

## Research Article

# Probability of Roadside Accidents for Curved Sections on Highways

Guozhu Cheng <sup>1</sup>, Rui Cheng <sup>1</sup>, Yulong Pei <sup>1</sup> and Liang Xu <sup>2</sup>

<sup>1</sup>School of Traffic and Transportation, Northeast Forestry University, 150040 Harbin, China

<sup>2</sup>School of Civil Engineering, Changchun Institute of Technology, 130012 Changchun, China

Correspondence should be addressed to Guozhu Cheng; guozhucheng@126.com

Received 1 April 2020; Revised 9 May 2020; Accepted 16 May 2020; Published 30 May 2020

Guest Editor: Yong Chen

Copyright © 2020 Guozhu Cheng 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.

To predict the probability of roadside accidents for curved sections on highways, we chose eight risk factors that may contribute to the probability of roadside accidents to conduct simulation tests and collected a total of 12,800 data obtained from the PC-crash software. The chi-squared automatic interaction detection (CHAID) decision tree technique was employed to identify significant risk factors and explore the influence of different combinations of significant risk factors on roadside accidents according to the generated decision rules, so as to propose specific improved countermeasures as the reference for the revision of the Design Specification for Highway Alignment (JTG D20-2017) of China. Considering the effects of related interactions among different risk factors on roadside accidents, path analysis was applied to investigate the importance of the significant risk factors. The results showed that the significant risk factors were in decreasing order of importance, vehicle speed, horizontal curve radius, vehicle type, adhesion coefficient, hard shoulder width, and longitudinal slope. The first five important factors were chosen as predictors of the probability of roadside accidents in the Bayesian network analysis to establish the probability prediction model of roadside accidents. Eventually, the thresholds of the various factors for roadside accident blackspot identification were given according to probabilistic prediction results.

## 1. Introduction

Roadside accidents occur when a vehicle leaves the travel line, crosses an edge line or a centre line, collides with trees, guardrails, utility poles, and other natural or man-made objects located on roadsides, or overturns or falls into deep ditches or rivers. According to the Fatal Accident Reporting System (FARS), these accident types account for more than 39% of fatal accidents in the United States [1]. In China, roadside accidents account for approximately 50% of the collisions in which more than three people perish [2]. A European study also shows that 20% of all traffic accidents every year are roadside accidents; however, the fatality rate is over 35% in these accidents, and approximately one-third of run-off-road (ROR) collision fatalities occurred on curved road sections [3], the road type upon which this study focuses.

There are several complex reasons a vehicle departs from the travelled path, such as an inappropriate avoidance

manoeuvre or inattention of a driver, crossing a curve segment with a high speed, or understeering. A variety of contributing factors to roadside accidents have been identified based on various collected data and data analysis methods. Numerous studies have confirmed that highway geometric design indexes (i.e., roadway characteristics and roadside characteristics) play a significant role in whether a crash occurs resulting from driver error [4], especially for curve sections on highways. In terms of roadway characteristics, a wider shoulder has been found to decrease the occurrence of ROR accidents on horizontal curves [3, 5], but the increase of the shoulder width is also associated with an increasing vehicle operating speed [6]. The frequency of ROR accidents will increase if vehicles travel in a narrower lane because the requirement for sharing the roadway with other vehicles increases the chance of conflicts, whereas driveway density has little impact on ROR accidents [3, 5]. Moreover, some research has confirmed that pavement edge

drop-off and low friction of pavement surfaces tend to cause a high frequency of single-vehicle accidents [7]. Sharp curves are also a key factor contributing to roadside accident occurrence and approximately 30% of the ROR events occur on road curves [8–10]. In an attempt to identify roadside design risk factors, a large number of studies have been implemented, involving analysis of the relationship between the frequency of roadside accidents and critical slope, fences, bridges, guardrail, ditches, utility pole density, distance to pole and distance to tree, and so on [3, 11–17].

Among the environmental factors, most ROR accidents tend to occur on weekends [3, 5]. Area type and lighting conditions are found to be significant factors contributing to the probability of roadside accidents [9]. A study investigated whether road type and local amenities are associated with single-vehicle accident frequency [18]. Additionally, local population density is also related to accident occurrence [19].

In terms of human factors, the National Highway Traffic Safety Administration (NHTSA) suggested that driver distraction, fatigue, driver's degree of familiarity with the roadway, blood alcohol presence, age, and gender were the most significant factors contributing to roadside accidents [18], and 30% of these accidents occurred due to driver inattention [8–10]. All of these factors have a direct or indirect effect on changes in vehicle speeds, and the risk of accidents increases, followed by an increase in vehicle speeds.

From a methodological perspective, different methods have been employed to determine these factors. Originally, Zegeer and Deacon [20] developed a lognormal regression model to investigate the relationship between ROR accident frequencies and various variables, such as average annual daily traffic (AADT), shoulder width, lane width, terrain type, and clear roadside recovery distance (CRRD). In a further study, they added some variables (i.e., density and lateral offset of the roadside object) to the previous model [21]. However, this conventional linear regression model has been demonstrated as inappropriate and to be often erroneous [22–24]. More appropriate prediction models for accident frequency (i.e., Poisson and negative binomial (NB) regression models) have been widely used in recent decades [22, 25–27]. To address the problem of zero-inflated counting processes in accident frequency analysis, the zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression models have gained considerable acceptance [28–32].

Although there have been a considerable number of roadside accident frequency studies, few studies have focused on the quantitative analysis of roadside accident probability. Various approaches (i.e., Poisson model, NB model, ZIP model, and ZINB model) are capable of predicting the number or frequency of roadside accidents based on mass accident data but cannot precisely calculate probability values under the effects of various variables. Moreover, the research results based on the prediction of accident frequency or number are often influenced by different traffic characteristics in various regions, which is not universal. Considering that accident probability is more able

to represent the degree of frequent accidents, it is better to carry out the prediction of accident probability than to carry out the prediction of accident frequency or number. To identify the roadside accidents blackspot and reduce the accidents probability, we therefore used a data mining technique (i.e., CHAID decision tree technique) to identify significant risk factors contributing to roadside accidents and another data mining technique (i.e., Bayesian network analysis) to establish the probability prediction model of roadside accidents. Additionally, we investigate the importance of various variables under the interactions of accident occurrence by developing a path analysis based on a logistic regression model. To the best of our knowledge, no research has used these three methods together in the study of the probability of roadside accidents.

## 2. Data and Methodology

*2.1. Data.* Substantial statistical analysis generally relies on historical accident data. However, the constantly changing traffic environment, the high cost of maintaining or collecting roadside accident data, and the long-term lack of detailed data have formed a barrier to developing a study of the relationship between road design and the probability and severity of accidents [25, 33]. Automobile dynamics simulation technology, regarded as an alternative approach, has been popularly applied to obtain accident data in recent years. Compared to collected accident data, the data from simulation software has the following advantages: (1) comprehensive accident information, (2) no consideration of the impact of time and traffic condition on accident data, (3) universal applicability in the absence of regional characteristics, (4) low cost of research, and (5) free choice of variables according to your interest. In the present study, we used accident data obtained from PC-crash simulation software. This software is primarily developed to take accident reconstruction and has been used for collisions between vehicles [34] and accidents involving vehicles and pedestrians [35], as well as single-vehicle accidents [36, 37]. It has been demonstrated that PC-crash has good performance in simulating single-vehicle (rollover) accidents [36–40].

We chose highway geometric design indexes (horizontal curve radius, hard shoulder width, longitudinal slope, superelevation slope, and width value of the curve), pavement condition (adhesion coefficient), and traffic characteristics (vehicle speed and vehicle type) as input variables, and vehicle final states as the output variable. In the present study, the final states of vehicles include departing from the roadway and not departing from the roadway. The former state refers to the circumstances of vehicle rollover or any of the vehicle wheels entering the slope represents the occurrence of a roadside accident (see Figure 1), and the latter state involves a vehicle running normally and represents no occurrence of a roadside accident.

Consider that the values of slope gradient and slope height mainly affect the severity of the roadside accident when the vehicle enters the slope and have little effect on the occurrence of the roadside accident. In addition, in

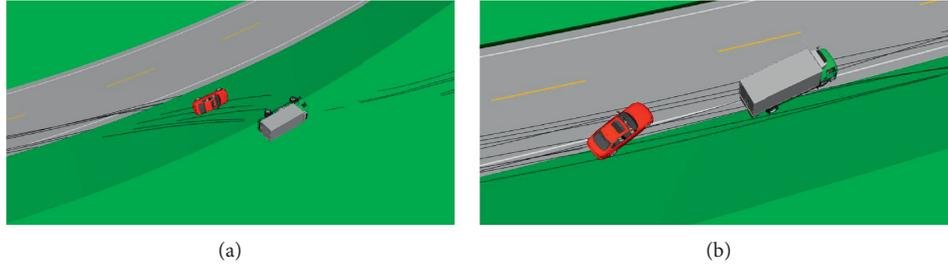


FIGURE 1: Occurrence of roadside accidents. (a) Vehicle rollover. (b) Wheel of vehicle entering slope.

combination with the provisions of carriageway width and crown slope in the Design Specification for Highway Alignment (DSHA) (JTG D20-2017) of China [41], in the PC-crash simulation software, we built a two-way two-lane road model with a carriageway width of 3.75 m, a crown slope of 2%, a slope gradient of 1 : 1 and a slope height of 5 m as a typical representative, and two rigid models for the car and truck. “BMW-116d autom” and “ASCHERSLEBEN KAROSS” were chosen as the represented car and truck model, respectively, and the initial position of the vehicles were set on the centre of the one-way lane; their parameters are shown in Table 1.

Notably, in the vehicle parameter setting, the steering of the vehicle was set ahead to match with different horizontal curve radii because we are unable to involve the driving behaviour factors considering the characteristics of the simulation software. For instance, when the horizontal curve radius is 200 m, the steering degree of the car is automatically updated to  $1.57^\circ$  and  $1.54^\circ$  to match the above radius by setting the turning radius of the vehicle as 200 m in the simulation software (see Figure 2(a)). In terms of the width value of the curve setting, according to the DSHA [41], the widened value was set only when the horizontal curve radius was no more than 250 m (see Figure 2(b)), and in case the horizontal curve radius was 200 m~250 m, the widened value was 0.4 m for car and 0.6 m for truck. Therefore, the corresponding widened values were set for different vehicle types in the simulation test.

Each variable value is shown in Table 2. Among these variables, horizontal curve radius, hard shoulder width, width value of the curve, adhesion coefficient, vehicle speed, and vehicle type can be set directly in the simulation software; however, longitudinal slope and superelevation slope need some complex operations to be set. For instance, the setting of the longitudinal slope can be achieved by adjusting the difference in height from the beginning to the end of the test section, and the difference in height  $h_1$  is calculated as follows:

$$h_1 = l_1 \times \sin(\arctan(i_1)), \quad (1)$$

where  $l_1$  represents the length of the test section (m) and  $i_1$  denotes the longitudinal slope (%) (the value of the downhill slope is positive).

Similarly, the setting of superelevation slope can be achieved by adjusting the difference in height from the outside to the inside of the test section, and the difference in height  $h_2$  is shown as follows:

$$h_2 = l_2 \times \sin(\arctan(i_2)), \quad (2)$$

where  $l_2$  is the width of the test section (including hard shoulder width) (m) and  $i_2$  denotes the superelevation slope (%), which is set in the middle of the test section and its value is positive when the outside height is greater than the inside height of the test section.

According to the value of each variable (excluding the width value of the curve) from the highway geometric design indexes and pavement condition (see Table 2),  $5 \times 4 \times 4 \times 4 \times 4 = 1280$  combinations were established, and then two kinds of the flat curve and curved slope combination sections were constructed corresponding to different hard shoulder widths, adhesion coefficients, and superelevation slopes. By applying 5 initial speeds to the vehicle and setting the width value of curve according to different vehicle types, simulation experiments were carried out for truck and car. Eventually,  $1280 \times 5 \times 2 = 12800$  simulation data were collected, in which the data of no roadside accidents occurrence was 9,973 (77.9%) and the data involving the occurrence of roadside accidents was 2,827 (22.1%).

**2.2. CHAID Decision Tree Technique.** The CHAID decision tree, as a data mining technique, has been widely applied in various fields, such as the airline industry and public transport management. However, few studies have investigated traffic risk, especially for roadside accidents.

The CHAID decision tree is a technique of database segmentation that is capable of extracting significant information from a large quantity of data [42, 43]. After a test order is conducted, the data are split by means of a statistical algorithm in CHAID. The original node on the independent variable is split into as many subgroups as possible, which are significantly different from binary variables. The process then splits these new nodes according to the variables that distinguish each of them. This process continues until no other splits are significant.

The CHAID analysis is generally called tree analysis, similar to a trunk (i.e., original node) being split into multiple branches; then, more branches until the trunk cannot be split any further in which case overfitting occurs. To identify optimal splits, the chi-square independence test is employed to examine and test the cross tabulations between each of the input variables (i.e., predictors of the occurrence of roadside accidents) and the outcome variables (i.e., occurrence of roadside accidents). The CHAID decision

TABLE 1: Vehicle parameter.

Parameter	Value	
	Car	Truck
Length (m)	4.325	6.370
Width (m)	1.765	2.500
Height (m)	1.420	3.100
Wheelbase (m)	2.690	3.700
Weight (kg)	1385	7200
Height of centre of gravity (m)	0.450	1.200
Distance of height of centre of gravity from front axle (m)	1.210	1.070
Tyre pattern	215/50 R 16 (621 mm)	7.50 R 16 (719 mm)
ABS	Yes	Yes
ESP	Yes	No



FIGURE 2: Setting (a) the steering degree for the vehicle and (b) the width value of the curve.

TABLE 2: Description of variables.

Variable	Value				
Highway geometric design indexes	Horizontal curve radius (m)	200	300	400	600
	Hard shoulder width (m)	0.75	1.5	2.25	3.00
	Longitudinal slope (%)	0	2	4	6
	Superelevation slope (%)	0	2	4	6
	Width value of curve (m)	0.4			0.6
Pavement condition	Adhesion coefficient	0.2	0.4	0.6	0.8
Traffic characteristics	Vehicle type	"Truck" = 0		"Car" = 1	
	Vehicle speed (km/h)	40	60	80	100
Output variable	Vehicle final state	"No departing from road" = 0		"Departing from road" = 1	

tree is, therefore, capable of providing detail that identifies the significant factors that result in the highest or lowest risk of roadside accidents using a series of if-then-else rules.

Furthermore, to prevent the occurrence of overfitting, CHAID uses  $P$  values with a Bonferroni correction as splitting criteria;  $P$  value criteria are sensitive to the number of data involved in the split and tend to avoid splitting into too small groups [44]; the smaller the  $P$  value is, the greater the goodness of tree model fitting. The  $P$  value of the  $F$  statistic for the difference in mean values is shown as follows [45]:

$$F = \frac{TSS - WSS(WSS/(g-1))}{WSS/(n-g)} \sim F_{(g-1),(n-g)}, \quad (3)$$

where TSS denotes the total sum of squares before the split, WSS is the variance,  $g$  represents the nodes generated by the split, and  $n$  is the number of categories of variables.

**2.3. Path Analysis.** Path analysis is a form of structural equation modelling (SEM), in which all the variables are observed variables. In the present study, SEM was used because the mediated and moderated relationships of a set of variables can be tested in SEM. In other words, SEM can not only test the direct impact of independent variables on dependent variables but also analyses the indirect effect on dependent variables through other variables (mediators). In path analysis, mediation, moderation, moderated mediation, and mediated moderation can all be tested [46], and mediation is a statistical approach applied to understand how a variable  $x$  delivers its effects to another variable  $z$ . In other words, whether the effect of  $x$  on  $z$  is direct only, indirect only or both direct, and indirect can be obtained in mediation analysis [47].

A simple mediation model describes a model in which the independent variable  $x$  has an impact on the dependent

variable  $z$  through a single mediator variable  $y$  (i.e.,  $x$  is assumed to have an impact on  $y$ , and this impact then transmits to  $z$ , apart from the direct relationship between  $x$  and  $z$ ). Two basic mediation models are built in equations (4) and (5). In particular, equation (4) represents the combination of the paths from  $x$  to  $z$  and  $y$  to  $z$ , and equation (5) represents the path from  $x$  to  $y$ :

$$z = \alpha_0 + \alpha_1 y_i + \alpha_2 x_i + \varepsilon_1, \quad (4)$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_2, \quad (5)$$

where  $z$  is the outcome variable,  $y_i$  is the mediator variable,  $x_i$  is the independent variable,  $\varepsilon_1$  and  $\varepsilon_2$  are the errors,  $\alpha_0$  and  $\beta_0$  are the intercepts of the models, and  $\alpha_1$ ,  $\beta_1$ , and  $\alpha_2$  are partial regression coefficients of the models.

However, these partial regression coefficients in the above models denote the direct effect of various variables but cannot reflect the magnitude of impact from these variables on the outcome variable due to the presence of their different units and standard deviation. For this purpose, a binary logistic regression model was fitted to obtain a standard regression coefficient that can meet the demand of testing the magnitude of direct effects from input variables on the outcome variable as follows:

$$\alpha'_i = \alpha_i \left( \frac{S_i}{S_Z} \right), \quad (6)$$

where  $\alpha'_i$  is the standard regression coefficient of  $x_i$ ;  $\alpha_i$  is the partial regression coefficient of  $x_i$ ;  $S_i$  is the standard deviation of  $x_i$ ; and  $S_Z$  is the standard deviation of the  $Z$  random variable in the logistic regression model, set as  $\pi/\sqrt{3}$  [48].  $\alpha'_i$  represents the magnitude of the direct effect of  $x_i$  on the outcome variable ( $z$ ).

Then, the indirect effect of  $x$  on  $z$  through all other mediator variables ( $y_i$ ) can be estimated using the product-of-coefficient estimator as in [46]

$$\gamma_i = \sum_{j=1}^n \beta_{ij} \alpha'_j, \quad (7)$$

where  $\gamma_i$  represents the magnitude of the indirect effect of  $x_i$  on  $z$  and  $\beta_{ij}$  is the correlation coefficient between  $x_i$  and  $y_j$ . Finally, the overall effects  $\chi_i$  (i.e., both the direct and indirect effects) of  $x_i$  on  $z$  can be computed as follows:

$$\chi_i = \gamma_i + \alpha'_i. \quad (8)$$

**2.4. Bayesian Network Analysis.** The Bayesian network became popular in the late 1990s and has been increasingly used since 2000. The Bayesian network, also known as the belief network, is regarded as one of the most effective theoretical models applied for representation and reasoning of uncertain knowledge. Bayesian nets and probabilistic directed acyclic graphs are technologies for graphically representing the joint probability distribution of a set of selected variables [49, 50]. The structure of the Bayesian network is a directed acyclic graph, in which node sets represent various variables and directed edges denote the

dependencies between variables. The confidence level or correlation strength between variables can be described using a conditional probability table (CPT). Tasks such as prediction, diagnosis, and classification can be realized through statistical inference functions and automatic learning of the Bayesian theorem. The structure of the Bayesian network can be regarded as the qualitative part of the model, while the added probability parameter represents a quantitative dimension to the model [51]. The Bayesian network represents various forms of uncertainty by using probability and applies probabilistic rules for achieving the training and reasoning processes, as shown in equations (9) and (10), respectively:

$$P(B_{ij} | A_j) = \frac{P(A_j | B_{ij})P(B_{ij})}{\sum_{j=1}^m P(A_j | B_{ij})P(B_{ij})}, \quad (9)$$

$$P(A_j | B_{1j} \dots B_{nj}) = \frac{P(B_{1j} | A_j)P(B_{2j} | A_j) \dots P(B_{nj} | A_j)P(A_j)}{\sum_{j=1}^m P(B_{1j} | A_j)P(B_{2j} | A_j) \dots P(B_{nj} | A_j)P(A_j)}, \quad (10)$$

where  $P(A_j | B_{ij})$  ( $i = 1, 2, \dots, n; j = 1, 2, \dots, m$ ) represents the prior probability of  $A_j$  (i.e., the final state of the vehicles in the accident simulation) under the effect of variable  $B_{ij}$  (i.e., the risk factor leading to roadside accident),  $P(B_{ij} | A_j)$  denotes the conditional probability of variable  $B_{ij}$  under the premise of  $A_j$  occurrence, and  $P(A_j | B_{1j} \dots B_{nj})$  is the posterior probability of  $A_j$  under the effects of a set of variables ( $B_1 \dots B_{nj}$ ). The above processes can also be achieved by efficient algorithms, such as the gradient descent (GD) algorithm in Netica software.

Compared to other theoretical models, the Bayesian network is suitable for traffic safety studies based on the following advantages: (1) combining data with expert experience and prior knowledge, (2) avoiding overfitting, (3) dealing with missing data, and (4) denoting causality by means of providing an understandable graph [52]. The Bayesian network, as an effective tool for developing an accident prediction model, has been widely used to predict accident injury [53–56] and frequency [57–59] and has demonstrated higher accuracy in predicting crash severity compared to regression models [60]. However, few studies have involved quantitative analysis of the probability of roadside accidents using the Bayesian network.

### 3. Results and Discussion

**3.1. Identification of Risk Factor.** For crossvalidation, we divided the accident data obtained from the simulation into a training dataset (70%) and a test dataset (30%). The training data were applied to fit the model and estimate the model parameters, while the test data were used to determine the model for its ability to generalize and confirm the model's applicability to independent variables. In the present study, we used exhaustive CHAID because it is superior in checking all possible splits [61]. To limit the growth of decision trees, we set the classification level to

four. Additionally, to minimize the intrinsic imbalanced nature of the data, a misclassification cost ratio of 100 : 1 was selected to promote CHAID to identify roadside accidents accurately more often.

CHAID provides the percentage of records with a particular value to the outcome variable, and the given value represents the confidence (accuracy) of the generated rules for the input variables. The overall classification accuracy of both the training set and testing set was 94% using the CHAID decision tree. Moreover, the  $p$  value in each node of both the training set and testing set was  $0.001 < 0.05$  (significance level), which indicates quite accurate classification with no overfitting.

CHAID analysis took 3,783 samples from the overall dataset for testing, and the percentage of roadside accident data was 22%. All data involving roadside accidents and nonroadside accident occurrences were divided into 67 subgroups from the parent node to child nodes through different branches. The percentage of roadside accidents varied from 0% to 100%. The decision tree included horizontal curve radius, hard shoulder width, longitudinal slope, adhesion coefficient, vehicle speed, and vehicle type in the final structure, which indicates that these variables are significant risk factors in determining the occurrence of roadside accidents. Other predictors not involved in the tree structure (i.e., superelevation slope and width value of the curve) only play a slight role in improving roadside safety performance.

Figure 3 only displays major tree structures that have a higher accuracy of generated classification rules due to the limitation of scope. The split at the first classification level was according to vehicle speed, which indicates that the influence of vehicle speed on the roadside accidents is relatively significant, while other risk factors are considered as nonsignificant risk factors at this classification level. By analogy, the classification of data at the second to four level could be obtained. Through the analysis of 3 783 test data, the generated decision rules were screened and sorted, as shown in Table 3.

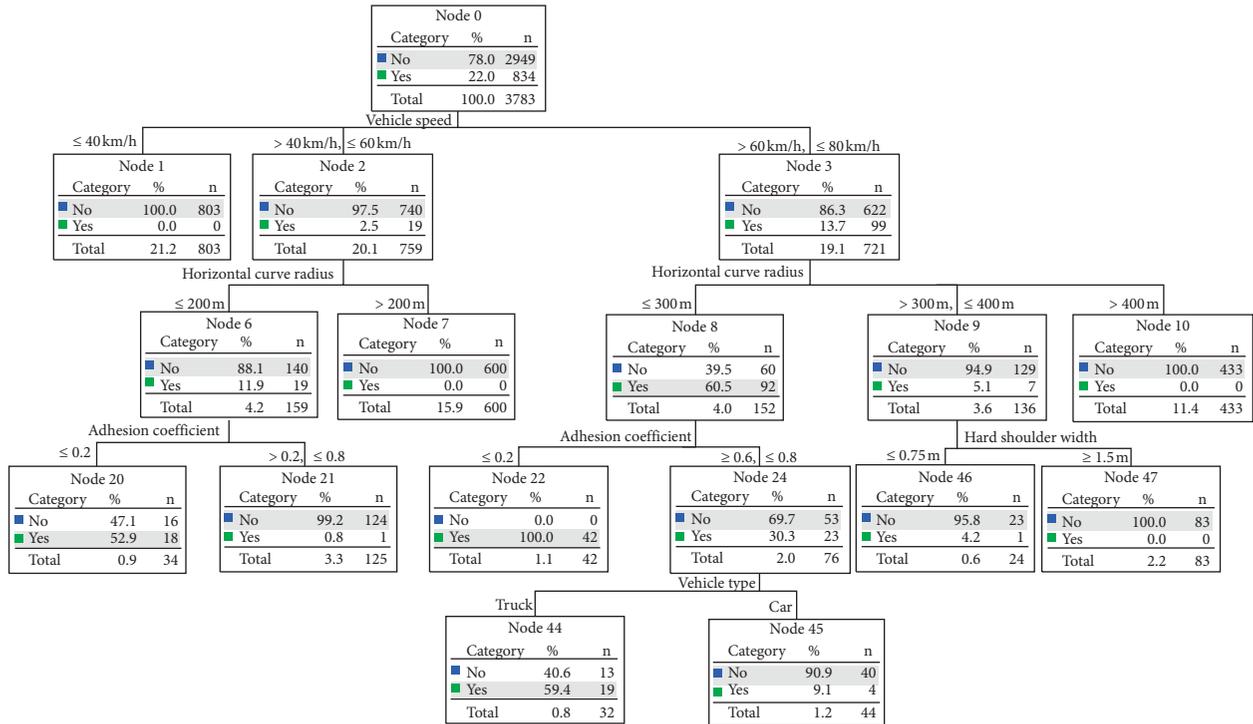
Each decision rule in Table 3 corresponds to different combinations of risk factors. By analyzing the influence of these combinations on the percentage of roadside accidents, some important conclusions and specific improved measure were obtained as follows:

- (1) According to decision rule 1, when  $V \leq 40$  km/h, other risk factors had no significant effect on the roadside accidents, and the percentage of roadside accidents was 0%. 40 km/h is, therefore, considered as the relatively safe speed to ensure the no occurrence of roadside accidents. Decision rules 2~12 presented that when  $V > 40$  km/h, there was a significant influence of horizontal curve radius on roadside accidents, and roadside accidents tend to decrease with an increase in the horizontal curve radius. Decision rules 12 showed that only when  $100 \text{ km/h} < V \leq 120 \text{ km/h}$  and  $200 \text{ m} < R \leq 300 \text{ m}$ , the longitudinal slope had a certain impact on the occurrence of truck roadside accidents, and in case of

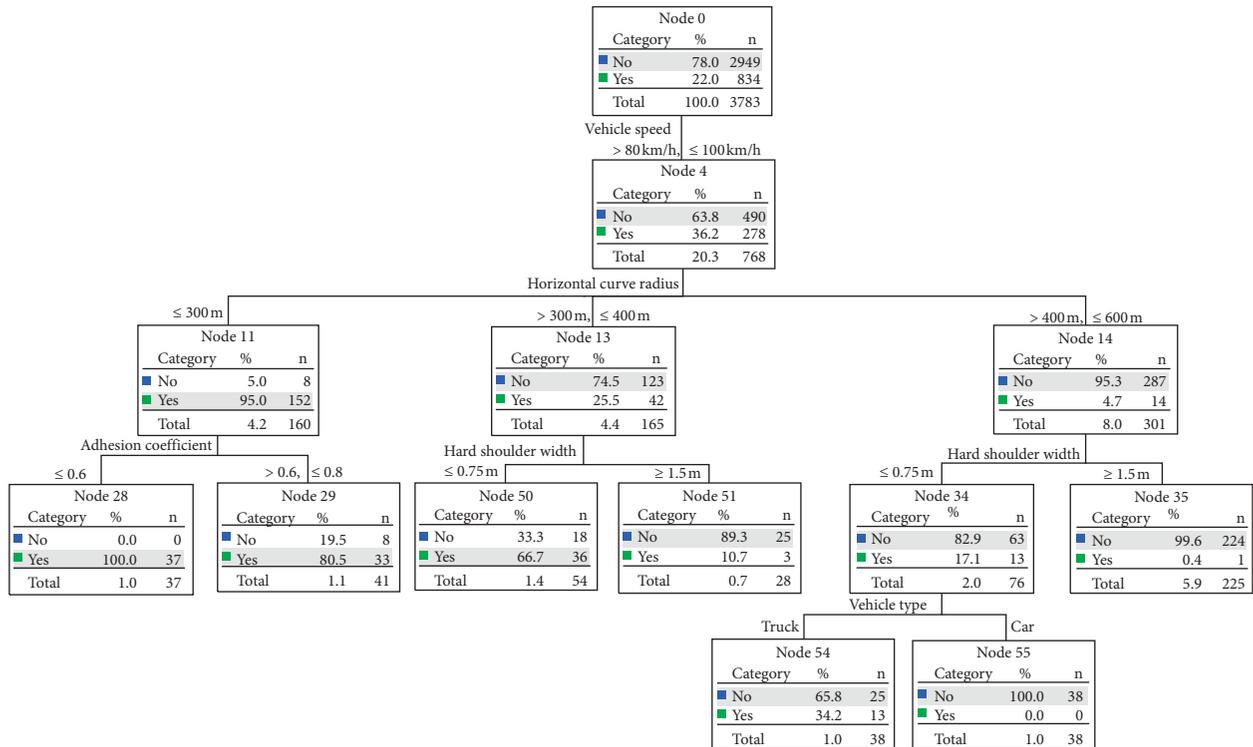
longitudinal slope  $\geq 4\%$ , the accidents percentage increased to 100%. This finding shows that the frequency of roadside accidents increases with a larger longitudinal slope.

- (2) Decision rules 6 and 9 presented that the percentage of roadside accidents for trucks was larger than that for cars under the same road condition, which can be concluded that trucks have a higher risk of roadside accidents compared to accidents involving cars because the higher centre of gravity for trucks cause them to be more likely to rollover than cars.
- (3) It can be seen from decision rules 2 and 5 that, in case of  $40 \text{ km/h} < V \leq 60 \text{ km/h}$  and  $R \leq 200 \text{ m}$ , as well as when  $60 \text{ km/h} < V \leq 80 \text{ km/h}$  and  $R \leq 300 \text{ m}$ , adhesion coefficient showed a significant impact on roadside accidents, and the percentage of roadside accidents gradually decreased as adhesion coefficient increased. Therefore, antislip measures should be strengthened for the highway with the above operating speed and horizontal curve radius. The abovementioned conclusion can be used as a supplement to the revision of the DSHA.
- (4) According to decision rule 7, when  $60 \text{ km/h} < V \leq 80 \text{ km/h}$  and  $300 \text{ m} < R \leq 400 \text{ m}$ , hard shoulder width played a certain role in reducing roadside accidents, but the improvement is not obvious. According to decision rules 8 and 9, in case of  $80 \text{ km/h} < V \leq 100 \text{ km/h}$  and  $300 \text{ m} < R \leq 600 \text{ m}$ , hard shoulder width had a significant impact on roadside accidents, and setting hard shoulder width  $\geq 1.5 \text{ m}$  could obviously reduce the percentage of roadside accidents. Therefore, for the highway with the above operating speed and horizontal curve radius, the width of hard shoulder should be set as  $\geq 1.5 \text{ m}$ .
- (5) Decision rule 9 showed that, in case of  $80 \text{ km/h} < V \leq 100 \text{ km/h}$  and  $400 \text{ m} < R \leq 600 \text{ m}$ , if the width of hard shoulder  $\leq 0.75 \text{ m}$ , the percentage of truck roadside accidents was 34.2% and that of car roadside accidents was 0%; if the hard shoulder width  $\geq 1.5 \text{ m}$ , the percentage of roadside accidents was only 0.4% for both trucks and cars. This finding adequately illustrates that the hard shoulder width has more significant impact on the frequency of roadside accidents involving trucks than cars.
- (6) It can be seen from decision rules 10 and 11 that when  $100 \text{ km/h} < V \leq 120 \text{ km/h}$  and  $300 \text{ m} < R \leq 600 \text{ m}$ , a setting of hard shoulder width  $\geq 2.25 \text{ m}$  can effectively avoid the occurrence of truck roadside accidents. Therefore, for freeway with the above operating speed and horizontal curve radius, the width of hard shoulder should be set as  $\geq 2.25 \text{ m}$  to ensure driving safety of trucks.

Using decision tree analysis, we discussed the relationship between different combinations of risk predictors and the occurrence of roadside accidents and identified the significant risk factors resulting in roadside accidents.



(a)



(b)

FIGURE 3: Continued.

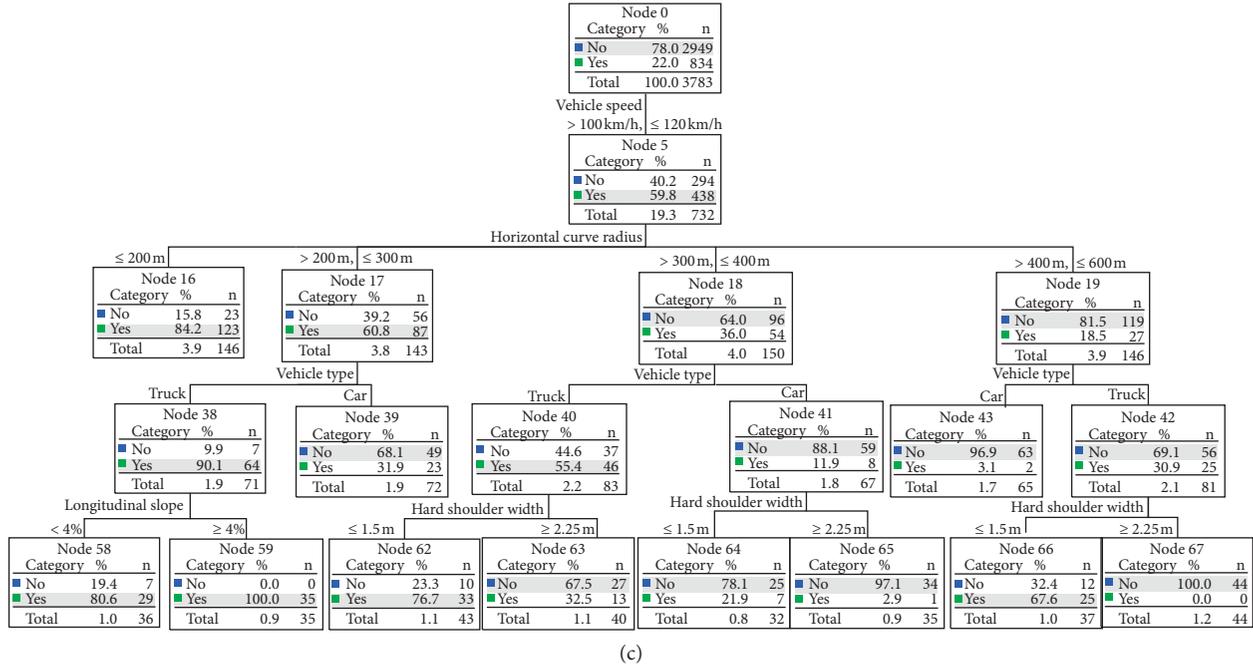


FIGURE 3: Decision tree for the identification of risk factors. (a) Tree 1. (b) Tree 2. (c) Tree 3. Note. “No” represents “no roadside accident occurrence” and “Yes” represents “roadside accident occurrence”.

TABLE 3: Decision rules.

No.	Classification level				Percentage (%)
	1	2	3	4	
1	$V \leq 40$	—	—	—	0
2	$40 < V \leq 60$	$R \leq 200$	$\mu \leq 0.2$	—	52.9
3	—	$200 < R \leq 300$	$0.2 < \mu \leq 0.8$	—	0.8
4	—	$300 < R \leq 400$	—	—	60.5
5	$60 < V \leq 80$	$R \leq 300$	$\mu \leq 0.2$	—	100
6	—	—	$0.6 \leq \mu \leq 0.8$	Truck	59.4
7	—	$300 < R \leq 400$	$0.6 \leq \mu \leq 0.8$	Car	9.1
8	$80 < V \leq 100$	$300 < R \leq 400$	$w \leq 0.75$	—	4.2
9	—	$400 < R \leq 600$	$w \geq 1.5$	—	0
10	$100 < V \leq 120$	$300 < R \leq 400$	$w \leq 0.75$	Truck	76.7
11	—	$400 < R \leq 600$	$w \geq 1.5$	—	32.5
12	—	$200 < R \leq 300$	Truck	$w \leq 1.5$	67.6
				$w \geq 2.25$	0
				$\alpha < 4$	80.6
				$\alpha \geq 4$	100

Note:  $V$  represents the vehicle speed (km/h),  $R$  is the horizontal curve radius (m),  $\mu$  represents the adhesion coefficient,  $w$  denotes the hard shoulder width (m),  $\alpha$  represents the longitudinal slope (%), and — indicates that the risk factors at the corresponding decision rules have no significant impact on roadside accidents.

However, the magnitude of the importance of these factors has not been investigated. To obtain a deeper insight into the interactions of factors and their impacts on roadside

accidents, a path analysis based on a logistic regression model was built.

**3.2. Importance of Risk Factors.** We input the risk factors (horizontal curve radius, hard shoulder width, longitudinal slope, adhesion coefficient, vehicle speed, and vehicle type) into the path analysis model and found that these factors were also statistically significant because they were all retained by the model. The coefficient of determination  $R^2 = 0.868$ , illustrating the model fit, is good. Table 4 shows the outcomes of the model and represents the direct effects of different variables on roadside accident occurrence. According to the magnitude of direct effects, the most important risk factors were in decreasing order of importance, vehicle speed (3.321), horizontal curve radius (−2.572), vehicle type (−1.005), adhesion coefficient (−0.827), hard shoulder width (−0.812), and longitudinal slope (0.314). As expected, vehicle speed and longitudinal slope were found to be positively correlated with the occurrence of roadside accidents. In contrast, horizontal curve radius, vehicle type, adhesion coefficient, and hard shoulder width were inversely related to roadside accidents.

It is important to note that unlike real accident data, there seemed to be no interaction between factors in the present study because the values of all these factors were set artificially in the simulation. However, to investigate the indirect effects caused by the interaction of variables on the occurrence of roadside accidents, we assumed that the correlation coefficient between variables could be regarded as their interaction.

A structural diagram of path analysis is shown as Figure 4. This figure represents that all risk factors are correlated

TABLE 4: Modelling results.

Variable	Parameter estimate	S.E. <sup>a</sup>	S.D. <sup>b</sup>	P value	Standard parameter estimate
Horizontal curve radius	-0.033	0.001	141.388	<0.05**	-2.572
Hard shoulder width	-2.527	0.108	0.583	<0.05**	-0.812
Longitudinal slope	0.430	0.027	1.325	<0.05**	0.314
Adhesion coefficient	-6.699	0.279	0.224	<0.05**	-0.827
Vehicle speed	0.213	0.006	28.282	<0.05**	3.321
Vehicle type	-3.645	0.136	0.5	<0.05**	-1.005

Note. <sup>a</sup>is standard error. <sup>b</sup>is standard deviation. \*\*indicates 95% confidence level is used (i.e., P value <0.05 is statistically significant).

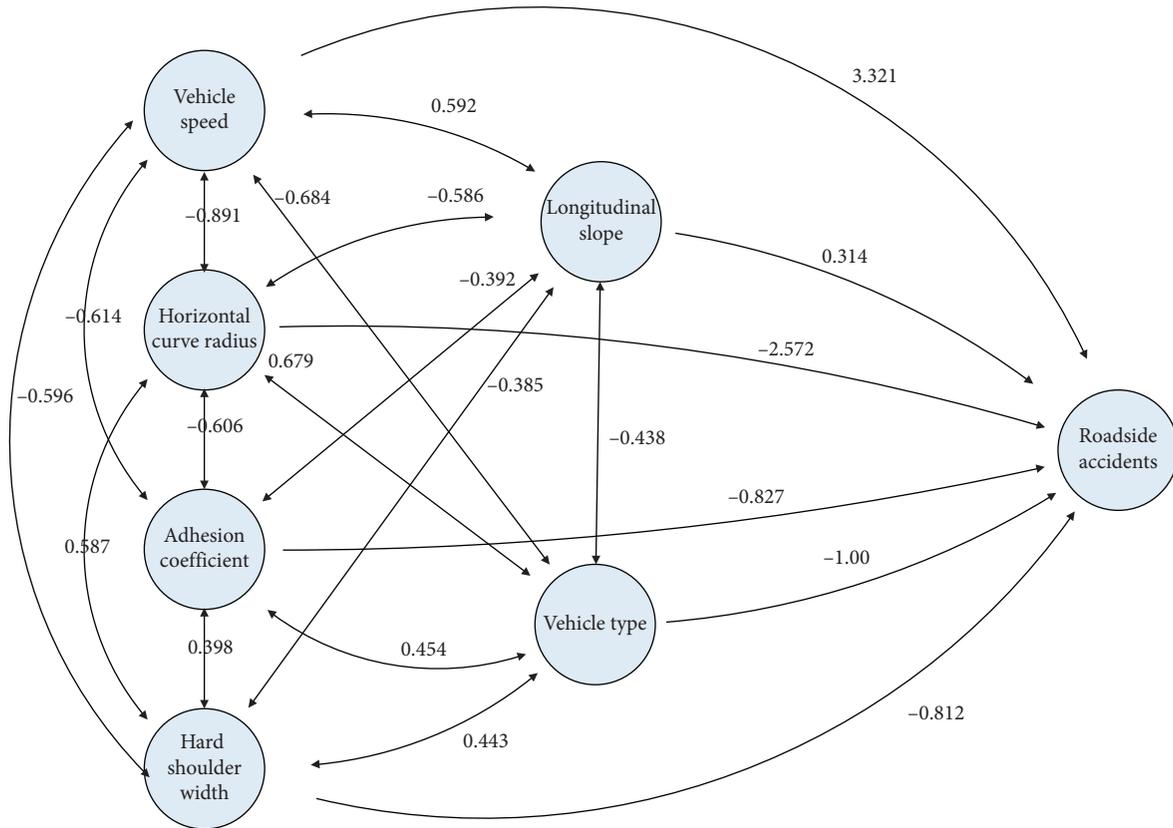


FIGURE 4: Structure diagram of path analysis.

and indicates that apart from direct effects, all risk factors had indirect effects on roadside accidents through other factors based on these correlations in the model. Among these interactions, the combination of vehicle speed-horizontal curve radius had the largest impact of interaction (-0.891) on roadside accident occurrence, and the negative interaction indicated that there was a mutually restricted relationship between these two factors. There were also other large interactions involved in Figure 4, including vehicle speed-vehicle type (-0.684), horizontal curve radius-vehicle type (0.679), vehicle speed-adhesion coefficient (-0.614), horizontal curve radius-adhesion coefficient (-0.606), and vehicle speed-hard shoulder width (-0.596).

Table 5 mainly shows the indirect effect of each risk factor through another mediating factor and the overall effect on roadside accident occurrence. It can be observed

that vehicle speed transmitted its largest indirect effect (2.292) on roadside accidents through the horizontal curve radius than other factors, while horizontal curve radius had the largest indirect effect (-2.960) on accidents by vehicle speed. In addition, it was interesting to note that all other factors also had their largest and second indirect effects on roadside accidents by means of vehicle speed and horizontal curve radius. These results emphasize that vehicle speed and horizontal curve radius are still the most significant risk factors causing roadside accidents.

According to the overall effect of each risk factor on roadside accidents shown in Table 5, the most important risk factors were in decreasing order of importance, vehicle speed (7.749), horizontal curve radius (-7.644), vehicle type (-6.086), adhesion coefficient (-5.496), hard shoulder width (-5.373), and longitudinal slope (2.607). It is significant to

TABLE 5: The magnitude of the impact of factors on roadside accidents.

Variable	Direct effect	Indirect effect						Overall effect
		Vehicle speed	Horizontal curve radius	Adhesion coefficient	Hard shoulder width	Longitudinal slope	Vehicle type	
Vehicle speed	3.321	—	2.292	0.508	0.484	0.461	0.684	7.749
Horizontal curve radius	-2.572	-2.960	—	-0.501	-0.477	-0.456	-0.679	-7.644
Adhesion coefficient	-0.827	-2.040	-1.559	—	-0.323	-0.305	-0.443	-5.496
Hard shoulder width	-0.812	-1.979	-1.510	-0.329	—	-0.300	-0.443	-5.373
Longitudinal slope	0.314	1.066	1.007	0.124	0.013	—	0.083	2.607
Vehicle type	-1.005	-2.272	-1.746	-0.366	-0.360	-0.341	—	-6.086

note that the order of importance of these risk factors was not changed by the overall effects compared to the direct effects. This finding indicates that the indirect effects of different factors are not expected to play an important role in the occurrence of roadside accidents.

**3.3. Probability of Roadside Accidents.** Given that Bayesian network performs best with a small set of variables [62] and the least impact was longitudinal slope on roadside accidents compared to other important factors, we input the first five important factors (i.e., vehicle speed, horizontal curve radius, vehicle type, adhesion coefficient, and hard shoulder width) into a Bayesian network analysis to establish the probability prediction model for roadside accidents.

In the present study, the Bayesian network structure was developed based on the results of path analysis, and the Bayesian network parameter learning of roadside accidents was performed using the GD algorithm in Netica software, in which the prior and conditional probability distribution of each node could be obtained. In addition, according to the sensitivity analysis (see Table 6), the order of nodes (variables) based on the magnitude of mutual information (impact on roadside accidents) was consistent with the order obtained from path analysis, indicating that an accurate Bayesian network model used to predict the probability of roadside accidents was built (see Figure 5).

The probability of roadside accidents (i.e., posterior probability) under different combinations of variables can be obtained in this prediction model. For instance, assuming that a road section was a dry asphalt pavement with a speed limit of 80 km/h, a horizontal curve radius of 235 m, and a hard shoulder width of 0.75 m, then the probability of roadside accidents for truck passing through above road section need be predicted. First, the state of  $60 \text{ km/h} < V \leq 80 \text{ km/h}$ ,  $200 \text{ m} < R \leq 300 \text{ m}$ ,  $0.6 \leq \mu \leq 0.8$ ,  $w \leq 0.75 \text{ m}$ , and vehicle type of 0 were as 100%, and after automatically updating the probabilities of the whole network, the calculated probability of roadside accidents for truck driving at a speed of  $60 \text{ km/h} < V \leq 80 \text{ km/h}$  was 38.7% (see Figure 6).

Furthermore, the developed prediction model can also predict probabilities under the effects of any number (from 1 to 5) of factors (i.e., in the absence of some factors). For example, given that the speed limit of a road section was 80 km/h and the width of hard shoulder was 0.75 m, but lack

TABLE 6: Sensitivity analysis result of the node “roadside accident.”

Node	Mutual info	Percent	Variance in beliefs
Roadside accident	0.84054	100	0.1968046
Vehicle speed	0.13101	15.6	0.0358728
Horizontal curve radius	0.07927	9.43	0.0220364
Vehicle type	0.01063	1.27	0.0028862
Adhesion coefficient	0.00959	1.14	0.0026972
Hard shoulder width	0.00731	0.87	0.0020842

of other indicators, and it could also be calculated that the probability of roadside accidents for car with a speed of  $60 \text{ km/h} < V \leq 80 \text{ km/h}$  was 18.2% (see Figure 7(a)).

For another example, assume a road section was a dry asphalt pavement and horizontal curve radius and hard shoulder width were unknown. If the speed limit of this section was 80 km/h, the probabilities of roadside accident were 3.52% for car and 14.9% for truck (see Figures 7(b) and 7(c)), whereas the same probabilities were 14.3% for car and 44.5% for truck if the speed limit was 100 km/h (see Figures 7(d) and 7(e)), which further indicates that trucks have a higher risk of rollover than cars, especially when the vehicle speed was great than 80 km/h. Of course, the more factors involved, the more precise the obtained probability.

It is important to note that when various variables were in extreme states tending to avoid roadside accidents, even if vehicle speed was set as 120 km/h, whether for car or truck, and the probability of roadside accidents was, not as expected, only 1.31% (see Figure 7(f)), which adequately illustrates the importance of reasonable road design in situations where the driver’s behaviour cannot be controlled. Therefore, in the purpose of further improving roadside safety and identifying the road conditions in which roadside accidents occur frequently, a variety of thresholds of horizontal curve radius, adhesion coefficient, and hard shoulder width corresponding to different vehicle speeds and vehicle types are given based on the Bayesian network prediction model, as shown in Table 7.

**3.4. Identification of Roadside Accident Blackspot.** We considered that there was a high frequency of roadside accidents (i.e., accident blackspot) when the probability of roadside

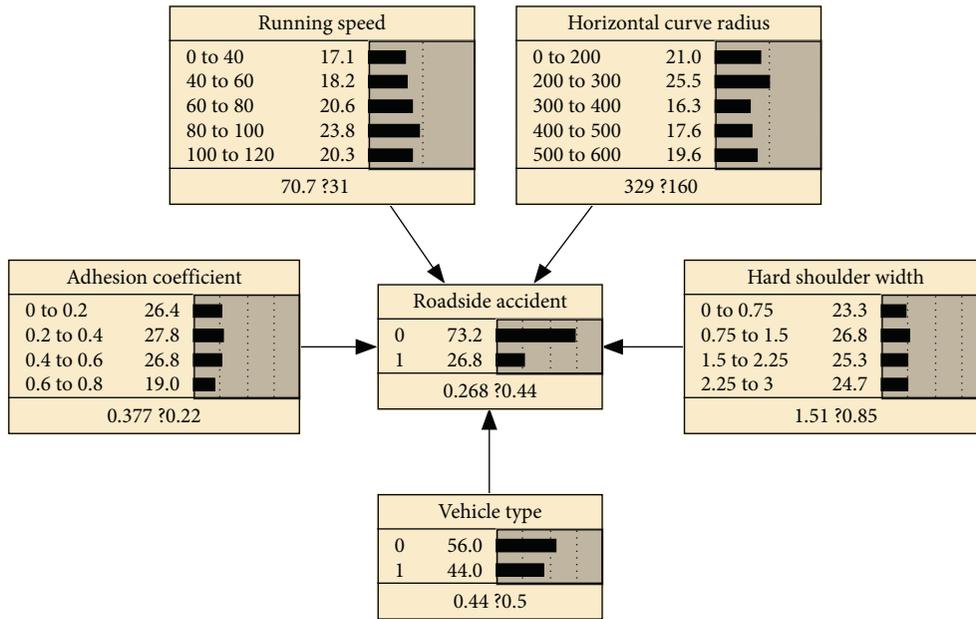


FIGURE 5: Bayesian network prediction model of roadside accidents. Note: 0 represents “truck” and 1 represents “car” in vehicle type; 0 denotes “no roadside accident occurrence” and 1 denotes “roadside accidents occurrence” in roadside accidents.

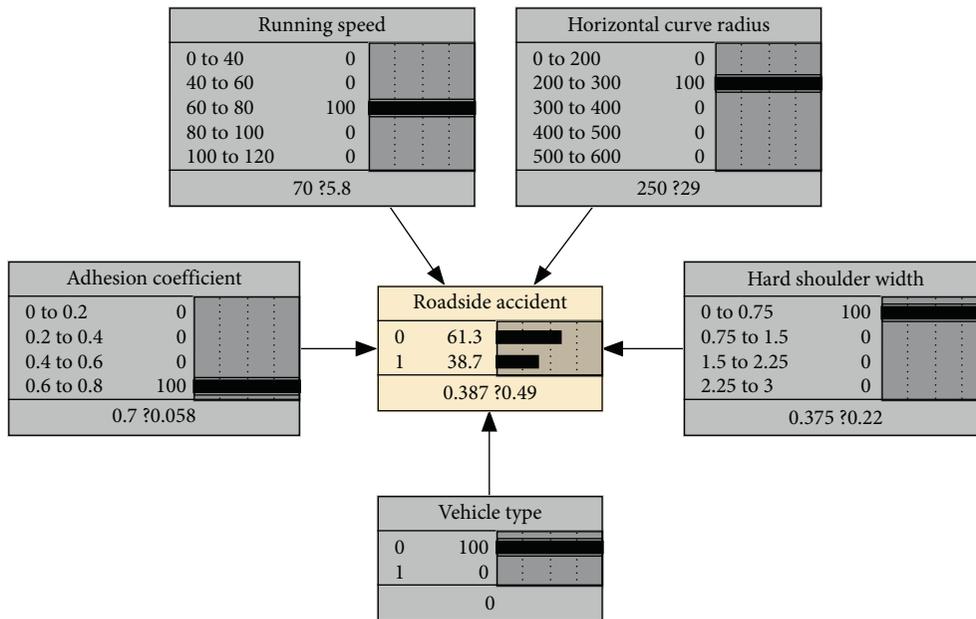


FIGURE 6: Probability calculation of roadside accidents.

accidents occurrence was greater than that of no roadside accidents occurrence (i.e., the probability of roadside accidents was greater than 50%). According to the results from Table 7, a range of vehicle speeds corresponds to 1 to 4 identification rules for roadside accident blackspots. When the value of each risk factor from a certain road section meets any of these 18 identification rules, this road section is then judged to be the road section with frequent occurrences of roadside accidents.

In this paper, a section (K2639 + 498.02 to K2679 + 170) from G105 was selected to confirm the effectiveness of the proposed method of identification. The G105 is a first-class road with a design speed of 80 km/h. By collecting road design documents and data of annual operating speed, the location of K2669 + 256.378 is determined to be the road section with frequent accidents according to the risk factor threshold, as shown in Table 7. The horizontal curve radius of this location is 280 m, the width of the hard shoulder is

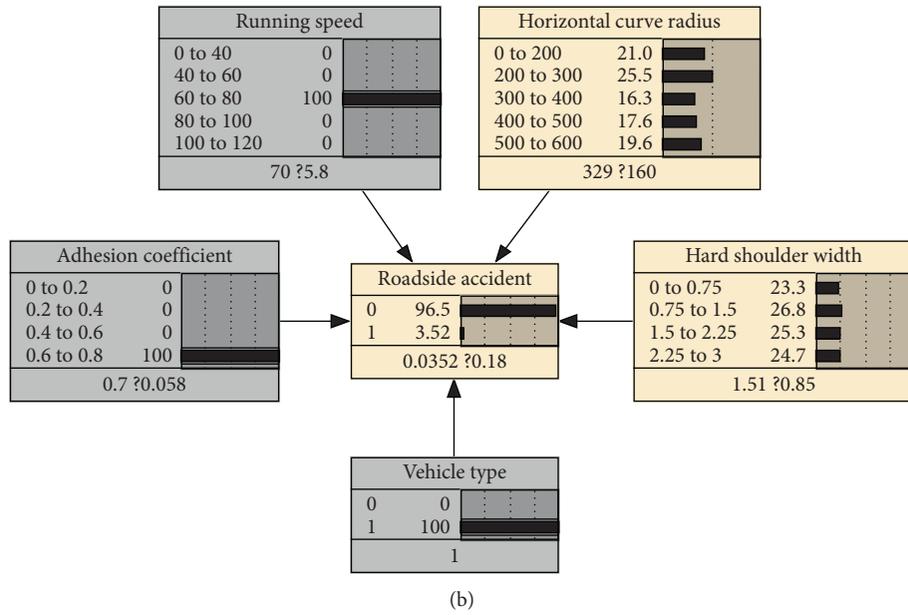
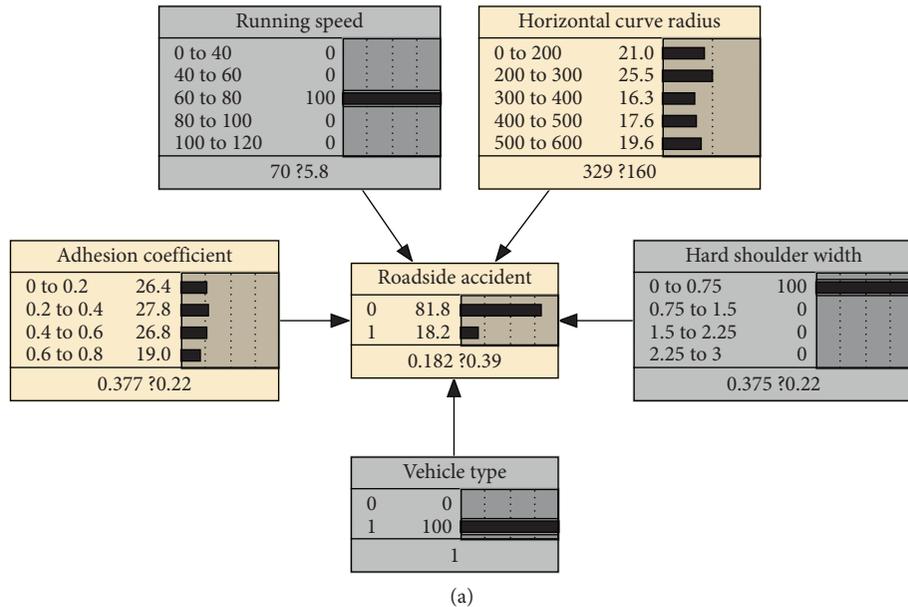


FIGURE 7: Continued.

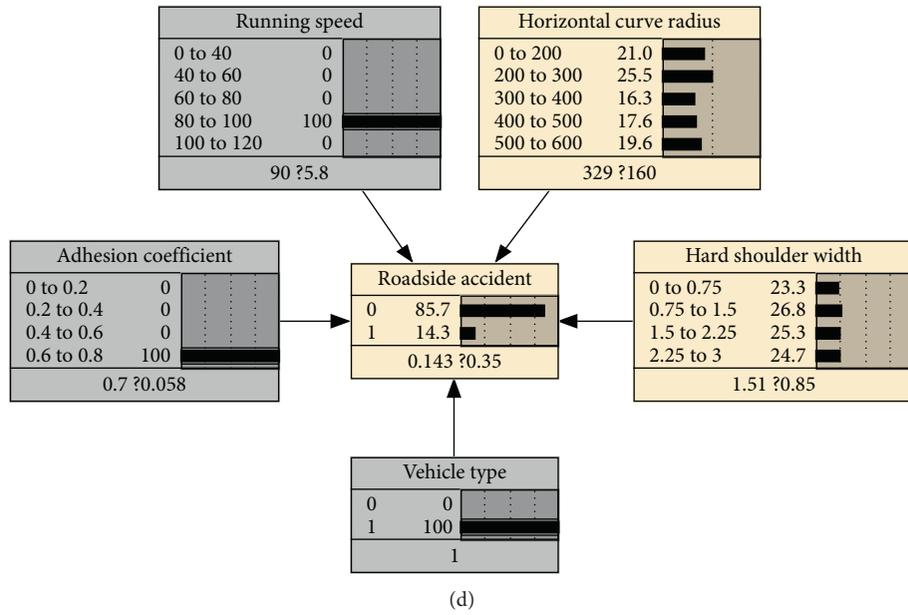
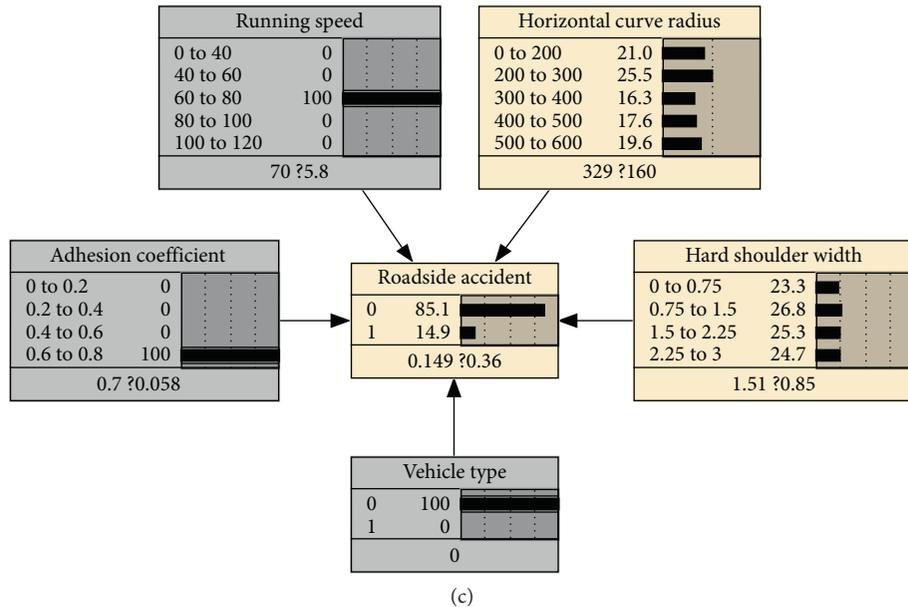


FIGURE 7: Continued.

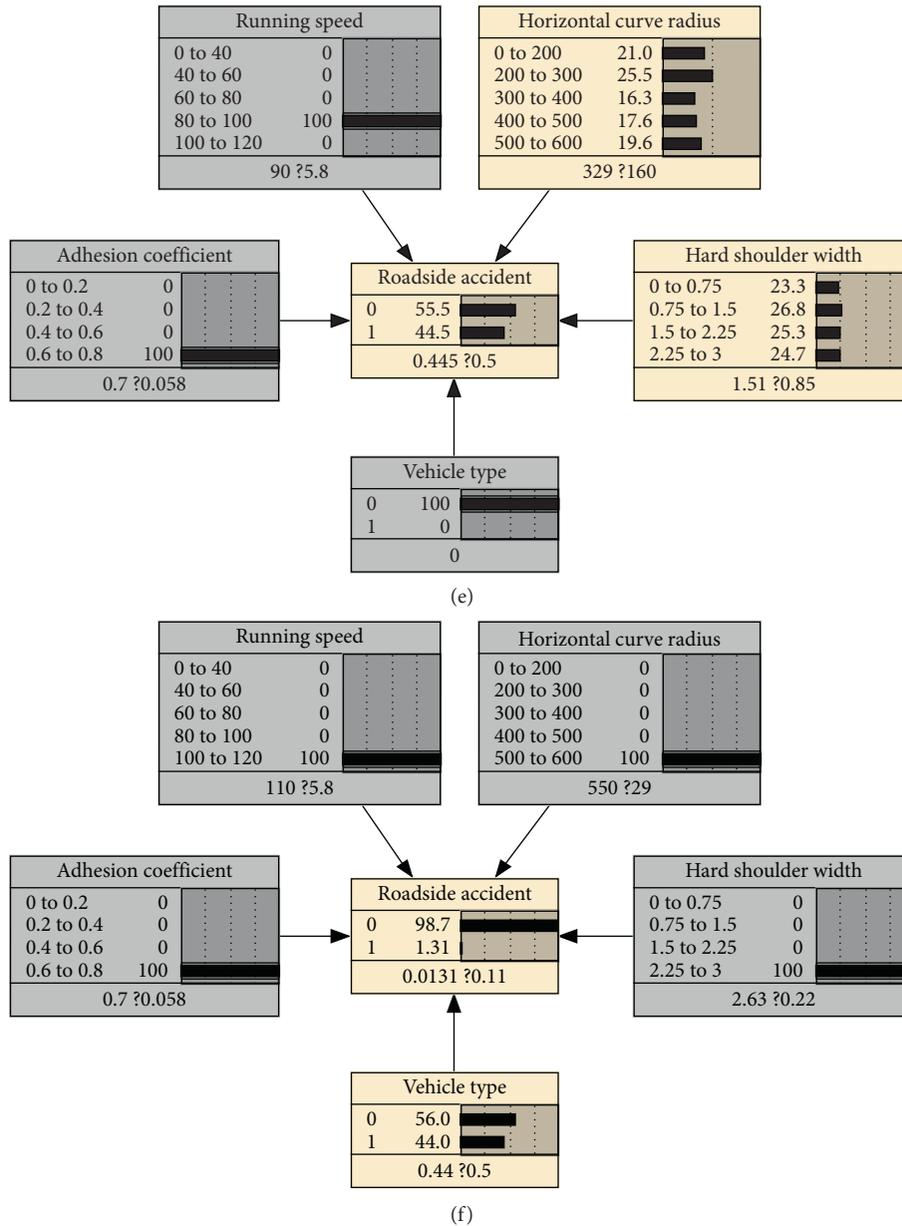


FIGURE 7: Probability calculation of roadside accidents. (a) Result 1. (b) Result 2. (c) Result 3. (d) Result 4. (e) Result 5. (f) Result 6.

1.5 m, the operating speed of cars are mainly distributed in  $120 \text{ km/h} > V \geq 100 \text{ km/h}$  (see Figure 8(a)), and that of trucks are mainly distributed in  $100 \text{ km/h} > V \geq 80 \text{ km/h}$  (Figure 8(b)). The above indicators, respectively, conform to the 7 and 17 identification rules. According to the traffic police department’s accident records, there were more than 80 roadside accidents in the above section from 2014 to 2018, which has been classified as the roadside accident-prone road. Based on the above analysis, the reliability of the proposed identification method for roadside accident blackspot in this paper is, therefore, verified.

The importance of such a study lies in the fact that it can help authorities identify significant risk factors that result in

frequent roadside accidents in small curve segments to implement effective countermeasures or optimize alignment design in the process of future road construction and reconstruction. For instance, most of the thresholds for trucks were larger than those for cars at the same vehicle speed in Table 7, which suggests that the higher designed standard of geometric design and pavement condition is required for truck driving safety. Furthermore, for curve sections with truck speeds of no less than 60 km/h or car speeds of no less than 80 km/h, some thresholds of adhesion coefficients had almost reached the maximum 0.8. Therefore, we can reduce the risk of roadside accidents by optimizing other factors according to their respective thresholds.

TABLE 7: Threshold of significant factors leading to frequent roadside accidents.

No.	Vehicle speed (km/h)	Vehicle type	Horizontal curve radius (m)	Hard shoulder width (m)	Adhesion coefficient	Probability (>50%)
1			$300 < R \leq 400$	$w \leq 0.75$	$\mu \leq 0.2$	$\geq 55.9$
2	$80 \geq V > 60$	Truck	$200 < R \leq 300$	$w \leq 0.75$	$\mu \leq 0.6$	$\geq 54.3$
3			$R \leq 200$	$w \leq 1.50$	$\mu \leq 0.8$	$\geq 54.9$
4			$500 < R \leq 600$	$w \leq 0.75$	$\mu \leq 0.2$	$\geq 54.2$
5	$100 \geq V > 80$		$400 < R \leq 500$	$w \leq 0.75$	$\mu \leq 0.8$	$\geq 55.0$
6			$300 < R \leq 400$	$w \leq 1.50$	$\mu \leq 0.8$	$\geq 54.8$
7			$R \leq 300$	$w \leq 2.25$	$\mu \leq 0.8$	$\geq 55.3$
8			$500 < R \leq 600$	$1.50 < w \leq 2.25$	$\mu \leq 0.2$	$\geq 55.3$
9	$120 \geq V > 100$			$w \leq 1.50$	$\mu \leq 0.8$	$\geq 54.3$
10			$R \leq 500$	$w \leq 2.25$	$\mu \leq 0.8$	$\geq 54.2$
11	$80 \geq V > 60$	Car	$R \leq 300$	$w \leq 0.75$	$\mu \leq 0.2$	$\geq 54.2$
12			$300 < R \leq 400$	$w \leq 1.50$	$\mu \leq 0.2$	$\geq 55.3$
13	$100 \geq V > 80$		$200 < R \leq 300$	$w \leq 1.50$	$\mu \leq 0.4$	$\geq 55.3$
14			$R \leq 200$	$w \leq 2.25$	$\mu \leq 0.8$	$\geq 54.2$
15			$300 < R \leq 400$	$w \leq 2.25$	$\mu \leq 0.4$	$\geq 54.5$
16	$120 \geq V > 100$		$200 < R \leq 300$	$w \leq 1.50$	$0.4 < \mu \leq 0.6$	$\geq 54.6$
17				$w \leq 2.25$	$\mu \leq 0.8$	$\geq 55.9$
18			$R \leq 200$	$w \leq 3.00$	$\mu \leq 0.8$	$\geq 54.3$

Note.  $V$  represents the vehicle speed,  $R$  represents the horizontal curve radius,  $\mu$  denotes the adhesion coefficient, and  $w$  is the hard shoulder width.

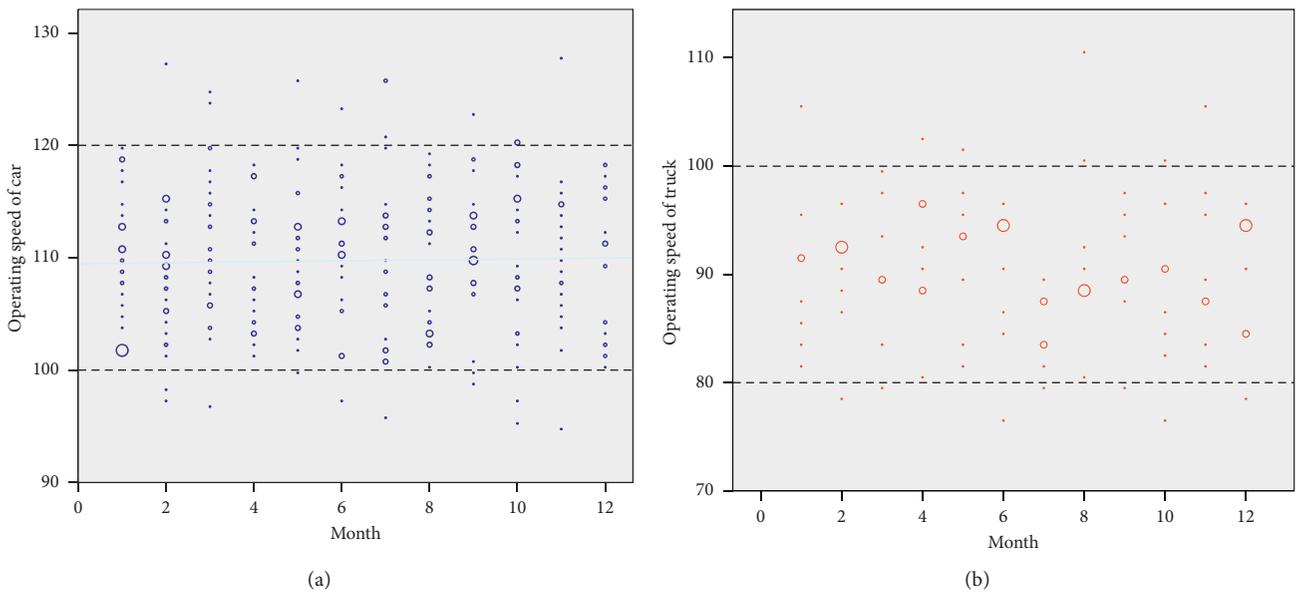


FIGURE 8: Operating speed distributions. (a) Car. (b) Truck.

### 4. Conclusions and Recommendations

The issue of roadside safety is crucial, especially for curve sections. In the present study, we employed CHAID decision tree analysis to identify significant risk factors resulting in the occurrence of roadside accidents, explored the impact of different combinations of risk factors on roadside accidents, and then used path analysis to determine the importance of these significant risk factors by investigating their direct and indirect effects on roadside accident occurrence. According to the results of the CHAID technique and path analysis, the significant predictors were in decreasing order of importance, vehicle speed, horizontal curve radius, vehicle type,

adhesion coefficient, hard shoulder width, and longitudinal slope. The first five important factors were included as predictors of the probability of roadside accidents in the Bayesian network analysis to establish the probability prediction model of roadside accidents. Based on the results of probabilities of roadside accidents, the thresholds of horizontal curve radius, adhesion coefficient, and hard shoulder width corresponding to different vehicle speeds and vehicle types for accident blackspot identification in curve section were given.

These findings contribute to improving roadside safety in curve sections with a small radius. For instance, we confirmed again that vehicle speed and horizontal curve

radius are still the most critical factors leading to roadside accidents, whether in this study or other previous literature [63–66], and road sections with a high running speed and small radius are usually regarded as accident blackspot areas. Furthermore, based on the results of CHAID analysis, some specific recommended countermeasures as a supplement or reference for the revision of the DSHA of China were proposed as follows:

- (i) For the highway with an operating speed of 60 km/h and a horizontal curve radius  $\leq 200$  m or an operating speed of 80 km/h and a horizontal curve radius  $\leq 300$  m, antislip measures should be strengthened
- (ii) For the highway with an operating speed of 100 km/h and a horizontal curve radius of  $300 \text{ m} < R \leq 600 \text{ m}$ , the width of hard shoulder should be set as  $\geq 1.5$  m
- (iii) For the freeway with an operating speed of 120 km/h and a horizontal curve radius of  $300 \text{ m} < R \leq 600 \text{ m}$ , the width of hard shoulder should be set as  $\geq 2.25$  m to ensure driving safety of trucks

Another important findings is that compared with cars, the width of the hard shoulder has a more significant influence on roadside accidents involving trucks, and trucks are more likely to have roadside accidents, especially in case of the vehicle speed  $> 80$  km/h. To ensure truck driving safety, the design standards of the horizontal curve radius, adhesion coefficient, and hard shoulder width should be further improved by decision makers in future highways construction. Additionally, limiting the load and running speed can be the most effective measures to mitigate the risk resulting from a higher centre of gravity. In recent years, a real-time monitoring system transmitting warning messages to truck drivers in cases of overload or overspeed has been designed by combining embedded technology and GPRS technology [67]. This system is expected to perform well in reducing truck roadside accidents. Another countermeasure is that regular maintenance of the truck, in case the brake failed in an emergency, also contributes to a decrease of accident rate [68].

The most remarkable result in this paper is that the developed Bayesian network prediction model can achieve the quantitative analysis of the probability of roadside accidents under the effects of any number (from 1 to 5) of factors. The resulting threshold of factors leading to accident blackspot can be a guide for authorities to identify and check roadside accidents prone areas located in small curve sections. In fact, if there are obstacles to promoting safe design standards for the horizontal curve radius, the adhesion coefficient, and the hard shoulder width due to high construction cost or unrealistic issues, many other effective countermeasures, such as setting deceleration strips in the pavement or related warning signs to control running speeds, widening the road in curve sections to provide a fault-tolerant space for drivers [69], and removing roadside hazards to reduce the loss of run-off-road accidents [70], could also be implemented.

Despite these promising results, some limitations exist in this paper. For example, this paper mainly predicts the roadside accident probability for two-way two-lanes or outer lanes of more than two lanes. Therefore, it remains to be further studied whether the prediction model is applicable to other road types (e.g., inner lanes of more than two lanes). In future studies, given the important impact of vehicle speed on roadside accidents, the limitation of maximum safe speed corresponding to different road geometric designs will be an additional research direction.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported by the National Key Research and Development Program of China (no. 2018YFB1600902), MOE Layout Foundation of Humanities and Social Sciences (no. 18YJAZH009), National Natural Science Foundation of China (no. 51778063), and Fundamental Research Funds for the Central Universities (no. 2572019AB26).

## References

- [1] National Highway Traffic Safety Administration (NHTSA), *Fatality Analysis Reporting System*, National Highway Traffic Safety Administration (NHTSA), Washington, DC, USA, 2018.
- [2] Traffic Management Bureau of the Ministry of Public Security, *Annual Road Traffic Accident Statistics*, China Communications Press, Beijing, China, 2019.
- [3] S. B. McLaughlin, J. M. Hankey, S. G. Klauer, and T. A. Dingus, *Contributing Factors to Run-Off-Road Crashes and Near-Crashes*, National Highway Traffic Safety Administration (NHTSA), Washington, DC, USA, 2009, <http://hdl.handle.net/10919/55073>.
- [4] C. Roque and M. Jalayer, “Improving roadside design policies for safety enhancement using hazard-based duration modeling,” *Accident Analysis & Prevention*, vol. 120, pp. 165–173, 2018.
- [5] D. Lord, M. A. Brewer, K. Fitzpatrick et al., *Analysis of Roadway Departure Crashes on Two Lane Rural Roads in Texas*, Texas Transportation Institute, College Station, TX, USA, 2011.
- [6] I. Cruzado and E. Donnell, “Models of vehicle operating speeds along two-lane rural highway transition zones: panel and multilevel modeling approaches,” *Transportation Letters*, vol. 3, no. 4, pp. 265–278, 2011.
- [7] C. Liu and T. J. Ye, “Run-off-road crashes: an on-scene perspective,” 2011.
- [8] C. Roque, F. Moura, and J. Lourenço Cardoso, “Detecting unforgiving roadside contributors through the severity analysis of ran-off-road crashes,” *Accident Analysis & Prevention*, vol. 80, pp. 262–273, 2015.

- [9] C. Liu and R. Subramanian, "Factors related to fatal single-vehicle run-off-road crashes," 2009.
- [10] M. H. A. Hussein, T. Sayed, K. Ismail, and A. Van Espen, "Calibrating road design guides using risk-based reliability analysis," *Journal of Transportation Engineering*, vol. 140, no. 9, 2013.
- [11] C. V. Zegeer and F. M. Council, "Safety relationships associated with cross-sectional roadway elements," *Transportation Research Record*, vol. 1512, 1995.
- [12] J. E. Hummer, W. Rasdorf, D. J. Findley, C. V. Zegeer, and C. A. Sundstrom, "Curve collisions: road and collision characteristics and countermeasures," *Journal of Transportation Safety & Security*, vol. 2, no. 3, pp. 203–220, 2010.
- [13] C. D. Fitzpatrick, C. P. Harrington, M. A. Knodler, and M. R. E. Romoser, "The influence of clear zone size and roadside vegetation on driver behavior," *Journal of Safety Research*, vol. 49, no. 6, pp. 91–97, 2014.
- [14] C. Roque and J. L. Cardoso, "SAFESIDE: a computer-aided procedure for integrating benefits and costs in roadside safety intervention decision making," *Safety Science*, vol. 74, pp. 195–205, 2015.
- [15] M. Jalayer and H. Zhou, "Evaluating the safety risk of roadside features for rural two-lane roads using reliability analysis," *Accident Analysis & Prevention*, vol. 93, pp. 101–112, 2016.
- [16] G. Cheng, R. Cheng, S. Zhang, and X. Sun, "Risk evaluation method for highway roadside accidents," *Advances in Mechanical Engineering*, vol. 11, no. 1, Article ID 754323857, 2019.
- [17] H. Li, F. Pang, H. Chen, and Y. Du, "Vibration analysis of functionally graded porous cylindrical shell with arbitrary boundary restraints by using a semi analytical method," *Composites Part B: Engineering*, vol. 164, pp. 249–264, 2019.
- [18] K. L. Stephan and S. V. Newstead, "Characteristics of the road and surrounding environment in metropolitan shopping strips: association with the frequency and severity of single-vehicle crashes," *Traffic Injury Prevention*, vol. 15, pp. S74–S80, 2014.
- [19] R. B. Noland, "Traffic fatalities and injuries: the effect of changes in infrastructure and other trends," *Accident Analysis & Prevention*, vol. 35, no. 4, pp. 599–611, 2003.
- [20] C. V. Zegeer and J. A. Deacon, "Effect of lane width, shoulder width, and shoulder type on highway safety," *State of the Art Report*, vol. 6, pp. 1–21, 1987.
- [21] C. Zegeer, R. Stewart, D. Reinfurt et al., "Cost effective geometric improvements for safety upgrading of horizontal curves," Final Report, vol. 6, Federal Highway Administration, McLean, VA, USA, 1990. <https://trid.trb.org/view/386551>.
- [22] S. Miaou, *Development of Relationship between Truck Accidents and Geometric Design: Phase I*, US Department of Transportation, Federal Highway Administration, McLean, Virginia, USA, 1993.
- [23] S. C. Joshua and N. J. Garber, "Estimating truck accident rate and involvements using linear and Poisson regression models," *Transportation Planning and Technology*, vol. 15, no. 1, pp. 41–58, 1990.
- [24] P. P. Jovanis and H. Chang, "Modeling the relationship of accidents to miles traveled," *Transportation Research Record*, vol. 1068, pp. 42–51, 1986.
- [25] S.-P. Miaou, "Estimating vehicle roadside encroachment frequencies by using accident prediction models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1599, no. 1, pp. 64–71, 1997.
- [26] R. Rusli, M. M. Haque, M. King, and W. S. Voon, "Single-vehicle crashes along rural mountainous highways in Malaysia: an application of random parameters negative binomial model," *Accident Analysis & Prevention*, vol. 102, pp. 153–164, 2017.
- [27] J. Milton and F. Mannering, "The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies," *Transportation*, vol. 25, no. 4, pp. 395–413, 1998.
- [28] M. Hosseinpour, A. S. Yahaya, A. F. Sadullah, N. Ismail, and S. M. R. Ghadiri, "Evaluating the effects of road geometry, environment, and traffic volume on rollover crashes," *Transport*, vol. 31, no. 2, pp. 221–232, 2016.
- [29] X. Jiang, X. Yan, B. Huang, and S. H. Richards, "Influence of curbs on traffic crash frequency on high-speed roadways," *Traffic Injury Prevention*, vol. 12, no. 4, pp. 412–421, 2011.
- [30] F. T. Kibar, F. Celik, and B. P. Aytac, "Statistical analysis of truck accidents for divided multilane interurban roads in Turkey," *KSCE Journal of Civil Engineering*, vol. 22, no. 5, pp. 1927–1936, 2018.
- [31] V. Shankar, J. Milton, and F. Mannering, "Modeling accident frequencies as zero-altered probability processes: an empirical inquiry," *Accident Analysis & Prevention*, vol. 29, no. 6, pp. 829–837, 1997.
- [32] H. Li, F. Pang, and H. Chen, "A semi-analytical approach to analyze vibration characteristics of uniform and stepped annular-spherical shells with general boundary conditions," *European Journal of Mechanics-A/Solids*, vol. 74, pp. 48–65, 2019.
- [33] J. Lee and F. Mannering, "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis," *Accident Analysis & Prevention*, vol. 34, no. 2, pp. 149–161, 2002.
- [34] Y. Peng, R. Li, G. B. Li, X. M. Yang, and D. Zhou, "Method for investigation of child occupant impact dynamics based on real-world accident," *International Journal of Automotive Technology*, vol. 16, no. 5, pp. 791–797, 2015.
- [35] J. Mandelík and M. Bundzel, "Application of neural network in order to recognise individuality of course of vehicle and pedestrian body contacts during accidents," *International Journal of Crashworthiness*, vol. 24, no. 2, pp. 221–234, 2019.
- [36] M. Gopal, K. Baron, and M. Shah, *Simulation and Testing of a Suite of Field Relevant Rollovers*, SAE Technical Paper, Pennsylvania, MI, USA, 2004.
- [37] H. Steffan and A. Moser, *How to Use PC-CRASH to Simulate Rollover Crashes*, SAE Technical Paper, Pennsylvania, MI, USA, 2004.
- [38] N. A. Rose and G. Beauchamp, *Analysis of a Dolly Rollover with PC-Crash*, SAE Technical Paper, Pennsylvania, MI, USA, 2009.
- [39] C. L. Naing, J. Hill, R. Thomson et al., "Single-vehicle collisions in Europe: analysis using real-world and crash-test data," *International Journal of Crashworthiness*, vol. 13, no. 2, pp. 219–229, 2008.
- [40] R. Ootani and C. Pal, "Effective numerical simulation tool for real-world rollover accidents by combining PC-crash and FEA," SAE Technical Paper, Pennsylvania, MI, USA, 2007.
- [41] Ministry of transport of the People's Republic of China (MOTO), *Design Specification for Highway Alignment*, Ministry of transport of the People's Republic of China (MOTO), Beijing, China, 2017.
- [42] C. Strobl, J. Malley, and G. Tutz, "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random

- forests," *Psychological Methods*, vol. 14, no. 4, pp. 323–348, 2009.
- [43] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2-3, pp. 131–163, 1997.
- [44] G. Ritschard, *CHAID and Earlier Supervised Tree Methods*, pp. 70–96, Routledge, Abingdon, UK, 2011.
- [45] G. V. Kass, "An exploratory technique for investigating large quantities of categorical data," *Applied Statistics*, vol. 29, no. 2, pp. 119–127, 1980.
- [46] A. F. Hayes, *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, Guilford Publications, New York, NY, USA, 2017.
- [47] S. A. Gargoum and K. El-Basyouny, "Exploring the association between speed and safety: a path analysis approach," *Accident Analysis & Prevention*, vol. 93, pp. 32–40, 2016.
- [48] C. Yu, *SPSS Statistical Analysis*, Publishing House of Electronics Industry, Beijing, China, 2007.
- [49] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Elsevier, Amsterdam, Netherlands, 2014.
- [50] A. Hadayeghi, A. Shalaby, and B. Persaud, "Development of planning-level transportation safety models using full Bayesian semiparametric additive techniques," *Journal of Transportation Safety & Security*, vol. 2, no. 1, pp. 45–68, 2010.
- [51] A. Darwiche, *Modeling and Reasoning with Bayesian Networks*, Cambridge University Press, Cambridge, UK, 2009.
- [52] L. Uusitalo, "Advantages and challenges of Bayesian networks in environmental modelling," *Ecological Modelling*, vol. 203, no. 3-4, pp. 312–318, 2007.
- [53] J. Zhao and W. Deng, "The use of Bayesian network in analysis of urban intersection crashes in China," *Transport*, vol. 30, no. 4, pp. 411–420, 2015.
- [54] M. Deublein, M. Schubert, and B. T. Adey, "Prediction of road accidents: comparison of two Bayesian methods," *Structure and Infrastructure Engineering*, vol. 10, no. 11, pp. 1394–1416, 2014.
- [55] Q. Zeng, H. Wen, H. Huang, X. Pei, and S. C. Wong, "A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity," *Accident Analysis & Prevention*, vol. 99, pp. 184–191, 2017.
- [56] S. H. Lee, Y. D. Lee, and M. Do, "Analysis on safety impact of red light cameras using the Empirical Bayesian approach," *Transportation Letters*, vol. 8, no. 5, pp. 241–249, 2016.
- [57] X. Ma, Y. Xing, and J. Lu, "Causation analysis of hazardous material road transportation accidents by bayesian network using genie," *Journal of Advanced Transportation*, vol. 2018, 2018.
- [58] C. Chen, X. Liu, H. Chen, M. Li, and L. Zhao, "A rear-end collision risk evaluation and control scheme using a Bayesian network model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 264–284, 2018.
- [59] X. Zhu, Y. Yuan, X. Hu, Y.-C. Chiu, and Y.-L. Ma, "A Bayesian Network model for contextual versus non-contextual driving behavior assessment," *Transportation Research Part C: Emerging Technologies*, vol. 81, pp. 172–187, 2017.
- [60] F. Zong, H. Xu, and H. Zhang, "Prediction for traffic accident severity: comparing the Bayesian network and regression models," *Mathematical Problems in Engineering*, vol. 2013, pp. 1–9, 2013.
- [61] G. Prati, L. Pietrantoni, and F. Fraboni, "Using data mining techniques to predict the severity of bicycle crashes," *Accident Analysis & Prevention*, vol. 101, pp. 44–54, 2017.
- [62] S. H. Chen and C. A. Pollino, "Good practice in Bayesian network modelling," *Environmental Modelling & Software*, vol. 37, pp. 134–145, 2012.
- [63] J. Xu, X. Luo, and Y. Shao, "Vehicle trajectory at curved sections of two-lane mountain roads: a field study under natural driving conditions," *European Transport Research Review*, vol. 10, no. 1, p. 12, 2018.
- [64] H. Farah, A. van Beinum, and W. Daamen, "Empirical speed behavior on horizontal ramp curves in interchanges in The Netherlands," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2618, no. 1, pp. 38–47, 2017.
- [65] S. Mavromatis, B. Psarianos, P. Tsekos, G. Kleioutis, and E. Katsanos, "Investigation of vehicle motion on sharp horizontal curves combined with steep longitudinal grades," *Transportation Letters*, vol. 8, no. 4, pp. 220–228, 2016.
- [66] S. Alkheder, "Learning from the past: traffic safety in the eyes of affected local community in Abu Dhabi City, United Arab Emirates," *Transportation Letters*, vol. 9, no. 1, pp. 20–38, 2017.
- [67] 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, Article ID 4017032, 2017.
- [68] B. L. Boada, M. J. L. Boada, M. Ramirez, and V. Diaz, "Study of van roadworthiness considering their maintenance and periodic inspection. The Spanish case," *Transportation Letters*, vol. 6, no. 4, pp. 173–184, 2014.
- [69] American Association of State Highway and Transportation Official (AASHTO), *Roadside Design Guide*, American Association of State Highway and Transportation Official (AASHTO), Washington, DC, USA, 4th edition, 2011.
- [70] J. M. Holdridge, V. N. Shankar, and G. F. Ulfarsson, "The crash severity impacts of fixed roadside objects," *Journal of Safety Research*, vol. 36, no. 2, pp. 139–147, 2005.