Journal of Safety Research 76 (2021) 176–183

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Risk factors affecting crash injury severity for different groups of e-bike riders: A classification tree-based logistic regression model

Zhengwu Wang [a,b], Shuai Huang [a,b], Jie Wang [a,b,*], Denisa Sulaj [b], Wei Hao [b], Aiwu Kuang [b]

[a] Key Laboratory of Highway Engineering of Ministry of Education, Changsha University of Science and Technology, Changsha 410114, China
[b] School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha 410114, China

ARTICLE INFO

Article history:
Received 12 June 2020
Received in revised form 10 September 2020
Accepted 10 December 2020
Available online 30 December 2020

Keywords:
E-bike crash
Injury severity
Classification tree-based logistic regression
Different riders groups

ABSTRACT

Introduction: As a convenient and affordable means of transportation, the e-bike is widely used by different age rider groups and for different travel purposes. The underlying reasons for e-bike riders suffering from severe injury may be different in each case. *Method*: This study aims to examine the underlying risk factors of severe injury for different groups of e-bike riders by using a combined method, integration of a classification tree and a logistic regression model. Three-year of e-bike crashes occurring in Hunan province are extracted, and risk factor including rider's attribute, opponent vehicle and driver's attribute, improper behaviors of riders and drivers, road, and environment characteristics are considered for this analysis. *Results*: E-bike riders are segmented into five groups based on the classification tree analysis, and the group of non-occupational riders aged over 55 in urban regions is associated with the highest likelihood of severe injury among the five groups. The logistics analysis for each group shows that several risk factors such as high-speed roads have commonly significant effects on injury severity for different groups; while major factors only have significant effects for specific groups. *Practical application:* Based on model results, policy implications to alleviate the crash injury for different e-bike riders groups are recommended, which mainly include enhanced education and enforcement for e-bike risky behaviors, and traffic engineering to regulate the use of e-bikes on high speed roads.

1. Introduction

As a convenient and affordable means of transportation, the electric bike (e-bike) is widely prevalent in China. Since the introduction of the legislation to deal with e-bikes in 1998, the number of e-bikes in China increased from 58,000 to more than 170 million in 2014 (World Health Organization, 2017). Accordingly, the growing popularity of e-bike entails significant safety concerns as observed in crash statistics. According to China's Road Traffic Crash Statistics (National Bureau of Statistics, 2015), 7,201 people died in e-bike related crashes in 2014, representing 11.4% of all road traffic deaths.

E-bike is officially defined as a non-motorized transportation mode in China. Because of a lack of safety awareness and strict traffic law enforcement, e-bike riders are prone to violate the traffic rules, such as disobeying traffic signals, riding on the motorized lane, and failing to give right of way (Wu, Yao, & Zhang, 2012; Du et al., 2013; Yang, Mei, Abdel-Aty, Peng, & Gao, 2015; Wang, Xu, Xia, & Qian, 2017a; Bai, Liu, Chen, Zhang, & Wang, 2015). These improper/illegal behaviors will increase the e-bike rider's own crash risk, as well as the crash risk of other road users (Li, Xing, Wang, Liang, & Wang, 2018). Moreover, compared to the traditional non-motorized modes such as bicycles, e-bikes have a higher injury risk in crashes because they often move at higher speeds. A study conducted by Hu, Lv, Zhu, and Fang (2014) found that the probability of severe injury for e-bike crashes was nearly two times that of bicycle crashes. The growing popularity of e-bikes and their high injury risk in crashes highlight urgent needs to examine the factors that affect injury severity of e-bike crashes in China.

Crash injury severity models associate the likelihood of crash injury level with various contributing factors such as driver/riders, traffic and road, vehicle, crash type characteristics, and weather environment; therefore it is an important quantitative tool to predict crash injury severity and identify high-risk factors that significantly aggravate injury severity. The choice of appropriate analytical methods and the selection of representative explanatory variables (also called contributing factors or risk factors) are two

* Corresponding author at: Key Laboratory of Highway Engineering of Ministry of Education, Changsha University of Science and Technology, Changsha 410114, China.

E-mail address: wjie334@163.com (J. Wang).

https://doi.org/10.1016/j.jsr.2020.12.009
0022-4375/

important considerations for establishing an accurate crash injury severity model. Up to date, studies on injury severity of e-bike crashes are relatively few. These studies mainly focus on injury-contributing factors related to rider age and gender, crash characteristics, and road environment (Hu et al., 2014, 2020b; Hu, Hu, Wan, Chen, & Cao, 2020a). In China, especially in middle and small sized cities, e-bikes are always used as occupational transportation modes (such as delivery-e-bikes and taxi-e-bikes). To the best of our knowledge, contributing factors on crash injury severity of occupational e-bike riders were not investigated in previous studies. In addition, the direct factors that contribute to e-bike crash occurrence--riders/drivers' improper/illegal behaviors effects on crash injury severity are still not fully examined.

In regard to analytical methods, discrete choice models such as logistic regression models are generally applied because injury severities in crash datasets are often modeled as discrete severity outcomes (such as fatal injury, serious injury, slight injury, no injury; Savolainen, Mannering, Lord, & Quddus, 2011). For instance, Hu et al. (2014) developed a logistics regression for e-bike crash injury severity using 146 cases of trauma-related patients involved in e-bike crashes in Hefei, China. Recently, based on China in-depth crash data, Hu et al. (2020b) used logistics regression model to examine the fatal/serious injury risk of e-bike riders in relation to the impact speed and rider's age.

Although logistic regression models have important superiority in measuring the marginal effects of the risk factors, it may appear a low prediction as they have shortcomings in dealing with interactions between various risk factors (Huang, Peng, Wang, Luo, & Li, 2018; Zeng, Wen, & Huang, 2016). In addition, because traffic crashes might occur under distinct conditions, specific risk factors may have different magnitude or even opposite direction effects on injury outcomes (Chang, Xu, Zhou, Chan, & Huang, 2019; Wang, Huang, & Zeng, 2017b). A general logistic regression model cannot account for this phenomenon. For this case, an effective measure is to separate the full sample crash data into homogeneous groups/segmentations. One common segmentation approach for crash data is based on expert judgments, such as crashes by collision types and crashes involving different age groups. However, expert judgments cannot guarantee a homogeneous group in each segment, because they may ignore the fact that crash injury severity is the result of the combination of multiple factors. On the other hand, data mining techniques (such as tree classification methods) refer to a statistical analysis aimed at searching hidden structures or patterns among "big data;" which have a better ability to identify homogeneous groups compared with expert judgments (Prati, Pietrantoni, & Fraboni, 2017; Wang et al., 2017a). Logistic regression models, combined with data mining techniques, could provide more powerful insights than applying a general logistic regression model to the entire data.

In this study, a classification tree-based logistic regression model is applied to identify the underlying risk factors affecting crash injury severity for different types of e-bike riders. Specifically, based on e-bike crash data collected from Hunan province of China from 2014 to 2016, the chi-squared automatic interaction detection (CHAID) tree is first employed to split the entire data into several homogeneous groups by considering multiple riders-related attributes, including rider's age, gender, occupation and living region. This tree analysis could help us understand the question of "who are the high risk riders associated with high likelihoods of severe injury." Then logistic regression models are used to examine significant factors and their marginal effects for each group. The results of logistic regression analysis will provide useful sights for solving the question of "how to reduce the crash injury for different types of e-bike riders, especially for high risk riders."

2. Method

2.1. Data preparation

Crash data are obtained from the Traffic Management Sector-Specific Incident Case Data Report, which is maintained by Hunan Provincial Department of Traffic Police. The report covers various aspects of a traffic crash, including demographic characteristics of the casualties, the cause of the crash, collision types, and vehicle types; and environmental factors such as weather conditions, the precise crash time, and location of the crash. Injury severity of crashes recorded by traffic police is categorized as fatal (i.e., immediate or subsequent death from injuries within 7 days after a crash), serious (i.e., disability injury), slight (i.e., non-disability injury) and property damage only (i.e., no injury). Three years (from 2014 to 2016) of e-bike crashes that occurred in Hunan province were extracted from the crash report. The database was rearranged and crash-related variables used for this analysis were shown in Table1.

2.2. Statistical analysis

Chi-squared Automatic Interaction Detection (CHAID) is a type of classification tree technology that is used to split the data into statistically significant homogeneous subgroups based on stepwise chi-square test. The chi-square test employed by CHAID is determined by the relationship between the independent variables (or output variables) and dependent variables (predictor variables). In the present study, the CHAID is employed to classify the e-bike riders who are involved in the crashes. We chose CHAID because of its superior ability to examine all possible splits and thus identify all subgroups with high homogeneity. The independent variables used for the classification include four rider-related attributes: rider's age, gender (male vs. female), occupation (occupational rider vs non-occupational rider), and living region (rural vs. urban). The dependent variable is the injury severity of the e-bike rider. The injury severity is re-categorized as severe injury (fatal and serious) and non-severe injury (slight and property damage only). This dichotomy classification is consistent with the view that converting target variable to a binary class could mitigate the bias by selection (Jung, Qin, & Oh, 2016).

Based on the above CHAID technology, e-bike riders are divided into several homogeneous groups with similar injury severity, defined by rider's age, gender, rider's occupation, and living region. After that, binary logistic regression models are employed to identify significant risk factors affecting the crash injury severity for each group. Specifically, the binary outcomes (i.e., severe and non-severe) represented by a dummy variable are used as the response variable; while 10 risk factors are specified as explanatory variables: (1) time, (2) week, (3) season, (4) weather, (5) lighting, (6) road type, (7) opponent vehicle type, (8) driver age of opponent vehicle, (9) pre-crash improper behavior of driver, and (10) pre-crash improper behavior of the rider. The significant risk factors identified by logistic regression models will provide us a full understanding of "how to alleviate the crash injury for different groups of e-bike riders."

3. Results

First, CHAID is used to classify all crash-involved e-bike riders into different homogeneous groups by considering multiple riders-related attributes, including rider's age, gender, rider's occupation, and living region. For the estimation of CHAID, the total crash data were randomly split into 50% for learning and 50% for

Table 1
Descriptive statistics for variables.

Variable	Coding and description	Frequency (proportion %)
Injury severity	1: Fatality & Serious injury	1464(83.2)
	2: Slight injury & Property damage only	296(16.8)
E-bike rider age	1: <25	112(6.4)
	2: 25–34	229(13.0)
	3: 35–44	339(19.3)
	4: 45–54	478(27.2)
	5: 55–64	372(21.1)
	6: 65–74	206(11.7)
	7: ≥75	24(13.0)
Rider gender	1: Male	986(56.0)
	2: Female	774(44.0)
Rider occupation[1]	1: Occupational	497 (28.3)
	2: Non-occupational	1263 (71.7)
Rider living region	Rural	515(29.3)
	Urban	1245(70.7)
Time	1: 22:00–6:59	240 (13.60)
	2: 7:00–8:59	237(13.5)
	3: 9:00–16:59	826(46.9)
	4: 17:00–18:59	258(14.7)
	5: 19:00–21:59	199(11.3)
Week	1: Weekday	1344(76.4)
	2: Weekend	416(23.6)
Season	1: Spring	472(26.8)
	2: Summer	486(27.6)
	3: Autumn	425(24.1)
	4: Winter	377(21.5)
Weather	1: Rainy	290(16.5)
	2: Others	1470(83.5)
Lighting	1: Daylight	1280(72.7)
	2: dawn/dusk	89(5.1)
	3: Dark but lighted	303(17.2)
	4: Dark	88(5.0)
Road type[2]	1: high-speed road	275(15.6)
	2: low-speed road	1485(84.4)
Opponent vehicle type	1: E-bike	702(39.9)
	2: Motorcycle	84(4.8)
	3: Truck	173(9.8)
	4: Car	716(40.7)
	6: Other vehicle	85(4.8)
Driver age of opponent vehicle	1: <25	150(8.5)
	2: 25–34	419(23.8)
	3: 35–44	453(25.7)
	4: 45–54	427(24.3)
	5: 55–64	191(10.9)
	6: 65–74	97(5.5)
	7: ≥75	23(1.3)
Driver gender of opponent vehicle	1: Male	1293(73.5)
	2: Female	461(26.5)
Pre-crash improper behavior of Riders	1: Riding on the wrong lane[3]	315 (17.9%)
	2: Riding against traffic signals at the intersection	311 (17.6%)
	3: While crossing the road, not to get off and push the e-bike	136 (7.7%)
	4: Reverse riding	108 (6.1%)
	5: Other improper behaviors of e-bike riders	95 (5.4%)
	6: Having no improper behaviors of e-bike riders	796 (45.3%)
Pre-crash improper behavior of opponent vehicle drivers	1: Driving on the non-motor vehicle lane	221(12.6)
	2: Driving against traffic signal at the intersection	111(6.3)
	3: Safety distance violation	35(2.0)
	4: Speeding & Overloaded	48(2.7)
	5: Approaching illegally & Changing lane illegally	124(7.0)
	6: Failing to give way	220(12.5)
	7: Reverse driving	74(4.2)
	8: Other improper behaviors of the driver	540(30.7)
	9: Having no improper behaviors of the driver	387(22.0)

[1] Occupational riders include delivery-purpose e-bike and taxi-purpose e-bike riders.
[2] High-speed road (refers to the design speed higher than 50 km/h).
[3] Riding on the wrong lane includes two types: a) on the road with non-motor vehicle lanes, not to ride within the non-motor vehicle lane, b) on the road without non-motor vehicle lanes, not to ride by the right side of the motor vehicle lane.

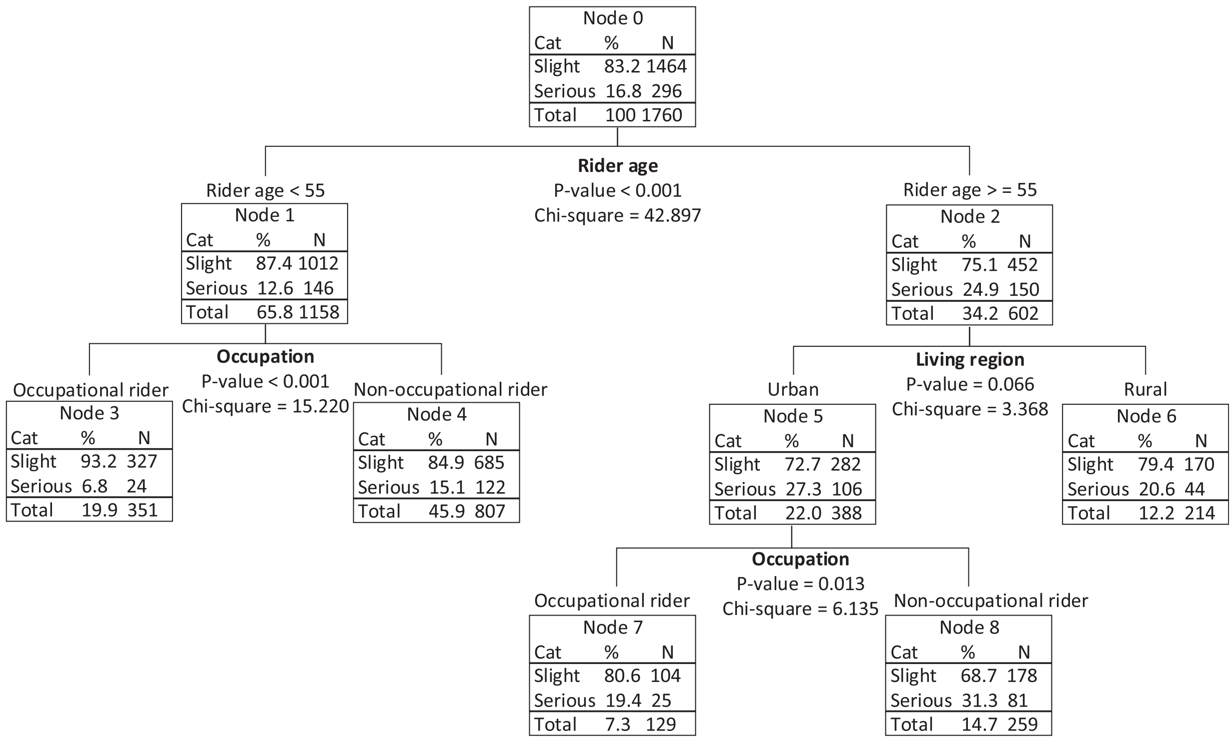

Fig. 1. CHAID classification tree for injury severity of e-bike rider.

testing data. The training dataset is used to build the model, while the test dataset is used to test the ability of the model for its applicability and generalization. Then, for each group of e-bike riders, a binary logistic regression model is built to examine risk factors that significantly affect the injury severity of e-bike riders. For the model estimation, backward step-wise is used to exclude insignificant explanatory variables at the significance level with p-value >0.1, and the final model is re-calibrated with only significant variables. Both CHAID and binary logistic regression models are calculated by using the econometric and statistical software SPSS 21.0.

3.1. Classification tree analysis

Fig. 1 displays the classification tree for crash injury severity of e-bike riders. The tree classification results in four splitters and five terminal nodes. It shows that rider's age, rider's occupation, and living region are important variables to classify the rider's groups, while rider's gender is not significant for the classification.

The tree classification is first split by the variable of rider's age. It implies that rider's age is the most important variable to differentiate injury severity of the e-bike rider among riders-related attributes. This split directs riders aged under 55 to the left forming node 1, and directs riders aged 55 or above to the right forming node 2. This indicates that older riders, aged 55 or above, have a higher percentage of severe injury (24.9%) than riders aged under 55 (12.6%). This result is similar to previous findings that older adults are more likely to suffer from severe injury in bicycle crashes (Kim, Kim, Ulfarsson, & Porrello, 2007; Yan, Ma, Huang, Abdel-Aty, & Wu, 2011), which may be associated with physical fragility and decrease in risk-avoiding ability of old riders.

On the left side of the tree, node 1 continued to be split based on the variable of the rider's occupation, forming terminal node 3 and 4. For riders aged under 55, non-occupational riders have a higher likelihood of severe injury (15.1%) than occupational riders (6.8%). In this study, occupational riders include delivery-purpose e-bike and taxi-purpose e-bike riders. The lower injury risk of occupational riders may be due to the fact that they are more familiar with the driving environment and have relatively richer driving skills (Chung, Song, & Yoon, 2014; Wu & Loo, 2016). Turning to the right side of the tree, based on the variable of living region, node 2 is split into child node 5 and terminal node 6. For riders aged 55 and above, riders in urban regions have higher likelihoods of severe injury (27.3%) than the ones in rural regions (22.6%). Further down to the tree, node 5 is split by the variable of improper behavior, forming terminal node 7 and 8. For rider's aged above 55 and riding in urban areas, non-occupational riders have higher risks of severe injury (31.3 %) than occupational riders (19.4%), which is similar to riders aged under 55.

Based on the results of the tree classification, the e-bike riders are segmented into five groups (i.e., five terminal nodes). Group 1: occupational riders aged under 55 (node3); group 2: non-occupational riders aged under 55 (node 4); group 3: riders aged above 55 in rural regions (node 6); group 4: occupational riders aged above 55 in urban regions (node 7); group 5: non-occupational riders aged above 55 in urban regions (node 8). Among these five groups, group 5 is associated with the highest likelihood of injury severity (31.3%), which is much higher than the average percentage (16.8%).

3.2. Logistic regression analysis

Tree-based logistic regression model for separated rider's groups and a general logistic regression model for the whole data are developed. Table 2 shows the goodness-of-fit measures for tree-based models and the general model. The likelihood ratio test comparing the tree-based models and the general model indicated that there is more than 99.9% confidence that the tree-base models are statistically superior in terms of goodness-of-fit.

Parameter estimates of both general and tree-based model are shown in Table 3. To directly assess the impacts of explanatory variables on injury severity probability of each e-bike rider group, marginal effects are also computed, as shown in Table 4. In this

Table 2
Goodness-of-fit measures for tree-based and general models.

Model statistics	General	Group 1	Group 2	Group 3	Group 4	Group 5
Number of observers	1760	351	807	214	129	259
Number of parameters	13	6	5	5	6	6
Log likelihood at convergence	724.385	68.044	327.7115	101.255	50.886	146.829
Log-likelihood ratio test						
$\chi^2 = -2[LL_{(general)} - LL_{(tree\text{-}based)}]$						−59.319
Degrees of freedom						15
P-value						<0.001

Table 3
Parameter estimates of both tree-based and general models.

Variables	General		Group 1		Group 2		Group 3		Group 4		Group 5	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Constant	−3.580	0.302	−4.070	0.450	−2.302	0.166	−2.031	0.486	−0.791	0.348	−0.982	0.240
Rider aged 35–44	0.472	0.267	–	–	–	–	–	–	–	–	–	–
Rider aged 45–54	0.778	0.245	–	–	–	–	–	–	–	–	–	–
Rider aged 55–64	1.282	0.244	–	–	–	–	–	–	–	–	–	–
Rider aged 65–74	1.583	0.270	–	–	–	–	–	–	–	–	–	–
Rider aged >=75	1.667	0.514	–	–	–	–	–	–	–	–	–	–
Non-occupational	0.761	0.175	–	–	–	–	–	–	–	–	–	–
Rider Living in urban	0.521	0.158	–	–	–	–	–	–	–	–	–	–
22:00–06:59	0.834	0.196	1.225	0.549	1.042	0.285						
19:00–21:59	0.384	0.188									0.877	0.326
Spring	−0.319	0.127										
Summer					0.454	0.215					0.617	0.253
Weekend									−1.245	0.691		
Daylight							0.796	0.496	−1.318	0.562		
Dawn/Dust			1.435	0.656								
High-speed road	0.747	0.161	2.189	0.530	0.654	0.240	0.814	0.386	1.432	0.803		
Having no improper behaviors of e-bike rider	−0.272	0.117					−0.820	0.385	−0.535	0.226	−0.649	0.305
While crossing the road, not to get off and push the e-bike			1.072	0.449								
Driver gender of opponent vehicle is female											−0.641	0.306
Type of opponent vehicle is truck	0.649	0.199			0.544	0.181	0.976	0.320			1.630	0.477
Driving on the wrong lane											1.037	0.429
Reverse driving			2.368	1.027								
Other improper behaviors of the driver			1.130	0.493	0.359	0.136						

"–"Rider attribute variables are included in the general model, while not included in the separated models since they have been considered in preliminary tree analysis.

study, the marginal effects represent the change in the resulting probability of severe injury due to one unit change (or change from 0 to 1 in the case of indicator variables) in an explanatory variable, while holding all other variables constant.

Several meaningful discoveries can be found in Table 3. (1) The results of the general model show that rider attribute variables of "old riders," "non-occupational riders," and "riders in urban areas" have significantly higher likelihoods of severe injuries, which further supported the preliminary result of CHAID tree. (2) While comparing the significant risk factors reveals between general model and separated models, some differences are discovered. For example, the variables of "summer," "weekend," "daylight," "dawn/dust" and several types of rider/driver behaviors are not statistically significant in the general model, but they have significant effects on injury severity of specific rider groups. This indicates that the tree-based separated model reveals new information that cannot be found by the general model using the entire data. (3) From the results of the tree-based model, we find that the significant variable sets for injury severity are not consistent for different rider groups. This confirms that it is very meaningful to use the tree-based model instead of the general model. The detailed interpretations for significant risk factors on injury severity of different rider groups are offered in the following.

For time and lighting related variables, the variable of 22:00–06:59 is positively related to severe injury of group 1 (occupational riders aged under 55) and group 2 (non-occupational riders aged under 55), with the marginal effects of 0.086 and 0.166. This indicates that, for crashes occurring at night, the probability of riders suffering from severe injury increases 8.6% for group 1 and 16.6% for group 2, compared to the crashes at other time periods. This is to say that the time of 22:00–06:59 only affects the injury severity of young and middle riders, which may be due to older people traveling less by e-bikes at night. The variable of 19:00–21:59 is associated with a high likelihood of severe injury of group 5 (non-occupational riders aged above 55 in urban regions). The variable of daylight has opposite direction effects on injury outcomes of riders group 3 (riders aged above 55 in rural regions) and group 4 (occupational riders aged above 55 in urban regions), which strongly supported the existence of heterogeneous effects in traffic safety (Chang et al., 2019; Wang et al., 2017a; Wang, Huang, Xu, Xie, & Wong, 2020). Crashes occurring at dawn or dust are more likely to associate with severe injury for these occupational riders aged under 55. In addition, crashes occurring in summer have higher probabilities of severe injury for two rider groups (i.e., non-occupational riders aged under 55 and non-occupational riders aged above 55 in urban region).

With regard to the road condition, crashes occurring at high-speed roads (the high-speed roads refer to the roads with design speed higher than 50 km) are commonly associated with higher likelihoods of severe injury except for rider group 5. Specially, the global marginal effect of "high-speed road" is 0.109 for the entire crashes-involved riders, and separated marginal effects of this variable are from 0.092 to 0.241 for different rider groups. This means that, for crashes occurring at high-speed roads, the proba-

Table 4
Marginal effects of risk factors on injury severity of different rider groups.

Risk factors	General	Group1	Group2	Group3	Group4	Group5
Rider aged 35–44	0.065	–	–	–	–	–
Rider aged 45–54	0.108	–	–	–	–	–
Rider aged 55–64	0.195	–	–	–	–	–
Rider aged 65–74	0.265	–	–	–	–	–
Rider aged >=75	0.299	–	–	–	–	–
Non-occupational	0.087	–	–	–	–	–
Rider Living in urban	0.063	–	–	–	–	–
22:00–06:59	0.126	0.086	0.166			
19:00–21:59	0.052					0.183
Weekend					−0.138	
Daylight				0.107	−0.194	
Dawn/Dust		0.116				
Spring	−0.039					
Summer			0.059			0.123
High-speed road	0.109	0.206	0.092	0.141	0.241	
Having no improper behaviors of e-bike rider	−0.034			−0.120	−0.115	−0.121
While crossing the road, not to get off and push the e-bike		0.075				
Driver gender of opponent vehicle is female						−0.116
Type of opponent vehicle is truck	0.095		0.077	0.179		0.357
Driving on the non-motor vehicle lane						0.218
Reverse driving		0.251				
Other improper behaviors of driver		0.065	0.046			

"–"Rider attribute variables are included in the general model, while not included in the separated models since they have been considered in the preliminary tree analysis.

bility of riders suffering from severe injury increases on average 10.9% compared with that at the low-speed road, and the probability of injury severe increases from 9.2% to 24.1% for different groups. This result agrees with the founding of our previous study (Hu et al., 2020b) that the fatality risk of e-bike riders sharply increases when vehicle impact speed[1] exceeds 50 km/h.

Concerning the pre-crash improper behaviors of e-bike riders, the variable of "having no improper behaviors of e-bike riders" is related to 3.4% lower probability of severe injury for the entire crashes-involved riders; specifically, 12.0% lower probability for group 3, 11.5% lower probability for group 4, and 12.1% lower probability for group 5. This is to say, improper behavior of e-bike riders could significantly increase the probability of severe injury, especially for the older riders. In addition, for non-occupational riders aged under 55, the improper behavior of "not to get off and push the e-bike" is significantly associated with a higher probability of severe injury. Regarding the pre-crash improper behaviors of the opponent vehicle, several types of behaviors have significant effects on severe injury for specific rider groups: driving on the non-motor vehicle lane for non-occupational riders aged above 55 in urban regions, reverse driving for occupational riders aged under 55, and other improper behaviors of driving (such as drunk driving, fatigue driving) for riders aged under 55.

For crashes colliding with the truck, the likelihood of riders suffering from severe injury increases 9.5% on average; specifically, 7.7% for the group 2, 17.9% for group 3, and 35.7% for group 5. These results are similar to previous studies on crash injury of traditional bicycles and motorcycles (Chang et al., 2019; Prati et al., 2017). In addition, for non-occupational riders aged above 55 in urban regions, they are associated with a lower likelihood of severe injury when the driver of the opponent vehicle is female.

4. Discussions

4.1. Policy implication

The popularity of e-bikes in China and their users' vulnerability to get fatal/severe injuries in crashes make it critical to identify factors influencing the injury severity of e-bike riders in crashes so as to provide guides for targeted e-bike crash countermeasures. For this purpose, a classification tree-based logistic regression model is implemented. Compared with general conventional model, the proposed model generates more reliable results in the model goodness-of-fit. More importantly, the estimated results from this model help identify important contributing factors that would be hidden if the whole dataset is used, which is very important for safety improvements and policy development.

Specifically, from the perspective of injury severity, the classification tree is first used to split the e-bike riders into five homogeneous groups based on multiple riders attributes including rider's age, gender, rider's occupation, and living region. By the tree analysis we understand the question of "who are the high risk riders associated with high likelihoods of severe injury." The highest risk group is non-occupational riders aged above 55 in urban regions, and then are the group of riders aged above 55 in rural regions and the group of occupational riders aged above 55 in urban regions. This implies that older riders should be considered a top priority for preventing fatal/severe e-bike crashes.

Separated logistic regression model by five homogeneous groups is then used to examine the differences of the contributing factors (such as rider behaviors, road type, weather) affecting crash injury severity of different e-bike riders groups. By logistic regression analysis we solve the question of "how to reduce the crash injury for different types of e-bike riders, especially for high risk riders." For example, the variable of "having no improper behaviors of e-bike rider" only has significant effects on injury severe for old rider groups. This is to say that the policy targeted at the prevention of e-bike rider risky behaviors is effective for reducing older riders suffering from fatal/severe injury in crashes. Main policy implications for e-bike safety improvement are recommended as follows.

The first implication involves the e-bike rider attributes and improper behaviors of e-bike riders. From the results of the tree classification, we find that riders aged 55 and above are associated with a higher likelihood of severe injury. Furthermore, logistical analysis results show that improper behaviors of e-bike riders could significantly increase the probability of severe injury, especially for the old riders (group3, group4, and group5). These results give us an important implication: if the rate of e-bike riders' improper behaviors could be reduced or controlled successfully, then the rate of serious injuries and fatalities would be reduced

[1] Vehicle impact speed refers to the vehicle speed at the moment of the vehicle contacting with the e-bike in the collision (Hu et al., 2020b).

accordingly--this is especially effective for the older riders. In addition, although occupational riders are associated with a relatively lower high likelihood of severe injury compared to non-occupational riders, the safety of occupational e-bike riders (i.e., delivery-e-bikes and taxi-e-bikes) still cannot be ignored because the number of occupational riders involved in the crashes account for a large proportion among the entire e-bikes crashes. Practical experiences in some developed countries have demonstrated the benefit of education and training systems to the alleviation of power-two-wheeler crash risks (Baldi, Baer, & Cook, 2005; Vlahogianni, Yannis, & Golias, 2012). Normalized education and licensing systems especially for high risk e-bike rider groups are recommend to increase awareness of unsafe behavior, encouraging the rider to behave safely. In addition, the prevention of improper/risky behaviors should be not only for e-bike riders but also for drivers, such as the driver behaviors of driving on the non-motor vehicle lane.

The second implication involves the important role of road and traffic engineering in enhancing e-bike safety. E-bike crashes occurring on high speed roads (refers to roads that of the design speed higher than 50 m/h) are associated with higher likelihoods of injury severity. In this study, we find that, compared to crashes on low speed road, the probability of severe injury crashes on high speed road has increased from 9.2% to 24.1% for different rider groups. In reality, our previous study (Hu et al., 2020b) has examined the relationship between the impact speed and injury severity of e-bike riders – the fatality risk of riders is approximately 2.9% at the vehicle impact speed of 30 km/h, 23% at 50 km/h, 50% at 60 km/h, and 90% at 80 km/h. Results of these two studies strongly imply that there is an urgent need to either regulate the use of e-bikes on high speed roads or separate e-bikes (e.g., by a physical barrier) from high-speed motor vehicles.

Lastly, without penalties and stricter enforcement for e-bikes improper/risky behaviors, the goal to effectively curb e-bike crashes is hard to reach (Bai, Liu, Chen, Zhang, & Wang, 2013). A comprehensive e-bike safety treatment should combine perfect laws and regulations for the use of e-bikes, sticker law enforcement for the rider and driver's risky behaviors, traffic engineering countermeasures, and safety education campaigns.

4.2. Study limitation

The limitation of our study is related to the source of the data: a police-based register of crashes. This database does not include detailed road characteristics (e.g., median width, number of lanes), detailed e-bike related facilities (e.g., size and mass of e-bike), physical/psychological status of riders and pre-crash traffic flow characteristics, though these factors are deemed important, their effects on injury severity cannot be analyzed. Thus, an extension of this paper is to incorporate police-reported data and other data sources (such as questionnaire surveys and field observations) to achieve a more comprehensive understanding of factors on injury severity of e-bikes.

Acknowledgments

This work was supported by the Key Research and Development Program of Hunan Province (No. 2019SK2171), the National Natural Science Foundation of China (No. 51678075), Hunan Province Key Laboratory of Smart Roadway and Cooperative Vehicle-Infrastructure Systems (No. 2017TP1016), Open Fund of Ministry of Education Engineering Research Center of Catastrophic Prophylaxis and Treatment of Road & Traffic Safety (Changsha University of Science & Technology) under research grant kfj130301 and Open Fund of Ministry of the Key Laboratory of Highway Engineering of Ministry of Education (Changsha University of Science & Technology) under research grant kfj190202.

References

Bai, L., Liu, P., Chen, Y., Zhang, X., & Wang, W. (2013). Comparative analysis of the safety effects of electric bikes at signalized intersections. *Transportation Research Part E: Logistics and Transportation Review, 20*, 48–54.

Baldi, S., Baer, J. D., & Cook, A. L. (2005). Identifying best practices states in motorcycle rider education and licensing. *Journal of Safety Research, 36*, 19–32.

Chang, F., Xu, P., Zhou, H., Chan, A. H. S., & Huang, H. (2019). Investigating injury severities of motorcycle riders: A two-step method integrating latent class cluster analysis and random parameters logit model. *Accident Analysis & Prevention, 131*, 316–326.

Chung, Y., Song, T. J., & Yoon, B. J. (2014). Injury severity in delivery-motorcycle to vehicle crashes in the Seoul metropolitan area. *Accident Analysis & Prevention, 62*, 79–86.

Du, W., Yang, J., Powis, B., Zheng, X., Ozanne-Smith, J., Bilston, L., & Wu, M. (2013). Understanding on-road practices of electric bike riders: An observational study in a developed city of China. *Accident Analysis & Prevention, 59*, 319–326.

Huang, H., Peng, Y., Wang, J., Luo, Q., & Li, X. (2018). Interactive risk analysis on crash injury severity at a mountainous freeway with tunnel groups in China. *Accident Analysis & Prevention, 111*, 56–62.

Hu, F., Lv, D., Zhu, J., & Fang, J. (2014). Related risk factors for injury severity of E-bike and bicycle vrashes in Hefei. *Traffic Injury Prevention, 15*(3), 319–323.

Hu, L., Hu, X., Wan, J., Chen, Y., & Cao, D. (2020a). The injury epidemiology of adult riders in vehicle-two-wheeler crashes in China, Ningbo, 2011–2015. *Journal of Safety Research, 72*, 21–28.

Hu, L., Hu, X., Wang, J., Kuang, A. W., Hao, W., & Lin, M. (2020b). Casualty risk of e-bike rider struck by passenger vehicle using China in-depth accident data. *Traffic Injury Prevention, 21*(4), 283–287.

Jung, S., Qin, X., & Oh, C. (2016). Improving strategic policies for pedestrian safety enhancement using classification tree modeling. *Transportation Research Part A: Policy and Practice, 85*, 53–64.

Kim, J. K., Kim, S., Ulfarsson, G. F., & Porrello, L. A. (2007). Bicyclist injury severities in bicycle–motor vehicle accidents. *Accident Analysis & Prevention, 39*, 238–251.

Li, Y., Xing, L., Wang, W., Liang, M., & Wang, H. (2018). Evaluating the impact of Mobike on automobile-involved bicycle crashes at the road network level. *Accident Analysis & Prevention, 112*, 69–76.

National Bureau of Statistics (2015). *China road traffic accident statistics 2014*. National Bureau of Statistics of China.

Prati, G., Pietrantoni, L., & Fraboni, F. (2017). Using data mining techniques to predict the severity of bicycle crashes. *Accident Analysis & Prevention, 101*, 44–54.

Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis & Prevention, 43*(5), 1666–1675.

Vlahogianni, E. I., Yannis, G., & Golias, J. (2012). Overview of critical risk factors in Power-Two-Wheeler safety. *Accident Analysis & Prevention, 49*, 12–22.

World Health Organization (2017). *Powered two- and three-wheeler safety: A road safety manual for decision-makers and practitioners*. World Health Organization.

Wu, C. X., Yao, L., & Zhang, K. (2012). The red light running behavior of electric bike riders and cyclists at urban intersections in China: An observational study. *Accident Analysis & Prevention, 49*(11), 186–192.

Wu, C., & Loo, B. (2016). Motorcycle safety among motorcycle taxi drivers and nonoccupational motorcyclists in developing countries: A case study of Maoming, South China. *Traffic Injury Prevention, 17*, 170–175.

Wang, C., Xu, C., Xia, J., & Qian, Z. (2017a). Modelling faults among e-bike related fatal crashes in China. *Traffic Injury Prevention, 18*(2), 175–181.

Wang, J., Huang, H., & Zeng, Q. (2017b). The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. *Accident Analysis & Prevention, 98*, 223–231.

Wang, J., Huang, H., Xu, P., Xie, S., & Wong, S. C. (2020). Random parameter probit analysis of pedestrian red-light violations and injury severity in pedestrian-motor vehicle crashes at signalized crossings. *Journal of Transportation Safety and Security, 12*, 818–837.

Zeng, Q., Wen, H., & Huang, H. (2016). The interactive effect on injury severity of driver-vehicle units in two-vehicle crashes. *Journal of Safety Research, 59*, 105–111.

Yang, X. B., Mei, H., Abdel-Aty, M., Peng, T. C., & Gao, Z. Y. (2015). A hazard-based duration model for analyzing crossing behavior of cyclists and electric bike riders at signalized intersections. *Accident Analysis & Prevention, 74*, 33–41.

Yan, X. P., Ma, M., Huang, H. H., Abdel-Aty, M., & Wu, C. Z. (2011). Motor vehicle–bicycle crashes in Beijing: Irregular maneuvers, crash patterns, and injury severity. *Accident Analysis & Prevention, 43*, 1751–1758.

Zhengwu Wang is currently a professor at Changsha University of Science and Technology, China. His research interests refer to traffic safety, transportation planning and management.

Shuai Huang is a PhD candidate in the School of Traffic and Transportation Engineering, Changsha University of Science and Technology. His research interest is road safety analysis.

Jie Wang is currently a post-doctoral fellow in the Key Laboratory of Highway Engineering of Ministry of Education, Changsha University of Science and Technology. He received his PhD degree at Central South University. His research interests include traffic safety, road user behavior, and econometric models for transport data analysis.

Denisa Sulaj is a master candidate in the School of Traffic and Transportation Engineering, Changsha University of Science and Technology. Her research interest is traffic safety statistical analysis.

Wei Hao is currently a professor at Changsha University of Science and Technology. His research interests refer to traffic safety, connected and automated vehicle controlling and management.

Aiwu Kuang is currently a associate professor at Changsha University of Science and Technology, China. His research interests refer to transportation planning and management.