Journal of Highway and Transportation Research and Development
2024,18(1):1-10
ISSN 2095-6215
DOI: 10.26599/HTRD.2024.9480001

Study on Accident Characteristics and Scenarios of Long-distance Passenger Lines over 800 Kilometers

Chang Liu*[1], Hongwen Xia**[1,2], Xingkai Meng[1], Xueran Wang[1], Wenhui Luo[1], Chu'na Wu[1]

(1. Automotive Transportation Research Center, Research Institute of Highway, Ministry of Transport, Beijing 100088, China;
2. Beijing Jiaotong University, Beijing 100091, China)

Abstract: Long-distance passenger lines exceeding 800 kilometers are a crucial component of highway passenger transport. This paper analyzes the causes of traffic accidents on long-distance passenger routes exceeding 800 kilometers from 2014 to 2020. This paper conducts a causal analysis and researches accident-prone scenarios for long-distance passenger lines exceeding 800 km. It selects key characteristics to establish an index system based on human factors, road conditions, and the environment. This paper analyzes the characteristics of humans, roads, and the environment in the traffic system when accidents occur. The Decision Tree model is utilized to investigate the internal relationships among key factors in accident scenarios, such as operating time, human factors, weather, quarter, accident region, holidays, spring transportation, road alignment, and road condition. The results show that the run time, human factors, and weather conditions are important influencing factors for accidents on long-distance passenger lines over 800 km. They have a coupling relationship with factors such as quarter, accident region, road alignment, and road condition. Holidays, spring transportation, highway technical grade, and highway administrative grade are not the important factors affecting the occurrence of accidents on long-distance passenger lines over 800 km. This paper provides a theoretical basis for reducing the potential for accidents and promoting safety within enterprises operating long-distance passenger lines over 800 km. In the future, road transport management departments and road transport enterprises should strengthen safety management, standardize drivers' safe driving behavior, and promote defensive driving knowledge in accident-prone situations.

Key words: traffic safety; accident characteristics and scenarios; decision tree model; long-distance passenger line over 800 km; accident cause analysis

0 Introduction

With the rapid growth of highways, railways, air transports, and other modes of transportation in China, Highway Passenger Transportation has become an essential part of passenger transport due to its speed and convenience. According to the mileage of the operation route, it can be divided into categories based on distance: less than 200 km, from 200 to 400 km, from 400 to 800 km, and over 800 km. Due to geographical constraints, some areas in China only have long-distance passenger lines that are over 800 km in length. These are referred to as "800 km Passenger Lines." Passenger lines covering 800 km, however, have characteristics such as long distances, extended travel times, complex road conditions, and a higher likelihood of drivers violating rules, all of which pose significant security risks. In addition, by the end of 2020, there were 2 855 lines in China with an average daily departure frequency of 4 166 times on 800 km Passenger Lines. These figures account for only 1.9% and 0.5%[1] of all lines, respectively. However, there were 1 424 deaths and 2 534 injuries in 273 major accidents involving passenger cars from 2014 to 2020. Among them, there were 48 accidents on 800 km of passenger lines, resulting in 265 deaths and 730 injuries, accounting for 18.6% and 28.8%, respectively. It can be seen that the number of casualties in accidents on 800 km Passenger Lines is higher compared to other lines. In fact, the number of accidents on 800 km passenger lines is 6.4 times higher than on other

Received 30 June 2023; Revised 22 October 2023; Accepted 30 October 2023

* *E-mail address*: c.liu@rioh.cn

** Corresponding author, hw.xia@rioh.cn

lines. The main reasons include long-distance driving, complex environments, safety supervision loopholes, and the predominant use of large and medium-sized buses with high passenger capacity. Therefore, it is easy to cause major casualties when danger occurs.

The Ministry of Transport ' s 2020 regulations require road transport management department to conduct risk assessments for applications to operate 800 km passenger transportation lines. Besides, the research shows that driving risks have significant influence in accident casualty. HSING[2] proved fatigue, not wearing seat belts, distraction, drunk, midnight, dawn, ramp and others which in driving will seriously affect the casualties of passenger car collision. CHANG[3] proved that using restraint equipment can prevent or reduce the injury, which caused by rollover accident in downhill section. It can be seen that the key to the safe operation of 800 km Passenger Lines is to ensure the safety in line.

Present studies of passenger transport risk mainly focuses on risk indicators and assessment methods. MORADI[4] describes the characteristics of long-distance cars accidents from passenger gender, road conditions, weather and accident forms. According to the four subsystems of participants, vehicles, environment and management in the transportation system of transportation buses. ZHU[5] established a hazard identification table of major traffic accidents for transportation buses with 30 kinds of dangerous sources and 7 attributes. WANG[6] constructs risk assessment indicators based on human, vehicles, roads and enterprises. In the risk assessment, ZHAO[7] constructs the driving risk assessment model to realize the risk management in vehicle operation. HU[8] evaluated the vehicle operation risk under traffic congestion based on BP neural network. JIN[9] used analytic hierarchy process-extension comprehensive evaluation method to evaluate the risk of operating vehicles in mountainous highway tunnels. While research on risk assessment indicators and methods of passenger transport lines has been extensive, both domestically and internationally, there is a significant gap in 800 km Passenger Lines. Therefore, it is necessary to analyze the causes and key factors of 800 km Passenger Lines accidents, thus to put forward targeted suggestions.

This paper analyzes the key factors of the accident causes on accident data for 800 km Passenger Lines. And then constructs an index system based on 4 contributing factors factors of the traffic system, which is human, vehicles, roads and environment. This paper takes the occurrence form of accidents as the dependent variable, because the force majeure characteristics of less numbers on accidents on 800 km Passenger Lines, and the elasticity and characteristics of the model are integrated. This paper selects decision tree model to analyze the key causes, coupling effects and importance on driving safety of 800 km Passenger Lines accidents. The research will be reducing the potential accident occurrence of 800 km Passenger Lines, which providing basis for enterprise safety prevention and control, and ensuring operation safety at the same time.

1 Data sources

The accident data of 800 km Passenger Lines were obtained for follow-up research. 48 road traffic accidents involving 800 km Passenger Lines over the years were obtained from the road traffic accident statistics data and relevant accident reports provided by the Ministry of transport, resulting in 265 deaths and 730 injuries. The details are seeing for Table 1. Moreover, 9 accident basic index data such as the region, season, run time and highway technical grade are obtained from the accident report. The information categories and examples are shown in Table 2.

Table 1 Accidents details of 800 km passenger lines from 2014 to 2020

Year	Number of accidents	Proportion of accidents (%)	Number of deaths	Number of injured
2014	13	27	125	188
2015	11	23	64	187
2016	9	19	69	66
2017	9	19	71	187
2018	3	6	25	62
2019	1	2	3	3
2020	2	4	8	37

Table 2 Indicators involved in 800 km passenger lines accident data and examples

Information category	Examples	Information category	Examples	Information category	Examples
Region	Northeast	Highway technical grade	I Highway	Road condition	Snow and ice
Season	Summer	Highway administrative grade	National Highway	Cause of the accident	Improper operation
Run time	00:00:00—00:59:59	Linear condition of road section	Straight sections of expressway	Holidays or spring festival	yes

2 Index selection

2.1 Accident causation index

Road traffic system is a dynamic system[10] composed of human, vehicles, roads and environment. Accidents are usually caused by the imbalance of dynamic system in a certain environment. This paper focusing on the situation of human, vehicles, roads and environment at the time of the accident, the accident characteristic index of 800 km Passenger Lines is selected. Based on the 4 first level indicators of human, vehicle, road and environment at the time of accident, the second level index and index system were established through the actual occurrence situation.

As the technical performance of vehicles is the material basis and directly affects the safe driving of vehicles, so the enterprises attach more importance to it. Only 8% of accidents are directly caused by vehicle factors, which does not have the significance of mathematical statistics. Therefore, this paper eliminates vehicle factors. The indicator system is established based on human, road and environment. The index system is shown in Table 3.

Table 3 Characteristic index system of accidents in 800 km passenger lines

Primary indicators	Secondary indicators	Category
Human	Human factors	Improper operation, not maintaining safe distances, overspeed, failure to cede in accordance with regulations, illegally driving on the road, lack of concentration, normal operation
Road	Region	Northeast China, North China, East China, South China, Central China, Northwest China, Southwest China
	Highway technical grade	Expressway, I highway, II highway, III highway
	Highway administrative grade	National highway, provincial highway, prefectural highway, country
	Linear condition of road section	Straight sections of expressway, curve, straight sections of ordinary roads, intersections, straight sections of tunnel, straight sections of bridges
	Road condition	Stagnant water, slippery, dry, wet, snow and ice
Environmental	Quarter	Spring, summer, autumn, winter
	Weather situation	Sunny, cloudy, overcast, foggy, rainy, snow and ice, others
	Holiday situation	Holidays, non-holidays
	Spring festival transport	Spring festival transport, non-spring festival transport.
	Run time	[0,1),[1,2),[2,3),[3,4),[4,5),[5,6),[6,7),[7,8),[8,9),[9,10),[10,11),[11,12),[12,13),[13,14),[14,15),[15,16),[16,17),[17,18),[18,19),[19,20),[20,21),[21,22),[22,23),[23,24)

It can be seen from table 3 that drivers are the main participants in the operation of 800 km Passenger Lines. The mistakes that caused by drivers' cognitive and behavioral errors and make them unable to complete tasks accurately, properly and fully are attributed to human errors. Therefore, Driver's mistakes in driving usually result in both cognitive and behavioral errors. Thus, the driver's errors in accidents are collectively referred to as human factors. As most of the accidents are caused by drivers, the human factors are classified into 7 categories, such as improper operation, not maintaining safe distances, etc.

The road index includes 5 secondary indexes, such as accident region, highway technical grade, highway administrative grade, linear condition of road section and road condition. According to the division of China's geographical regions, the accident places are divided into 7 categories, such as Northeast China, North China, East China, South China, Central China, Northwest China and Southwest China. The highway is divided into 4 categories according to the national standards and administrative standards. According to the accident coverage type, the section linearity can be divided into 6 categories, such as straight sections of expressway, curve, straight sections of ordinary roads, intersections, straight sections of tunnel and straight sections of bridges. According to the road surface state, the road condition can be divided into 5 categories, such as stagnant water, slippery, dry, wet, snow and ice.

Environmental indicators include 5 secondary indicators: season, holiday situation, spring festival transport, run time and weather condition. The seasons are divided into spring, summer, autumn and winter. Holiday situations are divided into holidays and non-holidays. Spring festival transports are divided into spring festival transport and non-spring festival transport. The run time is divided into 24 sections according to the hour, such as [0,1], which represents 00:00:00-00:59:59. The weather conditions are divided into 7 categories, including sunny, cloudy, overcast, foggy, rainy, others, snow and ice.

2.2 Form of accident

The form of traffic accident is the specific situation that the conflict among the participants of traffic accident or their own out of control. The accident is not only caused by the imbalance of an independent link, but also the result of multiple factors. Traffic accidents are formed by the coupling of many factors including human, vehicles, roads and environment through single or multiple factors. Each factor also contains different sub factors, leading to the final appearance of various traffic accidents.

According to the law of vehicle motion and contact, the accident forms are divided into 6 categories: rollover, scratch, collision, collision fixture, rear end collision and falling. According to the data in Table 4, among the accidents of 800 km Passenger Lines, the most frequent accidents are collision accidents, followed by rollover accidents. The largest number of deaths in a single accident is vehicle falling accidents, followed by collision accidents. Since the conflict between scratch accidents and collision with fixed objects is relatively simple, the casualties of single accident of them are the least. The largest number of humans injured in a single accident are rollover and collision. Once the vehicle falling accident occurs, the probability of survival is low, so the minimum number of injured is vehicle falling accident. It can be seen that different accident forms affect the severity of casualties. Therefore, when the accident cannot be avoided, the driver should choose the accident form with low casualties to minimize the loss.

Table 4 Casualties of different accident patterns

Accident pattern	Number of accidents	Death toll	Number of injured	Death toll of single accident	Number of injured in a single accident
Rollover	18	90	290	5	16.1
Scratch	1	3	5	3	5
Collision	23	226	368	9.8	16
Collision fixture	1	3	10	3	10
Rear end collision	4	17	53	4.3	13
Falling	1	26	4	26	4

3 Accident characteristic model

3.1 Model selection

Decision tree is a common classification regression algorithm, which represents the mapping relationship between object attributes and object values[11-12]. The model is composed of multi-layer nodes and branches. Through quantitative evaluation criteria, an object attribute is selected from the sample set as the current node splitting standard, and the sub nodes are generated recursively. The branches of each node represent the objects that meet the conditions of the node. Due to the difference of quantitative evaluation criteria for feature selection, there are different decision tree algorithms, such as ID3, CART and C4.5[13].

Among them, CART decision tree algorithm is widely used. It uses Gini to judge the possibility of unreasonable data classification results[14], which indicates the impure data set. The lower the Gini index is, the higher the purity is. The model finds the partition point with higher data purity by judging the Gini index. See Formula 1[15] for the calculation of Gini index. Compared with other algorithms, CART has high speed and high accuracy. In addition, its overall structure is obvious, easy to understand and extract important feature attributes. Therefore, CART algorithm is selected to analyze the key indicators and important scenarios of 800 km Passenger Lines.

$$Gini(p) = \sum_{k=1}^{K} p_k(1 - p_k) = 1 - \sum_{k=1}^{K} p_k^2, \qquad (1)$$

Where, p_k is the probability that the selected sample belongs to class k; $(1-p_k)$ is the probability that the sample is misclassified.

3.2 Model establishment and index input

3.2.1 Model establishment

Call *sksklearn tree* of *DecisionTreeClassifier* in Python to implements decision tree model algorithm. Call *compute_feature_importances* to judges the attribute importance of the index object in the model. See formula 2[16-18] for the judgment principle.

$$I_i = \frac{N_c}{N} \times \left[p_{im} - \frac{N_{cr}}{N_c} \times p_{im} - \frac{N_{cl}}{N_c} \times p_{iml}\right], \qquad (2)$$

Where, I_i is the importance of the attribute of the index object; N_c is the number of samples of the current node. N is the total number of samples; p_{im} is the impure of the current node, which represents the Gini index of the object attribute of the current node; N_{cr} and N_{cl} are the sample number of the right classification and the left classification of the current node; p_{imr} and p_{iml} are the impurity of the right classification and the left classification of the current node.

3.2.2 Parameter setting

Based on the input and output indexes and the characteristics of the algorithm model, the sample size and characteristics of the data are comprehensively considered to ensure the reliability. The parameters of the model are set as the following:

(1) Gini index is selected as the standard evaluation object.

(2) All features are used to find the best segmentation point of the model.

(3) Partition and reference all object attributes.

(4) The minimum number of samples per node is 4.

3.2.3 Input indicators

Based on CART decision tree algorithm model, 48 accident data are extracted. 11 accident causing indexes are taken as input of the model, including region, quarter, holiday situation, spring festival transport, run time, weather situation, highway technical grade, highway administrative grade, road alignment, road condition and human factors, and the accident occurrence form is taken as the output of the model. For this reason to analyze the direct, indirect or coupling relationship between the accident cause index and form, between the cause index and the cause index. The index name, category, input name and code are shown in Table 5.

3.3 Output

3.3.1 Index importance

The importance of input indexes is calculated by training model. The results are shown in Table 6. For the input indicators, holiday situation, spring festival transport, highway technical grade and highway administrative grade have no influence on accident form. The importance of other indicators on the accident form is run time, human factor, weather situation, quarter, region, road condition and road alignment. Therefore, the run time, human factors and weather situation are the most important factors. The future prevention should focus on the above 3 factors, coupled with the factors of quarter, region, road condition and road alignment.

3.3.2 Important accident scenarios

Based on the form of the accident, the coupling relationship between the 3 major accident causing factors (run time, human factor and weather situation) and other factors is analyzed. The important scenarios prone to traffic accidents are put forward. The study takes the run time, human factor and weather situation as the center, and combines with the season, region, road condition and road alignment to put forward the important accident scenarios of 800 km Passenger Lines.

Using Python to build decision tree model and call export_graphviz function. The result is visualized as Figure 1. Among them, the block diagram represents the judgment node of the data. The first row represents the judgment condition of the current node, the second line

Table 5 Model index details

Name	Input name	Input category (Code)
Region	Region	Northeast China (1), North China (2), East China (3), South China (4), Central China (5), Northwest China (6), Southwest China (7)
Season	Season	Spring (1), Summer (2), Autumn (3), Winter (4)
Whether holiday	Holiday	Yes (1), No (2)
Whether Spring Festival transport	Spring_Time	Yes (1), No (2)
Run time	Time	[0,1)(1),[1,2)(2),[2,3)(3),[3,4)(4),[4,5)(5),[5,6)(6),[6,7)(7),[7,8)(8),[8,9)(9),[9,10)(10),[10,11)(11),[11,12)(12),[12,13)(13),[13,14)(14),[14,15)(15),[15,16)(16),[16,17)(17),[17,18)(18),[18,19)(19),[19,20)(20),[20,21)(21),[21,22)(22),[22,23)(23),[23,24)(24)
Highway technical grade	Technical_level	I Highway (1), II Highway (2), III Highway (3), Expressway (4)
Highway administrative grade	Administrative_level	National highway (1), Provincial highway (2), Prefectural highway (3), Country (4)
Road alignment	Linear_type	Straight sections of expressway (1), Intersection (2), Straight sections of bridges (3), Straight sections of tunnel (4), Curve (5), Straight sections of ordinary roads (6)
Road Condition	Road	Snow and Ice (1), Dry (2), Stagnant Water (3), Snow (4), Wet (5)
Human factor	People	Improper Operation (1), Overspeed (2), Illegally Driving on the Road (3), Failure to Cede in Accordance with Regulations (4), Not Maintaining Safe Distances (5), Lack of Concentration (6), Normal Operation (7)
Weather situation	Weather	Sunny (1), Cloudy (2), Overcast (3), Foggy (4), Rainy (5), Ice and Snow (6), Others (7)
Accident form	Accident_type	Rollover (1), Scratch (2), collision (3), Collision fixture (4), Rear end collision (5), Falling (6)

Table 6 Significance results of accident characteristic indexes

Indicator name	Importance of indicators	Ranking
Run time	0.232	1
Human factor	0.168	2
Weather situation	0.149	3
Season	0.143	4
Region	0.130	5
Road condition	0.111	6
Road alignment	0.066	7
Whether holiday	0.000	—
Whether Spring Festival transport	0.000	—
Highway technical grade	0.000	—
Highway administrative grade	0.000	—

represents the Gini index of the judgment condition for the sample under this node, and the third line represents the sample size of the current node. For any node, when the sample characteristics of the entry node satisfy the current node judgment condition, the sample is classified into the left branch of the next layer set. Otherwise, it is included in the right branch of the next layer set. When the Gini index of the node is 0 or the sample size is less than 4, it means that the current node sample is of the same category, and the node is the end point of the branch.

It can be seen from Figure 1 that in the 800 km Passenger Lines accidents, the proportion caused by collision fixture and rear end collision is very large, which means that the traffic hazards caused by these two types of accidents are at the top, and prevention and control should be focused. In addition, the human factor is at the top of the whole model, because the superposition of any external conditions in driving will eventually point to the driver's decision and operation.

In order to ensure the representativeness of accident characteristics, the branches of decision tree model with classification number greater than 3 are extracted for subsequent analysis. Figure 2 is a branch graph of the left branch (human factor) of the decision tree. It can be seen from the figure that human factors can be divided

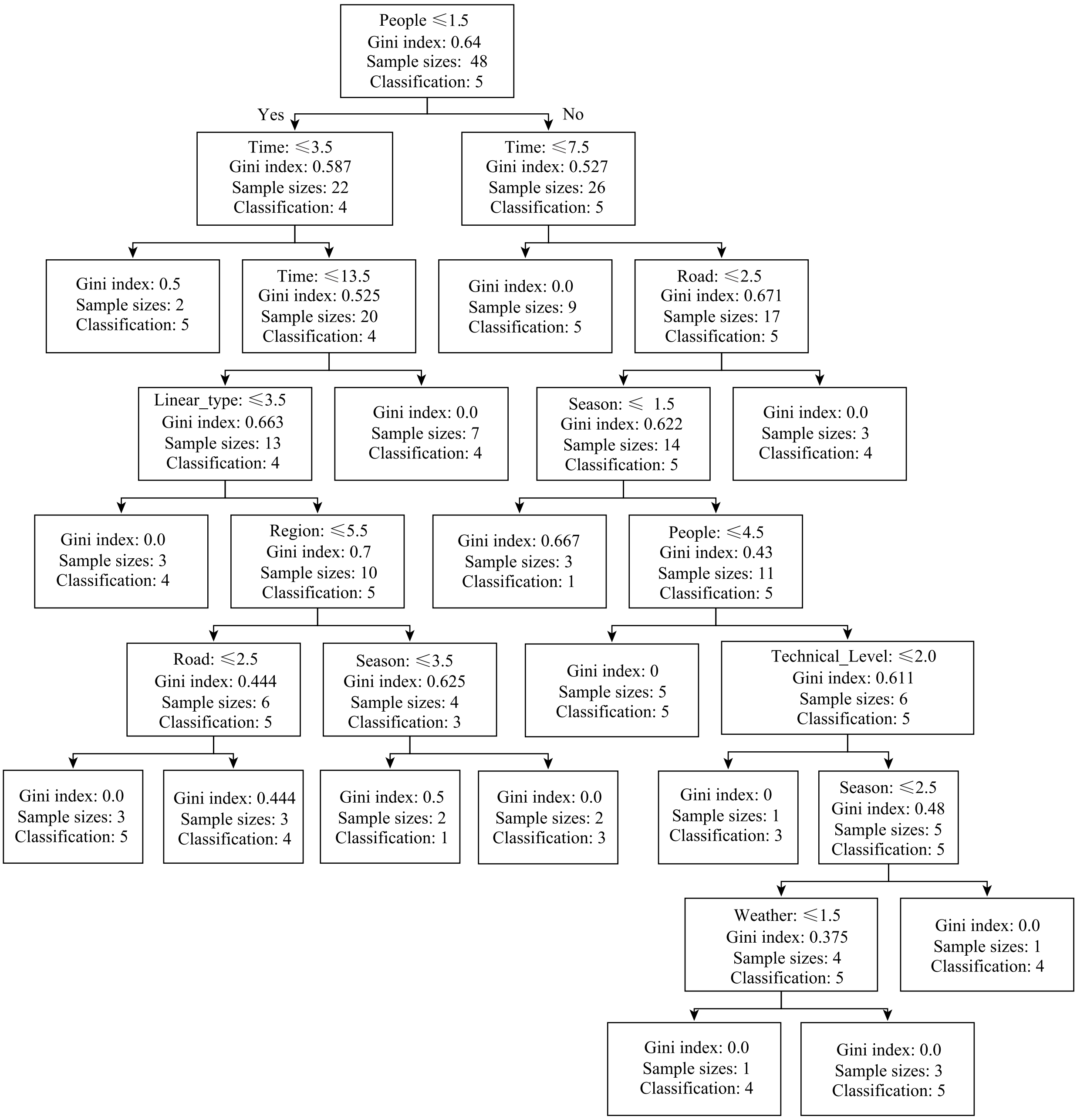

Fig. 1 Accident cause model output results of 800 km passenger lines

into 7 types, and the traffic accidents caused by improper operation (people: ≤1.5) account for about 50% of the total number (22 cases). This shows that China should focus on strengthening the driving skills, operation techniques and standards of drivers.

The analysis shows that there is a coupling relationship among weather, road alignment and road condition. 2 traffic accidents occurred under the weather conditions of snow, cloudy and sunny (weather: ≤3.5). The data shows that the accident occurred in snow, snow and ice, and the road condition was dry, which occurred in the curve section. It shows that the curve section is an important coupling cause of traffic accidents. There are many traffic accidents in foggy, cloudy, rainy and other weather conditions (weather: >3.5), which proves that weather factors are extremely important factors leading to traffic accidents.

There is a coupling relationship between weather and time. In foggy, overcast, rainy and other weather conditions (weather: >3.5), 7 traffic accidents occurred later than 13 o'clock (time: >13.5) were all caused by impact of collision fixture. In the accidents earlier than or equal to 13 o'clock (time: ≤13.5), there were 3 causes of impact on collision fixture. It shows that drivers are prone to collision fixture after 13 o'clock in foggy, overcast and rainy weather, which should be focused on prevention and control.

There are coupling relationships among region, road alignment, weather and road condition. In the not northwest and southwest regions (region: ≤5.5), the road alignment is straight sections of tunnel, curve and straight sections of ordinary roads (Linear_type: >3.5) of the traffic accidents. Among them, 3 accidents are caused by rear end collision, and they occur in overcast days, with stagnant water or slippery on the road surface, and 67% of them occur in the curve. It shows that the driver is prone to rear end collision accident when driving in the curve section of not northwest and southwest regions in overcast days. In the Northwest and Southwest regions (region: >5.5) 4 traffic accidents occurred, 75% of accidents occurred in overcast or foggy days with stagnant water or slippery road. It shows that drivers are prone to traffic accidents in Northwest and southwest China on overcast days.

Figure 3 is a branch graph of the right branch (not human factor) of the decision tree. It can be seen from the figure that 6 types of traffic accidents caused by non-improper operation (people: >1.5) only account for 50% (26 cases) of the total number. It can be seen that compared with improper operation, drivers in China have better performance in maintaining safe distance, speeding, giving way according to regulations, avoiding illegal road occupation and keeping attention, but it also needs to be strengthened.

In the traffic accidents not caused by improper operation (people: >1.5), there is a coupling relationship among run time, region and road alignment. Before 7 o'clock (time: ≤7.5), 7 accidents were caused by rear end collision, accounting for a large proportion, and most of them occurred in central and South China. 83% of the accidents were caused by drivers' failure to keep a safe distance from the vehicle in front. It shows that before 7:00 a.m., drivers in central and South China are prone to rear end collision due to failing to keep a safe distance from the vehicle in front.

There is a coupling relationship between weather

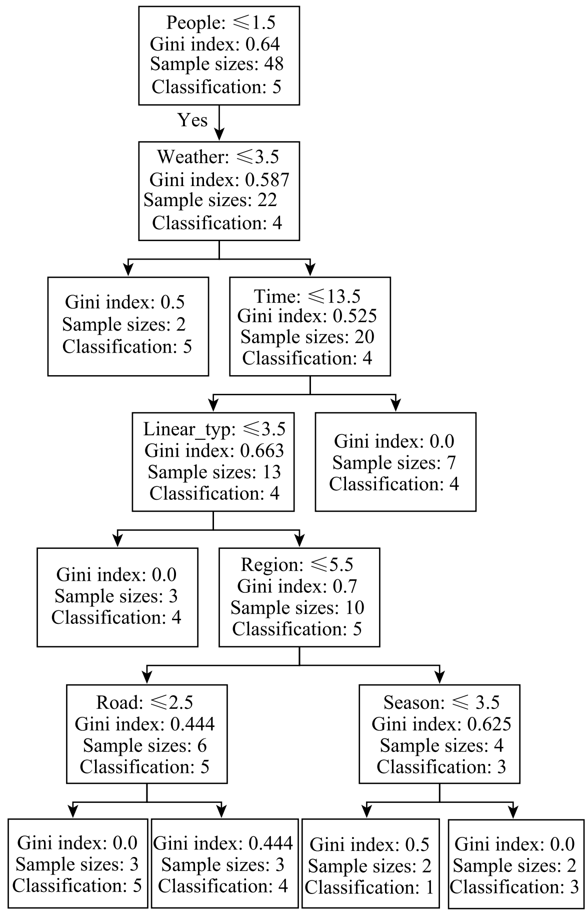

Fig. 2 Main branch of the left branch (human factors) of model

and road conditions. 3 collision accidents occurred under the condition of stagnant water, snow and slippery road (road: >2.5), which indicated that the driver was prone to traffic accidents on above roads. Among the traffic accidents occurred on snow and ice, dry road (road: ≤2.5), 5 accidents of rear end collision occurred due to bad driving behavior (people: ≤4.5), and 2 of them occurred in ice and snow weather. In addition, 3 accidents of rear end collision occurred in snow and ice weather when drivers were distracted or under normal operation (people: >4.5). It shows that the driver's bad driving behavior or inattention phenomenon is easy to cause rear end collision accident.

Based on different accident scenarios, road transport enterprises should focus on prevention and control. First, road transport enterprises should do a good job in passenger transport safety management, control potential safety hazards from the source, implement the main responsibility of safety production, and conduct real-time supervi-

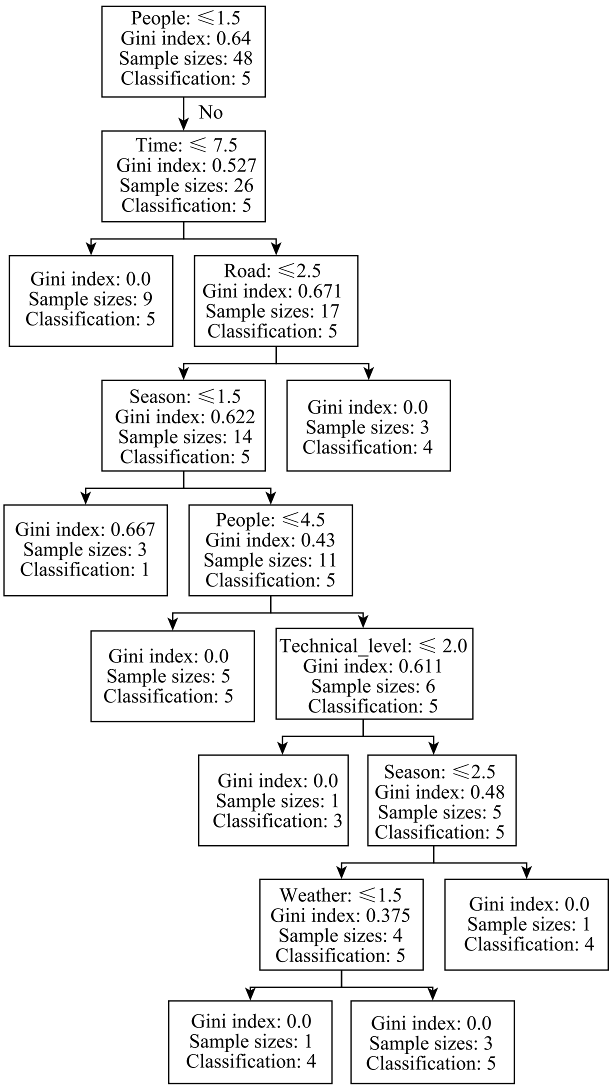

Fig. 3 Main branch of the right branch (non-human factors) of model

sion and punishment on drivers' illegal behaviors and bad behaviors. Second, road transport enterprises should cultivate drivers' emergency avoidance ability, pay special attention to dynamic supervision in fog, overcast, rain, snow and other special weather, focus on monitoring drivers' driving behavior and concentration, strengthen drivers' defensive driving methods in special weather and special periods, and appropriately introduce driving simulation technology to simulate driving scenes in weather and time. In addition, the training content should be strengthened to enhance the driver's ability to avoid accidents. And road transport enterprises should carry out targeted training and education. According to the accident scenarios, comprehensive consideration should be given to the characteristics of drivers and operation routes, and targeted training should be conducted for drivers involved in Northwest and southwest regions, and defensive driving methods under overcast conditions should be focused on. For the driving line involving central and South China, the safety distance from the vehicle in front should be monitored and the emergency measures for rear end collision should be trained.

4 Conclusions

(1) This paper analyzes the accident characteristics of 800 km Passenger Lines. Based on the contributing factors of human, vehicles, roads and environment in the accident, it extracts the accident occurrence form and cause index, and constructs the safety risk model of 800 km Passenger Lines through decision tree model. The results indicate that collision fixture and rear-end collisions account for the highest proportion of accidents on 800 km passenger lines, highlighting the need to prioritize the prevention and control of these accident types. Human factors are the most important factors leading to traffic accidents, so it is necessary to improve drivers' driving skills, operation techniques and specifications. In addition to whether holiday, whether Spring Festival Transport, highway technical grade and highway administrative grade, there are coupling relationships among human factors, weather situation, seasons, regions, road conditions and road alignment, among which human factors, weather and seasons are the 3 most important causative factors.

(2) In order to reduce the accident potential of 800 km Passenger Lines and provide theoretical basis for enterprise safety management, the existing data sample is limited. Although the accident characteristic index is representative, it can be further enriched, and the safety management of enterprise content has not been considered. In the future, it can be further optimized by combining with the human, vehicles, roads, environments, in transportation. The enterprise internal management and other factors, comprehensively build a passenger transport accident database of 800 km Passenger Lines to supplement the accident factors and important scenarios extracted from the current data group. To form a set of accurate, comprehensive and systematic accident characteristic indexes and prevention and control measures for

800 km Passenger Lines.

Acknowledgements

This work was supported by the Transportation Strategic Planning Policy Project [Grant number 2021/8/6].

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Department of Transport Services. Ministry of Transport. Statistical Analysis Report and Data Compilation of Road Transport in 2020 [R]. Beijing: Department of Transport Services, Ministry of Transport, 2021. (in Chinese)

[2] CHU H C. Assessing Factors Causing Severe Injuries in Crashes of High-deck Buses in Long-distance Driving on Freeways [J]. Accident Analysis and Prevention, 2014, 62 (1): 130-136.

[3] CHANG W H, GUO H R, LIN H J, et al. Association Between Major Injuries and Seat Locations in A Motorcoach Rollover Accident [J]. Accident Analysis and Prevention, 2006, 38 (5): 949-953.

[4] MORADI A, NAZARI S S H, RAHMANI K. Sleepiness and the Risk of Road Traffic Accidents: a Systematic Review and Meta-analysis of Previous Studies [J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2018, 34 (3): 194-203.

[5] ZHU Shou-sheng. Study on the Risk of Commercial Bus Traffic Safety Based on the Fault Tree Analysis [D]. Xi'an: Chang'an University, 2016. (in Chinese)

[6] WANG Zi-heng, MA Jiang-xiao, ZHOU Wen-zhi, et al. Risk Assessment of Passenger Line Transportation System Based on Bayesian Network [J]. Logistics Technology, 2019, 38 (5): 50-54, 87. (in Chinese)

[7] ZHAO W H, ZHANG Y H, ZHAI X X. Research on the Driving Risk Assessment and the Risk Control Methodology for the Road Transportation Businesses [J]. Journal of Safety and Environment, 2018, 18 (5): 1743-1748.

[8] HU Li-wei, YANG Jin-qing, HE Yue-ren, et al. Study on Vehicle Operational Risk Identification in Urban Traffic Congestion Beased on Improved BP Neural Network [J]. Journal of Highway and Transportation Research and Development, 2019, 36 (10): 105-113. (in Chinese)

[9] ZHENG Jing-long, YUAN Jia-wei, ZHU Chang-an, et al. Study on Safety Risk Assessment of Expressway Tunnel Operation [J]. Outhwest Highway, 2018, (2): 2-5, 12. (in Chinese)

[10] WANG Xiao. Research on Highway Alignment Design and Eevaluation Method Based on Traffic Safety [D]. Xi'an: Chang'an University, 2007. (in Chinese)

[11] ZHAO Ying-yan, ZHAN Wei-wei, LI Wei-wei, et al. Study on Central Compartment Lymph Node Metastasis of Papillary Thyroid Carcinoma By Using Decision Tree [J]. Chinese Journal of Ultrasound in Medicine, 2022, 38 (1): 5-8. (in Chinese)

[12] WU Xiao-yu, HE Jing-han, ZHANG-Pei, et al. Power System Short-term Load Forecasting Based on Improved Random Forest with Grey Relation Projection [J]. Automation of Electric Power Systems, 2015, 39 (12): 50-55. (in Chinese)

[13] LI Cang-bai, XIAO Ke-yan, LI Nan, et al. A Comparative Study of Support Vector Machine, Random Forest and Artificial Neural Network Machine Learning Algorithms in Geochemical Anomaly Information Extraction [J]. Acta Geoscientica Sinica, 2020, 41 (2): 309-319. (in Chinese)

[14] LUO Fen, WANG Xiao-qing, CAI Yan, et al. A New Dual-objective CD-CAT Item Selection Method Based on the Gini Index [J]. Acta Psychologica Sinica, 2020, 52 (12): 1452-1465. (in Chinese)

[15] SHANG Wen-qian, HUANG Hou-kuan, LIU Yu-ling, et al. Research on the Algorithm of Feature Selection Based on Gini Index for Text Cate-gorization [J]. Journal of Computer Research and Development, 2006, 43 (10): 1688-1694. (in Chinese)

[16] ZHAO Yun-bin, ZHANG Zhen, WANG Yan-peng, et al. Application of Decision Tree Analysis Method in the Governance Decision of Low Production and Low Efficiency Wells in BZ Oilfield [J]. Chemical Enterprise Management, 2021, 35 (12): 197-198. (in Chinese)

[17] YANG Quan. Research on Judgment Method of N1+N2 Structure Grammatical Relation Based on Decision Tree and Word Semantic Similarity [J]. Computing Technology and Automation, 2021, 40 (4): 36-41. (in Chinese)

[18] SUN Qi, GAO Yan-yu, XU Xin-yue, et al. Refined Passenger Flow Forecasting System for Urban Rail Transit [J]. Railway Computer Application, 2021, 30 (12): 70-78. (in Chinese)