Traffic & Transportation

Exploring Factors Influencing Injury Severity of Bicyclists in Crashes with Motor Vehicles – Combined Latent Class Clustering and MNL Model

Peikun LI[1], Jingyu DING[2], Quantao YANG[3], Wenbo LU[4]

Original Scientific Paper
Submitted: 7 Feb 2025
Accepted: 13 May 2025
Published: 30 Mar 2026

[1] lipeikun@imau.edu.cn, College of Energy and Transportation Engineering, Inner Mongolia Agricultural University, Hohhot, China
[2] dingjingyu613@163.com, School of Traffic and Transportation Engineering, Dalian Jiaotong University, Dalian, China
[3] Corresponding author, yqt@chd.edu.cn, Department of Public Security, Shaanxi Police College, Xi'an, China
[4] wenbo1996@seu.edu.cn, School of Transportation, Southeast University, Nanjing, China

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT
With the increase in travel costs brought about by motor vehicles and the enhancement of environmental awareness among travellers, the rise in bicycle usage has grown globally. However, cyclists are more vulnerable in the process of a crash with motor vehicles, and tend to have more serious injuries compared to motorists after a crash. Therefore, cycling safety needs to be taken seriously, and it is crucial to analyse the factors that influence the severity of injuries sustained by cyclists in collisions. In this paper, we investigated the bicycle-motor vehicle (BMV) crash statistics in North Carolina, U.S.A., conducted preliminary screening of the variables covered in the data, classified the data into clusters by applying the latent cluster analysis, and used the multinomial logit (MNL) model based on the latent category to explore the main influences on the severity of rider injuries in BMV crashes in each cluster and the mechanism of the factors' effects. The results of this paper explain the contribution of the influencing factors to the severity of cyclists' injuries under each cluster characteristic. The results can be used for the special management and optimisation of the causative factors affecting the severity of cyclists' injuries under different cluster categories.

KEYWORDS
bicycle riders; injury severity; latent class cluster analysis; MNL model; heterogeneity.

1. INTRODUCTION

As one of the low-carbon travel modes, bicycling has become a popular choice for people to travel due to its convenience, economy, health, simplicity and practicality [1, 2]. The number of cyclists has increased significantly in most countries around the world over the last decade [3]. However, it is concerning that the number of road traffic crashes involving bicycles has been increasing in recent years.

Bicycle crashes can be classified into single-vehicle crashes and bicycle-motor vehicle (BMV) crashes. Single-vehicle crashes are related to the bicycle itself (e.g. brake failure and poor vehicle manoeuvrability) and to infrastructure in the driving environment (e.g. slippery road surfaces and collisions with stationary objects) [4]. Compared to single-vehicle collisions, BMV crashes are more serious, and concerning the road safety report published by the International Transport Forum in 2022, the fatality rate of pedestrian-motor vehicle crashes decreased significantly from 2015 to 2021, but the fatality rate of BMV crashes showed an increasing trend [3]. As the number of cyclists continues to increase, it is crucial to ensure their safety [5]. Research on factors influencing BMV crashes needs to include not only various exogenous factors (e.g. age of the cyclist, gender of the cyclist, whether the driver is driving under the influence of alcohol, and whether the cyclist is wearing a helmet while riding) but also unobserved factors (i.e. heterogeneity of influencing factors).

Heterogeneity can be reflected in unobserved vehicle characteristics, rider characteristics, environmental characteristics, roadway characteristics, etc. [6]. Current models to study heterogeneity include the random parameter logit model [7], the latent category with random parameters [8] and the latent category model [9]. Unlike other models that account for individual-level heterogeneity in crashes, latent category models can account for population heterogeneity without assuming the form of parameter distributions in advance. Latent class clustering can uncover the impact of unobserved heterogeneity characteristics on accident severity, thereby enabling better categorisation and research of accident types [10].

The purpose of this paper is to analyse the severity of rider injuries in BMV crashes and to analyse the heterogeneity among the influencing factors through the potential categories in the potential category model, which will be more effective in deriving the main factors affecting the severity of rider injuries. Thus, theoretical support is proposed for reducing the risk of crashes, which is an important guide for the development of effective policy measures to improve cyclist safety.

This study provides two key innovations in BMV crash analysis. First, we introduce a latent class analysis combined with multinomial logit (MNL) modelling to address unobserved heterogeneity in crash severity factors. Our framework autonomously identifies three distinct crash clusters through environmental feature patterns. Second, this study refines prior conclusions and offers distinct insights by revealing the heterogeneous impacts of gender, environment and geography on BMV severity, and underscores the necessity of developing context-specific risk mitigation strategies.

2. LITERATURE REVIEW

Scholars in the field of transportation have studied the causes of BMV crashes, such as riders not wearing helmets, motorists not wearing seat belts, drunk driving and the driver's age, which can lead to collisions [11–13]. Overall, it can be summarised into five areas: rider characteristics, driver characteristics, environmental characteristics, roadway characteristics and management systems.

As far as cyclists are concerned, there are no regulations restricting the age of cyclists, and no driver's licences are issued for cyclists, so the age of a cyclist on the road can be lower than 16 years or higher than 70 years. Surveys have shown that older people are more likely to be involved in motor vehicle-bicycle collisions and are more likely to experience more serious crash consequences [14]. Drunk driving is also a cause of traffic crashes, and serious injuries are common among patients who are involved in collisions while riding under the influence of alcohol, with a predominance of serious head injuries [15, 16]. Head injuries sustained by cyclists in crashes are highly susceptible to serious crash consequences, and it has been shown that cyclists wearing helmets can reduce fatal injuries from collisions [17].

As far as drivers are concerned, driver overtaking is one of the main factors that lead to traffic crashes. According to statistics, 65% of road traffic crashes are due to speeding and overtaking [18]. The probability of fatal injuries is significantly higher when drivers are intoxicated. Salas-Wright et al. investigated the severity of crashes caused by drunk driving and came to the same conclusion, showing that alcohol consumption has an impact on the severity of injuries sustained by cyclists [15, 16, 19]. Driver gender has been an important influencing factor in research concerns. Male drivers, who drive more aggressively, are more likely to be involved in crashes, such as accelerating through yellow lights [20], but are not as likely to cause serious crashes as female drivers because male drivers tend to respond to emergencies more quickly than female drivers [21].

For road features, when road conditions are complex, the cyclist's ability to perceive safety will be weakened, making it harder to judge a safe driving speed, and motorised vehicles will also be less able to judge a safe distance, which will also lead to collisions between the bicycle and the vehicle [22]. Intersections are the places where vehicles converge, but it has been investigated that vehicles travelling to intersections pay extra attention to the road conditions around the roadway, and instead, the probability of traffic crashes is higher in non-intersection roadways [23].

Environmental characteristics can also have an impact on motor vehicle-bicycle crashes. Nawaf et al. also examined 3 years of BMV data in Florida and verified that crashes occurring at night cause more severe injuries than during the day [11, 23]. The probability of vehicle collisions on rural roads is higher than on urban roads, where there are stricter regulations and management systems, and therefore drivers do not violate the rules as much [22, 24]. Rural roads are poorly planned, and GPS can be biased in planning roads, and drivers can be involved in scrapes as a result [24]. In addition, the roads are prone to frost due to seasonal influences, which can lead to vehicle collisions due to skidding and uncontrolled speeds, and the cyclist is the most vulnerable party in a crash [4, 25, 26].

At this stage, there are many ways to study the factors affecting BMV crashes. Algurén and Rizzi used univariate statistics and Pearson chi-square tests to analyse the effects of gender and age on very injured persons [1]. Logistic regression analysis is also commonly used to analyse the effects of factors on injured persons [12, 18]. Hierarchical Bayesian estimation of Poisson lognormal random effects models can look for associations between environmental factors and crashes, thus studying the spatial correlation of crashes [27]. However, there is a large number of influencing factors that lead to a crash, and most studies do not take into account that unobserved potential influencing factors (i.e. heterogeneity among influencing factors) among the influencing factors may also have an impact on the occurrence of crashes [10]. Therefore, it is also necessary to study the effect of heterogeneity on crashes. Current research methods that take heterogeneity into account include stochastic parametric MNL models [21], mixed logit models [7], ordered logit models [28] and latent class analysis [28, 29]. Among them, latent category analysis has the advantage of explaining group heterogeneity without assuming the form of parameter distributions in advance and illustrating the possibility of common parameters among unobserved observation groups. Therefore, in this paper, latent class analysis and the MNL model were chosen as analytical models to investigate the main influencing factors affecting cyclists' injury severity.

In general, domestic and international researchers have conducted detailed studies on the influencing factors of BMV crashes. Traffic accident data need to be analysed for the impact factors in different regions, to study the association between collisions and the impact factors in the region, and to identify and correct the safety hazards in the infrastructure and management system. In this paper, the application of a latent class analysis model allows for a finer delineation of the severity of the impact factors in different environmental characteristics, providing more effective recommendations for localised governance.

3. DATA PREPARATION

The data analysed in this paper come from the North Carolina Department of Transportation's published Accident Statistics Library, which covers the years 2007–2019 with over 10,000 data entries. The data were selected to include the influencing factors such as the age group of the rider, whether the rider was driving under the influence of alcohol, the direction in which the rider was riding, the gender of the driver and the roadway environment, as shown in *Table 1*. Based on the crash records, the severity of cyclists' injuries was categorised into three groups: slight crash (53.16%), serious crash (44.40%) and fatal crash (2.44%).

As anticipated, there are numerous factors influencing accident severity, among which the complex interactions of correlated factors constitute a key research focus. Unobserved data heterogeneity may potentially interfere with the study findings [10]. *Table 1* contains the 12 impact factors from the North Carolina data, with data volume statistics for each impact factor, as well as showing the ratio of the severity caused by the impact factors in each subcategory to the total number of impact factor statistics for that broad category. Before model training, each influencing factor was digitised in terms of its rider's gender, where 1 represents male and 2 represents female.

Table 1 – Number and percentage of each influencing factor

Variables	Interpretation	Category	Slight		Serious		Fatal	
			frequency	proportion	frequency	proportion	frequency	proportion
Total	Total crash	——	5639	100%	4902	100%	260	100%
MV-Gender	Driver gender	1= Male	3095	54.87%	2829	57.71%	176	67.69%
		2= Female	2545	45.13%	2073	42.29%	84	32.31%
B-Gender	Rider gender	1= Male	4821	85.49%	4107	83.78%	230	88.46%
		2= Female	818	14.51%	795	16.22%	30	11.54%
Directions	Riding direction	1= With traffic	4117	73.01%	3221	65.71%	223	85.77%
		2= Facing traffic	1522	26.99%	1681	34.29%	37	14.23%

Variables	Interpretation	Category	Slight		Serious		Fatal	
			frequency	proportion	frequency	proportion	frequency	proportion
Weather	Weather at the time of the crash	1= Clear	4728	83.84%	3999	81.58%	208	80.00%
		2= Cloudy	676	11.99%	654	13.34%	39	15.00%
		3= Fog, smog, smoke	7	0.12%	8	0.16%	5	1.92%
		4= Other	5	0.09%	4	0.08%	0	0.00%
		5= Rain	217	3.85%	229	4.67%	7	2.69%
		6= Snow, sleet, hail, freezing rain/drizzle	6	0.11%	8	0.16%	1	0.38%
Urban_Rural	Crash area attribute	1= Rural	1630	28.91%	1278	26.07%	146	56.15%
		2= Urban	4009	71.09%	3624	73.93%	114	43.85%
Road condition	Crash road environmental conditions	1= Dry	5222	92.61%	4481	91.41%	236	90.77%
		2= Ice	3	0.05%	2	0.04%	0	0.00%
		3= Sand, mud, dirt, gravel	2	0.04%	3	0.06%	0	0.00%
		4= Snow	3	0.05%	3	0.06%	0	0.00%
		5= Water (standing, moving)	11	0.20%	14	0.29%	1	0.38%
		6= Wet	398	7.06%	399	8.14%	23	8.85%
Road surface	Road surface conditions where the crash occurred	1=Coarse asphalt	1869	33.14%	1616	32.97%	101	38.85%
		2=Concrete	131	2.32%	114	2.33%	2	0.77%
		3=Gravel	11	0.20%	9	0.18%	0	0.00%
		4=Grooved concrete	32	0.57%	27	0.55%	1	0.38%
		5=Other	5	0.09%	6	0.12%	0	0.00%
		6=Sand	2	0.04%	1	0.02%	0	0.00%
		7=Smooth asphalt	3587	63.61%	3125	63.75%	156	60.00%
		8=Soil	2	0.04%	4	0.08%	0	0.00%
Crash location	Crash location	1= Intersection	2895	51.34%	2648	54.02%	67	25.77%
		2= Non-intersection	2744	48.66%	2254	45.98%	193	74.23%
MV-DWI (driving while intoxicated)	Whether the motor vehicle driver was driving under the influence of alcohol	1= No	5548	98.39%	4836	98.65%	238	91.54%
		2= Yes	91	1.61%	66	1.35%	22	8.46%
B-DWI (driving while intoxicated)	Whether the cyclist was riding after drinking	1= No	5328	94.48%	4627	94.39%	208	80.00%
		2= Yes	311	5.52%	275	5.61%	52	20.00%

4. METHODOLOGY

From the selected data, each influencing factor is a nonlinear variable, but there may be potential covariance between the data, so the data need to be analysed for heterogeneity, then clustered and grouped to reduce the interference of potential associations between the influencing factors on the results of the study. In this study, a latent category analysis model was used to perform a heterogeneity analysis of the data, and then the categorised clusters were further examined with an MNL model to investigate the influencing factors affecting the severity of injuries in cyclists.

4.1 Latent class analysis

Latent class model includes latent profile analysis, which is used to analyse continuous experimental data, and latent class analysis, which is used to analyse discrete experimental data. Latent class analysis divides the observed data into mutually exclusive categories, which are used to explain the relationships among the exogenous variables [30]. By subjecting the data to latent category cluster analysis, the data were analysed by fitting the goodness-of-fit metrics Akaike information criterion (AIC), Bayesian information criterion (BIC), adjusted Bayesian information criterion (aBIC), Lo-Mendell-Rubin likelihood ratio test (LMRT), bootstrap likelihood ratio test (BLRT) and entropy information to obtain the optimal number of clusters.

The definitions used when categorising exogenous variables using the latent category analysis model are as follows:

$$P\left(y_i\right)=\sum_{m=1}^{M}\gamma_m\prod_{j=1}^{J}\prod_{a_j=1}^{A_j}\rho_{j,a_{j|m}}^{I\left(y_{ij}=a_j\right)} \quad (1)$$

Assuming that the crash dataset has M latent category clusters, γ_m denotes the probability that a crash is in latent category cluster m (m = 1, 2, ⋯, M), which is called the latent category probability. Each crash *i* contains *J* attributes, i.e. *J* epistemic variables. y_{ij} denotes the level number of the *j* th categorical variable for the *j* th crash, $y_{ij} = 1, 2, \cdots; a_i$. $\rho_{j,a_{j|m}}$ denotes the probability that the level number of the jth variable in cluster m is a_j, called the conditional probability. The conditional probability of each exogenous variable sums to 1; a_j is the total number of levels of the jth exogenous variable; and $I\left(y_{ij}=a_j\right)$ is an indicator function, which is equal to 1 when $y_{ij}=a_j$; otherwise, it is equal to 0 [28].

4.2 Multinomial logit model

The multinomial logit regression is appropriate for scenarios where the dependent variable *Y* is an unordered polytomous variable [10]. Assuming the dependent variable *Y* contains *J* mutually exclusive categories (*j*=1, 2, ..., *J*), its conditional probability formula is expressed as:

$$P(Y=j/\mathbf{X})=\frac{e^{\mathbf{X}^T\beta_j}}{1+\sum_{k=1}^{J-1}e^{\mathbf{X}^T\beta_k}},\ j=1,2,\ldots,J-1 \quad (2)$$

The probability for the baseline category (typically specified as *Y*=*J*) is given by:

$$P(Y=J/\mathbf{X})=\frac{1}{1+\sum_{k=1}^{J-1}e^{\mathbf{X}^T\beta_k}} \quad (3)$$

where *Y* is the unordered polytomous dependent variable (taking values 1, 2, ..., *J*), **X** is the vector of independent variables, β_j is the regression coefficient vector for the *j*-th category, *J* is the total number of categories of the dependent variable, and β_J is the baseline category coefficient (β_J defaults to 0).

The coefficients β_j are estimated using the maximum likelihood estimation method, with the likelihood function given by:

$$L(\beta)=\prod_{i=1}^{n}\prod_{j=1}^{J}\left[P(Y_i=j/\mathbf{X}_i)\right]^{I(Y_i=j)} \quad (4)$$

where $I(Y_i=j)$ is the indicator function (taking the value 1 if $Y_i=j$, and 0 otherwise).

5. RESULTS AND DISCUSSION

5.1 Cluster results

The influencing factors involved in *Table 1* were entered into the model as variables for potential category clustering analysis, and the estimation of the BMV crashes classification model was obtained by testing different numbers of clusters (1~10). The results of the analysis are shown in *Figure 1*, which includes the LMR test and BLR test, and when the P-value of both is less than 0.0001, it indicates that the result is significant; smaller values of AIC, BIC and aBIC, along with entropy values higher than 0.9, indicate better results [30].

From the values of AIC, BIC, aBIC and entropy, it shows that the 10th cluster is the optimal cluster. However, the values of LMRT and BLRT show that when the number of clusters is 9, the value of LMRT is no longer less than 0.0001, which means that 8 clusters have better results than 9 clusters. In terms of classification share, the classification is more even when the number of clusters is less than 3, and some of the classifications in other clusters have a share of only 7.5%. Combining the judgment of each index, 3 cluster models are selected as the optimal cluster models, and the BMV crash data are classified into 3 clusters in this paper.

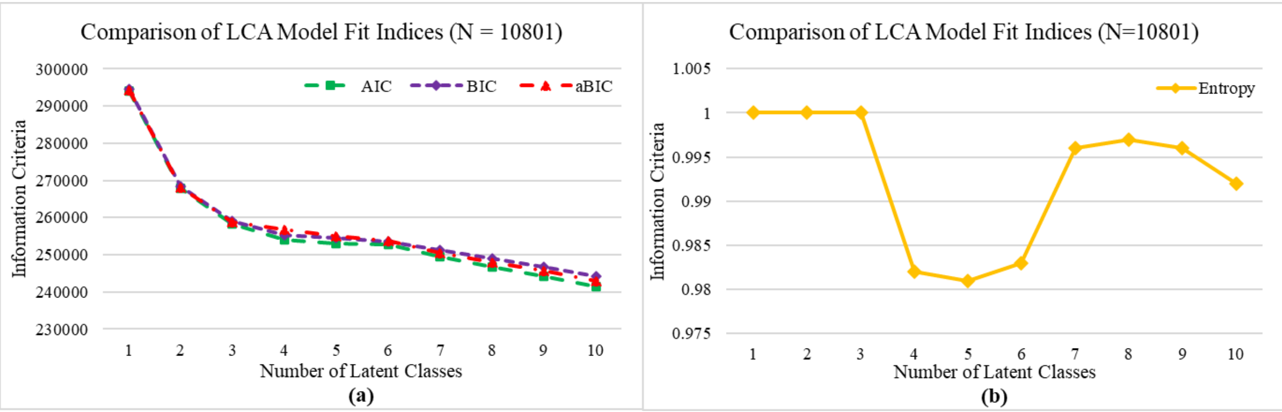

Figure 1 – Comparison of fitted metrics for potential category analysis models: (a) Trends of AIC, BIC and aBIC; (b) Trends of entropy

As shown in *Figure 1*, the horizontal coordinates of the table represent the number of classified clusters, and the right side of the table is the entropy value, which takes values between 0 and 1. When the value is closer to 1, the model's classification is more accurate. When the entropy value is 0.6, it means that about 20% of the individuals have classification errors; when the entropy value is 0.8, it means that the classification accuracy is more than 90%. From the data in the table, it can be seen that the accuracy rate of this classification is more than 90%, which means that the results of this classification are very accurate. The coordinates on the left side of the table are the values of AIC, BIC and aBIC, and the smaller values of these three indicators indicate that the model fits better. However, the values of the three metrics in *Figure 1* tend to stabilise after being divided into three clusters; therefore, dividing the three datasets into three clusters is the best classification method.

The sample sizes and percentages of the three clusters are shown in *Table 2*, 75.69% of the crashes in Cluster 1 occurred on urban roads, 99.25% of the crashes occurred on slippery roads, and the percentages of crashes on cloudy and rainy days were 39.22% and 54.89%, respectively, which can be defined as "BMV crashes on slippery roads in cloudy and rainy cities". Similarly, Cluster 2 can be defined as "Crashes occurring in dry road environments at rural non-intersections on sunny days". Cluster 3 can be defined as "Crashes occurring in dry road environments in urban areas on sunny days".

Table 2 – Cluster sample size and distribution of each variable

Variables	Cluster 1		Cluster 2		Cluster 3	
	Frequency	Occupy%	Frequency	Occupy%	Frequency	Occupy%
Number of incidents	798	100.00%	2656	100.00%	7171	100.00%
Slight injury crashes	412	51.63%	1024	38.55%	3847	53.65%
Serious crash	367	45.99%	1493	56.21%	3227	45.00%
Fatal crash	19	2.38%	139	5.23%	97	1.35%
Crash at an intersection	439	55.01%	752	28.31%	4339	60.51%
Crash occurred at a non-intersection	359	44.99%	1904	**71.69%**	2832	39.49%
Rural	194	24.31%	2649	**99.74%**	156	2.18%
City	604	**75.69%**	7	0.26%	7015	**97.82%**
Riders are male	705	88.35%	2266	85.32%	6040	84.23%
Riders are female	93	11.65%	390	14.68%	1131	15.77%
Rider not driving under the influence	731	91.60%	2434	91.64%	6833	95.29%
Bikers driving while intoxicated	67	8.40%	222	8.36%	338	4.71%
Riders travelling in the same direction	568	71.18%	2282	85.92%	4585	63.94%
Riders travelling in the opposite direction	230	28.82%	374	14.08%	2586	36.06%
Drivers are male	456	57.14%	1571	59.15%	3746	52.24%
Drivers are female	342	42.86%	1085	40.85%	3425	40.38%
The driver was not drunk	778	97.49%	2603	98.00%	7072	98.62%
The driver was drunk	20	2.51%	53	2.00%	99	1.38%
Road environment – dry	6	0.75%	2654	**99.92%**	7165	**99.92%**
Road environment – slippery	792	**99.25%**	2	0.08%	6	0.08%
Road surface – rough asphalt	272	34.09%	1178	44.35%	2112	29.45%
Road surface – concrete pavement	23	2.88%	7	0.26%	217	3.03%
Road surface – smooth asphalt	503	63.03%	1471	55.38%	4843	67.54%
Sunny	47	5.89%	2410	**90.74%**	6369	**88.82%**
Cloudy	313	**39.22%**	245	9.22%	802	11.18%
Rainy	438	**54.89%**	1	0.04%	0	0.00%

5.2 MNL model results

The estimation results of the MNL model are displayed as shown in *Tables 3*, *4* and *5*. The data include four influential features, such as rider characteristics, driver characteristics, road characteristics and environmental characteristics, and the data of slight injury crashes in each cluster are compared with the serious crashes and fatal crashes, respectively, to obtain the corresponding B-value, Exp(B) and significance of each cluster variable. Where B is the regression coefficient, a quantitative representation of the degree to which the independent variable affects the dependent variable. The change in B represents how much the dependent variable is expected to change when the independent variable is increased by one unit. A positive value of B indicates a positive effect of the independent variable on the dependent variable, and vice versa indicates a negative effect. Exp (B) value is the logarithm of the B value and refers to the dominance ratio. The dominance ratio is used as an indicator of the size of the effect and measures the magnitude of the effect of a particular independent variable on the dependent variable. Therefore, the effect of the independent variable on the dependent variable can be determined by looking at the B-value and Exp(B) value. The smaller the AIC and BIC values of the goodness-of-fit indicators obtained by the model, the better the model's fitting performance. The results show that the AIC in Cluster 1 is 1082.973 and 126.901; the BIC is 1166.818 and 200.091, the significance is 0.024 and 0.0001, which are less than 0.05, the cluster is classified accurately, and the results are reliable. The AIC in Cluster 2 is 3397.637, 799.321; the BIC is 3502.592, 890.379; the significance is 0.005 and 0.0001, which are less than 0.05; the cluster is classified accurately, and the results are reliable. In Cluster 3, AIC was 9635.877 and 818.136; BIC was 9752.568 and 924.895; the significance was 0.0001 and 0.0001, which were less than 0.05, respectively, and the cluster was categorised accurately, and the results were reliable.

The results of the MNL model indicate that drivers operating motor vehicles in Cluster 1 who are female and riders travelling in opposite directions to motor vehicles have a more significant effect on crashes, and that crashes are more likely to have fatal outcomes when the driver is female, relative to slight injury crashes. Cluster 2 cyclists driving under the influence of alcohol, cyclists travelling in the opposite direction, drivers driving under the influence of alcohol, drivers being female, and motor vehicles and bicycles travelling in non-intersections all had a significant effect on crashes. The four influences of travelling in a non-intersection, driver drunk driving, driver being female, and rider drunk driving were more likely to result in fatalities relative to slight injury crashes. Rider drunk driving, rider travelling in the opposite direction, driver drunk driving, driver being female, being on city streets, and travelling at a non-intersection all had a significant effect on crashes in Cluster 3. Relative to slight injury crashes, rider drunk driving, rider travelling in the opposite direction, driver drunk driving, and travelling at a non-intersection all resulted in more serious crashes.

Table 3 – Demonstration of B-value, Exp(B) and significance of model variables in Cluster 1

Categories	Variables	Serious crash			Fatal crash		
		B	Exp(B)	Significance	B	Exp(B)	Significance
Cyclist characteristics	Female	-0.082	0.921	0.72	-17.772	0.0001	0.997
	Driving under the influence	0.032	1.032	0.909	0.948	2.581	0.163
	Opposite direction of travel	-0.406	0.666	**0.015**	0.147	1.158	0.81
Driver characteristics	Driving under the influence	0.427	1.532	0.43	2.941	18.926	**0.001**
	Female	-0.442	0.643	**0.003**	-1.517	0.219	**0.032**
Road characteristics	Wetness	0.479	1.614	0.589	16.066	9494527.93	0.999
	Concrete	-0.076	0.927	0.865	-17.122	0.0001	0.999
	Smooth pitch	-0.156	0.856	0.32	-0.69	0.501	0.229
Environmental characteristics	Non-intersections	0.103	1.108	0.503	1.926	6.862	**0.012**
	City	-0.218	0.804	0.229	-0.604	0.546	0.343
	Cloudy	-0.243	0.784	0.455	0.02	1.02	0.987
	Rainy	-0.354	0.702	0.266	-0.805	0.447	0.53
Constant		568	0.284	1.328	0.774	-19.812	0.0001

Note: No/possible injuries are set as the reference category; - indicates not significant at the 95% confidence interval or not included in the calculation; the result is considered significant at significance (p-value) < 0.05.

Table 4 – Demonstration of B-value, Exp(B) and significance of model variables in Cluster 2

Categories	Variables	Serious crash			Fatal crash		
		B	Exp(B)	Significance	B	Exp(B)	Significance
Cyclist characteristics	Female	0.114	1.121	0.327	-0.069	0.933	0.813
	Driving under the influence	-0.106	0.899	0.5	0.731	2.077	**0.005**
	Opposite direction of travel	-0.331	0.718	**0.005**	-0.46	0.631	0.126
Driver characteristics	Driving under the influence	0.921	2.513	**0.013**	1.705	5.504	**0.001**
	Female	0.02	1.02	0.809	-0.602	0.548	**0.004**
Road characteristics	Wetness	21.181	1.58E+09	1	22.769	7734700727	1
	Concrete	-0.834	0.434	0.285	-19.363	0.0001	0.999
	Smooth pitch	0.076	1.079	0.36	0.013	1.013	0.946
Environmental characteristics	Non-intersections	0.158	1.171	0.081	0.985	2.677	**0.0001**
	City	-0.384	0.681	0.636	-21.31	0.0001	0.999
	Cloudy	-0.169	0.845	0.23	0.116	1.123	0.707
	Rainy	-21.287	0.0001	1	-17.405	0.0001	1
Constant		568	0.36	1.433	0.025	-3.247	0.039

Note: No/possible injuries are set as the reference category; - indicates not significant at the 95% confidence interval or not included in the calculation; the result is considered significant at significance (p-value) < 0.05.

Table 5 – Demonstration of B-value, Exp(B) and significance of model variables in Cluster 3

Categories	Variables	Serious crash			Fatal crash		
		B	Exp(B)	Significance	B	Exp(B)	Significance
Cyclist characteristics	Female	0.035	1.036	0.595	0.298	1.347	0.317
	Driving under the influence	0.401	1.494	0.001	1.299	3.664	0.0001
	Opposite direction of travel	-0.522	0.593	0.0001	-1.192	0.304	0.0001
Driver characteristics	Driving under the influence	0.367	1.444	0.094	2.131	8.422	0.0001
	Female	-0.131	0.878	0.007	-0.328	0.721	0.136
Road characteristics	Wetness	-0.309	0.734	0.724	-15.241	0.0001	0.999
	Concrete	0.061	1.063	0.677	-0.525	0.591	0.495
	Smooth pitch	0.018	1.019	0.729	-0.032	0.968	0.892
Environmental characteristics	Non-intersections	-0.084	0.92	0.277	0.114	1.121	0.727
	City	0.041	1.042	0.411	0.757	2.131	0.0001
	Cloudy	-0.329	0.72	0.049	16.004	8919877.43	0.997
	Rainy	0.542	1.72	0.004	-20.847	0.0001	0.996
Constant		568	0.035	1.036	0.595	0.298	1.347

Note: No/possible injuries are set as the reference category; - indicates not significant at the 95% confidence interval or not included in the calculation; the result is considered significant at significance (p-value) < 0.05.

5.3 Discussion of influencing factors results

Rider characteristics

The rider characteristics in this paper's data include the gender of the rider, whether the rider is driving under the influence of alcohol, and the direction in which the rider is travelling. In terms of rider gender, the significance of rider gender was greater than 0.05 in all three clusters and therefore did not have a significant effect on the outcome of crashes. It has been found that female cyclists are less responsive to the traffic environment and less able to deal with emergencies than male cyclists, and thus female cyclists have a higher percentage of serious crashes than male cyclists in the event of a crash due to factors such as untimely avoidance manoeuvres [31, 32].

The statistical significance of drunken cycling in Cluster 2 ($p = 0.5$ and $p = 0.005$) and Cluster 3 ($p = 0.001$ and $p = 0.0001$) demonstrates that this behaviour represents a significant distinguishing characteristic among the clusters. The B-values for the consequences of injury crashes caused by drunken driving of cyclists in the clusters were 0.032, -0.106 and 0.401, and the B-values for the consequences of fatal crashes were 0.948, 0.731 and 1.299, respectively. Crashes where the cyclist is driving while intoxicated are more likely to result in a fatality than slight crashes. The direction in which a cyclist travels, as shown in the table, has a significant effect on crashes and is more likely to result in an injury crash.

Driver characteristics

In terms of whether the driver is drunk driving or not, the p-value of the severity of the rider's injury in cluster 1 is 0.43 for "injury crash", and there is no significant effect of the driver's drunk driving on the result, and the p-value of the consequence of "death" is 0.001, which has a significant effect on the result of the crash. This suggests that drinking and driving has a significant positive effect on the severity of cyclists' injuries and that drivers are more likely to be "fatal" as opposed to "slight injury crashes" as a result of a crash. Driving under the influence of alcohol affects the driver's reaction time to unexpected situations, resulting in the driver's inability to accurately react to slowing down or turning, as well as affecting the driver's vision, which prevents them from correctly judging the safe distance to travel, making driving under the influence of alcohol more likely to result in serious consequences [21].

From the table, it can be seen that the p-value of "fatal" caused by female drivers in each cluster is less than 0.05, and the p-value of the consequences of "injury crashes" caused by female drivers in Cluster 2 is greater than 0.05, which indicates that female drivers have a significant impact on crashes and are more likely to cause serious consequences than "slight crashes". This indicates that female drivers have a significant effect on crashes and are more likely to cause serious consequences than "slight injury crashes". This is because female drivers are more emotional and unable to make rational judgments in a timely manner when encountering unexpected situations, thus increasing the risk of causing serious crashes [33].

Road characteristics

In all clusters, the exogenous variables of road characteristics do not have significant effects. Both cyclists and motorists will consciously judge the speed and distance according to the environmental conditions when travelling on the road to ensure the safety of their own driving. Driving in poor environmental conditions, such as cloudy and rainy days or slippery roads, will cause them to slow down [34].

Environmental characteristics

In Cluster 1, only the variable "non-intersection" has a significant effect, and the Exp(B) value of crashes resulting in "serious" is 1.108, and the Exp(B) value of crashes resulting in "death" is 6.862, which is six times higher than that of crashes resulting in "serious". The Exp(B) value for crashes resulting in "serious" was 1.108, and for serious resulting in "fatal" was 6.862, which is six times higher than for serious resulting in "serious", suggesting that "BMV crashes on wet roads in rainy cities" are more likely to occur at non-intersections. Most of the reported BMV crashes in most areas occurred in the middle of the roadway rather than at intersections because riders or motorists travelling at roadway intersections reduce their speeds and observe the vehicle's surroundings to be alert [18, 35].

In Cluster 3, the B-values for crashes in urban areas are -0.329 and 16.004, respectively, but the significance of "fatal" crashes in urban areas is not significant, so crashes in urban areas are more likely to result in "injury crashes". This is due to the fact that urban areas have a higher traffic volume and therefore a better and stricter enforcement system. Bicyclists tend to reduce their speed and travel with caution in these areas [33, 34]. Urban areas have higher land use and are more densely populated, so urban areas tend to have lower crash severity due to lower speeds of crash participants when crashes occur. There is also a very clear correlation between roadway speed limit regulations and speed limit signs on cyclist injury severity [27].

6. CONCLUSION

Based on the potential category analysis, this paper clusters the BMV crashes data into three clusters, which are Cluster 1, "BMV crashes occurring on slippery roads in urban areas on cloudy and rainy days", Cluster 2, "Crashes occurring on dry road environments in rural non-intersection roadways on sunny days" and Cluster 3 "Crashes occurring on dry road environments in urban areas on sunny days". MNLogit modelling was also performed to discuss the similarities and differences in the factors influencing the injury severity of the colliding cyclists in the different clusters.

The effect of female drivers on the severity of injuries to cyclists is significant, especially in Cluster 2, where a female driver results in a B-value of -0.602 for "fatal" and 0.02 for "serious crashes", suggesting that female drivers are more likely to cause serious consequences in BMV crashes. Therefore, female drivers must drive carefully and be more vigilant to react faster and more accurately when encountering unexpected situations.

Serious crashes at non-intersections have a significant effect in all three clusters, indicating that both cyclists and motor vehicle drivers can consciously reduce their speeds and observe the situation around the vehicles in the travelled roadway section when travelling to the intersection section, thus reducing the risk of a crash. Non-intersection is generally a roadway with fewer road conditions. Due to prolonged travel on roadways with limited traffic stimuli, cyclists or motorists may become less attentive to their surroundings; if unexpected situations are not responded to in a timely or appropriate manner, the resulting overreaction or delayed reaction may lead to more severe crash consequences.

Therefore, it is recommended to focus on strengthening traffic management in uncontrolled/non-road areas to reduce the probability of serious crashes. It is also recommended that the authorities set up warning signs at non-intersections and remind drivers to slow down on crash-prone roads, especially those with simple road conditions, where drivers are more likely to neglect to observe the surrounding road conditions, resulting in serious consequences. For the management of bicycle riders, traffic authorities should promulgate more detailed laws and regulations to regulate cycling behaviour and enhance road traffic safety.

Although this study reveals the influence of multiple factors on the severity of BMV crashes under latent class clustering, several limitations warrant further investigation. First, the varying proportions of data samples across different crash severity levels may affect clustering performance, potentially leading to suboptimal clustering results. Second, many studies have shown that accident severity classification exhibits an ordered nature, whereas this study does not explicitly account for this ordering in the MNL model [36, 37]. This study only used slight crashes as the reference category to examine the differential impacts of influential factors on serious and fatal crashes relative to slight crashes. Future research could incorporate ordered logit models or explore more complex interaction effects using machine learning approaches to capture nonlinear relationships.

ACKNOWLEDGEMENTS

This study is supported by the Natural Science Foundation of Shaanxi Province [Grant No. 2025JC-YBQN-773], the Natural Science Foundation of Inner Mongolia Autonomous Region [Grant No. 2025QN05123], the Inner Mongolia Agricultural University High-level and Outstanding PhD Talent Recruitment Research Startup Project [Grant No. NDYB 2023-67], the Program for Innovative Research Teams in Universities of Inner Mongolia Autonomous Region [Grant No. NMGIRT2304], and the Key Research and Achievement Transformation Program of Inner Mongolia Autonomous Region [Grant No. 2025YFHH0207].

REFERENCES

[1] Algurén B, Rizzi M. In-depth understanding of single bicycle crashes in Sweden - Crash characteristics, injury types and health outcomes differentiated by gender and age-groups. *Journal of Transport & Health*. 2022;24:101320. DOI: 10.1016/j.jth.2021.101320

[2] Ma L, Ye R, Wang H. Exploring the causal effects of bicycling for transportation on mental health. *Transportation Research Part D: Transport and Environment*. 2021;93:102773. DOI: 10.1016/j.trd.2021.102773

[3] Salmon PM, et al. Bicycle crash contributory factors: A systematic review. *Safety Science*. 2022;145:105511. DOI: 10.1016/j.ssci.2021.105511

[4] Utriainen R, et al. Single-bicycle crashes in Finland - Characteristics and safety recommendations. *Journal of Safety Research*. 2023;87:96-106. DOI: 10.1016/j.jsr.2023.09.008

[5] Ward NJ, et al. Traffic safety culture and prosocial driver behavior for safer vehicle-bicyclist interactions. *Journal of Safety Research*. 2020;75:24-31. DOI: 10.1016/j.jsr.2020.07.003

[6] Mannering FL, Shankar V, Bhat CR. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*. 2016;11:1-16. DOI: 10.1016/j.amar.2016.04.001

[7] Obaid I, et al. Factors associated with driver injury severity of motor vehicle crashes on sealed and unsealed pavements: Random parameter model with heterogeneity in means and variances. *International Journal of Transportation Science and Technology*. 2023;12(2):460-475. DOI: 10.1016/j.ijtst.2022.04.002

[8] Xiong Y, Mannering FL. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. *Transportation Research Part B: Methodological*. 2013;49:39-54. DOI: 10.1016/j.trb.2013.01.002

[9] Kim SH, Mokhtarian PL. Finite mixture (or latent class) modeling in transportation: Trends, usage, potential, and future directions. *Transportation Research Part B: Methodological*. 2023;172:134-173. DOI: 10.1016/j.trb.2023.03.001

[10] Afghari AP, et al. A home-based approach to understanding seatbelt use in single-occupant vehicles in Tennessee: Application of a latent class binary logit model. *Accident Analysis & Prevention*. 2020;146:105743. DOI: 10.1016/j.aap.2020.105743

[11] Alnawmasi N, Mannering F. An analysis of day and night bicyclist injury severities in vehicle/bicycle crashes: A comparison of unconstrained and partially constrained temporal modeling approaches. *Analytic Methods in Accident Research*. 2023;40:100301. DOI: 10.1016/j.amar.2023.100301

[12] Boufous S, et al. Risk factors for severe injury in cyclists involved in traffic crashes in Victoria, Australia. *Accident Analysis & Prevention*. 2012;49:404-409. DOI: 10.1016/j.aap.2012.03.011

[13] Chong S, et al. Relative injury severity among vulnerable non-motorised road users: Comparative analysis of injury arising from bicycle-motor vehicle and bicycle-pedestrian collisions. *Accident Analysis & Prevention*. 2010;42(1):290-296. DOI: 10.1016/j.aap.2009.08.006

[14] Stone M, Broughton J. Getting off your bike: cycling accidents in Great Britain in 1990-1999. *Accident Analysis & Prevention*. 2003;35(4):549-556. DOI: 10.1016/S0001-4575(02)00032-5

[15] Harada MY, et al. Bicycle trauma and alcohol intoxication. *International Journal of Surgery*. 2015;24:14-19. DOI: 10.1016/j.ijsu.2015.10.013

[16] Wallén Warner H. Beliefs about riding a bicycle when mildly alcohol intoxicated. *Transportation Research Part F: Traffic Psychology and Behaviour*. 2021;76:384-392. DOI: 10.1016/j.trf.2020.09.012

[17] Kim JK, et al. Bicyclist injury severities in bicycle-motor vehicle accidents. *Accident Analysis & Prevention*. 2007;39(2):238-251. DOI: 10.1016/j.aap.2007.05.001

[18] Atombo C, et al. Investigating the motivational factors influencing drivers intentions to unsafe driving behaviours: Speeding and overtaking violations. *Transportation Research Part F: Traffic Psychology and Behaviour*. 2016;43:104-121. DOI: 10.1016/j.trf.2016.09.029

[19] Salas-Wright CP, et al. Driving under the influence of cannabis and alcohol: Evidence from a national sample of young drivers. *Addictive Behaviors*. 2023;147:107816. DOI: 10.1016/j.addbeh.2023.107816

[20] Yan X, et al. Temporal analysis of crash severities involving male and female drivers: A random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research*. 2021;30:100161. DOI: 10.1016/j.amar.2021.100161

[21] Behnood A, Mannering F. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research*. 2017;16:35-47. DOI: 10.1016/j.amar.2017.08.001

[22] Rivera Olsson S, Elldér E. Are bicycle streets cyclist-friendly? Micro-environmental factors for improving perceived safety when cycling in mixed traffic. *Accident Analysis & Prevention*. 2023;184:107007. DOI: 10.1016/j.aap.2023.107007

[23] Samerei SA, et al. Using latent class clustering and binary logistic regression to model Australian cyclist injury severity in motor vehicle–bicycle crashes. *Journal of Safety Research*. 2021;79;246–256. DOI: 10.1016/j.jsr.2021.09.005

[24] Yared T, Patterson P, All ESA. Are safety and performance affected by navigation system display size, environmental illumination, and gender when driving in both urban and rural areas? *Accident Analysis & Prevention*. 2020;142:105585. DOI: 10.1016/j.aap.2020.105585

[25] Hussain Z, et al. Aberrant driving behaviors as mediators in the relationship between driving anger patterns and crashes among taxi drivers: An investigation in a complex cultural context. *Traffic Injury Prevention*. 2023;24(5):393-401. DOI: 10.1080/15389588.2023.2199898

[26] Strandroth J, et al. The effects of studded tires on fatal crashes with passenger cars and the benefits of electronic stability control (ESC) in Swedish winter driving. *Accident Analysis & Prevention*. 2012;45:50-60. DOI: 10.1016/j.aap.2011.11.005

[27] Chen P. Built environment factors in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach. *Safety Science*. 2015;79:336-343. DOI: 10.1016/j.ssci.2015.06.016

[28] Song L, Fan WD. Exploring truck driver-injury severity at intersections considering heterogeneity in latent classes: A case study of North Carolina. *International Journal of Transportation Science and Technology*. 2021;10(2):110-120. DOI: 10.1016/j.ijtst.2020.12.006

[29] Sivasankaran SK, Balasubramanian V. Exploring the severity of bicycle-vehicle crashes using latent class clustering approach in India. *Journal of Safety Research*. 2020;72:127-138. DOI: 10.1016/j.jsr.2019.12.012

[30] Liu P, Fan W. Exploring injury severity in head-on crashes using latent class clustering analysis and mixed logit model: A case study of North Carolina. *Accident Analysis & Prevention*. 2020;135:105388. DOI: 10.1016/j.aap.2019.105388

[31] Li Z, et al. Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accident Analysis & Prevention*. 2019;129:230-240. DOI: 10.1016/j.aap.2019.05.016

[32] Klassen J, El-Basyouny K, Islam MT. Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: A City of Edmonton case study. *Safety Science*. 2014;62:295-304. DOI: 10.1016/j.ssci.2013.09.007

[33] Li Y, et al. Injuries and risk factors associated with bicycle and electric bike use in China: A systematic review and meta-analysis. *Safety Science.* 2022;152:105769. DOI: 10.1016/j.ssci.2022.105769

[34] Shin EJ. Decomposing neighborhood disparities in bicycle crashes: A Gelbach decomposition analysis. *Transport Policy*. 2023;131:156-172. DOI: 10.1016/j.tranpol.2022.12.014

[35] Zhang Y, Li H, Ren G. Analyzing the injury severity in single-bicycle crashes: An application of the ordered forest with some practical guidance. *Accident Analysis & Prevention*. 2023;189:107126. DOI: 10.1016/j.aap.2023.107126

[36] Zhao L, et al. Exploring injury severity of non-motor vehicle riders involving in traffic accidents using the generalized ordered logit model. *Ain Shams Engineering Journal*. 2023;14(5):101962. DOI: 10.1016/j.asej.2022.101962

[37] Wang C, et al. Injury severity analysis of e-bike riders in China based on the in-vehicle recording video crash data: a random parameter ordered logit model. *International Journal of Injury Control and Safety Promotion*. 2024;31(4):568-578. DOI: 10.1080/17457300.2024.2385102