k). \quad (3)$$

3.2. Apriori Rules. Apriori rules can be presented as the form of $X \rightarrow Y$, where X and Y are disjoint item sets, i.e., $X \cap Y = \emptyset$ [46]. The key advantage of using Apriori is that it can be measured by its support and confidence. Support (s) can be used to determine the frequency of a given dataset, while confidence (c) determines the frequency of Y in transactions containing X . The forms of support and confidence are defined as follows [44]:

$$s(X \longrightarrow Y) = \frac{\sigma(X \cup Y)}{N}, \quad (4)$$

$$c(X \longrightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}. \quad (5)$$

In this paper, (5) would be used to calculate the probability of a certain type of traffic crash risk in a given month.

3.3. Parametric Model (LR Model). The LR model, as a representative of the discrete choice model, has been widely used in crash cause analyses in the field of traffic safety [47]. In type b crash, the risk probability of month i for this crash type being relatively high and relatively low can be, respectively, defined by the LR model as follows:

$$P(Y_{bi} = 1 | X_i) = \frac{1}{1 + e^{-g(X_i)}}, \quad (6)$$

$$P(Y_{bi} = 0 | X_i) = 1 - \frac{1}{1 + e^{-g(X_i)}} = \frac{1}{1 + e^{g(X_i)}}.$$

X_i is the sequence of explanatory variables, $Y_{bi} = 1$ and $Y_{bi} = 0$ indicate the months with relatively high crash risk and relatively low crash risk, respectively. The variable odds denotes the ratio of the observed “relatively high risk” to “relatively low risk” probability.

After taking a logarithm of odds, the linear function is as follows:

$$\begin{aligned} \ln(\text{odds}) &= \ln\left(\frac{P(Y_{bi} = 1 | X_i)}{P(Y_{bi} = 0 | X_i)}\right) = \ln\left(e^{g(X_i)}\right) \\ &= g(X_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \end{aligned} \quad (7)$$

Additionally, to calculate the intercept β_0 and coefficients in (7), the maximum likelihood estimation method will be used.

The LR model is a multivariate analysis method that can interpret the relationship between binomial observation results and their determinants. In this study, the months are selected as the independent variables, while different types of crashes in different months are converted into discrete dependent variables according to the results of K -means and Apriori, which indicates whether the frequency of the corresponding crash type is “relatively high” and “relatively low”.

4. Results

As illustrated in Figure 1, the histogram shows that all types of crashes have an observable trend over the month. In terms of the total number of crashes, March, November, and December are associated with lower possibilities of crashes, whereas January, June, July, August, and October tend to suffer from more crashes. The monthly distribution pattern of fixed-object hitting crashes is almost consistent with the total number of crashes, while the rear-end and overturning crashes display different changes, with the opposite trend in certain periods. Although some qualitative information can

be interpreted from the distribution trend of the histogram, the information may not be accurate. For example, a certain month may result in a surge in the number of crashes because of road construction or temporary special management policies in a certain year. It is conceivable that if the monthly distribution pattern of crashes is not investigated on some dangerous roads, these unobservable risks will be completely covered up by the annual statistics of crashes.

4.1. Risk Ranking under a Single Nonparametric Model. Two methods are used to determine the proper number of clusters: the sum of within-group squared error (SSE) and the Calinsky criterion. For SSE, it aims to search for a point that splits the SSE function domain into two parts. Until this point, each added cluster results in a substantial reduction in the value of variance, and after the given point, any increase in k leads to a less-and-less reduction in the value of variance [48]. The Calinsky criterion employs a ratio of between-cluster variance and the overall within-cluster variance. Well-defined clustering solutions yield high values of Calinsky criterion [49, 50].

As shown in Figures 3 and 4, the number of clusters for total crashes is 5. In the process of determining the number of clusters, the best cluster numbers of rear-end, fixed-object hitting, and overturning crashes identified by Calinsky criterion are 10, 10, and 9, respectively. However, the excessive number of clusters may lead to biased conclusions as there are only 12 months in total. Hence, combined with SSE, using the inflection point of decreasing rate associated with SSE as the standard, the proper number of clusters is set as 3, 5, and 5, respectively. As presented in Figure 4, although they are not the optimal numbers, except for fixed-object hitting, the number selected is also the inflection point using the Calinsky criterion.

It should be noted that SSE and Calinsky criterion have poor performance on fixed-object hitting crashes. In fact, although not obvious, there is an inflection point in the decline rate of SSE when the number of clusters is 5. After the inflection point, the decline of error becomes stable.

Table 1 further explains the differences between clusters through the average frequency. Since the subject of the paper is the same road, the number of a certain type of crash of a certain month may more reliably reflect crash risk levels and is reasonable enough to be the basis for classification. All clusters are divided into 3 levels, denoted as “3,” “2,” and “1,” respectively, indicating the risk level. Specifically, “3” represents high risk, “2” represents medium risk, and “1” represents low risk.

From Figure 5, the main findings can be concluded as follows:

- (i) Overall, January, May, June, July, and August may lead to more crashes, while the crash type distribution varies among these months. For example, January is associated with the decreased possibilities of overturning crashes, while May is found to sustain fewer fixed-object hitting crashes. Besides, July has a negative impact on the occurrence of rear-end crashes.

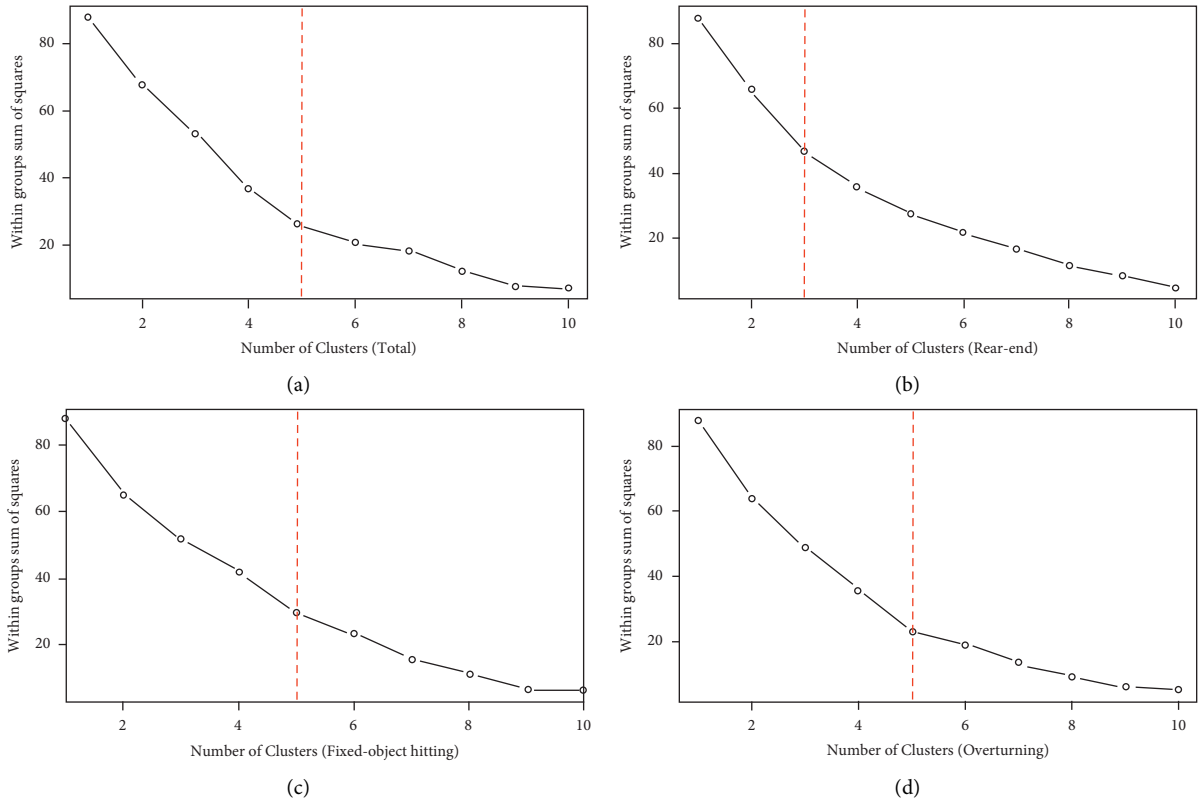


FIGURE 3: The number of clusters provided by the sum of squared error (SSE): (a) total number of crashes, (b) rear-end, (c) fixed-object hitting, and (d) overturning.

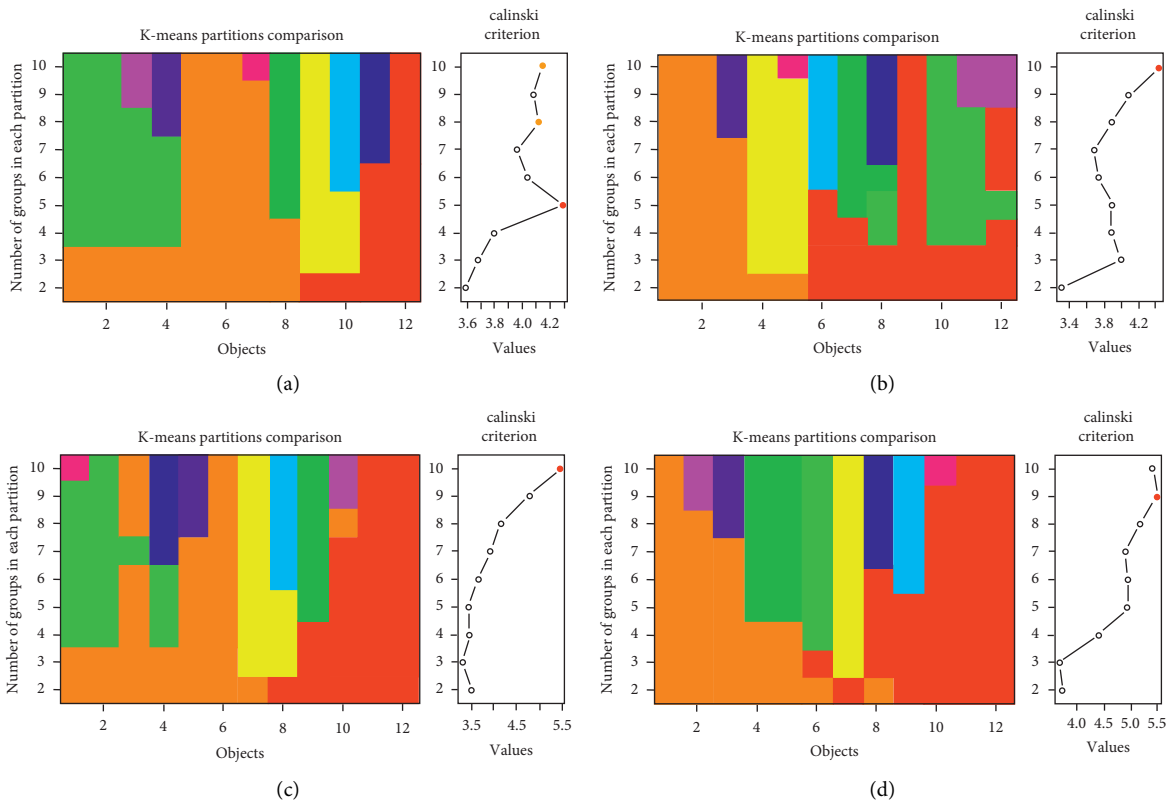
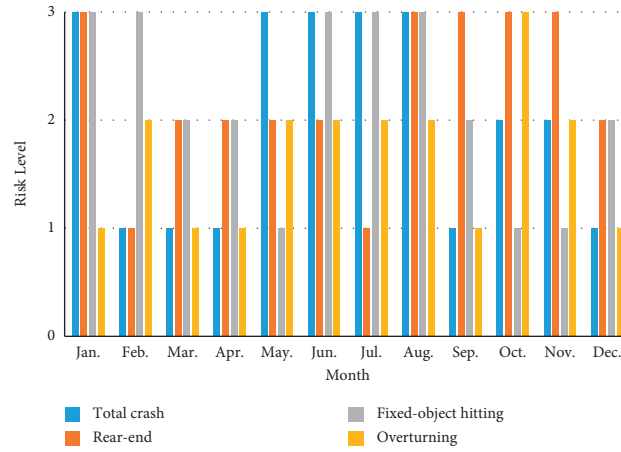


FIGURE 4: The number of clusters provided by Calinsky criterion: (a) total number of crashes, (b) rear-end, (c) fixed-object hitting, and (d) overturning.

TABLE 1: Results of K -means clustering.

Total crash number				Rear-end			
Cluster	Month	Avg. crash number	Risk level	Cluster	Month	Avg. crash number	Risk level
1	1	46.5	3	1	1, 8, 9, 10, 11	12.0	3
2	5, 6, 7, 8	44.8	3	2	3, 4, 5, 6, 12	9.2	2
3	10, 11	40.3	2	3	2, 7	8.1	1
4	3, 4, 9	38.9	1				
5	2, 12	38.4	1				

Fixed-object hitting				Overturning			
Cluster	Month	AVG crash number	Risk level	Cluster	Month	AVG crash number	Risk level
1	6, 7, 8	28.5	3	1	10	7.3	3
2	1, 2	26.1	3	2	5, 6, 7, 8, 11	4.8	2
3	3, 4, 9	21.6	2	3	2	4.5	2
4	12	21.3	2	4	3, 9, 12	3.7	1
5	5, 10, 11	20.2	1	5	1, 4	3.0	1

FIGURE 5: Monthly distribution pattern of various types of crashes based on K -means.

- (ii) Although February and September suffer from the fewest crashes, the frequency of fixed-object hitting crashes in February is significantly higher than that in other months, as well as the rear-end crashes in September.
- (iii) It is also observed that March, April, and December are less susceptible to higher crash frequencies.

Therefore, the status of traffic safety expenditures and crash prevention interventions in these months can be maintained.

Before implementing the Apriori algorithm, some definitions will be determined by (8), in which No. 1 indicated the high-risk level, and No. 3 indicated the low-risk level.

$$\text{Rank}_{ib}^y = \begin{cases} \text{No.1,} & \text{if } N_{ib}^y > \frac{2(N_{b,\max}^y - N_{b,\min}^y)}{3} + N_{b,\min}^y, \\ \text{No.2,} & \text{if } \frac{(N_{b,\max}^y - N_{b,\min}^y)}{3} + N_{ib,\min}^y < N_{ib}^y \leq \frac{2(N_{b,\max}^y - N_{b,\min}^y)}{3} + N_{b,\min}^y, \\ \text{No.3,} & \text{if } N_{ib}^y \leq \frac{(N_{b,\max}^y - N_{b,\min}^y)}{3} + N_{b,\min}^y, \end{cases} \quad (8)$$

where Rank_{ib}^y represents the risk level of each month for total crashes, rear-end crashes, fixed-object hitting crashes, and overturning crashes, respectively. $N_{b,\min}^y$ and $N_{b,\max}^y$ represent the minimum and maximum value of a certain type of crash in a certain year.

After conversion, the Apriori rules between each month and the risk level are shown in Table 2. The rules with lower confidence are not presented because they lack practical significance. As presented in Table 2, with respect to the risk level of total crashes, the confidence level of “No. 1” is 0.625

TABLE 2: The confidence distribution of Apriori rules.

Month	Total		Rear-end			Fixed-object hitting			Overturning		
	Rank	Conf	Month	Rank	Conf	Month	Rank	Conf	Month	Rank	Conf
Jan.	No. 1	0.625	Jan.	No. 2	0.625	Apr.	No. 2	0.625	Jan.	No. 3	0.750
Dec.	No. 3	0.625	Feb.	No. 3	0.625	Jun.	No. 1	0.625	Mar.	No. 3	0.750
Feb.	No. 3	0.500	Mar.	No. 3	0.625	Mar.	No. 3	0.500	Apr.	No. 3	0.625
Mar.	No. 3	0.500	Apr.	No. 2	0.625	Aug.	No. 2	0.500	Jun.	No. 2	0.625
Apr.	No. 2	0.500	Jun.	No. 3	0.625	Sep.	No. 3	0.500	Oct.	No. 1	0.625
Jun.	No. 1	0.500	Jul.	No. 3	0.625	Oct.	No. 3	0.500	May.	No. 2	0.500
Jun.	No. 2	0.500	Dec.	No. 3	0.625	Nov.	No. 3	0.500	Jul.	No. 2	0.500
Aug.	No. 1	0.500	Aug.	No. 1	0.500	Dec.	No. 3	0.500	Aug.	No. 3	0.500
Aug.	No. 2	0.500	Sep.	No. 2	0.500	Jan.	No. 1	0.375	Sep.	No. 3	0.500
Sep.	No. 2	0.500	Oct.	No. 1	0.500	Jan.	No. 2	0.375	Nov.	No. 2	0.500
Oct.	No. 1	0.500	Nov.	No. 2	0.500	Feb.	No. 1	0.375	Dec.	No. 3	0.500
Nov.	No. 3	0.500	May.	No. 1	0.375	Feb.	No. 2	0.375	Feb.	No. 2	0.375
Feb.	No. 1	0.375	May.	No. 2	0.375	Mar.	No. 2	0.375	Feb.	No. 3	0.375
Mar.	No. 2	0.375	Sep.	No. 3	0.375	Apr.	No. 3	0.375	Apr.	No. 2	0.375
Apr.	No. 3	0.375	Oct.	No. 2	0.375	May.	No. 1	0.375	May.	No. 3	0.375
May.	No. 1	0.375	Nov.	No. 3	0.375	May.	No. 1	0.375	Jul.	No. 3	0.375
May.	No. 2	0.375				Jun.	No. 2	0.375	Aug.	No. 1	0.375
Jul.	No. 1	0.375				Jul.	No. 1	0.375	Sep.	No. 2	0.375
Jul.	No. 2	0.375				Jul.	No. 2	0.375	Oct.	No. 2	0.375
Sep.	No. 3	0.375				Aug.	No. 1	0.375			
Oct.	No. 3	0.375				Sep.	No. 1	0.375			
						Oct.	No. 2	0.375			
						Nov.	No. 2	0.375			
						Dec.	No. 2	0.375			

Conf: indicates the confidence level.

in January. Confidence can be regarded as the conditional probability, which represents the possibility of rule implementation in the event of a precondition. Nevertheless, some months belong to different rules (for the total crash in April, the confidence level of “No. 2” is 0.500 and the confidence level of “No. 3” is 0.375). Thus, it is difficult to rank the risk level of a month according to these rules.

There is little research using Apriori rules to undertake the risk ranking of traffic crashes, which leads to difficulty in transforming the rules into the proper form. To overcome this, a simple linear model is introduced to transform the results of Apriori.

$$R_m = \sum c_{im} \cdot \alpha, \quad (9)$$

where R_m is the risk level of crash type m , c_{im} represents the confidence level of month i , α is the weight, and it means 1, 0, -1 for the rank of No. 1, No. 2, and No. 3, respectively.

As displayed in Figure 6, the monthly distribution patterns of various types of crashes are calculated for each month based on Apriori rules, reflecting some interesting findings as follows:

- (i) The overall patterns are similar to the results of K -means, however, the Apriori rules can diversify the value of different types of crashes, not just equal to 1, 2, and 3, which can reveal more information. For example, the results of K -means reveal that January, May, June, July, and August tend to witness the most crashes. Nevertheless, using the Apriori rules, it can

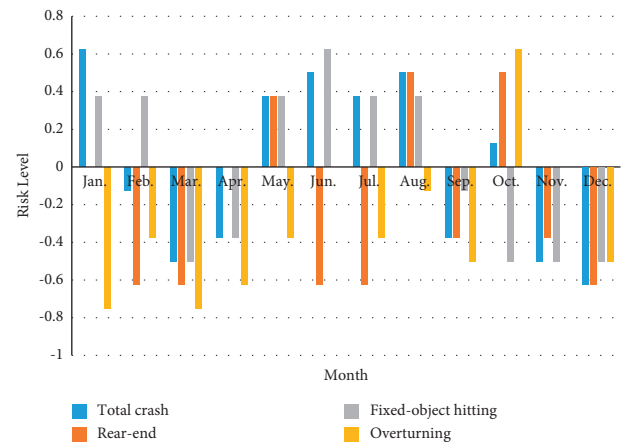


FIGURE 6: Monthly distribution pattern of various types of crashes based on Apriori rules.

be seen that although the total number of crashes occurring to these five months still ranks the top five, January suffers from more crashes than the other four months.

- (ii) The distribution of the same crash type varies significantly among different months. Likewise, the risk values of different types of crashes are also divergent in the same month. Thus, it is unscientific to simply rely on the total number of crashes to determine whether the month is a “black spot.”

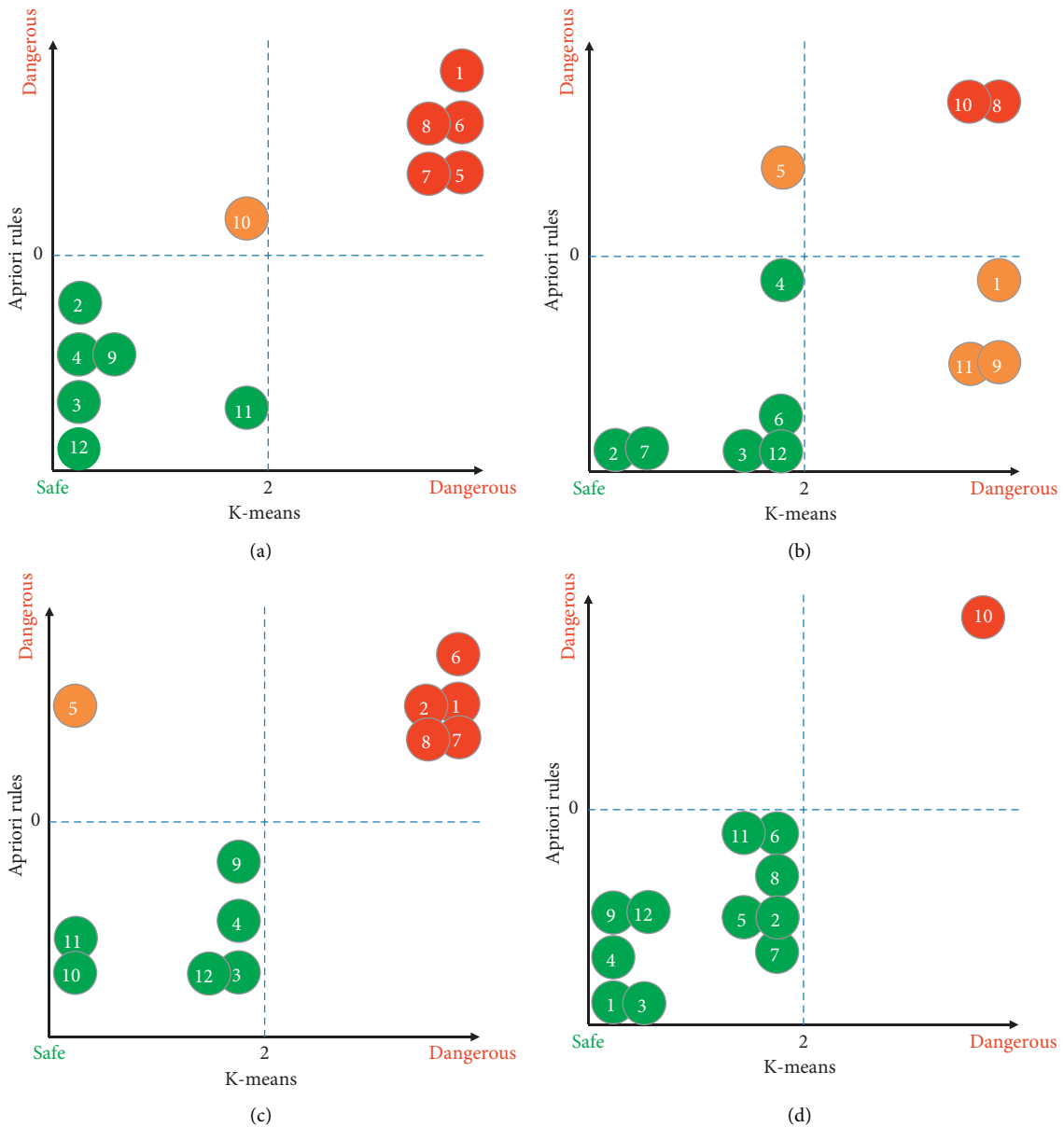


FIGURE 7: Monthly risk distribution: (a) total number of crashes, (b) rear-end, (c) fixed-object hitting, and (d) overturning (if two or more months overlap, it means they have the same risk value).

4.2. Crash Risk Analysis of the Combined Method.

Combining the results of *K*-means and Apriori rules, a monthly risk distribution plot is drawn to visually exhibit which crash type is more likely to occur in which month. Based on the results of *K*-means, the crash risk of all months can be divided into three levels: “3” represents high risk, “2” represents medium risk, and 1 represents low risk. Similarly, when calculating the risk index R_m based on the results of Apriori, “0” indicates the medium risk. The closer R_m is to 1, the higher the risk, and the closer R_m is to -1, the lower the risk. Consequently, the drawing is to use “2” and “0” as the risk

thresholds of *K*-means and Apriori, respectively, as shown in Figure 7.

Accordingly, the independent variables of the parametric model are defined as follows:

$$x_{ib} = \begin{cases} x_1, & \text{if month } i \text{ is green in Fig 8,} \\ x_2, & \text{if month } i \text{ is orange in Fig 8,} \\ x_3, & \text{if month } i \text{ is red in Fig 8.} \end{cases} \quad (10)$$

Next, the above three types of dummy variables will be used as independent variables to estimate LR models, as shown in Table 3. The estimated coefficient explains the

TABLE 3: Estimated results of the LR model.

Variables	Total		Rear-end		Fixed-object hitting	
	OR	<i>p</i> -value	OR	<i>p</i> -value	OR	<i>p</i> -value
x_1	Base		Base		Base	
x_2	2.429	0.252	2.200	0.094	3.333	0.128
x_3	5.667	0.000*	6.600	0.004*	5.273	0.000*
	AIC	123.909	AIC	127.980	AIC	127.744

* $P < 0.05$.

effects of each month on crashes compared to x_1 . The result suggests that the variable of x_3 significantly increases the probability of total crashes, rear-end crashes, and fixed-objected hitting crashes. Additionally, the higher risks of rear-end and fixed-object hitting crashes is observed in the months belonging to x_3 compared to months belonging to x_1 (6.6 times and 5.273 times, respectively). Therefore, for these months, it is necessary to adopt special traffic management policies and deploy the corresponding traffic signs along the road. Another noteworthy finding is that x_2 does not significantly increase the risk of crashes. The possible reason is that false-positive cases may occur in months that are identified as high risk only by a single method (*K*-means or Apriori), i.e., there is no significant difference between the high-risk months and low-risk months identified only by a single method. It also highlights the superiority of the combined modeling approach over the single method.

The LR model fails to estimate the overturning crashes. It may be because of the small sample size (only 422) of overturning crashes in this case. Therefore, Figure 8 is applied to visually present the difference in overturning crashes between low-risk clusters (x_1) and high-risk clusters (x_3). It can be seen that among all the 96 samples, the samples of the x_3 cluster are all of the high risk, while only 44% of the samples belonging to the x_1 cluster are high-risk. Similar findings also prove that the months belonging to x_3 are more likely to suffer overturning crashes.

5. Discussion

This paper analyses the monthly distribution patterns of crashes in mountainous areas. The results demonstrate that different types of crashes have inconsistent distribution patterns, which cannot be directly observed. The main findings and corresponding explanations are as given below.

5.1. Rear-End Crashes. August and October are associated with higher risks of rear-end crashes. Champahom et al. (2022) found that traffic volume may have a positive effect on the frequency of rear-end crashes [51, 52]. A large body of research has shown that the traffic volume on legal holidays is larger than that on weekdays. In China, August is the summer vacation, while October 1–7 is the national day, which is the peak of Chinese citizens' travel [53–55]. Intuitively, the reduced vehicle headway caused by the increased traffic volume may result in more rear-end crashes. Furthermore, the windy weather can aggravate the injury

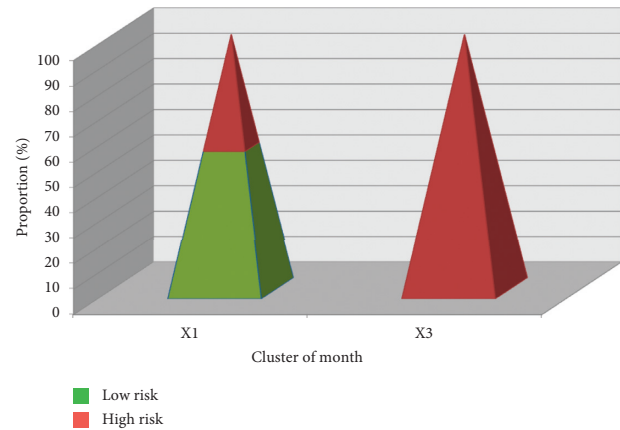


FIGURE 8: The difference in overturning between different clusters.

severity in rear-end crashes [35]. By manually observing the wind speed thermal map on the website of China Meteorological Administration [56], the monthly maximum wind speed range is obtained in the research area, as shown in Table 4. It can be seen that the wind speed in August is significantly higher than that in other months. It also explains why the risk of rear-end crashes in August is higher than that in October. In these two months, the safety management department should pay attention to helping the driver avoid rear-end crashes by reminding the driver to keep a safe distance from the leading vehicle through the on-board voice system or setting up warning signs to warn the driver to reduce the speed in a windy environment.

5.2. Fixed-Object Hitting Crashes. January, February, June, July, and August are the months that have a high risk of fixed-object hitting crashes. Generally, a fixed-object hitting crash refers to a single-vehicle crash, where the vehicle departs from the road and collides with roadside obstacles or fixed facilities (e.g., tree, utility pole, traffic sign, embankment, ditch, culvert, or barrier) [57]. The major reason for its occurrence is that the driver loses control of the vehicle. Compared with the dry road surface, the slippery road surface is more likely to cause the vehicle to go out of control and increase the injury severity of fixed-object hitting crashes [58, 59]. Since it is unrealistic to use the rainfall at one site to represent the rainfall of the whole freeway, the current study selects 7 main sites along the road (see Figure 9) to represent the overall rainfall of the Wenli Freeway. The monthly average rainfall and maximum rainfall from 1981 to 2010 are obtained, as shown in Tables 5 and 6. As presented in Figure 9, the average rainfall and maximum rainfall in June, July, and August are higher than those in other months. It suggests that the road surfaces will become more slippery because of excessive rainfall and water accumulation. Although there is less rainfall in January and February, the temperature is the lowest in the whole year, and it is generally below 0 degrees Celsius at night. Hence, even a small amount of rain or snow can cause the pavement to freeze,

TABLE 4: Monthly distribution of maximum wind speed of Wenli freeway.

Month	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Maximum wind speed (m/s)	10–15	10–15	15	15–20	15–20	15–25	20–25	30–35	15–25	10–15	15–20	10–15

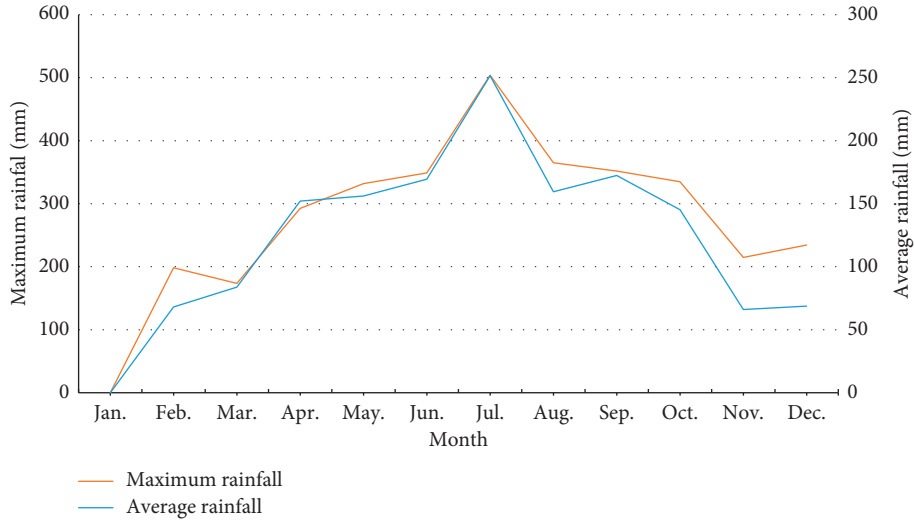


FIGURE 9: The monthly average rainfall and maximum rainfall of Wenli freeway.

TABLE 5: Monthly distribution of maximum rainfall of Wenli freeway.

	Jinhua (mm)	Wuyi (mm)	Yongkang (mm)	Jinyun (mm)	Lishui (mm)	Qingtian (mm)	Wenzhou (mm)
Jan.	75.3	76.2	73.5	69.2	62.9	55.1	64.5
Feb.	91.8	90.9	85.3	83.6	77.1	75.6	82.5
Mar.	160.7	160.5	154.4	152.2	143.8	143.7	149.7
Apr.	162.7	163.6	150.1	154.3	154.8	154.6	152.5
May.	169.8	171.5	157.2	164.8	160.6	173.4	188.9
Jun.	254.7	255.5	240.3	241.2	246.7	263.9	258.8
Jul.	139.5	146.4	142.5	153.8	129.5	210.5	194.3
Aug.	119.1	138.7	142.1	153.3	155.3	251.7	246.6
Sep.	99.6	127	123.2	130.7	113.8	190.2	232.1
Oct.	56.6	62.4	61.8	63.6	58.3	77.2	82.3
Nov.	67	74.5	71.1	68.6	60.7	60.4	78.7
Dec.	49	52.2	49.8	46.9	42.2	41.5	49.7

TABLE 6: Monthly distribution of average rainfall of Wenli freeway.

	Jinhua (mm)	Wuyi (mm)	Yongkang (mm)	Jinyun (mm)	Lishui (mm)	Qingtian (mm)	Wenzhou (mm)
Jan.	236.8	256.5	244.8	197.8	182.4	147	124.2
Feb.	217.9	189.3	175.6	175.5	165.7	149.6	142
Mar.	314.3	309.2	276.5	301.1	277.5	276.6	293.1
Apr.	410.9	376.2	340.5	349.4	332.9	268.9	244.1
May.	276.8	312.1	285.3	391.5	395.4	347.2	434.2
Jun.	481.7	474.7	450.1	476.6	550.6	545.2	548.8
Jul.	354.2	326.9	419.1	360.5	276.5	398.2	420.6
Aug.	290.1	285.7	232.3	300.4	331.6	532.1	491.1
Sep.	208.2	342.1	282.2	250.4	274	476.5	511.3
Oct.	149.6	215.6	193.1	219.3	192.4	260.5	273.8
Nov.	255.7	281.1	295.4	249.4	201.2	186.2	173.2
Dec.	186.3	199.9	186.7	168	169.5	165.1	144.9

making it more likely for vehicles to run out of control and hit roadside fixtures.

The overall crash frequency in February is low. It provides an interesting insight into developing appropriate safety interventions. The relevant departments may ignore the higher risk of fixed-objects hitting crashes in February without reference to the results of this study, subsequently leading to an increase in casualties and infrastructure damage. In addition, in January, February, June, July, and August, because of the relatively high incidence of fixed-object hitting crashes, more attention should be paid to road maintenance. A collision with the median barrier is the main form of fixed-object hitting [60]. Hence, the type of the median barrier and whether it has a buffer function will affect the crash injury severity. Additionally, the collision with the tunnel entrance should not be overlooked. It is necessary to set a warning sign in advance to inform the driver of driving in a safer lane.

5.3. Overtaking Crashes. October has a high risk of overturning crashes. An overturning crash occurs when the vehicle's lateral force or overturning moment generated by the road surface exceeds the counterbalancing [61–63]. As mentioned above, October is susceptible to heavy rainfall, making the road slippery. Water accumulation can further form a layer of water film between the tire and the road surface, reducing the friction resistance. Thus, the vehicle is likely to lose control and rollover when braking or turning suddenly. Considering the high risk of rear-end crashes and relatively large traffic volume in October, one possible explanation is that the potential collision avoidance demand leads to more braking behaviors, deceleration, and sharp steering, which somewhat increases the risk of overturning crashes. As such, to reduce the overturning frequency, drivers should be trained to pay more attention to the speed limit and distance maintenance in October.

Furthermore, although the risks of these three types of crashes in May are not significantly high, the overall crash frequency in May deserves the attention of relevant departments. In summary, considering the crash risks in months belonging to x_3 are significantly higher than those in the months belonging to x_1 , the road safety administration should consider adopting some targeted measures and policies in these months to reduce the occurrence of crashes. As mentioned in Section 1.1, most of the existing studies used a simple histogram to qualitatively describe the monthly distribution patterns of crashes. Therefore, the insights that can be referred to from the current study are still very limited. In addition, traffic-related data is not disclosed to the public in China. Hence, it is difficult to obtain the monthly traffic flow data of the Wenli freeway. The authors can only interpret and speculate the model's results from a macro perspective. In the future, on the basis of this study, additional efforts should be devoted to gaining a better understanding of safety factors and to providing more appropriate interventions.

6. Conclusion

The prevention of traffic crashes is always a hot topic in the field of traffic safety. This study takes a mountainous freeway in China as a case and develops a combined modeling framework to identify the monthly distribution patterns of crashes. The results indicate that different types of crashes have different tendencies in different months. In some months, the total number of crashes is not large, however, certain types of crash display a high-risk level, such as those in February and October. Similarly, although the total crash frequency in some months is high, a certain crash type may rarely occur, such as that in January, June, July, and August. In addition, the study also reveals that for a certain crash type, when a month is identified as high risk by the combined modeling approach, the probability of such a type of crash occurring in this month will be significantly increased.

The limitation of this study is that the data scale is relatively small, which cannot fully reflect the advantages of the developed method. In the fifth section of the book "SPSS. 11-0 statistical analysis tutorial (advanced part)," the sample size of the LR model is recommended [64]. The number of independent variables in the model should be approximately equal to one-tenth of the number of categories with the least number of cases. For instance, assuming that category I has sixty-seven samples and category II has fifteen samples, the number of independent variables in the model is recommended to be 1.5, which is approximately equal to 1 or 2. If there are too many independent variables, the interpretation may be biased or even ineffective. In this study, the number of independent variables is three and the number of x_2 cases in overturning crashes is zero, which is most likely the main cause of model failure. Therefore, because of the inherent randomness of crashes, a larger sample size can make the classification results more discrete and reduce the probability of model failure. In addition, because of the lack of monthly traffic volume data, the results of the combined model cannot be explained more reasonably. However, this kind of study is vital. The failure of the LR model of overturning crashes cannot mask the superiority of the proposed method framework. This study finally shows that using only one machine learning algorithm will lead to false-positive cases. The false-positive cases can result in an inefficient use of the resources applied to safety improvements and reduce the effectiveness of the safety management process. Results clearly illustrate the relationship between months and different types of crash risks, which helps the best use of the limited funds available. Next, more in-depth research will be further conducted to analyze the root causes that result in the emergence of these monthly patterns of crash risks and put forward some specific improvement suggestions for crash prevention.

Data Availability

The data used to support the findings of this study were supplied by the Transportation Department of Zhejiang Province under license, and hence, they cannot be made

freely available. The authors have not yet obtained permission from the agency to make the dataset available. If researchers are interested in this study, they can contact the corresponding author via email.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank the National Natural Science Foundation of China (Grant no. 52072069) and Scientific Research Foundation of Graduate School of Southeast University (YBPY2166).

References

- [1] World Health Organization, *Global Status Report on Road*, <https://www.who.int/publications/i/item/9789241565684>, 2018.
- [2] L. Wang, H. Lu, and Y. Zheng, "Cause analysis of traffic accidents on mountain expressway based on multistage microscope technology," *Acta Microsc*, vol. 28, no. 2, pp. 307–314, 2019.
- [3] H. Xue and G. Xue, "Exploring the relationships between single-vehicle traffic accident and driver's route familiarity on the mountainous highways," *Cognition, Technology & Work*, vol. 22, no. 4, pp. 815–824, 2020.
- [4] P. Rautela and S. S. Pant, "New methodology for demarcating high road accident risk-prone stretches in mountain roads," *Current Science*, vol. 92, no. 8, pp. 1157–1161, 2007.
- [5] J. He, M. King, Z. Liu et al., "A crash risk identification method for freeway segments with horizontal curvature based on real-time vehicle kinetic response," *Accident Analysis & Prevention*, vol. 150, Article ID 105911, 2021.
- [6] J. J. Cao and X. D. Cao, "System Analysis of potential accidents on mountain road based on rough set and quantitative theory," *KSCE Journal of Civil Engineering*, vol. 25, no. 3, pp. 1031–1042, 2021.
- [7] M. Yadollahi, "Time distribution of traffic accidents in Shiraz, Iran, during 2017-2018," *J. Heal. Manag. Informatics*, vol. 6, no. 4, pp. 156–162, 2019.
- [8] X. Ye, R. M. Pendyala, S. P. Washington, K. Konduri, and J. Oh, "A simultaneous equations model of crash frequency by collision type for rural intersections," *Safety Science*, vol. 47, no. 3, pp. 443–452, 2009.
- [9] R. Tay, J. Choi, L. Kattan, and A. Khan, "A multinomial logit model of pedestrian-vehicle crash severity," *International Journal of Sustainable Transportation*, vol. 5, no. 4, pp. 233–249, 2011.
- [10] N. N. Wong and S. C. Wong, "Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes," *Accident Analysis & Prevention*, vol. 39, no. 6, pp. 1267–1278, 2007.
- [11] J. Vorko-Jović, Z. Kern, and A. Biloglav, "Risk factors in urban road traffic accidents," *Journal of Safety Research*, vol. 37, no. 1, pp. 93–98, 2006.
- [12] L. Qu, M. Dai, F. Wang, and Z. Zeng, "Time distribution and analysis of traffic accident fatalities," *Applied Mechanics and Materials*, vol. 361, pp. 2308–2314, 2013.
- [13] 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.
- [14] H. Hao, Y. Li, A. Medina, R. B. Gibbons, and L. Wang, "Understanding crashes involving roadway objects with SHRP 2 naturalistic driving study data," *Journal of Safety Research*, vol. 73, pp. 199–209, 2020.
- [15] W. Ackaah, B. A. Apuseyine, and F. K. Afukaar, "Road traffic crashes at night-time: characteristics and risk factors," *International Journal of Injury Control and Safety Promotion*, vol. 27, no. 3, pp. 392–399, 2020.
- [16] N. S. S. Al-Bdairi, "Does time of day matter at highway work zone crashes?" *Journal of Safety Research*, vol. 73, pp. 47–56, 2020.
- [17] L. Lin, J. Jinhai, and Y. Yan, "Traffic crash characteristic analysis on mountain roads," in *Proceedings of the 2013 5th Conference on Measuring Technology and Mechatronics Automation*, pp. 1271–1274, Hong Kong, China, January 2013.
- [18] M. Abrari Vajari, K. Aghabayk, M. Sadeghian, and N. Shiwakoti, "A multinomial logit model of motorcycle crash severity at Australian intersections," *Journal of Safety Research*, vol. 73, pp. 17–24, 2020.
- [19] H. Seyed Taghi, H. Amin, S. Yaser et al., "Time analysis of fatal traffic accidents in Fars Province of Iran," *Chinese Journal Traumatol*, vol. 16, no. 2, pp. 84–88, 2013.
- [20] S. Zhao, K. Wang, C. Liu, and E. Jackson, "Investigating the effects of monthly weather variations on Connecticut freeway crashes from 2011 to 2015," *Journal of Safety Research*, vol. 71, pp. 153–162, 2019.
- [21] L. Xing, J. He, Y. Li, Y. Wu, J. Yuan, and X. Gu, "Comparison of different models for evaluating vehicle collision risks at upstream diverging area of toll plaza," *Accident Analysis & Prevention*, vol. 135, Article ID 105343, 2020.
- [22] C. Xu, P. Liu, W. Wang, and Z. Li, "Safety performance of traffic phases and phase transitions in three phase traffic theory," *Accident Analysis & Prevention*, vol. 85, pp. 45–57, 2015.
- [23] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207–216, 1993.
- [24] W. B. Zulfikar, A. Wahana, W. Uriawan, and N. Lukman, "Implementation of association rules with apriori algorithm for increasing the quality of promotion," in *Proceedings of the 2016 4th International Conference on Cyber and IT Service Management*, pp. 4–8, Bandung, Indonesia, April 2016.
- [25] S. Sun and X. Sun, "Association knowledge for fatal run-off-road crashes by Multiple Correspondence Analysis," *IATSS Research*, vol. 39, no. 2, pp. 146–155, 2016.
- [26] A. Verma, S. Khan, J. Maiti, and O. B. Krishna, "Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports," *Safety Science*, vol. 70, pp. 89–98, 2014.
- [27] M. Shaiba and H. Shaiba, "Apriori-based algorithm for dubai road accident analysis," *Procedia Computer Science*, vol. 163, pp. 218–227, 2019.
- [28] F. Nafie Ali, N. Mohamed Hamed, A. Ali, and M. Hamed, "Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents," *Journal of Information and Telecommunication*, vol. 2, no. 3, pp. 231–245, 2018.
- [29] X. Deng, D. Zeng, and H. Shen, "Causation analysis model: based on AHP and hybrid Apriori-Genetic algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 1, pp. 767–778, 2018.
- [30] J. MacQueen, "Some methods for classification and analysis of multivariate observation," in *Proceedings of the 5th Berkeley*

- Symposium on Mathematical Statistics and Probability*, pp. 281–297, California Press, Berkeley, 1967.
- [31] B. Dadashova, B. A. Ramírez, J. M. McWilliams, and F. A. Izquierdo, “The identification of patterns of interurban road accident frequency and severity using road geometry and traffic indicators,” *Transportation Research Procedia*, vol. 14, pp. 4122–4129, 2016.
- [32] A. Almjewail, A. Almjewail, S. Alsenaydi, H. AlSudairy, and I. Al Turaiki, “Analysis of traffic accident in riyadh using clustering algorithms,” *Advances in Intelligent Systems and Computing*, vol. 753, pp. 12–25, 2018.
- [33] C. Zhang, Y. U. E. Shu, and L. Yan, “A novel identification model for road traffic accident black spots: a case study in ningbo, China,” *IEEE Access*, vol. 7, pp. 140197–140205, 2019.
- [34] M. Osman, R. Paleti, S. Mishra, and M. M. Golias, “Analysis of injury severity of large truck crashes in work zones,” *Accident Analysis & Prevention*, vol. 97, pp. 261–273, 2016.
- [35] C. Chen, G. Zhang, R. Tarefder, J. Ma, H. Wei, and H. Guan, “A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes,” *Accident Analysis & Prevention*, vol. 80, pp. 76–88, 2015.
- [36] P. Fan and W. Fan, “Exploring injury severity in head-on crashes using latent class clustering analysis and mixed logit model: a case study of North Carolina,” *Accident Analysis & Prevention*, vol. 135, Article ID 105388, 2020.
- [37] M. A. Abdel-Aty and R. Pemmanaboina, “Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 167–174, 2006.
- [38] Z. Ci, C. Chen, G. Zhang et al., “Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models,” *Accident Analysis & Prevention*, vol. 124, pp. 219–229, 2019.
- [39] B. Dong, X. Ma, F. Chen, and S. Chen, “Investigating the differences of single-vehicle and multivehicle accident probability using mixed logit model,” *Journal of Advanced Transportation*, vol. 2018, pp. 1–9, 2018.
- [40] S. S. Ahmed, J. Cohen, and P. C. Anastasopoulos, “A correlated random parameters with heterogeneity in means approach of deer-vehicle collisions and resulting injury-severities,” *Analytic Methods in Accident Research*, vol. 30, Article ID 100160, 2021.
- [41] C. Zhang, J. He, X. Yan, Z. Liu, Y. Chen, and H. Zhang, “Exploring relationships between microscopic kinetic parameters of tires under normal driving conditions, road characteristics and accident types,” *Journal of Safety Research*, vol. 78, pp. 80–95, 2021.
- [42] C. Xu, W. Wang, and P. Liu, “A genetic programming model for real-time crash prediction on freeways,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 574–586, 2013.
- [43] L. Chien and J. Chien, “Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model,” *Safety Science*, vol. 51, no. 1, pp. 17–22, 2013.
- [44] X. Yan, J. Gong, J. He, H. Zhang, C. Zhang, and Z. Liu, “Integrated data mining and TOPSIS entropy weight method to evaluate logistics supply and demand efficiency of a 3PL company,” *Mathematical Problems in Engineering*, vol. 2020, Article ID 7057143, 12 pages, 2020.
- [45] S. R. Pollard and J. Siami-Irdemoosa, “Off-road truck-related accidents in U.S. mines,” *Journal of Safety Research*, vol. 58, pp. 79–87, 2016.
- [46] S. Raj, D. Ramesh, M. Sreenu, and K. S. Sethi, “EAFIM: efficient apriori-based frequent itemset mining algorithm on Spark for big transactional data,” *Knowledge and Information Systems*, vol. 62, no. 9, pp. 3565–3583, 2020.
- [47] M. A. Abdel-Aty, H. M. Hassan, M. Ahmed, and A. S. Al-Ghamdi, “Real-time prediction of visibility related crashes,” *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 288–298, 2012.
- [48] M. Török and Á. Török, “A comparative analysis of black spot identification methods and road accident segmentation methods,” *Accident Analysis & Prevention*, vol. 128pp. 1–7, February, 2019.
- [49] S. Lukasik, P. A. Kowalski, M. Charytanowicz, and P. Kulczycki, “Clustering using flower pollination algorithm and Calinski-Harabasz index,” in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation*, pp. 2724–2728, Vancouver, BC, Canada, July 2016.
- [50] T. Harabasz and J. Harabasz, “A dendrite method for cluster analysis,” *Communications in Statistics - Theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.
- [51] T. Champahom, S. Jomnonkwo, A. Karoonsoontawong, and V. Ratanavaraha, “Spatial zero-inflated negative binomial regression models: application for estimating frequencies of rear-end crashes on Thai highways,” *Journal of Transportation Safety & Security*, vol. 14, no. 3, pp. 523–540, 2022.
- [52] A. Abdel-Aty and M. A. Abdel-Aty, “A combined frequency-severity approach for the analysis of rear-end crashes on urban arterials,” *Safety Science*, vol. 49, no. 8-9, pp. 1156–1163, 2011.
- [53] X. Ji and W. Guoqing, “The tourism flow on network structure of shanghai based on tourism digital footprint,” *J. URBAN Stud.* vol. 40, no. 6, pp. 33–38, 2019.
- [54] J. Qian, D. Wang, and Y. Niu, “Changes in city Residents’Support for the national highway toll-free policy and their Travel behavior:the case of suzhou,” *Tourism Tribune*, vol. 30, no. 4, p. 50, 2015.
- [55] X. Ji, Y. Ge, and F. Chen, “Spatial and temporal differentiation characteristics of holiday tourism flow based on highway traffic flow big data: a case study of 7 holidays in yunnan Province,” *Tourism Tribune*, vol. 34, no. 6, pp. 37–47, 2019.
- [56] National Meteorological Science Data Center, “Schematic Diagram of Surface Climate in China,” <https://data.cma.cn/data/weatherBK.html>.
- [57] F. D. B. Albuquerque and D. M. M. Awadalla, “Roadside fixed-object collisions, barrier performance, and fatal injuries in single-vehicle, run-off-road crashes,” *Safety Now*, vol. 6, no. 2, pp. 27–15, Safety, 2020.
- [58] M. Shawky, H. M. Hassan, A. M. Garib, and H. A. Al-Harthei, “Examining the factors affecting the severity of run-off-road crashes in Abu Dhabi,” *Canadian Journal of Civil Engineering*, vol. 43, no. 2, pp. 132–138, 2016.
- [59] R. Abdel-Aty and M. Abdel-Aty, “Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data,” *Safety Science*, vol. 63, pp. 50–56, 2014.
- [60] N. Haghghi, X. C. Liu, G. Zhang, and R. J. Porter, “Impact of roadway geometric features on crash severity on rural two-lane highways,” *Accident Analysis & Prevention*, vol. 111, pp. 34–42, 2018.
- [61] B. Hassan and Y. Hassan, “Probabilistic, safety-explicit design of horizontal curves on two-lane rural highways based on reliability analysis of naturalistic driving data,” *Accident Analysis & Prevention*, vol. 123, pp. 200–210, 2019.

- [62] K. Jiang, D. Yang, S. Xie, Z. Xiao, A. C. Victorino, and A. Charara, "Real-time estimation and prediction of tire forces using digital map for driving risk assessment," *Transportation Research Part C: Emerging Technologies*, vol. 107, pp. 463–489, 2019.
- [63] A. Alrejfal, A. Farid, and K. Ksaibati, "Investigating factors influencing rollover crash risk on mountainous interstates," *Journal of Safety Research*, vol. 80, pp. 391–398, 2022.
- [64] W. Zhang, *SPSS. 11-0 Statistical Analysis Tutorial (Advanced Part)*, Hope Electronic Press, Beijing, 2002.