WARSAW UNIVERSITY OF TECHNOLOGY Index 351733
FACULTY OF CIVIL ENGINEERING
COMMITTEE FOR CIVIL AND WATER ENGINEERING
POLISH ACADEMY OF SCIENCES ISSN 1230-2945

DOI: 10.24425/ace.2021.138520
ARCHIVES OF CIVIL ENGINEERING
Vol. LXVII ISSUE 4 2021

pp. 591 –604

Research paper

Analyzing rear-end crash severity for a mountainous expressway in China via a classification and regression tree with random forest approach

Yonggang Wang[1], Xianyu Luo[2]

Abstract: To understand the contributory factors to rear-end accident severity on mountainous expressways, a total of 1039 rear-end accidents, occurring on G5 Jingkun Expressway from Hechizhai to Qipanguan in Shaanxi, China over the period of 2012 to 2017, were collected, and a non-parametric Classification and Regression Tree (CART) model was used to explore the relationship between severity outcomes and driver factors, vehicle characteristics, roadway geometry and environmental conditions. Then the random forest model was introduced to examine the accuracy of variable selection and rank their importance. The results show that driver's risky driving behaviours, vehicle type, radius of curve, angle of deflection, type of vertical curve, time, season, and weather are significantly associated with rear-end accident severity. Speeding and driving while drunk and fatigued are more prone to result in severe consequences for such accidents and driving while fatigued is found to have the highest fatality probability, especially during the night period (18:00–24:00). The involvement of heavy trucks increases the injury probability significantly, but decreases the fatality probability. In addition, adverse weather and sharp curve with radius less than 1000 m are the most risk combination of factors. These findings can help agencies more effectively establish stricter regulations, adopt technical measures and strengthen safety education to ensure driver's driving safety on mountainous expressways for today and tomorrow.

Keywords: rear-end accidents, mountainous expressway, accident severity, CART model, random forest

[1]Prof., PhD., Eng., Chang'an University, College of Transportation Engineering, Middle Section of South 2 Ring Rd., Xi'an 710064, Shaanxi, China, e-mail: wangyg@chd.edu.cn, ORCID: 0000-0002-9365-1851

[2]MSc., Eng., Chang'an University, College of Transportation Engineering, Middle Section of South 2 Ring Rd., Xi'an 710064, Shaanxi, China, e-mail: luoxianyu2021@sina.com, ORCID: 0000-0003-2626-4007

1. Introduction

During the past two decades, the number of registered motor vehicles has increased dramatically in China – from about 9.6 million in 2003 to more than 310 million in 2017, i.e., an almost thirty-two-fold increase (China Statistical Yearbook, 2018). Likewise, the amount of traffic accidents has also significantly grown, especially on mountainous expressways. In China, a number of alarming statistics show that mountainous expressways are subjected to have a high frequency of severe accidents, which, in turn, result in more injuries and fatalities [1–4], and this is, in part, due to the adverse traffic environment (i.e., small curve, steep slope, existence of bridges and tunnels, and changeable climatic condition), compared with those in plain areas. These facts clearly illustrate the urgent needs of understanding how the accidents occur on mountainous expressways to provide proper countermeasures for traffic safety improvement.

There is no doubt that a great number of contributory factors, ranging from human factors and vehicle characteristics to roadway geometry and environmental conditions, correlated with the severity of traffic crashes on mountainous expressways. Considerable research efforts have been devoted to investigate driver's demographic characteristics (i.e., age, gender, educational background, etc.) and risky driving behaviours that may contribute to the accident occurrence on mountainous expressways [1–4]. Expressway geometric design elements (i.e., length of road segments, number of vertical curves in a road section, horizontal curve and distance to the nearest access point) are also identified to have significant influence on accident severity on mountainous expressways [5].

According to the National Highway Traffic Safety Administration (NHTSA), rear-end collisions are the most frequent type of road traffic accidents in the United States, accounting for 29% of all car accidents (source: https://www.kffjlaw.com/library/a-few-facts-about-rear-end-collisions.cfm). The undesirable social impact induced by rear-end accidents on mountainous expressway has become a big problem that arouses much social concerns. A large number of previous studies reported that driving in mountainous terrains is found to be associated with more fatal and severe injury probabilities in rear end crashes [6, 7]. Specially, number of vehicles involved, large truck involvement, poor lighting conditions, windy driving conditions could significantly increase the injury severities of drivers involved in rear-end crashes [8].

Few previous researches, however, have specially investigated the cause mechanism of rear-end accidents on mountainous expressways. Classification and regression trees (CART) is a non-parametric model with no presumed relationships between the dependent and explanatory variables, which can determine the variable correlation and automatically eliminate the useless variables while dealing with large-scale dataset. Recently, CART model has been widely used to explore the determinants of accident occurrence and injury prevention [9, 10]. Therefore, CART model is proposed to handle the interactions between the rear-end accident severity and potential contributing factors, and the random forest approach is subsequently employed to verify the variable selection and determine the variables' ranking via importance.

2. Data

Since this study focused on the rear-end accidents on mountainous expressways, a total of 1039 policy reported rear-end accidents between 2012 and 2017, accounting for 56.7% of the total accidents, were originally collected from a typical four-lane segment of G5 Jingkun

Expressway from Hechizhai to Qipanguan (K1102 + 608 – K1463 + 451) in Shaanxi, China, as shown in Figure 1. The data include: accident information (i.e., location, number of casualties, property damage, etc.); driver factors (i.e., gender, age, driving behaviours, etc.); vehicle characteristics (i.e., vehicle type, status of vehicle, etc.); environmental conditions (i.e., month, time, weather, etc.); roadway geometry (i.e., curve radius, deflection angle, vertical curve type, etc.). Here the driving behaviour variables came from traffic police records and roadway geometric variables were extracted from the original design documents and updated through Google Earth.

Fig. 1. A typical mountainous expressway segment in Shaanxi, China

The injury severity of rear-end accidents is considered as the dependent variable and divided into three levels: property damage only (PDO), injury and fatality [2,3], which account for 48.8%, 32.7% and 18.5%, respectively, of the total sample collected for this study. Fourteen explanatory variables ranging from driver factors, vehicle characteristics and roadway geometry to environmental conditions were listed in Table 1.

Table 1. Sample description

Variable	Code	Frequency	%
Dependent variable			
Injury Severity	PDO	502	48.8
	Injury	337	32.7
	Fatality	200	18.5
Independent variables			
Driver factors			
Gender	Male	943	90.8
	Female	96	9.2

continued . . .

Table 1 [cont.]

Variable	Code	Frequency	%
Age	Young (Age ≤ 30)	280	26.9
	Adult (30 < Age ≤ 50)	713	68.6
	Old (Age > 50)	46	4.4
Risky riving behaviours	Fatigued driving (fatigue)	83	8.0
	Drunk driving (drinking)	48	4.6
	Driving in the wrong lane (lane)	334	32.1
	Speeding	118	11.4
	Overtaking on the right (overtaking)	68	6.5
	Risk following distance (distance)	159	15.3
	Turning round and crossing the central reservation (turn round)	38	3.7
	Other risky driving behaviours (Others)	191	18.4
Vehicle characteristics			
Type of vehicle	Car	563	54.2
	Truck	416	40.0
	Coach	17	1.6
	Other types	43	4.1
Status of vehicle	Break failure	45	4.3
Roadway geometry			
Radius of curve (Rad) /m	Rad ≤ 1000	406	39.1
	1000 < Rad ≤ 2000	283	27.2
	Rad > 2000	350	33.7
Angles of deflection (Ang) / °	Ang = 0	423	40.7
	0 < Ang ≤ 30	290	27.9
	Ang > 30	326	31.4
Longitudinal gradient (LGR) /%	|LGR| ≤ 1	299	28.8
	1 < |LGR| ≤ 2	296	28.5
	2 < |LGR| ≤ 3 |LGR| > 3	288 156	27.7 15.0
Types of vertical curve (TOV)	Concave	333	32.1
	Convex	233	22.4
	Line	473	45.5

continued . . .

Table 1 [cont.]

Variable	Code	Frequency	%
Superelevation (Sup) /%	Sup = 0	296	28.5
	0 < Sup ≤ 2	467	44.9
	Sup > 2	276	26.6
Environmental conditions			
Season	Spring (January to March)	277	26.7
	Summer (April to June)	176	16.9
	Autumn (July to September)	303	29.2
	Winter (October to December)	283	27.2
Day of week	Working days	708	68.1
	Weekends	331	31.9
Time of crash	Daytime (6:00–18:00)	513	49.4
	Nighttime (18:00–24:00)	284	27.3
	Early morning (24:00–6:00)	242	23.3
Weather	Fine (sunny, cloudy)	839	80.8
	Adverse (rainy, snowy, foggy)	200	19.2

3. Methodology

3.1. CART model

CART is a nonparametric model with no pre-defined relationships between the independent variable and dependent variable. Since the severity outcome of rear-end crashes is categorical (PDO, injury and fatality), a classification tree is developed, which consists of three steps [9].

The first step is tree growing and the principle is to recursively partition the target variable to minimize "impurity" in the terminal nodes. The most common measure of node impurity is the Gini criterion which is used to quantify the homogeneity based on computing the proportion of data that belong to a specific class as:

$$Gini(t) = \sum_{i \neq j} p(j|t)p(i|t) \tag{3.1}$$

where i and j are categories of the target field, which satisfy:

$$\begin{aligned} p(j|t) &= \frac{p(j,t)}{p(t)} \\ p(j,t) &= \frac{\pi(j)N_j(t)}{N_j} \\ p(t) &= \sum_j p(j,t) \end{aligned} \tag{3.2}$$

where $\pi(j)$ is the prior probability value for category j, $N_j(t)$ is the number of records in category j of node t, and N_j is the number of records of category j in the root node. Note that when the Gini index is used to find the improvement for a split during tree growth, only those records in node t and root node (top node of the tree) with valid values for the split-predictor are used to compute $N_j(t)$ and N_j, respectively [10].

Then the data can be split according to the Gini index of each independent variable. The root node is divided into two child nodes based on an independent variable that maximizes the purity. Then the child nodes can be considered as new parent nodes on each branch of the tree. If the classification condition is still satisfied, this process would repeat for each child node until all data obtain the optimal possible purity.

The algorithm subtracts some subtrees from the bottom of the fully growing decision tree. The pruning process starts with the maximal tree and selectively prunes upward to produce a sequence of sub-trees of the maximal tree, and eventually collapses to the tree of the root node. The outcome of this pruning process is tested by cross-validation method on an independent dataset and the best tree can then be selected.

3.2. Random forests

In order to study the degree of each variable affecting the injury severity, previous researches ranked the variables according to the relative importance of variables (VIM). However, one tree structure (or the final tree structure) usually could not unveil the variables' importance ranking since it could be completely masked by another correlated input [11]. Random forests belongs to the bagging algorithm in integrated learning, which uses the random resampling technique (bootstrap) and node random splitting technique to construct multiple classification regression trees [12]. The method is one of the most promising developments in extracting the variables' importance ranking and has been widely used in recent years. During the tree growing procedure, about one-third of the training data were left out from the training trees and became the OOB (out-of-bag) data. The OOB data are utilized to achieve unbiased estimate of variable importance as trees are added to the forest. The Mean Decrease Gini (MDG) criterion of each variable can rank the importance, and the relative more important variable has higher MGD value. Such an approach can be implemented using the "Random forest" package in the R program.

4. Results

4.1. Prediction accuracy of models

The CART model was established using SPSS 20.0 and the corresponding results are shown in Tab. 2. 829 samples (80%) were randomly selected from the overall data as the training data and the remaining 210 samples (20%) were used as the test data. The overall prediction accuracy is approximately 88.5% for the learning data, while that for the testing data is about 87.0%. The rediction results are shown in Table 2.

Table 2. Prediction results of CART model

Category	Learning data (N = 829)				Testing data (N = 210)			
	PDO	Injury	Fatality	Precision (%)	PDO	Injury	Fatality	Precision (%)
PDO	374	18	3	94.7	100	2	1	97.1
Injury	16	233	27	84.4	5	42	10	73.7
Fatality	5	26	127	80.4	1	6	26	78.8
Overall (%)	47.6	33.4	18.9	88.5	54.9	25.	19.2	87.0

4.2. Variables importance

Since it is essential to determine whether the number of trees is large enough to obtain stable results, a sensitivity analysis is conducted to measure the performance of the random forests with the gradual increasing number of trees. Fig. 2 displays the relationship between the OOB error rate of accident severity and the number of decision trees (from ntree = 1 to 1000), in which the dotted line represents the weighted error of all trees. When the value of 'ntree' is larger than 700, clearly, the OOB error rates relatively stabilize indicating that mtry = $\sqrt{14}$ and ntree = 700 could be accepted.

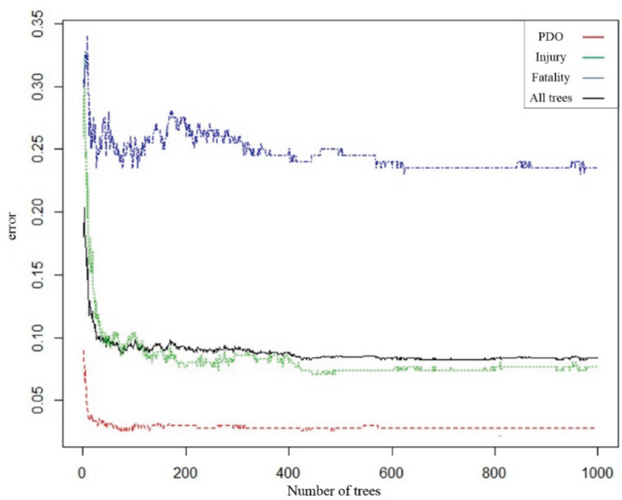
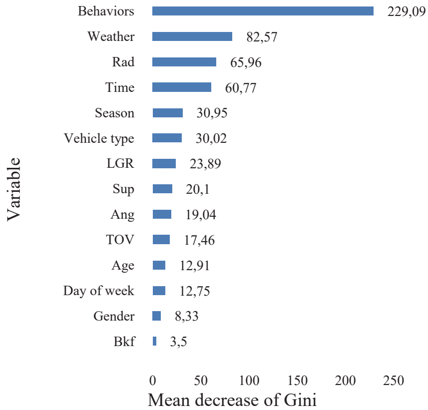
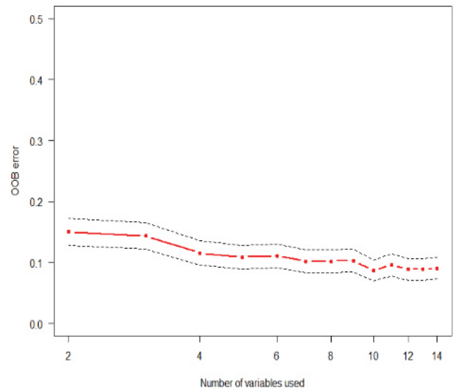


Fig. 2. Relationship between number of trees and prediction error

Figs. 3a and 3b show the importance ranking of variables and variable selection by the iterative calculation, which indicate that using 10 independent variables can minimize the out-of-bag error (9.34%) of random forest. Obviously, the importance ranking of each variable judged by the MDG is as followedŁşdriving behaviours, weather, curve radius, time period, season, vehicle type, longitudinal gradient, superelevation, deflection angle and types of vertical curve.



(a) Variable importance ranking



(b) Optimal number of variables

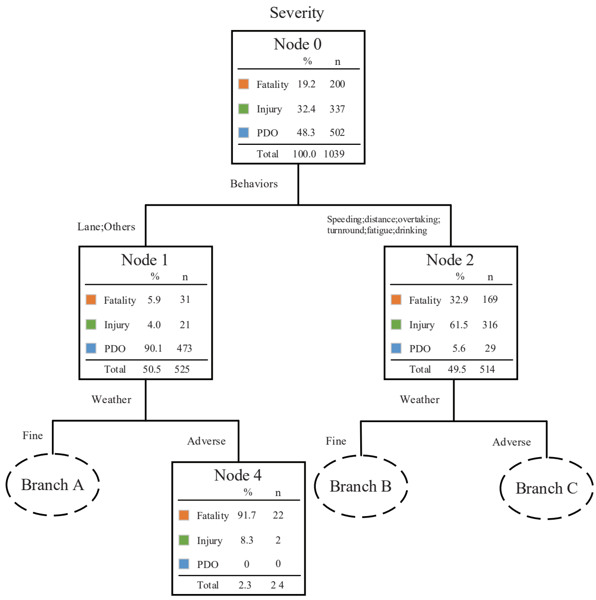
Fig. 3. Results of variable selection

Clearly, both roadway's longitudinal gradient and superelevation have significant influence on the injury severity in the random forest approach, but neither of them appear in the CART model. The reason is that superelevation and radius of curve are correlated with each other in roadway geometry design, and a similar relationship also exists between the longitudinal gradient and vertical curve. Additionally, the variables of gender and age of drivers and day of the week are not identified to have significant influence on the rear-end accident serverity, which are consistent with the previous results [13] and inconsistent with others [14].

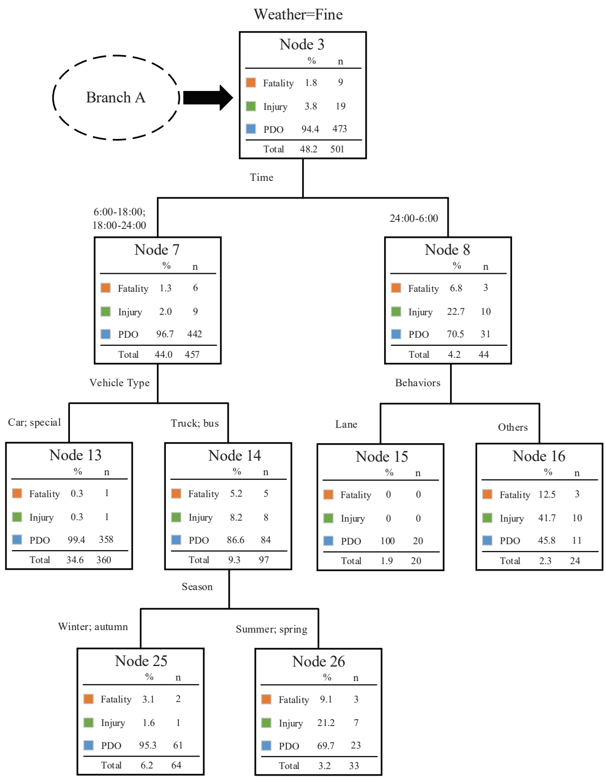
4.3. Classification tree

As shown in Fig. 4, the maximum depth of the tree is 5, and eight factors, including behaviour, weather, radius of curve, type of vehicle involved, time of day, season, angles of deflection and type of vertical curve, are the main splitter in the CART model. Driving behaviour is used to create the first split and generate two internal child nodes. Node 1 is composed of two variables (e.g., lane & others), and the accident consequences are mainly property loss. Node 2 contains six variables including speeding, not maintaining a safe following distance, drunk driving, driving while fatigued, making an illegal u-turn and dangerous overtaking, indicating that if the accident is related to these six risky driving behaviours, then there will be a higher probability of severe injury or fatality outcomes (94.4% vs. 9.9%).

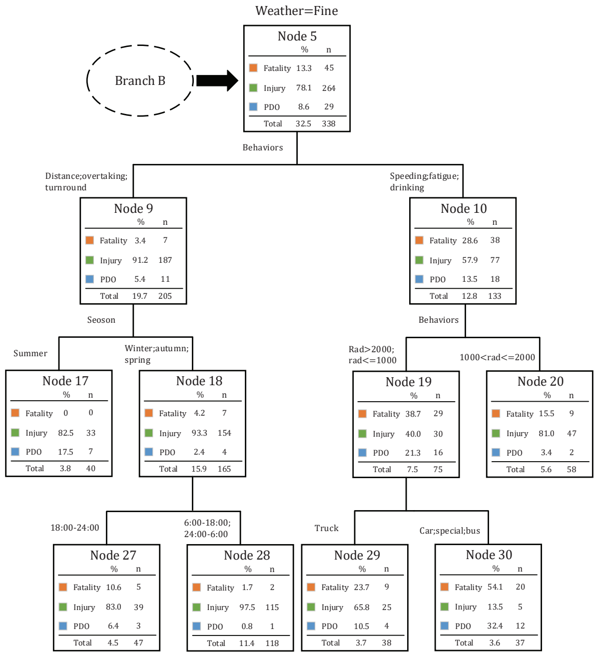
On the left side of the tree, node 1 is divided into a parent node 3 (see branch A of CART model) and a terminal node 4 representing the fine and adverse weather, respectively. Obviously, the adverse weather condition is associated with the lower injury and fatality probabilities in rear-end accidents. Specially, driver's risky driving behaviours under adverse weather conditions are more likely to cause the fatalities in such accidents (91.7 vs. 1.8%). Next, node 3 is split into nodes 7 and 8 by the time of crash (see Fig. 4b, implying that the probability of fatal and injury crashes increases significantly at midnight under the fine weather conditions, but it is more likely to cause the injury crashes than the fatal ones (20.7 vs. 6.8%). Once again, driving



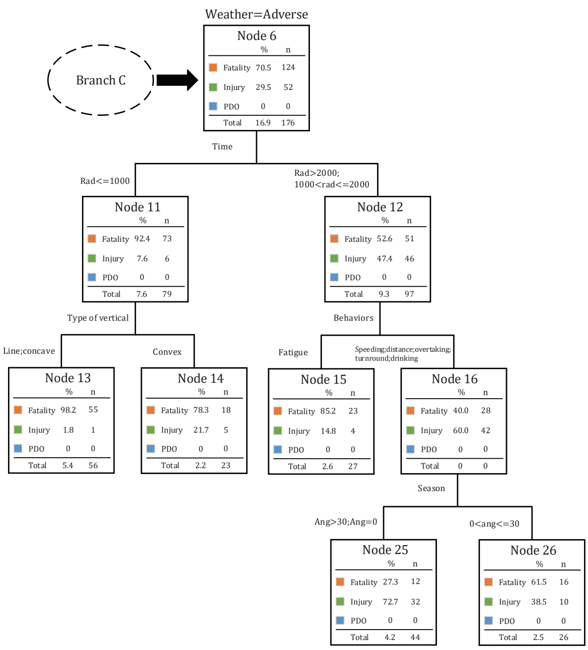
(a) Model framework



(b) Branch A



(c) Brach B



(d) Branch C

Fig. 4. CART model

behaviours are the reason of being divided for node 8, which means that the crashes due to driving in the wrong way may result in the slightest consequences. Finally, node 7 is split into nodes 13 and 14 by vehicle type, considering the factor of season, the two terminal nodes 25 and 26 indicate that trucks and coaches are more susceptible to cause injury and fatality in rear-end accidents in the daytime under fine weather conditions, especially during the spring and summer periods.

Turning to the right side of tree, weather is the second most important factor contributing to severe rear-end accidents, forming nodes 5 (see branch B of CART model) and 6 (see branch C of CART model) to next layer of the tree. Then, CART directs three variables including distance, overtaking and turning round to the left and forming node 9 (see Fig. 4c). The consequences of three types of driving behaviours are injury, while the probability of fatality accidents is much lower (3.4 vs. 28.6%). Node 9 continues to split, then nodes 17 and 18 is created by the variable of season, indicating that the injury and fatality probabilities of rear-end accidents are lower in summer season. Further down to the tree, nodes 27 and 28 indicate that it is more likely to have the fatality outcomes in such accidents during nighttime (10.6 vs. 1.7%).

To the right, node 10 (see Fig. 4c) consists of the rest improper driver behaviours. Clearly, such risky driving behaviors as speeding, fatigue driving and drunk driving are more likely to induce severe accidents (28.6 vs. 3.4%). According to the different radius of curves, node 10 splits into nodes 19 and 20. Node 20 represents the radius of curve between 1000 and 2000m. At this invetigated segment of expressway, the probability of injury accidents is 81%, and the probability of fatal accidents is lower than other types of radius (15.5 vs. 38.7%). The results from the designated radius of nodes 19, 29 and 30 indicate that the overall proportion of trucks in injury and fatal accidents is higher than other vehicles (89.5 vs. 67.6%).

Node 6 (see Fig. 4d), on the right branch of the CART, is divided into nodes 11 and 12 by the radius of curve. For the sharp curves with radius less than 1000m, the fatality probability in rear-end accidents can reach as high as 92.4%, especially under the influence of adverse weather conditions. This consquence can be worse at the straight-line or concave curve segments, as can be seen from node 13 with the fatality probability by 98.2%. For the curves with radius more than 1000 m, the node 12 is split into nodes 15 and 16 by the risky driving behaviours. Under this condition, driving while fatigued is correlated with a higher probability of fatal accidents (85.2%). Nodes 25 and 26 indicating fatal accidents are more likely to occur at the sgements with deflection angle between 0 and 30° (61.5 vs. 27.3%), especially while drivers are performing the risky driving behaviours under the adverse weather conditions.

5. Discussion

The importance of variables determined by random forests can facilitate the interpretation of the contributory factors in the CART model. Drivers can keep calm while performing the unsafe or dangerous driving behaviours like driving in the wrong lane, etc., thus they can take some avoidance manners in time to avoid the accidents or reduce the consequence severity. However, the high injury and fatality probabilities indicate that driver's worse reaction and perception ability impaired by alcohol consumption or fatigue, insufficient response time and visual field deterioration due to speeding or lacking safe distance have gradually impaired drivers' driving performance, resulting in frequent driving errors and, in extreme cases, traffic accidents [1–3, 3, 4, 6, 8]. Besides, the results of CART model show that the effects of driver's risky driving behaviours on injury severity are interactive with other potential factors including weather, time and road geometry in accordance with some previous findings [2–5, 8, 10] and in discordance with others [13].

For the negative influence of adverse weather conditions, there is no doubt that it is more likely to cause severe rear-end accidents on slippery segments under low visibility conditions [2, 4, 10]. Summer is identified to be significantly correlated with the higher probability of rear-end accident severity, and the possible reason lies in the more rainfall in this season in line with previous findings [2–4]. Furthermore, there are no enough lighting equipments on mountain expressways, thus driving in the night without lighting conditions always results in poor vision for drivers, especially for these older drivers. In this study, the effect of time on rear-end accident severity is interactive with other potential risk factors such as risky driving behaviours, weather and season, and such accidents occuring at night (24:00–6:00) are significantly associated with a higher probability of severe consquence in line with previous results [2], but not in line with others [3, 13]. By contrast, driving in the evening (18:00–24:00) also impacts on rear-end accident severity in accordance with some previous findings [3] and in discordance with others [2, 4, 13].

Crashes involving trucks and coaches are associated with higher likelihood of fatality and injury, in line with many previous studies [1, 6, 8, 11]. An interesting finding in this study is that when dealing with different types of curve, the fatality probability caused by other type vehicles is much higher than that of trucks (54.1 vs. 23.7%). One possible reason is due to the existence of many scenic tourist attractions, attracting foreign drivers who are not familiar with the local environment to be involved in serious accidents.

In view of the roadway geometrics, the radius of a horizontal curve between 1000 and 2000 m is expected to be much safer than others. While driving on sharp curves with radius less than 1000 m or large radius curves of more than 2000 m, it is more likely to suffer the severe rear-end accidents in accordance with previous reports [2] and in discordance with others [5, 13]. These important results suggest that the radius of the curve should not be too small when designing a mountain highway, but also not too large to avoid the distracted driving. while driving on the curves with deflection angle between 0 to 30°, however, drivers tend to be more relaxed with their increased vigilance and speeding choice, and thus are easier to cause severe accidents. The modeling results also show that the convex curve can reduce the fatality probability of such accidents, which can be taken in consideration in designing new mountainous expressways.

The importance ranking results shows that driver behavior is the most important variable affecting the severity of rear-end accidents in line with many previous findings [2–4, 10, 13]. Compared with other driving behaviours, speeding and driving while drunk and fatigued driving are more likely to cause injury and fatal accidents [2, 8, 13], which should be seriously banned by stricter traffic regulations through the increased speeding fines and penalty point premium, etc. In addition, drivers should be educated to comply with the traffic rules and those deliberately violating traffic rules and regulations should be heavily punished. Currently, the monitoring of fatigue driving is really a challenge. Because many vehicles are not equipped with positioning systems, it is impossible to track their specific driving time. In particular, the current definition of driving fatigue also lacks a scientific basis, and so is worthwhile to study this issue in depth.

Adverse weather has pernicious effect on the severity of rear-end accidents, especially while driving on the sharp curves (radius < 1000 m). Therefore, the appropriate enforceable speed limits on dynamic message signs should be posted in advance before the segments with poor geometrics and sight clearance on mountainous expressways. The consequences of rear-end accidents occurring during the nighttime period (18:00–24:00) are often more severe due to the inadequate lighting on mountainous expressways, which also illustrates the need to control the traveling speed of vehicles in low-light conditions [2, 4, 8, 10].

6. Conclusion

This study tries to identifiy the potential factors contributing to the severity of rear-end accidents on mountainous expressways via CART and random froest approach using the police-reported accidents from G5 Jingkun Expressway between Hechizhai and Qipanguan in Shaanxi, China during the period of 2012 to 2017. The modeling results exhibit that eight factors, including type of vehicle involved, risky driving behaviors, radius of curve, angles of deflection, type of vertical curve, time of day, season, and weather are significantly correlated with rear-end accident severity. Considering the interactive effects of risk factors, several countermeasures have been recommended to prevent and decrease the occurence of severe rear-end accidents, including strengthening the supervision of vehicle positioning devices, handing out stricter punishment to certain improper driving behaviours, enhancing the performance of traffic signs, and improving the frequency of driver safety education, etc., on mountainous expressways. Of course, this study has several methological limitations, such as small size of accident sample from one expressway segment and lack of traffic volume, which should be taken into cossideration in the near future research.

This study was not without important methodological limitations, however. First, the crash sample was only selected from one expressway segment in Shaanxi, China, which may be not representative of the overall traffic safety situation of mountainous expressways in the country as a whole. Second, the original data may contain some missing, incomplete, or possibly incorrect points due to unreported crashes or injuries and errors involved in manual data entry. Third, this study used the crash data from 2012 to 2017, and during such a large time period that many influencing factors, such as the roadway network, population distribution, speed limit standards, etc., have changed significantly, so the affecting parameters on injury severity outcomes in rear end crashes may have significantly temporal instability across different time periods. All these should be seriously considered in the future studies.

References

[1] L. Wang, et al., "Driver injury severity analysis of crashes in a western China's rural mountainous county: taking crash compatibility difference into consideration". Journal of Traffic and Transportation Engineering (English Edition), 2020, DOI: 10.1016/j.jtte.2020.12.002.

[2] Y. Wang and C.G. Prato, "Determinants of injury severity for truck crashes on mountain expressways in China: a case-study with a partial proportional odds model". Safety Science, vol. 119, pp. 100–107, 2019, DOI: 10.1016/j.ssci.2019.04.011.

[3] Y. Wang, Y. Luo, and F. Chen, "Interpreting risk factors for truck crash severity on mountainous freeways in Jiangxi and Shaanxi, China". European Transport Research Review, vol. 11, article 26, 2019, DOI: 10.1186/s12544-019-0366-4.

[4] Y. Wang, H. Zhang, and N. Shi, "Factors contributing to the severity of heavy truck crashes: a comparative study of Jiangxi and Shaanxi, China". Jordan Journal of Civil Engineering, vol. 15, no. 1, pp. 41–51, 2021.

[5] M. I. Sameen and B. Pradhan, "Assessment of the effects of expressway geometric design features on the frequency of accident crash rates using high-resolution laser scanning data and GIS". Geomatics, Natural Hazards and Risk, vol. 8, no. 2, pp. 733–747, 2017, DOI: 10.1080/19475705.2016.1265012.

[6] A. Ahmadi, et al., "Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods". Journal of Transportation Safety & Security, vol. 12, no. 4, pp. 522–546, 2020, DOI: 10.1080/19439962.2018.1505793.

[7] Z. Sun, et al., "Crash analysis of mountainous freeways with high bridge and tunnel ratios using road scenario-based discretization". PLoS ONE, vol. 15, no. 8, article e0237408, 2020, DOI: 10.1371/journal.pone.0237408.

[8] C. Chen, at al., "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, DOI: 10.1016/j.aap.2015.03.036.

[9] W.Y. Loh, "Classification and regression trees". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 1, no. 1, pp. 14–23, 2011, DOI: 10.1002/widm.8.

[10] A.T. Kashani, R. Rabieyan, and M.M. Besharati, "A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers". Journal of Safety Research, vol. 51, pp. 93–98, 2014, DOI: 10.1016/j.jsr.2014.09.004.

[11] R. Harb, et al., "Exploring precrash maneuvers using classification trees and random forests". Accident Analysis & Prevention, vol. 41, no. 1, pp. 98–107, 2009, DOI: 10.1016/j.aap.2008.09.009.

[12] L. Breiman, "Random forests". Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, DOI: 10.1023/A:1010933404324.

[13] Y. Peng, et al., "Investigation on the injuries of drivers and copilots in rear-end accidents between trucks based on real world accident data in China". Future Generation Computer Systems, vol. 86, pp. 1251–1258, 2018, DOI: 10.1016/j.future.2017.07.065.

[14] X. Li, et al., "A rear-end collision risk assessment model based on drivers' collision avoidance process under influences of cell phone use and gender – a driving simulator based study". Accident Analysis & Prevention, vol. 97, pp. 1–18, 2016, DOI: 10.1016/j.aap.2016.08.021.

Received: 2021-04-23, Revised: 2021-08-23