

## Article

# Latent Class Cluster Analysis and Mixed Logit Model to Investigate Pedestrian Crash Injury Severity

Arsalan Esmaili <sup>1</sup>, Kayvan Aghabayk <sup>1,\*</sup>  and Nirajan Shiwakoti <sup>2</sup> <sup>1</sup> School of Civil Engineering, College of Engineering, University of Tehran, Tehran 1417935840, Iran<sup>2</sup> School of Engineering, RMIT University, Melbourne 3001, Australia

\* Correspondence: kayvan.aghabayk@ut.ac.ir

**Abstract:** Traffic crashes involving pedestrians have a high frequency in developing countries. Among road users, pedestrians are the most vulnerable, as their involvement in traffic crashes is usually followed by severe and fatal injuries. This study aims to identify pedestrian crash patterns and reveal the random parameters in the dataset. A three-year (2015–2017) pedestrian crash dataset in Mashhad, Iran, was employed to investigate the influence of a rich set of factors on pedestrian injury severity, some of which have been less accounted for in previous studies (e.g., the vicinity to overpasses, the existence of vegetated buffers, and park lanes). A two-step method integrating latent class cluster analysis (LCA) and the mixed logit model was utilized to consider unobserved heterogeneity. The results demonstrated that various factors related to the pedestrian, vehicle, temporal, environmental, roadway, and built-environment characteristics are associated with pedestrian injuries. Furthermore, it was found that integrated use of LCA and mixed logit models can considerably reduce the unobserved heterogeneity and uncover the hidden effects influencing severity outcomes, leading to a more profound perception of pedestrian crash causation. The findings of this research can act as a helpful resource for implementing effective strategies by policymakers to reduce pedestrian casualties.

**Keywords:** pedestrian; traffic crashes; injury severity analysis; unobserved heterogeneity; latent class cluster analysis; mixed logit model; developing country



**Citation:** Esmaili, A.; Aghabayk, K.; Shiwakoti, N. Latent Class Cluster Analysis and Mixed Logit Model to Investigate Pedestrian Crash Injury Severity. *Sustainability* **2023**, *15*, 185. <https://doi.org/10.3390/su15010185>

Academic Editor: Armando Carteni

Received: 14 November 2022

Revised: 18 December 2022

Accepted: 20 December 2022

Published: 22 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Traffic crashes result from human activities interacting with diverse cultural, socio-economic, and geographic contexts, which disrupt the health system. Traffic crashes not only injure people, they can also result in deaths. Every year, traffic crashes are responsible for over 1.2 million deaths and 50 million injuries worldwide and are the eighth leading cause of death [1]. Road safety advancements are associated with socioeconomic factors, such as education, motorization level, and economic growth, which vary considerably between developing and developed countries [2]. Moreover, road safety is much more appreciated in developed countries, and they have a long history of applying road safety measures [3]. Owing to these differences, the number of casualties and financial expenses resulting from traffic crashes are significantly higher in developing countries.

Traffic crashes are the leading cause of mortality in developing countries, where 93% of all road traffic fatalities occur, while they own only 54% of the world's vehicles [3,4]. The financial repercussions of traffic crash injuries are also substantial. In developing countries, traffic crash consequences account for 2–7% of gross domestic product (GDP) [1,4].

The high rate of fatalities in traffic crashes has been a long-standing issue in Iran. In 2011, 20,068 people died, and 297,252 were injured in traffic crashes in Iran [5]. However, due to the coordinated efforts from road agencies and other relevant stakeholders, traffic casualties reached around 20.5 people per 100,000 population (17,000 people in total) annually in recent years [6]. However, this number is still considerably higher than the world average (18.2), European countries' average (10.3), and the U.S. (11.1) [7,8]. The

burden of traffic-related fatalities and injuries in Iran also created a number of economic difficulties. For instance, the cost of traffic-related fatalities accounts for more than 5% of Iran's gross national product (GNP) each year, costing the nation about USD 6 billion [9].

Meanwhile, around half of the fatalities in traffic crashes are related to vulnerable road users. Among them, pedestrians are the most vulnerable since they have the least protection during a vehicle–pedestrian collision [10]. As a result of continued investments in road safety research programs and countermeasures, pedestrian-related crashes in developed countries are presently declining [11]. However, pedestrian deaths are increasing in Iran, with the Iranian Legal Medicine Organization [12] reporting a 2.1% increase between 2018 and 2019. This increasing trend due to the high share of pedestrian fatalities needs attention. On average, 23 percent of all traffic fatalities are pedestrians in Iran, which is higher than the world average (22%) or developed countries such as Australia (13%) and the U.S. (17%) [13]. This number is much worse in populated cities, where it is around 40 percent [12]. For example, in Mashhad, the second most populated city of Iran, this number was more than 55% in 2018 [14].

For more than 30 years, considerable research has been conducted on pedestrian safety in developed countries. In contrast, in developing countries, the literature on vulnerable road users is at an early stage, and the number of studies in this field is limited [15,16]. Moreover, the focus of researchers and authorities is usually narrowed to motorized traffic instead of pedestrians in developing countries [15,17]. The situation is similar in Iran, and only a handful of studies have been conducted on pedestrian safety (e.g., [18–21]). Hence, there is a need for comprehensive studies on pedestrian safety in Iran, especially in urban environments that have a high share of pedestrian casualties.

Studies in pedestrian safety can generally be classified into two groups: some researchers concentrated on the frequency of pedestrian crashes (e.g., [22–25]), while others have studied the contributing factors affecting the injury severity of pedestrians involved in crashes (e.g., [26–31]). Moreover, various types of datasets have been used to investigate pedestrian safety. Many studies used crash databases collected by authorities, such as police, and emergency services reports. These databases are then aggregated into macro, meso, or micro geographic units [32]. This is the traditional type of data for pedestrian crash analysis that covers crash reports in a certain area and is limited in the number of observations. A more recent line of research employs naturalistic driving methodology, which is an experimentation model that allows recognition of driving modes by observing the driver's behavior at the wheel of a group of people in natural conditions over extended observation periods. This methodology aims to increase the representativeness of the acquired data, as opposed to data stemming from laboratory tests that are highly controlled [33,34]. Naturalistic driving provides extensive data with various variables such as vehicle kinematics, roadway geometry, traffic conditions, and environmental variables compared to police and emergency report [32].

In the following paragraphs, methodological approaches of studies that mainly focused on the severity of pedestrian injuries using traditional crash datasets are specifically reviewed.

It was found that several factors, including pedestrian's age (e.g., [26,35,36]), pedestrian's gender (e.g., [36,37]), driver's age and gender (e.g., [28,36,38]), blood alcohol consumption (e.g., [39,40]), vehicle type (e.g., [41,42]), road type (e.g., [27]), number of lanes (e.g., [43]), time of crash (e.g., [26,40]), weather condition (e.g., [28,44]), speed (e.g., [36,45]), pedestrian red light violation (e.g., [46]), traffic control (e.g., [28,41]), light condition (e.g., [27,47]), and land use (e.g., [36,43]), can significantly affect the severity of pedestrian crashes. Various modelling approaches have been used over the years to understand the effect of these factors.

Most of the early studies in this field analyzed the crash data by descriptive analysis. In these studies, variables extracted from the crash dataset, such as pedestrian's and driver's age, gender, alcohol consumption, or crash occurrence time, were compared across the crash severity categories (see [48–50]).

Over time, the use of multivariate analysis became prevalent among researchers. Therefore, models such as logistic regression (e.g., [26,37,51]) and ordered response models (e.g., [36,52]) have been widely adopted in many studies over the years.

In the social sciences, many researchers examine human dynamics and social behaviors by considering very particular scales and settings. Moreover, in a networked society, the complexity and spatial heterogeneity of these dynamics and behaviors are obscured by the absence of a holistic perspective in many studies [53]. Addressing heterogeneity is crucial in social sciences, and there are various methodologies and approaches for dealing with this issue. However, the main aspect of that is understanding the heterogeneity in the right context. For example, differences in environmental settings (urban or rural), culture, and financial level (developed or developing country) can trace very different social behaviors [53].

Traffic crashes are highly heterogeneous due to their spatiotemporal nature and because they are outcomes of human activities interacting with diverse cultural, socio-economic and geographic contexts. In other words, crashes can occur in different circumstances, and the influence of various factors could be hidden in the chain of events that lead to a crash. Thereby, some latent relationships may exist between crashes and influential factors that are hard to detect [29]. These hidden relationships show themselves as heterogeneity in crash datasets. The unobserved heterogeneity has been identified as a critical problem for traffic safety modelling [54]. For instance, it can cause the same factor to affect crash severity inconsistently or even oppositely under different conditions [55,56].

Accordingly, researchers in this area have tried to consider heterogeneity issues by using different approaches to investigate pedestrian crash datasets. Some studies concentrated on distinct influential variables or crash patterns. For example, specific pedestrian age groups such as children [57,58], pedestrian crashes with specific vehicles such as trucks [50,59], buses [60], and taxis [61], specific location such as intersections [36,62], and so on. The segmentation of data by focusing on specific parameters provides beneficial information and may lower the heterogeneity, but they do not necessarily lead to complete homogeneity in each segment [63].

One of the measures to reduce heterogeneity is to use data mining approaches such as cluster analysis [63,64]. The cluster analysis has been used in crash data analysis with different frameworks, such as k-means clustering [65], kernel density estimation [66], and latent class clustering [29,30].

Some studies used cluster analysis to reach homogenous groups, and, for each sub-data group, severity models were then applied to evaluate the risk factors associated with pedestrian crash severity. Severity models such as ordered probit [28], binary logit [29], and multinomial logit [30] were used after clustering to investigate the explanatory factors affecting pedestrian crash severity. For instance, Sasidharan et al. [29] first applied a latent class analysis to classify pedestrian crash datasets into homogenous clusters. A binary logistic regression model was then used to investigate variables associated with severity in each subgroup. The results revealed that the binary logistic regression model in terms of clusters is more accurate than utilizing a single binary logistic regression model to the whole data [29]. It was also found that some variables that are not significant in the whole data severity analysis could become significant in specific crash patterns [29].

The cluster analysis can decrease the heterogeneity of the data to some extent, yet it is still expected to remain within each identified cluster [56,67]. Other than clustering methods, another alternative for capturing the unobserved heterogeneity is to use discrete choice models that can address heterogeneity. The mixed (random parameter) logit model is one of the approaches that can meet the requirements mentioned above. This model is a more flexible version of multinomial logistic regression that allows parameters to differ across observations. With the development of computer power for modelling, it became easier to use such time-consuming models. Therefore, traffic safety researchers have been able to conduct the mixed logit model in recent years to investigate the injury severity of pedestrian-involved crashes [27,38,68]. The traditional mixed logit model evaluates

random parameters individually and independently without considering the correlations among random parameters. A more recent line of research in crash analysis extended the approach to the correlated mixed logit models [69–71]. The correlation between the variables could offer more profound insights into how some combinations can influence safety [70]. Moreover, the correlated RPL model was found to be statistically superior compared to the contemporary ones, such as random-effects models and uncorrelated random parameters models [71]. However, correlated random parameter logit models are computationally more complex than uncorrelated random parameter logit models. RPL models with correlated random parameters require the estimate of a much larger number of parameters, which increases not only the calculation time but also the probability of local maxima [72]. Correlated RPL model has still been used less frequently in severity analysis due to its computational complexity [72].

In general, using random parameter models has its own limitations. For example, the pre-specified parameter distributions in mixed logit models may not reveal the unobserved heterogeneity for some parameters across observations [54]. In other words, some parameters could show different influences in some specific crash patterns, which may remain unseen when working with the whole dataset. One way to overcome this problem is to categorize data into sub-datasets with maximum homogeneity.

According to the explanations given, it can be concluded that the integrated use of the mixed logit models and cluster analysis can significantly alleviate the drawbacks of both approaches and minimize the heterogeneity. A two-step method incorporating both LCA and the uncorrelated mixed logit model has been used lately on crash datasets [54,56,67,73].

#### *Objectives and Scope of the Study*

This study aims to identify pedestrian crash patterns, reveal the random parameters in the dataset, and investigate the influence of various factors on the injury severity of pedestrian-involved crashes. This is achieved by utilizing pedestrian crash data from Iran, a developing country with a considerably high rate of pedestrian casualties.

To analyze the pedestrian crash dataset, a two-step method using latent class cluster analysis (LCA) and the mixed logit model is followed. In the first step, the LCA was conducted to classify pedestrian crash data into homogenous clusters. Afterward, the mixed logit model was utilized as a severity model to determine risk factors in each sub-dataset by considering the possible remaining heterogeneity within each identified cluster. The integrated use of the mixed logit model and cluster analysis can substantially alleviate the limitations of both approaches and minimize the heterogeneity. To test this, the effectiveness of the integrated model on segmentation and reduction of the heterogeneity of pedestrian crash data was also evaluated.

Given that very few studies have been conducted on the comprehensive analysis of pedestrian injury severity in Iran, the results of this study can be a valuable reference to help policymakers prioritize designated investments in traffic safety. It can also assist in executing effective strategies to decrease the high rate of pedestrian casualties.

The rest of the paper is organized as follows. In the next section, the pedestrian crash dataset analyzed for this study is presented. Section 3 explains the methodology adopted in this study. Results and discussion are presented in Section 4. It is then followed by the final section, which contains conclusions and recommendations.

## **2. Data**

The injury severities of pedestrians in traffic crashes in Mashhad were investigated in this study. Mashhad is the second most populated city in Iran, with a population of approximately 3 million [74]. The three-year pedestrian crash data from 2015–2017 was obtained from the Mashhad Department of Transportation. This department is responsible for gathering police, emergency, and forensic medicine organization crash-related reports. The data includes crashes with fatal injuries, major injuries, minor injuries, and no injuries. It should be noted that crashes with property damage only (no injury) were removed due

to their very small share and potential underreporting cases. Further, crashes with missing information were also excluded for further analysis in this study.

Finally, a total of 6215 pedestrian-involved crashes was obtained from crash data, with 3192 (51.36%) minor injury crashes, 2517 (40.5%) major injury crashes, and 506 (8.14%) fatal crashes. The extracted data contains various factors such as pedestrian, vehicle, temporal, environmental, roadway, and built environment characteristics. To illustrate contributing factors and heterogeneity more efficiently, numeric variables were categorized based on previous research experience in this field (e.g., [27–29]). An overview of variables used in this study and descriptive statistics of the dataset are presented in Table 1.

**Table 1.** Descriptive analysis of the pedestrian crash dataset.

Variables	Description	No. of Crashes	Injury Severity (%)		
			Minor	Major	Fatal
Pedestrian crashes		6215	51.36%	40.50%	8.14%
<i>Pedestrian characteristics</i>					
Pedestrian gender	Women (ref.)	2489	55.16%	38.37%	6.47%
	Men	3726	48.82%	41.92%	9.26%
Pedestrian age	<15	1115	58.83%	36.95%	4.22%
	15–30 (ref.)	1847	56.96%	40.44%	2.60%
	30–45	1292	54.80%	39.24%	5.96%
	45–65	1239	43.99%	46.17%	9.85%
	>65	722	31.99%	38.64%	29.36%
<i>Involved party characteristics</i>					
Involved vehicle type	Motorcycle	1710	59.30%	36.73%	3.98%
	Heavy vehicle, bus	261	21.46%	43.30%	35.25%
	Minibus, van	62	17.74%	54.84%	27.42%
	Pickup	210	31.90%	45.24%	22.86%
	Bicycle	32	68.75%	25.00%	6.25%
	Passenger car (ref.)	3940	51.32%	41.60%	7.08%
Hit and run	No (ref.)	5332	54.08%	37.79%	8.12%
	Yes	883	34.88%	56.85%	8.26%
<i>Temporal characteristics</i>					
Time of crash	6–10	988	49.80%	39.27%	10.93%
	10–14	1525	56.79%	38.62%	4.59%
	14–18	1570	53.31%	40.51%	6.18%
	18–22	1498	50.20%	43.39%	6.41%
	22–6 (ref.)	634	38.64%	40.06%	21.29%
Day type	Weekday(ref.)	2533	50.02%	41.33%	8.65%
	Weekend	3682	52.28%	39.92%	7.79%
<i>Environmental characteristics</i>					
Weather	Adverse	506	41.90%	47.83%	10.28%
	Clear (ref.)	5709	52.20%	39.85%	7.95%
Season	Spring	1531	52.12%	39.91%	7.97%
	Summer	1878	51.65%	40.73%	7.60%
	Autumn	1540	51.33%	40.36%	8.31%
	Winter (ref.)	1266	50.08%	41.00%	8.93%
<i>Roadway and built-environment characteristics</i>					
Posted speed	40–60 km/h	2043	50.56%	41.16%	8.27%
	<40 km/h	1937	55.14%	37.07%	7.80%
	>60 km/h (ref.)	2235	48.81%	42.86%	8.32%
Junction	No (ref.)	4588	50.00%	40.80%	9.20%
	Yes	1627	55.19%	39.64%	5.16%
Traffic control	None (ref.)	1792	48.05%	42.91%	9.04%
	Signal	861	54.01%	40.30%	5.69%
	Sign	3562	52.40%	39.33%	8.28%



Table 1. Cont.

Variables	Description	No. of Crashes	Injury Severity (%)		
			Minor	Major	Fatal
Road type	Divided two-way	3885	52.48%	40.18%	7.34%
	Undivided two-way	1736	48.50%	41.19%	10.31%
Road width	One-way (ref.)	594	52.36%	40.57%	7.07%
	<20 m (ref.)	3366	49.35%	42.96%	7.69%
	>20 m	2849	53.74%	37.59%	8.67%
Sidewalk	No (ref.)	1956	46.27%	45.30%	8.44%
	Yes	4259	53.69%	38.30%	8.01%
Vegetation	No (ref.)	2750	49.60%	39.85%	10.55%
	Yes	3465	52.76%	41.01%	6.23%
Park lane	No (ref.)	2744	54.05%	37.76%	8.19%
	Yes	3471	49.24%	42.67%	8.10%
Overpass/underpass (in 300 m)	No (ref.)	3825	56.31%	36.24%	7.45%
	Yes	2390	43.43%	47.32%	9.25%
AADT	High (>30,000)	1656	52.72%	39.98%	7.31%
	Medium (15,000–30,000)	1987	50.63%	42.93%	6.44%
	Low (<15,000)(ref.)	2572	51.05%	38.96%	9.99%
Population density	<100 person/km <sup>2</sup> (ref.)	1164	44.16%	41.32%	14.52%
	>200 person/km <sup>2</sup>	3738	54.25%	40.10%	5.64%
	100–200 person/km <sup>2</sup>	1313	49.50%	40.90%	9.60%
Land use	Other	1207	45.15%	41.51%	13.34%
	Residential	3088	53.21%	39.73%	7.06%
	Commercial (ref.)	1920	52.29%	41.09%	6.61%

### 3. Methodology

#### 3.1. Latent Class Clustering Analysis

In the present study, the latent class cluster analysis (LCA) was conducted to address the heterogeneity issue. LCA is a probability model based on the cluster analysis method that can be considered an unsupervised learning and data mining technique due to the unknown number and form of clusters (e.g., [63,75]). In this method, the entire dataset is classified into exclusive latent classes, each representing specific traits of that data [30,76]. The main goal of cluster analysis is to maximize the similarity within each cluster and minimize the similarity between clusters [77].

LCA assumes that there exists a latent categorical variable that classifies the dataset into mutually exclusive and comprehensive subgroups [29,76,78]. In LCA, the likelihoods of each crash to be included in different clusters are specified based on different models developed for various values of clusters defined. After specifying the likelihood of a crash in every cluster, the cluster with the highest probability of including that crash is labeled as the best index cluster [29,67]. LCA has several advantages in comparison with the conventional clustering approaches (e.g., k-means clustering). These include: (a) In latent class clustering, there is no need to predetermine the number of clusters. Different statistical measures can be applied to identify the optimized number of clusters [29]; (b) LCA can handle various types of variables, including nominal, continuous, counts, and categorical variables without a standardization process [28,63,64].

The LCA plugin for Stata developed by the Penn State methodology center is applied to conduct latent class analysis in this study [79]. In light of Lanza and Rhoades's [79] study, a latent class analysis was carried out to divide the pedestrian crash dataset into several clusters.

The LCA Stata plugin estimates two sets of parameters ( $\gamma$  and  $\rho$ ). Gamma ( $\gamma$ ) parameters are the basis for interpreting the latent classes, which are latent class membership probabilities. Rho ( $\rho$ ) parameters indicate the item-response probabilities conditional on latent class membership. The assumption is there are C latent classes, and each crash i consists of M crash characteristics. The vector  $Z_i = (Z_{i1}, \dots, Z_{iM})$  presents crash i's response

to  $M$  attributes. Accordingly, crash  $i$ 's attribute of characteristic  $m$  is represented by  $Z_{im}$ , which is a categorical variable with one possible value of  $1, \dots, r_m$ .  $L_i = 1, 2, \dots, C$  is crash  $i$ 's latent class membership. The indicator function,  $I(Z_m = r_m)$ , equals 1 if the attribute of the characteristic  $m$  equals to  $r_m$  and equals zero, otherwise. Accordingly,  $\rho_{m,r_m|c}^{I(Z_m=r_m)}$  indicates the likelihood that a crash has the attribute  $r_m$  of characteristic  $m$ , conditional on membership in the latent class  $c$ .  $\gamma_c$  represents the probability of membership in latent class  $c$ . Consequently, the probability of observing a particular vector of responses is [76]:

$$P(Z = z) = \sum_{c=1}^C \gamma_c \prod_{m=1}^M \prod_{r_m=1}^{R_m} \rho_{m,r_m|c}^{I(Z_m=r_m)} \quad (1)$$

Equation (1) shows how the likelihood of observing a particular vector of responses is a function of both the probabilities of each latent class' membership (the  $\gamma$ 's) and the probabilities of observing each response conditional on latent class membership (the  $\rho$ 's) [78].

To fit a latent class model to a dataset, it is necessary to estimate  $\gamma$  and  $\rho$  parameters based on the data. The LCA Stata plugin estimates parameters by expectation-maximization (EM) algorithm. This algorithm is an iterative method to find maximum likelihood estimates of parameters in statistical models, where the model depends on unobserved latent variables.

By using Bayes' theorem (Equation (2)), posterior probabilities or, in other words, classification probabilities of each crash in LCA (Equation (3)) can be obtained [78]:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

$$P(L = c|Z = z) = \frac{(\prod_{m=1}^M \prod_{r_m=1}^{R_m} \rho_{m,r_m|c}^{I(Z_m=r_m)}) \gamma_c}{\sum_{c=1}^C \gamma_c \prod_{m=1}^M \prod_{r_m=1}^{R_m} \rho_{m,r_m|c}^{I(Z_m=r_m)}} \quad (3)$$

After obtaining classification probabilities, each observation could then be allocated to the cluster with the highest probability.

However, the most appropriate number of clusters ( $C$ ) is not known in LCA from the beginning. The most appropriate number can be found by testing different models with a different number of clusters. The assignment error of assigning a crash to a latent class can be minimized by choosing the optimal number of clusters [78]. This number can be decided by some methods to measure assignment accuracy. A popular method is to use information criteria, including the Akaike information criterion (AIC), consistent Akaike information criterion (CAIC), Bayesian information criteria (BIC), and entropy-based measures [30].

The cluster number that minimizes the score of AIC, BIC, and CAIC is considered as the most appropriate one. It was suggested that BIC is a more reliable criterion compared to AIC and CAIC when the data is large [80]. Furthermore, increasing the number of clusters might not always cause information criteria to reach the lowest value [81]. So, computing the percentage reduction in BIC values between different models is preferred [30,82].

An entropy measure ranging between 0 and 1 demonstrates the quality of the clustering solution. It is essentially a weighted average of posterior membership probabilities for each case. Closer values of entropy to 1 indicate a better clustering [83].

### 3.2. Mixed Logit Model

In this study, the uncorrelated mixed logit model is conducted for the whole sample and each sub-sample generated by LCA to identify the contributing factors and understand their effect on pedestrian injury severity. The utility function is linear in the mixed logit (random parameter) model with crash severity as the dependent variable. This function determines the pedestrian severity injury level  $k$  for observation  $n$  and is defined as:

$$U_{kn} = \beta_k X_{kn} + \varepsilon_{kn} \quad (4)$$

where  $X_{kn}$  is the vector of explanatory (independent) variables,  $\beta_k$  is the vector of estimable parameters, which could vary among observations,  $\varepsilon_{kn}$  represents the error term and shows the unobservable effects on severity.  $\varepsilon_{kn}$  is assumed to follow generalized extreme value distribution [84]. Consequently, the probability of  $n$ th individual to be involved in the severity of  $k$  can be given as follows:

$$P_{kn} = \frac{\exp(\beta_k X_{kn})}{\sum_{\forall k \in K} \exp(\beta_k X_{kn})} \quad (5)$$

where  $K$  represents the set of injury severity levels (in this study  $K = 3$ ). In order to let the parameters vary across individuals and capture unobserved heterogeneity, a mixing distribution is introduced [85]:

$$P_{kn}|\varphi = \int \frac{\exp(\beta_k X_{kn})}{\sum_{\forall k \in K} \exp(\beta_k X_{kn})} f(\beta|\varphi) d\beta \quad (6)$$

$f(\beta|\varphi)$  denotes the probability density function (PDF) of randomly distributed parameter  $\beta$ , and  $\varphi$  indicates a vector for describing parameters of the probability density function (mean and variance).

The normal distribution is set for the density function. In a normal distribution,  $\beta_i$  is specified as [86]:

$$\beta_k = \beta_i + \sigma_i v_i, \quad v_i \sim N(0, 1) \quad (7)$$

where  $\beta_i$  is the mean,  $\sigma_i$  is the standard deviation of the distribution, and  $v_i$  is the individual-specific heterogeneity, with mean equal to zero and standard deviation equal to one [86]. Considering the above formulation, if  $\sigma_i$  is not significantly different from zero,  $\beta_k$  will become equal to  $\beta_i$ , and the variable can be considered fixed.

A simulation-based maximum likelihood distribution estimation method was carried out for mixed logit model estimation regarding computational cost-efficiency. It has been shown that 200 Halton draws can provide adequate distribution of draws [85]; therefore, 200 Halton draws were used for model estimation. Following previous studies (e.g., [87,88]), a stepwise variable selection process was used. In the first step, it was considered that variables have random effects. In the next step, the insignificant random variables (variables with insignificant SD) were considered fixed parameters for the next run. This process was repeated until all of the random parameters in the model had significant standard deviations. This stepwise process was conducted for the overall sample and each of the four clusters exclusively.

### 3.3. Marginal Effects

Many studies suggested that, in the case of models with multinomial severity outcomes, estimated parameters might not illustrate the accurate association of the independent variables on the severity [54,87,89]. The marginal effect analysis is utilized to estimate the impact of significant explanatory variables in the mixed logit models on pedestrian injury severity probabilities. Considering that in this study, all explanatory variables are coded in binary form (dummy variables); marginal effects are obtained as follows:

$$E_{X_{kni}}^{P_{kn}} = P_{kn}(X_{kni} = 1) - P_{kn}(X_{kni} = 0) \quad (8)$$

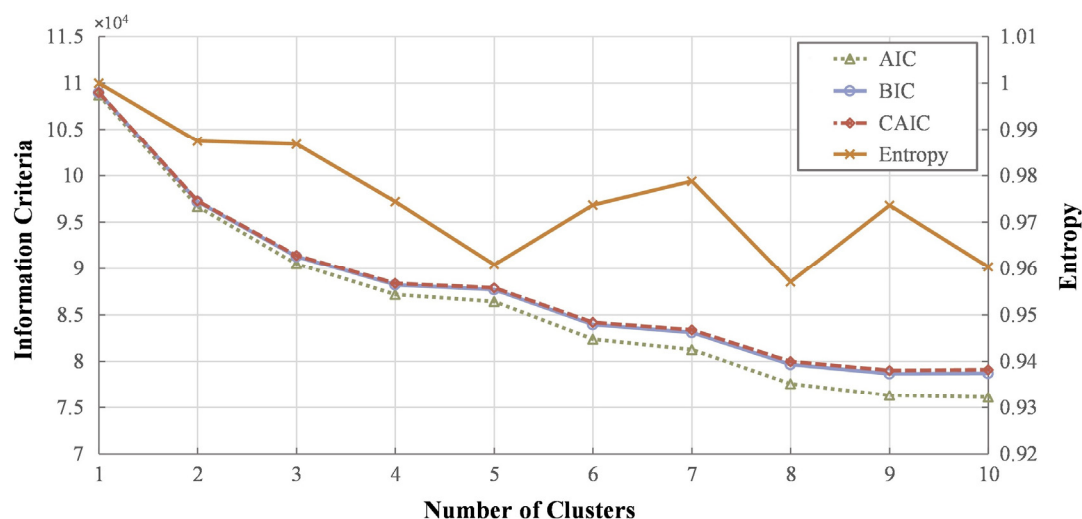
The probabilities ( $P_{kn}$ ) for the  $n$ th individual having the  $k$ th injury severity level are calculated when  $X$  equals 1 and 0. To calculate the marginal effect of random parameter variables, the mean value of the coefficients is applied in the utility function. By averaging the marginal effects of all observations, the marginal effects for each parameter are then estimated.



## 4. Results and Discussions

### 4.1. Latent Class Clustering Results

Pedestrian crashes were divided into clusters by LCA using variables in Table 1. In order to identify the most appropriate number of clusters, models with different numbers of clusters (ranging from one to ten) were conducted. AIC, BIC, CAIC, and entropy measures of these models are shown in Figure 1. It can be seen that all of the information criteria decreased by an increase in the number of clusters (Figure 1). As mentioned before, increasing the number of clusters does not necessarily minimize the information criteria values [81], so the percentage decrease in these criteria is computed in this study. The results indicate that the percentage difference of AIC, BIC, and CAIC declines below 1% from the four clusters. Furthermore, the entropy measure of the four clusters is 0.97, which indicates the clear separation between clusters and satisfactory fitness of the model. Accordingly, four clusters were selected for dividing the pedestrian crash dataset. The number of observations is 646, 2320, 1675, and 1574 for clusters 1–4, respectively. Table 2 shows the size of each cluster.



**Figure 1.** Information criteria and entropy for different number of clusters. Information criteria axis shows AIC (Akaike information criterion), BIC (Bayesian information criteria), and CAIC (consistent Akaike information criterion) values.

**Table 2.** Cluster summary.

Dataset	Proportion of Whole Dataset	Number of Observations
Cluster 1	10.39%	646
Cluster 2	37.32%	2320
Cluster 3	26.96%	1675
Cluster 4	25.33%	1574
Overall sample	100%	6215

Following previous works (e.g., [29,30,54,56,63,67]), the univariate skewed distribution of variables within clusters was calculated. The differences in variable proportions between clusters can reveal new information and show crash patterns for each cluster. Moreover, each cluster can be characterized and named by the skewed distribution of certain variables [63]. For instance, if in one cluster, crashes that occurred at night are over-represented, while the other clusters have a more balanced distribution over this variable, one can describe this cluster as the ‘nighttime crashes’ cluster.

Important variables with an unbalanced and skewed distribution that were used for profiling clusters are presented in Table 3. In cluster 1, 98% of crashes occurred in commercial land uses, and 99.7% of them occurred in high-density areas. From the results,

it can be concluded that this cluster also includes crashes in specific road characteristics. All crashes in this cluster happened on undivided two-way roadways wider than 20 m with vegetated buffers. So, cluster 1 can refer to crashes that occurred in the mentioned type of road design in high-density commercial areas.

**Table 3.** Distribution of featured variables describing cluster characteristics.

Variables	C 1	C 2	C 3	C 4	O.S.
Aged 15–30	19.81%	25.43%	<b>44.18%</b>	24.71%	29.72%
Motorcycle	15.79%	<b>47.46%</b>	18.03%	13.02%	27.51%
No traffic control	15.63%	0.26%	<b>93.43%</b>	7.62%	28.83%
Undivided two-way	<b>100.00%</b>	2.84%	59.88%	1.33%	27.93%
Road width: >20 m	<b>100.00%</b>	41.72%	28.12%	48.54%	45.84%
Without usable sidewalk	15.48%	12.93%	22.81%	<b>74.59%</b>	31.47%
Vegetated buffer	<b>99.38%</b>	40.91%	57.37%	58.01%	55.75%
Without park lane	49.23%	15.69%	<b>91.82%</b>	33.29%	44.15%
With park lane	50.77%	<b>84.31%</b>	8.18%	66.71%	55.85%
Near overpass/underpass	60.53%	16.72%	<b>71.88%</b>	25.86%	38.46%
Low AADT	32.66%	47.76%	3.22%	<b>76.18%</b>	41.38%
High density (>200 person/km <sup>2</sup> )	<b>99.69%</b>	<b>96.29%</b>	51.34%	0.00%	60.14%
Low density (<100 person/km <sup>2</sup> )	0.31%	0.00%	9.85%	<b>63.34%</b>	18.73%
Commercial land use	<b>97.99%</b>	24.09%	22.63%	22.17%	30.89%
Other land uses	0.00%	0.00%	20.66%	<b>54.70%</b>	19.42%

Note: O.S. stands for the overall sample and C1 to C4 for clusters 1 to 4.

Cluster 2 overlaps with cluster 1 on high population density, but it differs in overrepresentation of motorcyclist vehicle type. Moreover, 84.31% of the crashes in cluster 2 happened in places that park lanes existed. So, considering the skewed distribution of these three variables, this cluster can be identified as motorcycle-involved crashes in densely populated areas on roadways with park lanes. Cluster 3 is overrepresented in terms of young pedestrians (15–30), while other clusters have a more balanced distribution over this variable. Moreover, most crashes in this cluster occurred in places with no traffic control. Furthermore, over 70% of crashes in this cluster happened near pedestrian overpasses/underpasses, and most crashes occurred in paths without park lanes. Therefore, this cluster can be described as young pedestrian crashes near pedestrian overpasses/underpasses with no traffic control and park lanes. Cluster 4 contains crashes that are overrepresented in other land uses (e.g., recreational, vacant, industrial). Crashes in this cluster mainly occurred in low-density areas. In addition, around 75% of crashes happened in places with no usable sidewalks, and 76.2% of crashes were in low traffic volume days. Cluster 4 can, therefore, be identified as crashes that occurred in low density, low traffic volume, not commercial or residential land uses, in paths with no proper sidewalks.

From these results, it can be concluded that each cluster shows a certain crash pattern, and investigating cluster characteristics can provide important information. An interesting point here is that the trace of some hidden factors in the causation of crashes can be perceived by focusing on the crash patterns of each cluster. For example, in cluster 3, the trace of pedestrians' risky behavior (violations in crossing from an inappropriate place) can be noticed. As such, in cluster 4, the weakness of pedestrian facilities is more evident compared to other clusters. Moreover, it can be understood that in some regions of the city (low population density, not commercial or residential land uses, low AADT), the necessary measures to ensure the safety of pedestrians have not been sufficiently deployed.

Profiling clusters and assigning specific crash patterns to each cluster were only proceeded to reach a more profound perception of each cluster. This understanding can further help to describe severity analysis results for each cluster more precisely. However, it should be noted that the variables used to define the clusters are not necessarily associated with the significant parameters in the crash severity model.

#### 4.2. Mixed Logit Models Results

In this study, the mixed logit model for whole data and all of the clusters were conducted to explore the influence of different factors on pedestrian injury severity. These models were developed with Nlogit software. Minor injury severity is selected as the referent outcome category in all models. The modelling estimation results of significant variables for all five mixed logit models are shown in Table 4, while the marginal effects of these estimations are presented in Table A1 (see Appendix A). In Table 4, standard deviations (SDs) show the random distribution of the variables. In correspondence with some previous studies [28–30,63,89], a confidence level of 90% is used in the present study. In the following sub-sections, the impact of contributing factors on pedestrian injury severity, mainly the fatal injury, will be discussed thoroughly.

**Table 4.** Mixed logit estimation results for whole data and clusters.

Variables	Severity	Overall Sample		Cluster1		Cluster2		Cluster3		Cluster4	
		Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Constant	Major	−0.17	0.40	−4.77 ***	1.71	−5.02 **	1.96	−0.90	0.62	0.70	0.78
	Fatal	−3.28 **	1.45	−2.28 *	1.28	−3.88 **	1.56	−3.59 ***	1.03	−2.80 **	1.37
Gender (ref. female)											
Male	Major	0.31 ***	0.11					0.34 **	0.13		
	Fatal	0.76 **	0.36			1.49 ***	0.54				
Age (ref. 15–30)											
<15	Fatal	2.27 ***	0.82							2.65 ***	0.69
	SD	2.37 ***	0.71								
30–45	Major							0.28 *	0.17		
	Fatal					2.22 **	0.87	1.01 ***	0.39		
45–65	Major	0.71 ***	0.15	1.60 **	0.71	1.32 ***	0.48	0.51 ***	0.17		
	Fatal	4.25 ***	0.71	3.06 ***	1.10	3.25 ***	0.96	1.69 ***	0.39	2.19 ***	0.66
>65	Major	0.88 ***	0.20	3.78 ***	0.96	0.98 **	0.46	0.70 **	0.27		
	Fatal	4.39 ***	0.95	4.49 ***	1.28	6.08 ***	1.57	2.85 ***	0.44	4.63 ***	0.74
Vehicle (ref. passenger car)											
Motorcycle	Major	−0.40 ***	0.12					−0.37 **	0.15	−0.42 **	0.19
	Fatal	−1.04 **	0.48	−2.28 **	0.98	−1.25 **	0.61				
Heavy vehicle, bus	Major	1.62 ***	0.32			1.91 **	0.91	1.30 ***	0.39	1.24 ***	0.42
	Fatal	5.89 ***	0.78			4.75 ***	1.45	2.94 ***	0.47	3.77 ***	0.67
Minibus, van	Major	2.50 ***	0.66			2.69 *	1.55	1.76 *	0.93	1.79 **	0.78
	Fatal	5.47 ***	1.17			5.54 ***	2.01	3.63 ***	1.03		
Pickup	Major	0.79 ***	0.30			1.68 **	0.84	0.93 ***	0.36		
	Fatal	3.47 ***	0.75			2.55 **	1.01	2.24 ***	0.48	1.65 **	0.79
Bicycle	Major	−1.36 *	0.79	−2.91 **	1.46						
Time (ref. 22–6)											
6–10	Major	−0.41 *	0.20					−0.45 *	0.24		
	Fatal	−2.36 ***	0.59	−2.55 **	1.27	−2.89 ***	0.92	−0.76 **	0.38		
10–14	Major	−0.51 ***	0.19	3.36 **	1.51			−0.49 **	0.22		
	Fatal	−3.27 ***	0.64	−3.87 ***	1.29	−3.59 ***	1.03	−1.94 ***	0.43	−1.80 ***	0.65
14–18	Major	−0.38 **	0.19	2.88 *	1.51						
	Fatal	−4.62 ***	0.86	−3.44 **	1.46	−3.80 ***	1.11	−1.02 ***	0.39	−2.37 ***	0.67
	SD	2.87 ***	0.61								
18–22	Major			3.67 **	1.55						
	Fatal	−2.81 ***	0.60	−2.18 *	1.12	−3.62 ***	1.07	−1.62 ***	0.40		
Day type (ref. weekday)											
Weekend	Fatal	−0.79 **	0.34							−0.92 **	0.38
Weather (ref. clear)											
Adverse	Major	0.65 ***	0.19			1.83 **	0.79				
	Fatal					1.30 *	0.77	0.91 **	0.42		
Season (ref. winter)											
Spring	Major			−1.40 *	0.77						
Summer	Fatal							−0.67 *	0.38		
Junction (ref. no)											
Yes	Major	−1.14 ***	0.18					−0.75 ***	0.18		
	SD	1.94 ***	0.48								
	Fatal	−1.15 ***	0.20					−1.97 ***	0.40	−6.43 ***	0.67
Hit and run (ref. no)											
Yes	Major	1.47 ***	0.21	3.58 *	1.99	3.19 **	1.38	1.01 ***	0.18	1.19 ***	0.30
	SD	3.30 ***	0.69	10.04 *	5.93	6.39 *	3.55				
	Fatal	1.52 ***	0.50			1.33 **	0.65	0.90 **	0.36	1.05 *	0.55
Posted speed (ref. 60 km/h)											
40–60 km/h	Major	−0.31 ***	0.12	−2.30 ***	0.79					−0.35 *	0.20
	Fatal					−1.24 **	0.63				
<40 km/h	Major	−0.88 ***	0.20			−1.02 **	0.45	−0.69 ***	0.22	−1.33 **	0.56
	SD	2.37 ***	0.65					1.71 **	0.70	3.76 **	1.46
	Fatal	−0.89 **	0.44			−2.03 ***	0.74	−0.85 **	0.37		

Table 4. Cont.

Variables	Severity	Overall Sample		Cluster1		Cluster2		Cluster3		Cluster4	
		Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Traffic control (ref. none)											
Signals	Major	−1.47 ***	0.30							−0.98 **	0.49
	Fatal	−2.64 ***	0.49								
	SD	3.44 ***	1.01								
Signs/Surface markings	Major	−0.67 **	0.33								
	SD	3.64 ***	1.04								
	Fatal	−2.32 ***	0.74			−1.47 *	0.89				
	SD	2.79 ***	0.66								
Road type (ref. one-way)											
Divided two-way	Major					1.38 *	0.71				
	Fatal	0.98 *	0.54					−1.32 **	0.62		
Undivided two-way	Fatal	3.34 ***	0.85							5.30 ***	1.57
Road width (ref. <20 m)											
>20 m	Fatal					0.99 *	0.60	2.77 ***	0.66		
Sidewalk (ref. no)											
Yes	Major	−0.94 ***	0.25							−1.13 *	0.62
	Fatal	−1.66 ***	0.63							−1.70 **	0.74
Vegetation (ref. no)											
Yes	Major							0.60 ***	0.18		
	Fatal	−0.96 **	0.39			−2.09 ***	0.65				
Park lane (ref. no)											
Yes	Major	0.31 *	0.17			2.58 **	1.01	1.61 ***	0.54		
	Fatal							5.08 ***	0.79		
Overpass/Underpass (ref. no)											
Yes	Major	1.49 ***	0.23	2.10 ***	0.68	3.23 ***	1.02	1.77 ***	0.19	1.32 ***	0.33
	Fatal	1.74 ***	0.45			3.11 ***	0.85	2.23 ***	0.46	1.47 ***	0.44
AADT (ref. low)											
High (>30,000)	Major							−1.32 **	0.51	1.47 **	0.63
	Fatal					−2.59 **	1.11	−2.37 ***	0.67	3.93 ***	0.85
Medium (15,000–30,000)	Major							−0.98 *	0.51		
	Fatal	−1.68 ***	0.50					−1.83 ***	0.61		
Density (ref. <100 person/km <sup>2</sup> )											
100–200 person/km <sup>2</sup>	Fatal	−1.12 *	0.61					−1.35 ***	0.49	0.83 *	0.48
>200 person/km <sup>2</sup>	Fatal							−0.96 **	0.40		
Land use (ref. commercial)											
Other	Major	0.40 **	0.19								
	Fatal	2.41 ***	0.65					0.88 *	0.50		
Residential	Major					1.64 **	0.67			−0.59 *	0.31
	Fatal			3.44 **	1.38					−3.11 ***	0.76
	SD									2.05 ***	0.71
Model performance											
Restricted log likelihood		−6827.87		−709.70		−2548.78		−1840.18		−1729.22	
Log likelihood at convergence		−4572.91		−435.63		−1708.99		−1191.74		−1030.65	
AIC		9327.83		969.30		3578.00		2525.50		2225.30	
Pseudo r2		0.33		0.39		0.33		0.35		0.40	

Note: Only significant variables shown in the table. SD = Standard deviation. S.E. = Standard error. \* Significant at 90% confidence level. \*\* Significant at 95% confidence level. \*\*\* Significant at 99% confidence level.

#### 4.2.1. Pedestrian Characteristics

The results indicate that pedestrian gender significantly affects the severity in the overall sample and certain clusters. Based on the results in Tables 4 and A1, male pedestrians are more likely to be involved in crashes with major injuries in the overall sample and cluster 3. Male pedestrians are also more exposed to fatal injuries in the overall sample and cluster 2. According to marginal effects, the probability of major injury can be increased by 3.6% and 5.4% in the overall sample and cluster 3, respectively. The probability of fatal injuries can be increased by 1.24% and 3.24% in the overall sample and cluster 2.

This result is in line with some previous studies [26,39,48,90,91], and it can be explained through behavioral differences between men and women. According to previous findings, men are more likely to engage in risky behaviors, while women have more sensitivity toward traffic safety [92,93].

The estimation results reveal that age can significantly impact injury severities. For pedestrians under 15 years old, there is a significant increase in fatal crash probabilities for the overall sample and cluster 4 by 1.61% and 1.89%, respectively. This is in line with previous findings [26,41]. This result can be justified by the weaker physical strength of this age group compared to young adult pedestrians (15 to 30 years old). Additionally, children

have lower risk perception and experience, which can lead to more severe crashes. For instance, it was found that children are more likely to dart and dash into the streets [94] or be solely at fault in crashes [47].

As reported in Table 4, it was found that in the overall sample, age under 15 has random effects on fatal injury level. This may reflect differences in that age group and the impact of many observed or unobserved factors on the injury outcome of this age group. For instance, pedestrians under 15 can vary considerably in terms of factors such as physical strength, behaviors, and risk perception. The presence of children's parents at the time of the crash can also affect the injury severity of this age group. This variable has a fixed effect on all other clusters, indicating that the latent class analysis effectively diminished the corresponding heterogeneity for this variable.

The results show that pedestrians aged 30–45 are significantly more likely to suffer from fatal injuries in cluster 2 and cluster 3 by 1.28% and 1.32%. For pedestrians aged between 45 and 65, a significant increase in fatal severity was observed for whole data and clusters 1–4 by 2.43%, 3.47%, 2.33%, 2.52%, and 2.00%, respectively. Similarly, for pedestrians aged over 65 years, the probability of fatal pedestrian crashes was increased in whole data and clusters 1–4 (4.29%, 3.33%, 5.52%, 2.12%, and 6.02%, respectively). Similar results have been found in many previous studies [27,28,35,42,43,90]. Although older pedestrians are more risk-averse and have higher risk perception in roadway networks [92], their weakened physical attributes can cause critical injuries for this age group. Physical attributes include higher reaction time and lower elasticity of muscles and bones, leading to more severe injuries in traffic crashes than younger pedestrians [95,96].

#### 4.2.2. Involved Party Characteristics

As illustrated in Table 4, with regard to passenger cars, the other types of involved vehicles are significantly associated with injury severity. Pedestrian crashes with motorcycles tend to have a significantly lower probability of fatal injury severity in the overall sample (−0.48%), cluster 1 (−1.47%), and cluster 2 (−0.62%). The lighter weight, lower speed, and higher maneuverability of motorcycles than passenger cars lead to lower injury severities. This justification can be applied to bicycles, which also have a lower probability of major injuries in the overall sample and cluster 1. The estimation results of the overall sample and clusters 2–4 showed that heavy vehicles, busses, and pickup crashes with pedestrians have a significantly higher probability of fatal injuries than passenger cars. Similar results were also observed for minibuses and vans in the overall sample and clusters 2 and 3. Heavier weight and lower maneuverability endanger pedestrians to severe injuries in these types of vehicles compared to passenger cars [35,38]. Moreover, these vehicles are more likely to injure vital organs such as the head because of their larger dimensions and higher bumper altitude [10,41,42].

The severity of pedestrian injuries when the driver leaves the crash scene (hit and run) was investigated in the present study. In correspondence with previous studies, the results confirmed that injuries are more severe when the crash is hit and run [31,40,97]. As presented in Table A1, crashes where drivers leave the crash scene without reporting tend to have more likelihood of major injuries in the overall sample, and clusters 1–4 by 4.31%, 1.36%, 2.04%, 3.81%, and 2.63%, respectively. Furthermore, hit and run crashes have a higher probability of fatal injuries in the overall sample and clusters 2–4 by 0.48%, 0.32%, 0.20%, and 0.21%, respectively. The reason could be that, in these types of crashes, the much-needed crucial medical attention can be delayed [97]. This variable has a random effect through observations specific to major injuries in the overall sample and clusters 1 and 2. This result denotes that this parameter has unobserved heterogeneity across observations and its effect on injury severity is not always constant. Many factors, including crash characteristics, place of injury, distance from emergency departments, etc., can affect the injury outcome of this crash type.



#### 4.2.3. Temporal Characteristics

Regarding the time of day, injury severities are higher in mid-nights (10 p.m.–6 a.m.) than other times of the day. Marginal effects show that, in the overall sample and clusters 1–3, the probability of pedestrian involvement in fatal crashes significantly decreases when the crash occurs at 6–10, 10–14, 14–18, and 18–22, and in cluster 4, the probability of fatal crashes significantly decreases when the crash occurs at 10–18 (Table A1). This result is in line with previous findings (e.g., [37,98]).

The estimation result for the time between 14–18 (evening peak) indicates that this variable has random effects on fatal injuries in the overall sample, which specify unobserved heterogeneity in this period. Many unobserved conditions could affect crash severity and cause heterogeneity in observations at this time frame. For instance, at this time of the day on weekdays, most trips are from work to home, and road users often encounter traffic blockage; hence, factors such as fatigue and aggression can appear among some drivers, which lead to crashes with more severe injuries [99]. Additionally, the time of traffic peak and traffic volume can alter drastically on holidays and weekends in this period, and, as a result, the behavior pattern of road users can also change. Moreover, sun glare can occur in certain circumstances during this time and reduce visibility [100]. It was found that sun glare can lead to more severe injuries in pedestrian-related crashes [101]. This variable has a fixed effect in all other clusters, which shows the clustering method's efficacy in eliminating the heterogeneity for this variable.

It was found that injury severity can be affected by the type of day. As reported in Table A1, weekend crashes have a significantly lower probability of fatal injuries than weekday crashes by  $-0.72\%$  in the overall sample and  $-2.01\%$  in cluster 4. The differences in trip characteristics between weekdays and weekends may explain this result. On weekdays, trips are often commuting, while trips on weekends are usually related to recreational purposes. Previous findings denote that commuter trips are expected to be followed by drowsiness and inattentiveness, which may cause severe accidents [102,103]. Unlike weekdays, trips are usually for shopping or entertainment and away from stress and fatigue after work on weekends. In addition, on weekends, pedestrians usually spend time in recreational areas (such as parks, malls), and, in these places, where pedestrian activity is high, the traffic speed is lower. A similar result was also observed in previous studies [10,49].

#### 4.2.4. Environmental Characteristics

With regard to the weather condition, the results imply that, when the crash occurs in adverse weather, the likelihood of major injuries significantly increases in the overall sample and cluster 2 (marginal effects being  $0.97\%$  and  $1.16\%$ , respectively). Similarly, adverse weather can significantly increase the probability of fatal injuries in clusters 2 and 3 by  $0.24\%$  and  $0.57\%$ , respectively. Furthermore, investigating the effect of seasonal changes shows an association between this variable and pedestrian injury severity. Compared to crashes in winter, crashes in spring significantly decrease the probability of major injuries in cluster 1 by  $-3.08\%$ . Furthermore, crashes in summer can significantly lower the probability of fatal crashes in cluster 3 by  $-0.96\%$ .

The result is in line with previous studies, which confirm that the risk of severe and fatal injuries increases in winter [28,41]. Poor weather conditions can decrease visibility and pavement friction. Moreover, pedestrians can underestimate the vehicle speed in rainy weather [104]. Therefore, these aggravating factors can reduce the ability to appropriately react in dangerous situations and increase crash injury severity. This result is reasonable because Mashhad has the highest precipitation rate and the highest number of days with adverse weather in winter [105].

#### 4.2.5. Roadway and Built-Environment Characteristics

Based on the modelling results, posted speed is associated with pedestrian injury severity. Compared to the reference category ( $>60$  km/h), crashes that occurred in roadways

with the 40–60 km/h speed limit can decrease the probability of major crashes in the overall sample, cluster 1, and cluster 4 by  $-2.87\%$ ,  $-9.18\%$ , and  $-1.76\%$ , respectively, and decrease the probability of fatal crashes by  $-1.16\%$  in cluster 2. For speed limits below 40 km/h, the probability of major injuries decreased in the overall sample and clusters 2–4 by  $-2.61\%$ ,  $-2.52\%$ ,  $-2.85\%$ , and  $-0.17\%$ , respectively. Moreover, a posted speed below 40 km/h can decrease the probability of fatal injuries in the overall sample, cluster 2, and cluster 3 (marginal effects being  $-0.40\%$ ,  $-1.26\%$ , and  $-0.95\%$ , respectively). Previous international findings support this finding [26,27,37,38,61,95,106–108]. This result was expected because even a slight increase in speed can result in longer stopping distance, less decision time, less control on vehicles, and much more intense kinematic energy [106].

According to Table 4, for the speed limit below 40 km/h, there is a random effect on fatal crashes in the overall sample analysis. Moradi et al. [109] showed that drivers in Iran are far more likely to exceed the speed limit in the roadways with the lowest posted speed. This risky behavior could be an unobserved factor that can alter injury severities in areas with low posted speed and cause heterogeneity for this variable. Other factors such as the path's geometry (e.g., slope, curvature) can also change the severity of crashes at this speed. As illustrated in Table 4, this variable has fixed effects in all other clusters, confirming that the clustering analysis explicitly eliminated this variable's heterogeneity.

The results show that pedestrian crashes at midblock tend to generate more severe injuries than crashes at the junction of two or more paths. Crashes at junctions can decrease the probability of fatal crashes in the overall sample and clusters 3 and 4 by  $-3.44\%$ ,  $-8.21\%$ , and  $-13.23\%$ , respectively. This result aligns with previous studies [26–28,110]. The result was expected since drivers are more likely to reduce speed and pay more attention when they are in the vicinity of an intersection [10,28]. As shown in Table 4, this variable was found to have random effects across observations in the overall sample, specific to major injury severity. Various elements, such as junction geometry, presence of parked cars near the junction, lack of proper walkway and median at the junction, type of traffic control, and violation of the right of way and red-light, can affect pedestrian crashes and cause random effects for intersection crashes.

Based on marginal effects reported in Table A1, the presence of traffic signals can significantly reduce the probability of fatal crashes by  $-2.72\%$  in the overall sample and can lower the probability of major injuries in the whole data and cluster 4 (by  $-2.31\%$  and  $-1.58\%$ , respectively). In the presence of traffic signs and/or surface markings (crosswalks), crashes tend to have a lower probability of fatal crashes in the overall sample and cluster 2 (marginal effects being  $-0.46\%$  and  $-0.99\%$ ). Similar findings could be found in the study by Pour-Rouholamin and Zhou [96] that the presence of traffic control and crossing at crosswalks are associated with lower probabilities of severe injuries.

The variable traffic signal is a random parameter specific to the fatal severity level in the overall sample. Factors such as pedestrian crossing volume, traffic volume, and proximity of crossings to schools, bus stops, or subway stations can increase the pedestrian crash risk [106,111]. Moreover, risky behaviors such as pedestrian and driver red-light violations can influence pedestrian injury severity in signalized intersections [112].

Furthermore, the variable traffic sign also has a random effect on both major and fatal severities. Similar to signalized crossing, unobserved factors, such as traffic volume, pedestrian crossing volume, vicinity of crossings to school zones, bus stops, or other facilities, can affect crash risk. In these types of crossings, ignoring the right of way can also increase injury severities. Additionally, the severity of crashes in this type of traffic control can vary depending on the crash location (midblock crossing or intersection crossing). These variables are fixed on other clusters, which show that the random effects of these variables are minimized by clustering.

The severity analysis for road type shows that, compared to one-way roadways, divided two-way have a significantly higher probability of fatal injuries in the overall sample by  $1.52\%$ . This result was not expected because the medians are supposed to be a refuge for pedestrians. However, several studies have confirmed it [96,113]. This result

shows that, because the overall sample is highly aggregated, many other factors can also affect injuries besides the median presence in this type of roadway. For instance, the median is generally utilized in roadways with higher speeds [96,113] or places with a higher volume of a pedestrian crossing. This leads to a higher risk of severe injuries when a crash occurs. However, the contradictory result can be seen in cluster 3. In this cluster, the probability of fatal crashes in divided roadways decreases by  $-3.09\%$ . This result reveals the advantage of LCA in describing the crash data and illuminating the hidden effect of some factors by maximizing the heterogeneity between clusters. As discussed above, for this type of roadway in the overall sample, the aggregation of other aggravating factors (such as speed and volume) made the effect of median existence inconspicuous. The explanation here could be the vicinity of crashes in this cluster to pedestrian overpasses that are usually built on roadways with higher speed and higher pedestrian crossing demand. Therefore, the changes in pedestrian volume and vehicle's speed are largely restricted in this cluster, and the impact of the median's presence can be seen more clearly.

According to obtained results, undivided two-way roadways can increase the probability of fatal crashes in the overall sample and cluster 4 by  $2.42\%$  and  $0.74\%$ . Similar results were found by Zhai et al. [44]. The more complex traffic pattern in two-way paths can raise the chance of interaction between vehicles and pedestrians and may also increase pedestrian lapse or distraction, which can explain this finding [114].

Road width is associated with pedestrian injury severity. Accordingly, in roadways wider than 20 m, there is an increase in the probability of fatal crashes in cluster 2 ( $0.96\%$ ), and cluster 3 ( $4.79\%$ ). This result is supported by previous articles, which suggest that an increase in the number of lanes can increase the risk of more severe crashes [26,28,37,96]. Designing wider roadways is linked with higher vehicular demand and higher speed. Additionally, wider roadways need extra time to cross, raising the risk of crash for pedestrians, especially older ones [96,115]. Moreover, the chance of improper crossing increases in wider roadways [45].

The presence of usable sidewalks can significantly decrease the risk of major injuries by  $-8.71\%$  and  $-14.01\%$  and decrease the risk of fatal injuries by  $-2.34\%$  and  $-2.87\%$  in the overall sample and cluster 4. Researchers have reported similar results in previous articles [23,61,116]. The lack of essential facilities for pedestrians, such as sidewalks, forces pedestrians to use the street for walking, which can increase the risk of crashes.

In the present study, the influence of vegetated buffers in sideways or medians on pedestrians' injury severity was examined. Vegetated buffers can decrease the probability of fatal crashes in the overall sample and cluster 2 by  $-0.86\%$  and  $-2.4\%$ . In Iran, vegetation is often set up in places with better infrastructure for pedestrians and a safer road system, which may explain the lower injury severity of crashes that occur in roadways with vegetation. As suggested by Hanson et al. [116], pedestrian crash injuries were reduced by providing proper walkways and buffers. Increasing walkability by providing a safe walking environment through appropriate pedestrian facilities is an essential policy to reduce traffic casualties in modern urban design [116]. However, unexpectedly, the vegetated buffer can increase the risk of major injuries by  $9.27\%$  in cluster 3. This opposite result can be explained by possible visibility obstruction caused by vegetation in certain situations. According to some of the specifications of cluster 3 (vicinity to pedestrian bridges, no traffic signal), it can be concluded that this cluster includes routes that are not adapted for pedestrians. Therefore, vegetation in areas that are not pedestrian-friendly can act as a visibility obstruction for drivers and cause more severe injuries when pedestrians are involved in crashes. In line with this conclusion, Yue et al. [117] also suggested that vegetation in sidewalks or medians can prevent drivers from seeing pedestrians in time.

The effect of the park lane presence was explored in the severity analysis. It can be observed from Table A1 that the presence of parking lanes can significantly increase the probability of major injuries in the overall sample, cluster 2, and cluster 3 by  $4.03\%$ ,  $13.34\%$ , and  $0.38\%$ . The presence of parking lanes can significantly increase the risk of fatal injuries in cluster 3 by  $1.92\%$ . Parking lanes can be considered as temporary visibility

obstructions [117]. Accordingly, parked cars can cause drivers not to notice pedestrians until the last moment in certain situations.

Since there are few studies about visibility obstructions, it is recommended that further studies in this area explore the association of visibility obstructions, such as vegetation and park lanes, with pedestrian severity.

In Iran, constructing pedestrian bridges or underpasses is a common precautionary measure to increase safety in places with high pedestrian casualties. Accordingly, the number of pedestrian bridges in cities of Iran is relatively high and can be a burden for a developing country, considering the high cost of building these facilities. For instance, there are around 200 overpasses/underpasses constructed near pedestrian crash hotspots throughout the city of Mashhad.

Unfortunately, despite all these efforts, the number of crashes near pedestrian bridges is still high and thought-provoking. Therefore, the effect of the vicinity on pedestrian overpass/underpasses (300 m), which can be an influential variable considering the urban environment of Iran, was explored in this study. Estimation results show that crashes in the vicinity of pedestrian overpasses/underpasses are strongly associated with more severe injury for pedestrians. For major injuries, vicinity to pedestrian bridges or underpasses can increase the probabilities in the overall sample and clusters 1–4 by 13.87%, 14.48%, 5.63%, 20.14%, and 5.82%, respectively. This factor can also significantly increase the probability of fatal injuries in the overall sample and clusters 2–4 by 1.10%, 1.62%, 3.62%, and 0.86%, respectively.

At first glance, this result may seem contradictory, considering that these facilities were constructed to provide a safer environment for pedestrians. The potency of these facilities depends on the pedestrians' compliance with using them [118]. Many people may take risks and cross the main road to avoid the extra effort of using bridges or underpasses [119]. This decision can expose pedestrians, especially older ones, who may find it more challenging to use the stairs, to a high risk of crashes.

Therefore, to confront this issue and decrease the risk of crashes, there is a need to take extra preventive measures along with building pedestrian overpasses/underpasses. These countermeasures include changes in the geometry of the road environment, such as channelizing and segregating sidewalks from the main road by using barriers or fences and utilizing ramps and escalators for the convenient use of the elderly or physically disabled pedestrians. Likewise, informing people about the importance of using safe crossings through targeted educational campaigns and applying more attractive and engaging designs to encourage pedestrians to use them may also influence pedestrians to use the overpasses or underpasses. Additionally, it is important to prevent possible breakdowns of these facilities through constant maintenance and monitoring. In some situations, other preventive solutions such as traffic calming, may be even more efficient than constructing an expensive overpass or underpass, so safety agencies and experts should evaluate different preventive scenarios beforehand.

Despite the importance of this facility, particularly in developing countries, its effect on pedestrian crashes has not been sufficiently investigated in previous studies of this field. Therefore, it is suggested that future studies investigate this effect more precisely.

Traffic characteristics such as traffic volumes can strongly affect pedestrian injuries in crashes [31]. The time variable, which was investigated earlier in this study, provided a helpful perception of the influence of temporal changes in traffic volumes on pedestrian crash severities. In order to explore the effect of traffic volume on pedestrian injury severities more profoundly, roadways' annual average daily traffic (AADT) was obtained from the Mashhad transportation department and allocated to each crash location for further analysis. The severity analysis shows that, compared to low AADT, pedestrian crashes that occurred in roadways with medium AADT tend to have a significantly lower likelihood of fatal injuries by  $-0.78\%$  in the overall sample and  $-4.17\%$  in cluster 3. For high AADT, the influence on injury severity is varied between clusters. This variable can decrease the likelihood of fatal injuries in clusters 2 and 3 by  $-1.01\%$  and  $-4.13\%$ ,

respectively, and increase the likelihood of fatal injuries in cluster 4 by 1.25%. It was found in previous literature that higher volume can increase the frequency and risk of pedestrian–vehicle crashes [22,120]. However, for injury severity, opposing results were reported by previous studies, and the effect of higher traffic volumes was found to both aggravate [121] and alleviate [97] pedestrian injury severities.

The explanation could be that, in higher volumes, vehicle speed is lower. Moreover, when traffic volume is high, pedestrians tend to be more alert and cross the road more cautiously [122]. For instance, in cluster 3, considering some of the characteristics such as no traffic control or vicinity to the pedestrian bridge, pedestrians are probably aware that they are crossing from the wrong and high-risk place. Therefore, with the increase in passing vehicles, pedestrians are extra careful in crossing the street. Careful crossing, along with the slower speed of vehicles, can lead to lower injury severities of pedestrian–vehicle crashes, as observed in cluster 3. However, this effect is the opposite in cluster 4. In this cluster, most crashes occurred in places with no proper sidewalks. Therefore, a lack of sidewalks can lead to improper crossing and lapses, and, considering the higher number of cars, the risk of severe crashes can increase. Additionally, in high traffic volumes, drivers may use the sides of the roadway and shoulder to overtake and may hit the pedestrians standing or walking at the side of the roadway. This difference in the direction of estimation results shows that, depending on the situation, the impact of traffic volume can be very inconsistent, and many unobserved and observed factors (such as pedestrian volume, road geometry, traffic violations) can affect traffic volume impact on injury severities.

The city of Mashhad comprises 13 municipal districts. Demographic information is available for each district. In this study, the population density was assigned to each observation according to the district where the crash occurred. The results indicate that, compared to crashes in low population density areas, areas with medium population density (100–200 person/km<sup>2</sup>) have a lower probability of fatal injuries in the overall sample and cluster 3 by −0.39% and −2.12%. Likewise, crashes in areas with high population density (>200 person/km<sup>2</sup>) have a lower probability of fatal injuries in cluster 3 by −2.15%. A similar result was also obtained in previous studies [29,52,123]. The population and population density were considered a surrogate for pedestrian exposure and activity in many studies (e.g., [22,24,124]). Accordingly, it can be assumed that higher density represents higher pedestrian activity. In places with higher pedestrian activity, drivers are more cautious and drive slower, which may explain the lower injury severity in places with higher density.

However, a different result was estimated in cluster 4, which indicates that crashes in medium-density areas have a higher likelihood of fatal injuries by 1.55% than those in low-density areas. Lack of sidewalks and low volume roadways, which lead to higher speeds, can explain the result for this cluster. This shows that higher pedestrian activity in places with inadequate pedestrian facilities can lead to more severe injuries in pedestrian–vehicle crashes.

It has been found that land use can significantly impact pedestrian injury severities in crashes [27,28,38,43,114]. In this study, Mashhad land uses were classified into three categories of residential, commercial, and other (e.g., recreational, industrial, vacant). Each crash was then assigned to one of these categories according to the adjacent land use to the crash location. Marginal effect analysis indicates that crashes in residential land use have a higher likelihood of major injuries in cluster 2 by 13.85% and a higher likelihood of fatal injuries in cluster 1 by 0.83%. Similarly, crashes in other land uses (e.g., recreational, industrial, vacant) have a significantly higher likelihood of major and fatal injuries by 0.97% and 1.72% in the overall sample and a significantly higher probability of fatal injuries (1.56%) in cluster 3.

Compared to residential and other land uses, in areas with commercial land use, pedestrian activity is higher, and, accordingly, drivers are more careful, and traffic speed is slower. Furthermore, lack of parking space, mainly for commercial land use, is always a problem in populated cities of Iran. Therefore, the lack of parking space causes more traffic



near commercial areas, which reduces vehicle speeds. Additionally, commercial areas benefit from better lighting condition at night, which provides better visibility for drivers.

However, these justifications can be highly affected by other factors, including road geometry, time of day, day type, quality of pedestrian facilities, type of travel (commuter or not), and so on. As we can see in cluster 4, crashes in residential areas, tend to have a significantly lower likelihood of major and fatal injuries by  $-1.53\%$  and  $-2.58\%$ . In addition, this variable is found to have a random effect specific to fatal injuries in this cluster.

#### 4.3. Model Evaluation

Besides investigating the factors affecting pedestrian-involved crash severity, one of the objectives of this study was to evaluate the effectiveness of the performed model on segmentation and reduce the heterogeneity of pedestrian crash data. Reviewing and comparing the results of the models conducted on the overall sample and each cluster revealed some important observations that will be discussed in this section. Table A2 (see Appendix A) summarizes the differences between the two-step approach and the single mixed logit model.

The following information about contributing variables of each cluster can be perceived from this table: (a) The case that variable effect is significant in both overall and the cluster model, (b) the case that variable effect is just significant in the cluster model, (c) The case that variable effect in the cluster is opposite in the overall sample, (d) The case that variable effect is random in the overall sample and fixed in the cluster (e) The case that variable effect is fixed in the overall sample and random in the cluster. Cases a, b, and c have been shown in previous studies (e.g., [28–30,67]). Additionally, cases d and e are observed in the articles of Chang et al. [67] and Li et al. [86].

As illustrated in Table A2, some variables that are not significant in the pooled data model are significant in cluster-based models. These results confirm this assumption that some significant relations can be obscured because of the heterogeneous nature of crash datasets [63]. Moreover, the results of each model have some differences in terms of variables' impact on injury severity levels, which implies that variables' impact can be changed in different or even opposite specific crash patterns.

In addition, while some variables have random effects in the overall sample, their effect is fixed in segmented data. These results indicate that clustering eliminated the heterogeneity that caused random effects in the whole data. Furthermore, the clustering analysis reduced the heterogeneity of crash data, but some variables with random effects can still be observed in sub-datasets. This finding suggests that unobserved heterogeneity can still exist within each sub-dataset even after clustering, which shows the importance of using mixed logit after clustering analysis.

Other than the above explanations, comparing goodness of fit measures indicates that the proposed two-step approach performed better than the single mixed logit model in terms of AIC and log-likelihood values (Table 4).

Overall, it can be said that the approach used in this study can appropriately minimize the heterogeneity and provide more information about the contributing factors. It is robust, considering that many unobserved factors can impact the outcome in such datasets. Therefore, it provides a more reliable insight into the cause of crashes, and, accordingly, effective strategies and policies can be implemented to reduce casualties in traffic crashes.

## 5. Conclusions and Recommendations

Pedestrians are the most vulnerable road users, and the high rate of severe injuries and fatalities in pedestrian-involved crashes made pedestrian safety a public health concern. Therefore, many researchers have attempted to understand the contributing factors to pedestrian injury severities. However, considering the high number of pedestrian casualties in developing countries, the number of studies in this field is insufficient, mainly due to the lack of comprehensive and detailed data. Therefore, there is a need to conduct a comprehensive analysis of pedestrian injury severity in developing countries.

The present study explores the interaction between pedestrian injury severities and a rich set of determinants representing the environment, demographics, geometric design, traffic, pedestrian, and involved vehicle characteristics. Moreover, this study investigates some variables that have been less accounted for in previous studies (e.g., the vicinity to overpasses, the existence of vegetated buffers, and park lanes). Investigating the effect of these variables helps to better understand the road traffic interactions in an environment such as the urban context of Iran.

For this purpose, a two-step approach by combining latent class clustering analysis and the mixed logit model was carried out. Subsequently, the performance of this approach was compared with the single mixed logit model on the overall sample. The results indicated that this method could reduce the heterogeneity in pedestrian crash data to a great extent. Furthermore, several contributing factors were found to have an impact on the probability of injury outcomes in pedestrian-involved crashes. Recognizing the underlying risk factors and understanding their interaction in pedestrian-involved crashes is very important from a practical viewpoint. Moreover, categorizing the data and determining the random parameters led to a deeper understanding of the unseen factors and hidden relationships. The robust approach provides reliable information for authorities in policy implications and road safety agencies to enhance pedestrian safety with more effective countermeasures.

The findings from the study can assist authorities in developing specific strategies and action plans to reduce pedestrian-involved crashes and improve the safety of pedestrians. For example, certain promotional campaigns and educational interventions could be implemented to increase awareness about traffic safety among pedestrians who are more prone to severe crashes, such as male pedestrians and children. Furthermore, the results indicate that the urban environment is not properly adapted for older pedestrians. Therefore, actions such as installing pedestrian signal countdowns [27], constructing medians in wide crossings, and improving sidewalk conditions should be deployed to increase the safety of this vulnerable group in the urban environment.

The educational campaigns could also target drivers of certain vehicles, such as heavy vehicles, buses, and pickups, to make them more cautious and decrease their risky behaviors. Furthermore, the exposure of pedestrians to heavy vehicles should be reduced as much as possible [27,28]. For this purpose, heavy vehicles movement could be restricted in certain areas and periods of time with high pedestrian activity. Additionally, increasing the safety of bus routes for pedestrians, providing safe access to bus stops, and using safer buses can reduce the risk of pedestrian–bus crashes.

Furthermore, providing consistent street lighting, especially at midnight in paths with poor lighting conditions, can reduce the severity of crashes by enhancing visibility in dark hours and adverse weather [10,28].

The higher injury severities in roadways with higher posted speeds reveal the importance of speed reduction and traffic calming policies, especially in places with high pedestrian activity and near crossings. Improving geometric design by installing facilities such as chicanes and refuge islands or even reducing the speed limit in these areas could result in lower injury severities.

In addition, as explained in the discussion section, the random effect of posted speed lower than 40 km/h could show a possibility of speed violation in certain areas, especially in paths with the lowest speed limits. Therefore, appropriate measures must be taken to prevent this violation, such as educational and training campaigns for delinquent drivers and stricter law enforcement and monitoring.

Crossing in places with no traffic control could lead to more severe injuries. By warning pedestrians about the dangers of crossing from places with no traffic control, they can be encouraged to cross from safer places. Furthermore, increasing the number of crosswalks in midblock with high crossing demand is one of the preventive actions in this regard. Moreover, restricting pedestrians from crossing in roadway sections with no traffic controls (e.g., installing barriers or fencing) can reduce the number of casualties caused by crossing from places with no traffic control. Moreover, it is recommended that

authorities install signs and marked crosswalks or, if possible, signals at junctions with no traffic control. Additionally, as the random effect of signal control was observed in this study, policies and measures should be employed to prevent pedestrians and drivers from red-light violations.

According to the results, the safety of unseparated two-way roadways and wide roadways should be increased with geometric corrections such as constructing raised medians or refuge islands and reducing the roadway width in the vicinity of crossing points. The absence of usable sidewalks plays a vital role in increasing pedestrian accidents. Therefore, installation of sidewalks in places without them, increasing the width of narrow sidewalks, complete segregation of the sidewalks from the roadway by installing sidewalk buffers, and proper paving of sidewalks are the measure that can considerably reduce the severity of accidents [23,116].

Drivers' inadequate information about subsequent legal consequences of crashes could push their decision to leave the crash scene. In dealing with hit and run crashes, increasing awareness of drivers through targeted educational campaigns could effectively decrease this type of crash [125].

Regarding pedestrian overpasses/underpasses, some strategies, such as restricting access to the roadway by fencing, utilizing ramps and escalators, etc., can be implemented. It was found that the existence of parking lanes could increase the risk of more severe injuries. Therefore, removing the parking space in places that cause sight obstruction, especially near the midblock crossing locations and intersections, and replacing it with curb extensions can provide a better sightline for both pedestrians and drivers and shorten pedestrians' crossing distance. Additionally, the results showed that, in certain situations, the presence of vegetation could increase injury severities, so in places where vegetation restricts vision, the plants should be relocated or pruned.

Overall, this study presented some insightful results concerning pedestrian crashes in a developing country such as Iran. These findings can act as a resource for employing safety strategies by policymakers to reduce pedestrian casualties. However, some limitations exist in this research, which is related to the crash database. The impact of some factors, such as driver characteristics, was not considered in this study due to the unavailability of these factors in the pedestrian crash dataset. The possibility of under-reporting cases for crashes with minor injury severities is another limitation. Moreover, future studies can use correlated random parameter models to investigate pedestrian crash injury severities. Using correlated RPL models with clustering analysis may provide a deep understanding of how some combinations can influence pedestrian safety. Furthermore, future research can focus on temporal correlations or temporal instabilities for investigating pedestrian injury severities as they could provide insight into the time trend of crashes over time.

**Author Contributions:** Conceptualization, A.E. and K.A.; methodology, A.E. and K.A.; software, A.E.; validation, A.E., K.A., and N.S.; formal analysis, A.E.; investigation, A.E. and K.A.; data curation, A.E.; writing—original draft preparation, A.E.; writing—review and editing, A.E., K.A. and N.S.; supervision, K.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data can be made available by contacting the first coauthor, Arsalan Esmaili (arsalan.esmaili@ut.ac.ir).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

Table A1. Marginal effects of all models.

Variables		Sev.	O.S. (%)	C1 (%)	C2 (%)	C3 (%)	C4 (%)
Gender (ref. women)	Men	Min.	−4.85		−1.19	−6.02	
		Maj.	3.61		−2.06	5.40	
		Fat.	1.24		3.24	0.62	
Age (ref. 15–30)	<15	Min.	−0.51				−1.53
		Maj.	−1.09				−0.36
		Fat.	1.61				1.89
	30–45	Min.			−0.08	−2.30	
		Maj.			−1.20	0.96	
		Fat.			1.28	1.32	
	45–65	Min.	−4.29	−5.72	−4.44	−3.69	−1.74
		Maj.	1.87	2.27	2.12	1.17	−0.26
		Fat.	2.43	3.47	2.33	2.52	2.00
	>65	Min.	−5.38	−5.36	−6.45	−3.18	−3.50
		Maj.	1.08	2.03	0.93	1.07	−2.52
		Fat.	4.29	3.33	5.52	2.12	6.02
Vehicle (ref. passenger car)	Motorcycle	Min.	2.66	1.52	1.61	2.39	1.76
		Maj.	−2.17	−0.03	−0.99	−2.12	−2.15
		Fat.	−0.48	−1.49	−0.62	−0.27	0.39
	Heavy vehicle, bus	Min.	−1.87		−1.14	−1.31	−1.59
		Maj.	0.23		0.08	−0.59	0.03
		Fat.	1.64		1.07	1.89	1.55
	Minibus, van	Min.	−0.88		−0.35	−0.31	−0.50
		Maj.	0.22		0.03	−0.11	0.47
		Fat.	0.66		0.30	0.42	0.03
	Pickup	Min.	−0.76		−0.90	−1.04	−0.24
		Maj.	0.22		0.50	−0.08	−0.23
		Fat.	0.54		0.41	1.11	0.45
	Bicycle	Min.	0.09	0.44			
		Maj.	−0.33	−0.42			
		Fat.	0.24	−0.02			
Crash time (ref. 22–6)	6–10	Min.	1.93	−0.95	1.70	2.07	
		Maj.	−1.11	4.17	−0.03	−1.37	
		Fat.	−0.82	−3.23	−1.67	−0.71	
	10–14	Min.	3.60	−7.85	3.18	3.63	1.83
		Maj.	−2.53	9.93	−1.01	−2.06	−0.45
		Fat.	−1.08	−2.09	−2.18	−1.58	−1.38
	14–18	Min.	2.93	−5.58	2.69	1.61	0.36
		Maj.	−1.59	8.33	−0.23	−0.44	1.83
		Fat.	−1.34	−2.75	−2.46	−1.17	−2.19
	18–22	Min.	1.37	−7.88	0.48	2.94	
		Maj.	−0.19	10.70	1.67	−1.34	
		Fat.	−1.18	−2.82	−2.15	−1.61	
Day type (ref. weekday)	Weekend	Min.	2.47				3.83
		Maj.	−1.75				−1.82
		Fat.	−0.72				−2.01
Weather (ref. clear)	Adverse	Min.	−1.36		−1.40	−0.80	
		Maj.	0.97		1.16	0.23	
		Fat.	0.40		0.24	0.57	
Season (ref. winter)	Spring	Min.		3.95			
		Maj.		−3.08			
		Fat.		−0.86			
	Summer	Min.				1.14	
		Maj.				−0.18	
		Fat.				−0.96	

Table A1. Cont.

Variables		Sev.	O.S. (%)	C1 (%)	C2 (%)	C3 (%)	C4 (%)
Junction (ref. no)	Yes	Min.	5.11			16.14	19.88
		Maj.	−1.67			−7.94	−6.65
		Fat.	−3.44			−8.21	−13.23
Hit and run (ref. no)	Yes	Min.	−4.79	−1.58	−2.37	−3.99	−2.84
		Maj.	4.31	1.36	2.04	3.81	2.63
		Fat.	0.48	0.21	0.32	0.20	0.21
Posted speed (ref. 60 km/h)	40–60 km/h	Min.	2.53	7.67	1.94		1.14
		Maj.	−2.87	−9.18	−0.77		−1.76
		Fat.	0.36	1.52	−1.16		0.62
	<40 km/h	Min.	3.01		3.78	3.80	0.63
		Maj.	−2.61		−2.52	−2.85	−0.17
		Fat.	−0.40		−1.26	−0.95	−0.45
Traffic control (ref. none)	Signal	Min.	5.02				1.53
		Maj.	−2.31				−1.58
		Fat.	−2.72				0.05
	Signs and surface marking	Min.	0.23		1.37		
		Maj.	0.25		−0.36		
		Fat.	−0.46		−0.99		
Road type (ref. one-way)	Divided two-way	Min.	−3.39		−10.53	−0.86	
		Maj.	1.87		11.79	3.95	
		Fat.	1.52		−1.26	−3.09	
	Undivided two-way	Min.	−0.47				−0.30
		Maj.	−1.95				−0.44
		Fat.	2.42				0.74
Road width (ref. <20 m)	>20m	Min.			−5.12	−4.62	
		Maj.			4.16	−0.17	
		Fat.			0.96	4.79	
Sidewalk (ref. no)	Yes	Min.	11.04				16.89
		Maj.	−8.71				−14.01
		Fat.	−2.34				−2.87
Vegetation (ref. no)	Yes	Min.	1.29		4.70	−9.00	
		Maj.	−0.43		−2.30	9.27	
		Fat.	−0.86		−2.40	−0.26	
Park lane (ref. no)	Yes	Min.	−3.51		−13.26	−2.30	
		Maj.	4.03		13.34	0.38	
		Fat.	−0.52		−0.08	1.92	
Overpass/underpass (ref. no)	Yes	Min.	−14.97	−13.13	−7.23	−23.77	−6.68
		Maj.	13.87	14.48	5.63	20.14	5.82
		Fat.	1.10	−1.35	1.62	3.62	0.86
AADT (ref. low)	High (>30,000)	Min.			3.53	17.13	−0.92
		Maj.			−2.52	−13.01	−0.35
		Fat.			−1.01	−4.13	1.25
	Medium (15,000–30,000)	Min.	2.09			14.84	
		Maj.	−1.31			−10.67	
		Fat.	−0.78			−4.17	
Population density (ref. <100 person/km <sup>2</sup> )	100–200	Min.	0.11			4.52	−2.48
		Maj.	0.28			−2.39	0.92
		Fat.	−0.39			−2.12	1.55
	>200	Min.				3.65	
		Maj.				−1.52	
		Fat.				−2.15	



Table A1. Cont.

Variables		Sev.	O.S. (%)	C1 (%)	C2 (%)	C3 (%)	C4 (%)
Land use (ref. commercial)	Other	Min.	−2.68			0.50	
		Maj.	0.97			−2.06	
		Fat.	1.72			1.56	
	Residential	Min.		−0.95	−14.88		4.13
		Maj.		0.12	13.85		−1.53
		Fat.		0.83	1.04		−2.58

Table A2. Summary of significant variables in each cluster.

Cluster #	Effect on Probability	Major Injuries		Fatal Injuries	
		Significant in Both Overall and Cluster Models	Significant Just in Cluster Models	Significant in Both Overall and Cluster Models	Significant Just in Cluster Models
Cluster1	Increase	Aged 45–65; aged > 65; crash time 10–14 *; 14–18 *; hit and run; near overpass	Crash time 18–22	Aged 45–65; aged > 65	Residential
	Decrease	Bicycle; posted speed 40–60 km/h	Spring	Motorcycle; crash time at 6–10; crash time 6–10; 10–14; 14–18 **; 18–22	
Cluster2	Increase	Aged 45–65; aged >65; heavy vehicle, bus; minibus, van; pickup; hit and run; with park lane; near overpass; adverse weather	Divided two-way; residential	Men; aged 45–65; aged > 65; heavy vehicle, bus; minibus, van; pickup; hit and run; near overpass	Aged 30–45; road width > 20 m; adverse weather
	Decrease	Posted speed < 40 km/h **		Motorcycle; crash time at 6–10; 10–14; 14–18 **; 18–22; posted speed < 40; control with signs and surface markings **; with vegetated buffer	Posted speed 40–60; high AADT
Cluster3	Increase	Men; aged 45–65; >65; heavy vehicle, bus; minibus, van; pickup; hit and run **; with park lane; near overpass	Aged 30–45; with vegetated buffer	Aged 45–65; aged > 65; heavy vehicle, bus; minibus, van; pickup; hit and run; near overpass; other land uses	Aged 30–45; road width > 20 m; with park lane; adverse weather
	Decrease	Motorcycle; crash time at 6–10; 10–14; at junction **; posted speed < 40 km/h	Medium AADT; high AADT	Crash time at 6–10; 10–14; 14–18 **; 18–22; at junction; posted speed < 40 km/h; divided two-way *; medium AADT; medium density	Summer; high AADT; high density
Cluster4	Increase	Heavy vehicle, bus; minibus, van; hit and run **; near overpass	High AADT	Aged < 15 **; 45–65; >65; heavy vehicle, bus; pickup; hit and run; undivided two-way; near overpasses; medium density *	High AADT
	Decrease	Motorcycle; posted speed 40–60 km/h; <40 km/h; control with signals; with sidewalk	Residential	Crash time at 10–14; 14–18 **; weekend; at junction; with sidewalk	Residential ***

\* Variable effect is opposite in the overall sample. \*\* Variable is random in the overall sample and fixed in the cluster. \*\*\* Variable is fixed in the overall sample and random in the cluster.

## References

- World Health Organization. *Global Status Report on Road Safety 2018*; World Health Organization (WHO): Geneva, Switzerland, 2019.
- Kayani, A.; King, M.J.; Fleiter, J.J. Fatalism and road safety in developing countries, with a focus on Pakistan. *J. Australas. Coll. Road Saf.* **2011**, *22*, 41–47.
- Jadaan, K.; Al-Braizat, E.; Al-Rafayah, S.; Gammoh, H.; Abukahlil, Y. Traffic safety in developed and developing countries: A comparative analysis. *J. Traffic Logist. Eng.* **2018**, *6*, 1–5. [\[CrossRef\]](#)
- Safarpour, H.; Khorasani-Zavareh, D.; Mohammadi, R. The common road safety approaches: A scoping review and thematic analysis. *Chin. J. Traumatol.* **2020**, *23*, 113–121. [\[CrossRef\]](#) [\[PubMed\]](#)
- Salamati, P.; Moradi, A.; Soori, H.; Amiri, M.; Soltani, M. High crash areas resulting in injuries and deaths in Tehran traffic areas from november 2011 through february 2012: A geographic information system analysis. *Med. J. Islam. Repub. Iran* **2015**, *29*, 214.
- Shabanikiya, H.; Hashtarkhani, S.; Bergquist, R.; Bagheri, N.; VafaeiNejad, R.; Amiri-Gholanlou, M.; Akbari, T.; Kiani, B. Multiple-scale spatial analysis of paediatric, pedestrian road traffic injuries in a major city in North-Eastern Iran 2015–2019. *BMC Public Health* **2020**, *20*, 722. [\[CrossRef\]](#)
- Delaney, P.G.; Eisner, Z.J.; Bustos, A.; Hancock, C.J.; Thullah, A.H.; Jayaraman, S.; Raghavendran, K. Cost-effectiveness of lay first responders addressing road traffic injury in sub-Saharan Africa. *J. Surg. Res.* **2022**, *270*, 104–112. [\[CrossRef\]](#)

8. Spencer, M.R.; Hedegaard, H.; Garnet, M. Motor vehicle traffic death rates by sex, age group, and road-user type: United States, 1999–2019. *NCHS Data Brief* **2021**. [CrossRef]
9. Safaei, B.; Safaei, N.; Masoud, A.; Seyedekrami, S. Weighing criteria and prioritizing strategies to reduce motorcycle-related injuries using combination of fuzzy TOPSIS and AHP methods. *Adv. Transp. Stud.* **2021**, *54*, 217–234.
10. Nasri, M.; Aghabayk, K.; Esmaili, A.; Shiwakoti, N. Using ordered and unordered logistic regressions to investigate risk factors associated with pedestrian crash injury severity in Victoria, Australia. *J. Saf. Res.* **2022**, *81*, 78–90. [CrossRef]
11. Mukherjee, D.; Mitra, S. Investigating the fatal pedestrian crash occurrence in urban setup in a developing country using multiple-risk source model. *Accid. Anal. Prev.* **2021**, *163*, 106469. [CrossRef]
12. Iranian Legal Medicine Organization. *National Status Report on Pedestrian Fatalities*; Iranian Legal Medicine Organization: Tehran, Iran, 2019.
13. Sheykhfard, A.; Haghighi, F.; Papadimitriou, E.; Van Gelder, P. Analysis of the occurrence and severity of vehicle-pedestrian conflicts in marked and unmarked crosswalks through naturalistic driving study. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *76*, 178–192. [CrossRef]
14. Mashhad Transport and Traffic Organization. 13th Statistical Report on Mashhad Traffic. 2019. Available online: <https://traffic.mashhad.ir/> (accessed on 13 November 2022).
15. Chakraborty, A.; Mukherjee, D.; Mitra, S. Development of pedestrian crash prediction model for a developing country using artificial neural network. *Int. J. Inj. Control Saf. Promot.* **2019**, *26*, 283–293. [CrossRef] [PubMed]
16. Mesa-Arango, R.; Valencia-Alaix, V.G.; Pineda-Mendez, R.A.; Eissa, T. Influence of socioeconomic conditions on crash injury severity for an urban area in a developing country. *Transp. Res. Rec. J. Transp. Res. Board* **2018**, *2672*, 41–53. [CrossRef]
17. Gupta, U.; Tiwari, G.; Chatterjee, N.; FAzio, J. Case study of pedestrian risk behavior and survival analysis. In Proceedings of the 8th International Conference of Eastern Asia Society for Transportation Studies, Surabaya, Indonesia, 16–19 November 2009; Volume 7, p. 389.
18. Sheykhfard, A.; Haghighi, F.; Nordfjærn, T.; Soltaninejad, M. Structural equation modelling of potential risk factors for pedestrian accidents in rural and urban roads. *Int. J. Inj. Control Saf. Promot.* **2020**, *28*, 46–57. [CrossRef]
19. Jamali-Dolatabad, M.; Sadeghi-Bazargani, H.; Sarbakhsh, P. Predictors of fatal outcomes in pedestrian accidents in Tabriz Metropolis of Iran: Application of PLS-DA method. *Traffic Inj. Prev.* **2019**, *20*, 873–879. [CrossRef]
20. Kashani, A.T.; Besharati, M.M. Fatality rate of pedestrians and fatal crash involvement rate of drivers in pedestrian crashes: A case study of Iran. *Int. J. Inj. Control Saf. Promot.* **2017**, *24*, 222–231. [CrossRef]
21. Payam, P.; Seyed, T.H.; Amin, H.; Yaser, S.; Arya, H.; Mohammad, Z.; Ghasem, M.; Mohammad, R.A.; Najmeh, M.; Ali, F.; et al. Epidemiological characteristics of fatal pedestrian accidents in Fars Province of Iran: A community-based survey. *Chin. J. Traumatol.* **2012**, *15*, 279–283.
22. Miranda-Moreno, L.F.; Morency, P.; El-Geneidy, A.M. The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. *Accid. Anal. Prev.* **2011**, *43*, 1624–1634. [CrossRef]
23. Wang, Y.; Kockelman, K.M. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accid. Anal. Prev.* **2013**, *60*, 71–84. [CrossRef]
24. Lee, J.; Abdel-Aty, M.; Choi, K.; Huang, H. Multi-level hot zone identification for pedestrian safety. *Accid. Anal. Prev.* **2015**, *76*, 64–73. [CrossRef]
25. Su, J.; Sze, N.; Bai, L. A joint probability model for pedestrian crashes at macroscopic level: Roles of environment, traffic, and population characteristics. *Accid. Anal. Prev.* **2021**, *150*, 105898. [CrossRef]
26. Sze, N.N.; Wong, S.C. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accid. Anal. Prev.* **2007**, *39*, 1267–1278. [CrossRef]
27. Aziz, H.A.; Ukkusuri, S.V.; Hasan, S. Exploring the determinants of pedestrian-vehicle crash severity in New York City. *Accid. Anal. Prev.* **2013**, *50*, 1298–1309. [CrossRef]
28. Mohamed, M.G.; Saunier, N.; Miranda-Moreno, L.F.; Ukkusuri, S.V. A clustering regression approach: A comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US and Montreal, Canada. *Saf. Sci.* **2013**, *54*, 27–37. [CrossRef]
29. Sasidharan, L.; Wu, K.-F.; Menendez, M. Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. *Accid. Anal. Prev.* **2015**, *85*, 219–228. [CrossRef]
30. Sun, M.; Sun, X.; Shan, D. Pedestrian crash analysis with latent class clustering method. *Accid. Anal. Prev.* **2019**, *124*, 50–57. [CrossRef]
31. Li, Y.; Song, L.; Fan, W. Day-of-the-week variations and temporal instability of factors influencing pedestrian injury severity in pedestrian-vehicle crashes: A random parameters logit approach with heterogeneity in means and variances. *Anal. Methods Accid. Res.* **2021**, *29*, 100152. [CrossRef]
32. Wang, X.; Liu, Q.; Guo, F.; Fang, S.; Xu, X.; Chen, X. Causation analysis of crashes and near crashes using naturalistic driving data. *Accid. Anal. Prev.* **2022**, *177*, 106821. [CrossRef]
33. Balsa-Barreiro, J.; Valero-Mora, P.M.; Berné-Valero, J.L.; Varela-García, F.-A. GIS mapping of driving behavior based on naturalistic driving data. *ISPRS Int. J. Geo-inf.* **2019**, *8*, 226. [CrossRef]
34. Balsa-Barreiro, J.; Valero-Mora, P.M.; Menéndez, M.; Mehmood, R. Extraction of naturalistic driving patterns with geographic information systems. *Mob. Netw. Appl.* **2020**, 1–17. [CrossRef]

35. Eluru, N.; Bhat, C.R.; Hensher, D.A. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accid. Anal. Prev.* **2008**, *40*, 1033–1054. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Lee, C.; Abdel-Aty, M. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. *Accid. Anal. Prev.* **2005**, *37*, 775–786. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Tay, R.; Choi, J.; Kattan, L.; Khan, A. A multinomial logit model of pedestrian–vehicle crash severity. *Inr. J. Sustain. Transp.* **2011**, *5*, 233–249. [\[CrossRef\]](#)
38. Kim, J.-K.; Ulfarsson, G.F.; Shankar, V.N.; Mannering, F.L. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accid. Anal. Prev.* **2010**, *42*, 1751–1758. [\[CrossRef\]](#)
39. Harruff, R.C.; Avery, A.; Alter-Pandya, A.S. Analysis of circumstances and injuries in 217 pedestrian traffic fatalities. *Accid. Anal. Prev.* **1998**, *30*, 11–20. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Jang, K.; Park, S.H.; Kang, S.; Song, K.H.; Kang, S.; Chung, S. Evaluation of pedestrian safety: Pedestrian crash hot spots and risk factors for injury severity. *Transp. Res. Rec.* **2013**, *2393*, 104–116. [\[CrossRef\]](#)
41. Sarkar, S.; Tay, R.; Hunt, J.D. Logistic regression model of risk of fatality in vehicle–pedestrian crashes on national highways in Bangladesh. *Transp. Res. Rec. J. Transp. Res. Board* **2011**, *2264*, 128–137. [\[CrossRef\]](#)
42. Tarko, A.; Azam, M.S. Pedestrian injury analysis with consideration of the selectivity bias in linked police-hospital data. *Accid. Anal. Prev.* **2011**, *43*, 1689–1695. [\[CrossRef\]](#)
43. Ukkusuri, S.; Miranda-Moreno, L.F.; Ramadurai, G.; Isa-Tavarez, J. The role of built environment on pedestrian crash frequency. *Saf. Sci.* **2012**, *50*, 1141–1151. [\[CrossRef\]](#)
44. Zhai, X.; Huang, H.; Sze, N.N.; Song, Z.; Hon, K.K. Diagnostic analysis of the effects of weather condition on pedestrian crash severity. *Accid. Anal. Prev.* **2019**, *122*, 318–324. [\[CrossRef\]](#)
45. Peng, H.; Ma, X.; Chen, F. Examining injury severity of pedestrians in vehicle–pedestrian crashes at mid-blocks using path analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6170. [\[CrossRef\]](#)
46. Wang, J.; Huang, H.; Xu, P.; Xie, S.; Wong, S.C. Random parameter probit models to analyze pedestrian red-light violations and injury severity in pedestrian–motor vehicle crashes at signalized crossings. *J. Transp. Saf. Secur.* **2020**, *12*, 818–837. [\[CrossRef\]](#)
47. Ulfarsson, G.F.; Kim, S.; Booth, K.M. Analyzing fault in pedestrian–motor vehicle crashes in North Carolina. *Accid. Anal. Prev.* **2010**, *42*, 1805–1813. [\[CrossRef\]](#)
48. Holubowycz, O.T. Age, sex, and blood alcohol concentration of killed and injured pedestrians. *Accid. Anal. Prev.* **1995**, *27*, 417–422. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Kong, L.B.; Lekawa, M.; Navarro, R.A.; McGrath, J.; Cohen, M.; Margulies, D.R.; Hiatt, J.R. Pedestrian-motor vehicle trauma: An analysis of injury profiles by age. *J. Am. Coll. Surg.* **1996**, *182*, 17–23.
50. Lefler, D.E.; Gabler, H.C. The fatality and injury risk of light truck impacts with pedestrians in the United States. *Accid. Anal. Prev.* **2004**, *36*, 295–304. [\[CrossRef\]](#)
51. Chen, Z.; Fan, W. A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. *Int. J. Transp. Sci. Technol.* **2019**, *8*, 43–52. [\[CrossRef\]](#)
52. Zajac, S.S.; Ivan, J.N. Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. *Accid. Anal. Prev.* **2003**, *35*, 369–379. [\[CrossRef\]](#)
53. Balsa-Barreiro, J.; Menendez, M.; Morales, A.J. Scale, context, and heterogeneity: The complexity of the social space. *Sci. Rep.* **2022**, *12*, 9037. [\[CrossRef\]](#)
54. Li, Z.; Wu, Q.; Ci, Y.; Chen, C.; Zhang, G. Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accid. Anal. Prev.* **2019**, *129*, 230–240. [\[CrossRef\]](#)
55. Pai, C.-W.; Saleh, W. An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK. *Saf. Sci.* **2007**, *45*, 832–847. [\[CrossRef\]](#)
56. 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. *Accid. Anal. Prev.* **2020**, *135*, 105388. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Abdel-Aty, M.; Chundi, S.S.; Lee, C. Geo-spatial and log-linear analysis of pedestrian and bicyclist crashes involving school-aged children. *J. Saf. Res.* **2007**, *38*, 571–579. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Wheeler-Martin, K.C.; Curry, A.E.; Metzger, K.B.; DiMaggio, C.J. Trends in school-age pedestrian and pedalcyclist crashes in the USA: 26 states, 2000–2014. *Inj. Prev.* **2020**, *26*, 448–455. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Rahimi, A.; Azimi, G.; Asgari, H.; Jin, X. Injury severity of pedestrian and bicyclist crashes involving large trucks. In Proceedings of the International Conference on Transportation and Development 2020, Seattle, WA, USA, 26–29 May 2020; pp. 110–122.
60. Nasri, M.; Aghabayk, K. Assessing risk factors associated with urban transit bus involved accident severity: A case study of a Middle East country. *Int. J. Crashworthiness* **2021**, *26*, 413–423. [\[CrossRef\]](#)
61. Chung, Y. Injury severity analysis in taxi-pedestrian crashes: An application of reconstructed crash data using a vehicle black box. *Accid. Anal. Prev.* **2018**, *111*, 345–353. [\[CrossRef\]](#)
62. Rifaat, S.M.; Tay, R.; Raihan, S.M.; Fahim, A.; Touhidduzzaman, S.M. Vehicle-Pedestrian crashes at Intersections in Dhaka city. *Open Transp. J.* **2017**, *11*, 11–19. [\[CrossRef\]](#)
63. Depaire, B.; Wets, G.; Vanhoof, K. Traffic accident segmentation by means of latent class clustering. *Accid. Anal. Prev.* **2008**, *40*, 1257–1266. [\[CrossRef\]](#)

64. De Ona, J.; López, G.; Mujalli, R.; Calvo, F.J. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. *Accid. Anal. Prev.* **2013**, *51*, 1–10. [\[CrossRef\]](#)
65. Kim, K.; Yamashita, E.Y. Using ak-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii. *J. Adv. Transp.* **2007**, *41*, 69–89. [\[CrossRef\]](#)
66. Anderson, T.K. Kernel density estimation and K-means clustering to profile road accident hotspots. *Accid. Anal. Prev.* **2009**, *41*, 359–364. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Chang, F.; Xu, P.; Zhou, H.; Chan, A.H.; Huang, H. Investigating injury severities of motorcycle riders: A two-step method integrating latent class cluster analysis and random parameters logit model. *Accid. Anal. Prev.* **2019**, *131*, 316–326. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Behnood, A.; Mannering, F.L. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. *Anal. Methods Accid. Res.* **2016**, *12*, 1–17. [\[CrossRef\]](#)
69. Caliendo, C.; De Guglielmo, M.L.; Russo, I. Analysis of crash frequency in motorway tunnels based on a correlated random-parameters approach. *Tunn. Undergr. Space Technol.* **2019**, *85*, 243–251. [\[CrossRef\]](#)
70. Caliendo, C.; Guida, M.; Postiglione, F.; Russo, I. A Bayesian bivariate hierarchical model with correlated parameters for the analysis of road crashes in Italian tunnels. *Stat. Methods Appl.* **2022**, *31*, 109–131. [\[CrossRef\]](#)
71. Wang, K.; Shirani-Bidabadi, N.; Shaon, M.R.R.; Zhao, S.; Jackson, E. Correlated mixed logit modeling with heterogeneity in means for crash severity and surrogate measure with temporal instability. *Accid. Anal. Prev.* **2021**, *160*, 106332. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Mariel, P.; Artabe, A. Interpreting correlated random parameters in choice experiments. *J. Environ. Econ. Manag.* **2020**, *103*, 102363. [\[CrossRef\]](#)
73. Song, L.; Fan, W.; Li, Y.; Wu, P. Exploring pedestrian injury severities at pedestrian-vehicle crash hotspots with an annual upward trend: A spatiotemporal analysis with latent class random parameter approach. *J. Saf. Res.* **2021**, *76*, 184–196. [\[CrossRef\]](#)
74. Statistical Center of Iran. *Detailed Results of the General Census of Population and Housing in the Country Iran*; Statistical Center of Iran: Tehran, Iran, 2016.
75. Kaplan, S.; Prato, C.G. Cyclist–motorist crash patterns in Denmark: A latent class clustering approach. *Traffic Inj. Prev.* **2013**, *14*, 725–733. [\[CrossRef\]](#)
76. Lanza, S.T.; Rhoades, B.L. Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prev. Sci.* **2013**, *14*, 157–168. [\[CrossRef\]](#)
77. Hair, J.F.; Anderson, R.; Tatham, R.; Black, W.C. *Multivariate Data Analysis*, 5th ed.; Prentice Hall: Hoboken, NJ, USA, 1998; p. 730.
78. Collins, L.M.; Lanza, S.T. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*; John Wiley & Sons: New York, NY, USA, 2009; Volume 718.
79. Lanza, S.T.; Dziak, J.J.; Huang, L.; Wagner, A.T.; Collins, L.M. *LCA Stata Plugin Users' Guide (Version 1.2)*; The Methodology Center, Penn State: University Park, PA, USA, 2015.
80. Biernacki, C.; Govaert, G. Choosing models in model-based clustering and discriminant analysis. *J. Stat. Comput. Simul.* **1999**, *64*, 49–71. [\[CrossRef\]](#)
81. Bijmolt, T.H.; Paas, L.J.; Vermunt, J. Country and consumer segmentation: Multi-level latent class analysis of financial product ownership. *Int. J. Res. Mark.* **2004**, *21*, 323–340. [\[CrossRef\]](#)
82. Samerei, S.A.; Aghabayk, K.; Mohammadi, A.; Shiwakoti, N. Data mining approach to model bus crash severity in Australia. *J. Saf. Res.* **2021**, *76*, 73–82. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Peel, D.; McLachlan, G. Robust mixture modelling using the t distribution. *Stat. Comput.* **2000**, *10*, 339–348. [\[CrossRef\]](#)
84. Manski, C.F.; McFadden, D. *Structural Analysis of Discrete Data with Econometric Applications*; MIT press: Cambridge, MA, USA, 1981.
85. Train, K.E. *Discrete Choice Methods with Simulation*; Cambridge University Press: Cambridge, UK, 2009.
86. Li, Z.; Ci, Y.; CheCamn, C.; Zhang, G.; Wu, Q.; Qian, Z.S.; Prevedouros, P.D.; Ma, D.T. Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. *Accid. Anal. Prev.* **2019**, *124*, 219–229. [\[CrossRef\]](#)
87. Wu, Q.; Chen, F.; Zhang, G.; Liu, X.C.; Wang, H.; Bogus, S.M. Mixed logit model-based driver injury severity investigations in single- and multi-vehicle crashes on rural two-lane highways. *Accid. Anal. Prev.* **2014**, *72*, 105–115. [\[CrossRef\]](#)
88. Liu, P.; Fan, W. Modeling head-on crash severity on NCDOT freeways: A mixed logit model approach. *Can. J. Civ. Eng.* **2019**, *46*, 322–328. [\[CrossRef\]](#)
89. Kim, J.-K.; Ulfarsson, G.F.; Kim, S.; Shankar, V.N. Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accid. Anal. Prev.* **2013**, *50*, 1073–1081. [\[CrossRef\]](#)
90. Onieva-García, M.Á.; Martínez-Ruiz, V.; Lardelli-Claret, P.; Jiménez-Moleón, J.J.; Amezcua-Prieto, C.; Luna-del-Castillo, J.d.D.; Jiménez-Mejías, E. Gender and age differences in components of traffic-related pedestrian death rates: Exposure, risk of crash and fatality rate. *Inj. Epidemiol.* **2016**, *3*, 14. [\[CrossRef\]](#)
91. Olszewski, P.; Szagała, P.; Wolański, M.; Zielińska, A. Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland. *Accid. Anal. Prev.* **2015**, *84*, 83–91. [\[CrossRef\]](#)
92. Esmaili, A.; Aghabayk, K.; Parishad, N.; Stephens, A.N. Investigating the interaction between pedestrian behaviors and crashes through validation of a pedestrian behavior questionnaire (PBQ). *Accid. Anal. Prev.* **2021**, *153*, 106050. [\[CrossRef\]](#)



93. Sullman, M.J.M.; Gras, M.E.; Font-Mayolas, S.; Masferrer, L.; Cunill, M.; Planes, M. The pedestrian behaviour of Spanish adolescents. *J. Adolesc.* **2011**, *34*, 531–539. [\[CrossRef\]](#)
94. Preusser, D.F. Reducing pedestrian crashes among children. *Bull. New York Acad. Med.* **1988**, *64*, 623–631.
95. Haleem, K.; Alluri, P.; Gan, A. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accid. Anal. Prev.* **2015**, *81*, 14–23. [\[CrossRef\]](#)
96. Pour-Rouholamin, M.; Zhou, H. Investigating the risk factors associated with pedestrian injury severity in Illinois. *J. Saf. Res.* **2016**, *57*, 9–17. [\[CrossRef\]](#)
97. Jahangeer, A.A.; Anjana, S.S.; Das, V.R. A hierarchical modeling approach to predict pedestrian crash severity. In *Transportation Research*; Springer: Singapore, 2020; pp. 355–366.
98. Hu, L.; Wu, X.; Huang, J.; Peng, Y.; Liu, W. Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China. *Saf. Sci.* **2020**, *127*, 104710. [\[CrossRef\]](#)
99. Zhang, T.; Chan, A.H.; Zhang, W. Dimensions of driving anger and their relationships with aberrant driving. *Accid. Anal. Prev.* **2015**, *81*, 124–133. [\[CrossRef\]](#)
100. Mitra, S. Sun glare and road safety: An empirical investigation of intersection crashes. *Saf. Sci.* **2014**, *70*, 246–254. [\[CrossRef\]](#)
101. Ma, H.-P.; Chen, P.-L.; Chen, S.-K.; Chen, L.-H.; Linkov, V.; Pai, C.-W. Population-based case-control study of the effect of sun glare on pedestrian fatalities in Taiwan. *BMJ Open* **2019**, *9*, e028350. [\[CrossRef\]](#)
102. Williamson, A.; Lombardi, D.A.; Folkard, S.; Stutts, J.; Courtney, T.K.; Connor, J.L. The link between fatigue and safety. *Accid. Anal. Prev.* **2011**, *43*, 498–515. [\[CrossRef\]](#)
103. Caponecchia, C.; Williamson, A. Drowsiness and driving performance on commuter trips. *J. Saf. Res.* **2018**, *66*, 179–186. [\[CrossRef\]](#)
104. Sun, R.; Zhuang, X.; Wu, C.; Zhao, G.; Zhang, K. The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment. *Transp. Res. Part F Traffic Psychol. Behav.* **2015**, *30*, 97–106. [\[CrossRef\]](#)
105. Iran Meteorological Organization. *Monthly Total Precipitation in Mashhad by Month 1951–2010*; Iran Meteorological Organization: Tehran, Iran, 2018.
106. Zegeer, C.V.; Bushell, M. Pedestrian crash trends and potential countermeasures from around the world. *Accid. Anal. Prev.* **2012**, *44*, 3–11. [\[CrossRef\]](#)
107. Tulu, G.S.; Washington, S.; Haque, M.M.; King, M.J. Injury severity of pedestrians involved in road traffic crashes in Addis Ababa, Ethiopia. *J. Transp. Saf. Secur.* **2017**, *9*, 47–66. [\[CrossRef\]](#)
108. Prato, C.G.; Kaplan, S.; Patrier, A.; Rasmussen, T.K. Considering built environment and spatial correlation in modeling pedestrian injury severity. *Traffic Inj. Prev.* **2018**, *19*, 88–93. [\[CrossRef\]](#)
109. Moradi, A.; Motevalian, S.A.; Mirkoohi, M.; McKay, M.P.; Rahimi-Movaghar, V. Exceeding the speed limit: Prevalence and determinants in Iran. *Int. J. Inj. Control Saf. Promot.* **2013**, *20*, 307–312. [\[CrossRef\]](#)
110. Zafri, N.M.; Prithul, A.A.; Baral, I.; Rahman, M. Exploring the factors influencing pedestrian-vehicle crash severity in Dhaka, Bangladesh. *Int. J. Inj. Control Saf. Promot.* **2020**, *27*, 300–307. [\[CrossRef\]](#)
111. Xin, C.; Guo, R.; Wang, Z.; Lu, Q.; Lin, P.-S. The effects of neighborhood characteristics and the built environment on pedestrian injury severity: A random parameters generalized ordered probability model with heterogeneity in means and variances. *Anal. Methods Accid. Res.* **2017**, *16*, 117–132. [\[CrossRef\]](#)
112. Cinnamon, J.; Schuurman, N.; Hameed, S.M. Pedestrian injury and human behaviour: Observing road-rule violations at high-incident intersections. *PLoS ONE* **2011**, *6*, e21063. [\[CrossRef\]](#)
113. Kim, J.-K.; Ulfarsson, G.F.; Shankar, V.N.; Kim, S. Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accid. Anal. Prev.* **2008**, *40*, 1695–1702. [\[CrossRef\]](#)
114. Li, Y.; Fan, W. Mixed logit approach to modeling the severity of pedestrian-injury in pedestrian-vehicle crashes in North Carolina: Accounting for unobserved heterogeneity. *J. Transp. Saf. Secur.* **2022**, *14*, 796–817. [\[CrossRef\]](#)
115. Dommes, A.; Cavallo, V.; Dubuisson, J.-B.; Tournier, I.; Vienne, F. Crossing a two-way street: Comparison of young and old pedestrians. *J. Saf. Res.* **2014**, *50*, 27–34. [\[CrossRef\]](#)
116. Hanson, C.S.; Noland, R.B.; Brown, C. The severity of pedestrian crashes: An analysis using Google Street View imagery. *J. Transp. Geogr.* **2013**, *33*, 42–53. [\[CrossRef\]](#)
117. Yue, L.; Abdel-Aty, M.; Wu, Y.; Zheng, O.; Yuan, J. In-depth approach for identifying crash causation patterns and its implications for pedestrian crash prevention. *J. Saf. Res.* **2020**, *73*, 119–132. [\[CrossRef\]](#)
118. Fitzpatrick, K.; Iragavarapu, V.; Brewer, M.; Lord, D.; Hudson, J.G.; Avelar, R.; Robertson, J. Characteristics of Texas Pedestrian Crashes and Evaluation of Driver Yielding at Pedestrian Treatments. 2014. Available online: <http://tti.tamu.edu/documents/0-6702-1.pdf> (accessed on 13 November 2022).
119. Rankavat, S.; Tiwari, G. Association between built environment and pedestrian fatal crash risk in Delhi, India. *Transp. Res. Rec. J. Transp. Res. Board* **2015**, *2519*, 61–66. [\[CrossRef\]](#)
120. Morency, P.; Gauvin, L.; Plante, C.; Fournier, M.; Morency, C. Neighborhood social inequalities in road traffic injuries: The influence of traffic volume and road design. *Am. J. Public Health* **2012**, *102*, 1112–1119. [\[CrossRef\]](#)
121. Pour, A.T.; Moridpour, S.; Tay, R.; Rajabifard, A. Modelling pedestrian crash severity at mid-blocks. *Transp. A Transp. Sci.* **2017**, *13*, 273–297.
122. Gårder, P.E. The impact of speed and other variables on pedestrian safety in Maine. *Accid. Anal. Prev.* **2004**, *36*, 533–542. [\[CrossRef\]](#)



123. Goel, R.; Jain, P.; Tiwari, G. Correlates of fatality risk of vulnerable road users in Delhi. *Accid. Anal. Prev.* **2018**, *111*, 86–93. [[CrossRef](#)]
124. Cai, Q.; Abdel-Aty, M.; Lee, J. Macro-level vulnerable road users crash analysis: A Bayesian joint modeling approach of frequency and proportion. *Accid. Anal. Prev.* **2017**, *107*, 11–19. [[CrossRef](#)]
125. Tay, R.; Rifaat, S.M.; Chin, H.C. A logistic model of the effects of roadway, environmental, vehicle, crash and driver characteristics on hit-and-run crashes. *Accid. Anal. Prev.* **2008**, *40*, 1330–1336. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.