Priority Order for Improvement of Intersections using Pedestrian Crash Prediction Model

Ali Karami[1], Mansour Hadji Hosseinlou*[2], Mohammadhossein Abbasi[3], Monteiro Figuerira[4].

Received: 09.03.2019 *Accepted: 25.09.2019*

Abstract

One of the most important needs of pedestrians is safety at crossing points, especially at intersections. Intersections are important parts of the urban road network because any disruption in them reduces the capacity of the entire network. The main objective of this research is to propose an appropriate method for prioritizing urban intersections with considering the important factors affecting pedestrian crashes to promote the safety level of pedestrians at intersections of the 11th district of Tehran. In this paper, after comparing different models, finally, the negative binomial model was developed to predict the effects of a set of factors expected to the frequency of pedestrian crashes. According to the proposed model, a larger volume of pedestrians and vehicles reduced the safety of intersections. Also imposing traffic restrictions in the central business district causes increasing motorcycle flows and has led to more dangerous area. Also, according to the results of prioritization using this method showed that the intersection of Imam Khomeini and Valiasr with an improvement potential of 6.93 has the most potential of improvement. Based on crash reduction factor, a method for estimating the effect of a variable on crash frequencies, one-unit increasing in natural logarithm of average pedestrian and vehicle volumes, commercial land use and number of public transport stations will increase the crash frequencies by 29.82, 83.49, 56.99 and 14.34 percents, respectively. Also, when sidewalk effective widths increased by one-unit, the probability of pedestrian crashes at intersections will reduced by 14.87 percent.

Keywords: Prioritizing, pedestrian crashes, negative binomial model, urban intersection, safety**.**

* Corresponding author Email: mansour@kntu.ac.ir

[1] MSc. Grad. Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran.

[2] Associate Professor, Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran.

[3] PhD. Student, Department of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran.

[4] Professor, Department of Civil Engineering, University of Lisbon, Lisbon, Portugal.

1. Introduction

Walking is the most fundamental form of human movement as well as the most unprotected mode of transportation. According to the report published by the European Union in 2015, pedestrian crashes account for about 23% of fatality crashes in European countries [National Highway Traffic Safety Administration, 2017]. In Australia, pedestrians also account for 17% of total transport damage and 13% of total deaths [European Commission, 2016]. In the United States, it was estimated that pedestrian crashes caused 5,376 fatalities and 70,000 injuries, accounting for 15% of overall deaths caused by crashes [BITRE, 2013]. According to a report published in March of 2018 by the Tehran Municipality, pedestrian fatalities in Tehran account for 50% of the total deaths caused by traffic crashes.

According to the data including a large number of pedestrian crashes, it was necessary to investigate the prioritization of crash prone intersections by pedestrian crash prediction models. Also nowadays, traffic attributes are usually the more preferred than safety status in the design of an intersection [Mirbaha, Saffarzadeh, and Noruzoliaee, 2013; Ticali and Corriere, 2018; Li et al.2019].

This paper is structured as follows: Section 2 is about literature reviews while section 3 describes the methodology while in section 4 conducted a pedestrian crash data analysis and modeling process was introduced in section 5, and section 6 outlines the major conclusions.

2. Literature Review

It is obvious that the better solution for eliminating pedestrian crashes at the intersections is to use pedestrian overpass or underpass and banning on entering pedestrians to streets, but this is not social and economically feasible at most of the intersections, or it encounters to budget constraints. Financial constraints make the point that investments should take place where the return on capital occurs at the right time [Qureshi et al.2003]. Therefore, prioritization of urban intersections improvement seems to help to allocate appropriate funding for increasing pedestrian safety. Using prediction models can be a good way to prioritize urban intersections improvement, which has also been considered in previous studies [Xie et al. 2017; Ayati, Zaker, and Sadeghi, 2010]. Estimating an appropriate model for prioritizing based on pedestrian safety requires the identification of factors affecting pedestrian crashes and how each of these factors affects the frequency of crashes. In the following has been provided several previous articles that focused on developing pedestrian crash prediction models at intersections or somehow the safety of pedestrians at intersections was taken into consideration.

Siddiqui, Abdel-Aty and Choi. (2012) focused on investigating the effect of spatial correlation on pedestrian and bicycle crashes at Traffic Analysis Zones (TAZ) in two counties of Florida. They used Bayesian Poisson-lognormal model that including roadway characteristics and various demographic and socio-economic factors. They found that average household income has been significant with a negative sign in pedestrian crashes at TAZs. Also, total number of intersections per TAZ, total number of dwelling units, logarithm of population per square mile of a TAZ, and logarithm of the total employment in a TAZ were positively associated with pedestrian crashes in TAZs. [Siddiqui, Abdel-Aty and Choi, 2012].

Miranda-Moreno, Morency and El-Geneidy, (2011) focused on the effects of built environment – including land use types, road network connectivity, transit supply and demographic characteristics – on pedestrian activity and pedestrian-vehicle crashes frequency at signalized intersections. They applied standard negative binomial model, generalized negative binomial model, and latent class negative binomial model, using Montreal Transportation Department and Montreal Public Health Department databases. They found to decrease in traffic volume would be associated with great improvements in pedestrian safety [Miranda-Moreno, Morency and El-Geneidy, 2011].

Pulugurtha and Sambhara (2011) used negative binomial models for examining pedestrian crashes at signalized intersections in the City of Charlotte, North Carolina. Variables including Pedestrian volume, vehicle volume, number of bus stations, land use categories and population found to be significant in three proposed models. Population, number of bus stations, number of approaches at an intersection, and the pedestrian volume found to increase the number of pedestrian crashes [Pulugurtha and Sambhara, 2011].

Bennet and Yiannakoulias (2015) focused on finding out, and contrast crashes involving child pedestrian risks at mid-block and intersection in Hamilton, Ontario, Canada. They developed conditional logistic regression to predict the log-odds of collision risk at intersections and mid-blocks. Intersection control type, traffic volume, and land use characteristics found to be statistically significant in the proposed model [Bennet and Yiannakoulias, 2015].

In 2016, Kim investigated the impact of intersection features on pedestrian and vehicles crashes. The impact of several features of intersections such as various locations of constructed intersections, geometric design factors, traffic control devices, and socio-geographic features of adjacent neighborhood intersections on pedestrian crashes. In this study, first it was reviewed the archives of all pedestrian crashes then dangerous intersections identified by using a negative binomial regression model. Finally, found that some factors such as number of lanes, sequence of the entrance streets to the intersection and the presence of on-street parking near the intersections were negatively associated with pedestrian safety [Kim, 2016].

Mooney et al. (2016) have applied negative binomial model for examining the relationship between intersections' characteristics and frequency of pedestrian injuries at intersections in New York City. They concluded that marked crosswalks, pedestrian signals, nearby billboards, and bus stops were positively associated with frequency of pedestrian injuries at intersections [Mooney et al. 2016].

Murphy, Levinson and Owen (2017) developed log-linear regression models for estimating risk factor (either crashes per pedestrian, or crashes per car) at urban intersections for automobile-pedestrian crash in Minneapolis. They used 6-hour pedestrian traffic flow and AADT as dependent variables. Based on the findings of this article, pedestrians were at a lower risk of being hit by a car at intersections with higher pedestrian traffic, and individual cars were at a lower risk of hitting pedestrians at intersections with more car traffic [Murphy, Levinson and Owen, 2017].

Xie et al. (2018) had an effort to identify the effect of potential factors contributing to the occurrence of pedestrian crashes. They used Poisson lognormal models for the pedestrian crash frequency at signalized intersections in Hong Kong. So, the number of crossing pedestrians, the number of passing vehicles, presence of curb parking and presence of ground-floor shops were found to increase pedestrian crash frequencies [Xie et al.2018].

According to the previous studies mentioned above, in most countries, models have been proposed to identify the factors affecting pedestrian crashes at intersections. These articles have significantly differed in their goals. In most of them, the purpose is to identify the factors affecting the frequency of pedestrian crashes at intersections. Considering the very limited studies in determining the factors affecting pedestrian crashes at urban intersections in Iran, this study attempted to calibrate a model for predicting the frequency of pedestrian crashes at intersections with the aim of prioritizing intersections in conditions of Tehran, Iran.

3. Methodology

At this section, an appropriate model was developed to predict the frequency of crashes at intersections. A current approach to model crash frequency is to develop univariate models, such as Poisson or negative binomial models, to predict the number of crashes for different categories independently [Hauer et al.2004; Noland and Quddus, 2005; Chang, 2005; Kim et al.2006; Jonsson et al.2007; Hosseinpour et al.2014; Mohammadi et al.2014; Kim, 2016; Mothafer et al.2017]. In this paper, the number of pedestrian crashes at intersections of the 11th district and their boundary conditions (50 m) was modeled by negative binomial regression model. Pedestrian crash statistics for at least two years are required.
After identifying the intersections, an appropriate model was developed based on the distribution of the dependent variable. An important issue in calibration of a suitable model for a dependent variable is to examine independent variables that affect the dependent variable. By reviewing previous studies and consulting with experts, suitable variables that affect the occurrence of crashes were identified. Finally, it was concluded that traffic characteristics, geometric design and crash-related characteristics are effective factors. Therefore, in this research various variables including: average daily traffic volume, average pedestrian volumes, presence of central business distinct, land use categories, number of entry and outlet lanes to the intersection, number of exclusive lanes, sequence of input streets to the intersection, sidewalk average width, hawker presence and their sidewalk blockages, sidewalk average effective width, number of bus and metro stations in vicinity of the intersection, intersection control type were considered as an independent variables in the modeling. Finally, after the modeling process, concerning the potential for improvement, the intersections were prioritized.

3.1 Statistical Models

Various methods have been used in recent years to analyze the frequency of crashes. Using a simple linear regression model to predict the frequency of crashes has some limitations. In this model, it is assumed that the dependent variable (the frequency of pedestrian crashes) follows a normal distribution. Also, given that the model's prediction range from $-\infty$ to $+\infty$, it is likely to predict a negative value for the frequency of crashes [Abedini et al. 2013]. The dependent variable in this research was pedestrian crash frequencies and is a non-negative value. Poisson and negative binomial regression models are among methods that have been used for modeling of crash frequencies by regarding the distribution of data. accordingly, Poisson and negative binomial regression models have been described.

3.1.1 Poisson Regression Model

Poisson model is sometimes used for the discrete and non-negative nature of crashes [Ayati et al. 2010]. Accordingly, the average number of predicted pedestrian crashes at each intersection during a given time period is defined as Equation 1 [Washington, Karlaftis and Mannering, 2010]
.

$$(\lambda_i) = \mathrm{EXP}(\beta X_i) \tag{1}$$

Where λ_i is the average number of expected pedestrian crashes at each intersection, X_i is a vector of explanatory variables at intersection i in a given time period and β is a vector of estimable parameters. In Poisson model, it is assumed that crash frequencies of crashes follow the Poisson distribution. In this case, the probability of an intersection experiences pedestrian crashes with a set of certain characteristics is expressed as Equation 2 [Washington, Karlaftis and Mannering, 2010].

$$P(Y_i) = \frac{e^{-\lambda_i} * \lambda_i^{y_i}}{y_i!} \quad (2)$$

The coefficients of β are estimated using the maximum log of the likelihood function, LL(β) as Equation 3 [Washington, Karlaftis and Mannering, 2010]:

$$LL(\beta) = \sum_i [-EXP(\beta X_i) + y_i \beta X_i - LN(y_i!)] \quad (3)$$

Where y_i is the number of observed pedestrian crashes at intersection i.

3.1.2 Negative Binomial Regression Model

Poisson model in contrast to the negative binomial model, it is assumed that the mean and variance of dependent variable are equal. A negative binomial distribution is a discrete distribution that provides a model for high-dispersion data, such as crash data [Abedini et al. 2013]. The negative binomial model is derived by rewriting Equation 1, the average number of predicted pedestrian crashes at each intersection during a given time period is defined as Equation 4 [Washington, Karlaftis and Mannering, 2010].

$$(\lambda_i) = EXP(\beta X_i + \varepsilon_i) \quad (4)$$

where $EXP(\varepsilon_i)$ is a Gamma-distributed disturbance term with mean 1 and variance α.

Unlike Poisson distribution, the distribution of negative binomial has two parameters which general form of that is as Equation 5 [Washington, Karlaftis and Mannering, 2010].

$$P(y_i) = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{y_i!\,\Gamma\left(\frac{1}{\alpha}\right)} * \left(\frac{\lambda_i}{(\frac{1}{\alpha}) + \lambda_i}\right)^{y_i} * \left(\frac{1}{(1+\alpha\lambda_i)}\right)^{\frac{1}{\alpha}} \quad (5)$$

Where α is overdispersed parameter, the mean and variance of negative binomial models are as Equations 6 and 7.

$$Var(Y) = k\alpha + k\alpha^2 = \mu_i + \frac{\mu_i^2}{k} \quad (6)$$

$$E(Y) = \mu_i = k\alpha \quad (7)$$

The vector of estimated coefficients is obtained using the maximum likelihood $L(\lambda_i)$ as Equation 8 [Washington, Karlaftis and Mannering, 2010].

$$L(\lambda_i) = \prod_i \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{y_i!\,\Gamma\left(\frac{1}{\alpha}\right)} * \left(\frac{\lambda_i}{(\frac{1}{\alpha}) + \lambda_i}\right)^{y_i} * \left(\frac{1}{(1+\alpha\lambda_i)}\right)^{\frac{1}{\alpha}} \quad (8)$$

3.2 Model Evaluation

3.2.1 Over Dispersion Evaluation

Deciding whether a Poisson model is appropriate or a negative binomial model can be made based on one of the statistical tests such as Deviance or Pearson's chi-squared goodness of fit. Deviance of the model is as Equation 9:

$$D^m = 2(L^f - L^m) \quad (9)$$

Where L^f and L^m obtained from Equation 3 and 8 respectively. D^m distributed with "n-p" degree of freedom, where 'n' is number observations and 'p' equals the observed variables in the model. If deviance is significantly greater than n-p, it means that the model is overdispersed [Ayati and Ghasemi Noqabi, 2010]. Also, Pearson's chi-squared stated as Equation 10 with the degree of freedom equals to 'n-p' [Washington, Karlaftis and Mannering, 2010].

$$\text{Pearsons } \mathcal{X}^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} \quad (10)$$

Where y_i and $\hat{y}_i$ are the number of observed and estimated crashes respectively.

3.2.2 Goodness of Fitness

The general criteria for the Poisson model and the negative binomial goodness of fit are the deviation and Pearson's chi-squared. If the statistical model is appropriate, both values are distributed as χ 2 with n-p degree of freedom, n is the number of samples and p is the number of variables entered in the model. Therefore, if the regression model is appropriate, both the Scaled Deviance and Scaled Pearson Chi-Square are one or nearly one [Ayati, Zakeri and Sadeghi, 2010].

To determine that the proposed model is appropriate for other intersections and validation of proposed models, T-test has been used and for overall verification of the models, R^2 has been used as Equation 12 [Vaziry et al. 2014].

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y_i})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \qquad (12)$$

Where $\bar{y}$ is the mean of observed crashes and another variables defined previously.

3.3 Crash Reduction Factors

A method for estimating the effect of a variable on crash frequencies is to assume that other variables are constant and calculates the average change in crashes for one-unit increase in a variable. The crash reduction factor is defined as Equation 13 [Ayati and Ghasemi Noqabi, 2010].

$$\text{ARF} = -100\left(\frac{\exp\big(\beta_i(x_i + 1)\big)}{\exp(\beta_i x_i)} - 1\right) \qquad (13)$$

3.4 Prioritizing by Prediction Model

Here, the number of predicted crashes subtracts from the number of observed crashes to calculate the potential of improvement in each location. One of the key features of this ranking method is to report the potential of improvement of all sites in a list. The potential of improvement for each intersection calculated as Equation 14 [AASHTO, 2010].

$$\text{P.I.} = f_j - f_{pj} \qquad (14)$$

Where f_j and f_{pj} are the numbers of observed and estimated crashes at each location.

4. Data Analysis

In this research, 85 dangerous intersections were selected for priority improvement in Tehran's 11th district and crash data obtained from Tehran's police which are related to 2015 and 2016. According to the statistics, more than 30% of the total deaths were due to pedestrian crashes at the intersections so intersection modifications will increase pedestrian safety at the intersection. A summarized of data illustrated in Figures 1 through 5 which Fig. 1 shows that the number of pedestrian crashes in April is at the lowest and at the highest level in December. According to Fig. 2, an ascending increase in the number of crashes has been seen in the study area. Fig. 3 shows that the highest frequency of pedestrian crashes is at 11-14 and 18-21, respectively. In the preliminary analysis of crash data as Fig. 4, it became clear that motorcycles are a serious threat to pedestrian safety. Fig. 5 shows the number of pedestrian crashes based on weekdays. Accordingly, the frequency of pedestrian crashes is higher on weekdays than weekends, and it has reached to the highest level on Mondays.

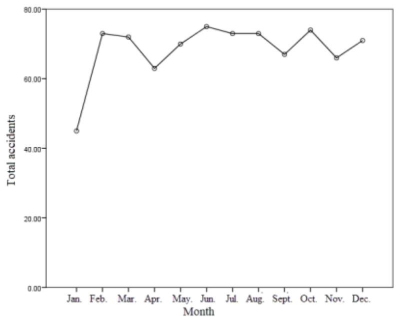

Figure 1. Total pedestrian crashes per month [Tehran municipality, 2017]

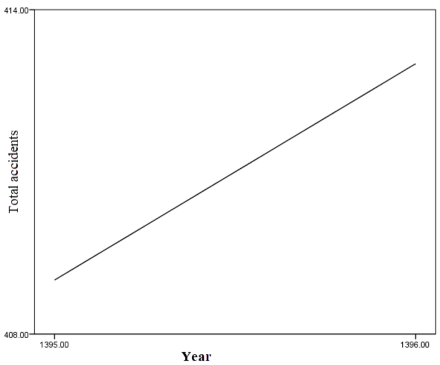

Figure 2. Number of pedestrian crashes per year [Tehran municipality, 2017]

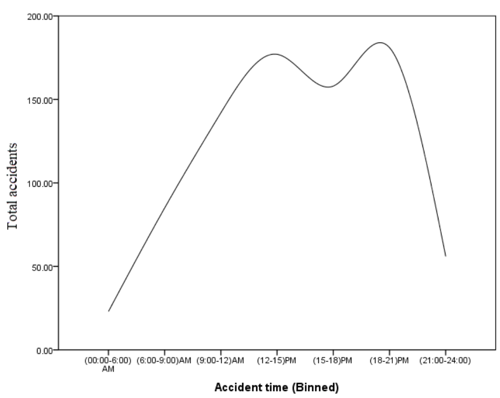

Figure 3. Number of pedestrian crashes with vehicles at different time intervals [Tehran municipality, 2017]

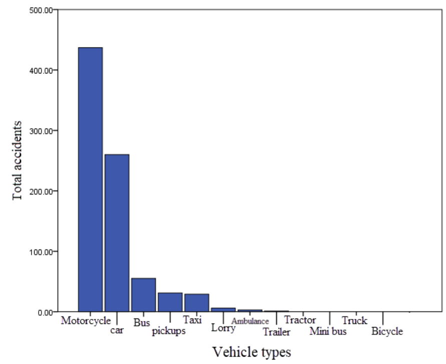

Figure 4. Type of vehicle collision with pedestrians [Tehran municipality, 2017]

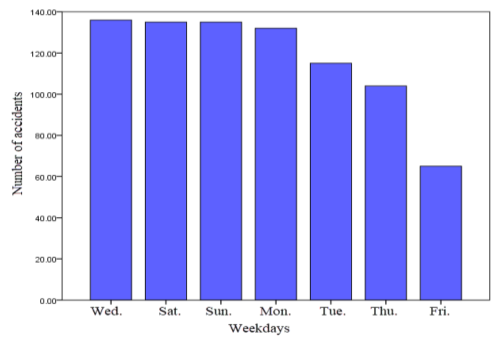

Figure 5. Number of pedestrian crashes with vehicles on weekdays [Tehran municipality, 2017]

4.1 Prioritization by crash prediction model

At this section, an appropriate model was developed to predict crash frequencies at the intersections. According to the previous sections, 85 intersections in Tehran's 11th distinct were used for modeling. For validation of models, the data were divided into two groups of 70 for creating a model and 15 for validating the model. The first step in constructing a suitable model, is the determination of effective independent variables on the dependent variable. Then a list of all independent variables was compiled, prepared and their values were collected for all sample members. After collecting various variables (characteristics of the geometric design, traffic flow characteristics) with the help of statistical tools, the effect of each of these variables was evaluated with the number of crashes. Finally, after creating the model and validating it, they were prioritized concerning the potential for improvement of these 85 intersections. A characteristic of these intersections is given in Table 1.

4.2 Choosing Appropriate Distribution

The dependent variable in this research is pedestrian crash frequencies which is a non-negative value that has been showed in Table 1. According to the comparison of mean and variance of the dependent variable, it was found that its variance is significantly higher than the mean (mean = 5.94 and the variance = 20.07), which indicates the dispersion of pedestrian crash data. Chi-squared test was used to select a Poisson or negative binomial distribution. The results are presented in Tables 2 and 3. According to Table 2, the estimated chi-squared estimates for fitting the Poisson distribution to pedestrian crash data were 9.3735, and is less than $\chi^2_{\propto.k-p-1}$=18.307 so the hypothesis of choosing Poisson distribution hasn't been rejected

Table 1. Frequency analysis of research data in Tehran 11th distinct

Variable name	Description	Mean	SD.	Max.	Min.
Dependent variable					
PVC	Pedestrian Vehicle crashes frequencies	5.94	4.48	17.00	0.00
Land use					
LC	Presence commercial areas =1, otherwise =0	0.76	0.43	1.00	0.00
LR	Presence residential areas =1, otherwise =0	0.71	0.45	1.00	0.00
LE	Presence educational areas =1, otherwise =0	0.32	0.47	1.00	0.00
Transit characteristics					
M	Number of metro stations up to 400 meters	0.23	0.53	2.00	0.00
B	Number of bus stations up to 100 meters	1.19	1.14	4.00	0.00
MB	Number of bus and metro stations up to 200 meters	1.57	1.34	5.00	0.00
Traffic variables					
VT	Average daily traffic volume	32391.70	15622.50	90568.00	8520.00
Ln(VT)	Ln (Average daily traffic volume)	10.91	0.53	11.41	9.05
VP	3-hour pedestrian volumes at pedestrian peak hour	1819.33	1252.53	6470.00	390.00
Ln(VP)	Ln (Average pedestrian volume)	7.29	0.65	8.77	5.97
Intersection geometric design and traffic control variables					
N_L	Number of entry and outlet lanes to the intersection	13.97	4.33	32.00	8.00

S	Presence of exclusive lanes =1, otherwise =0	0.56	0.49	1.00	0.00
D_{cont}	Control type (signalized =1, uncontrolled =0)	0.32	0.48	1.00	0.00
D_{RC}	Sequence (main-main and secondary-secondary =0, main-secondary =1)	0.40	0.49	1.00	0.00
W_{cw}	Average width of sidewalk (Meter)	2.72	0.53	5.10	1.90
E	Percentage of sidewalk utilization	0.15	0.07	30.00	0.00
EW	sidewalk effective width($EW=W_{CW} - E*W_{CW}$)	2.40	0.48	3.40	1.90

According to table 3, the chi-square estimated for negative binomial distribution fitted to pedestrian crashes data were 4.6445 and is less than $\chi^2_{\propto.k-p-1}$=16.919 with a confidence level of 95%. results show that the hypothesis that the distribution of crash data follows a negative binomial distribution has not rejected. Considering that in both chi-square tests for the fitting of the negative binomial and Poisson distribution, the assumption of the distribution of Poisson and negative binomials weren't rejected, so the distribution with the minimum calculated value of chi-square was selected as the appropriate distribution. Therefore, the negative binomial distribution is chosen for the frequency of pedestrian crash data.

4.3 Model Calibrations

To identify effective variables in pedestrian crashes, the negative binomial model has been used. After calibration of many models, it was found that among the 18 variables in Table 1, some of them such as average pedestrian volumes, average vehicle volumes, number of bus and railway (Metro) stations, number of lanes at intersection, entry and outlet street sequence to intersection, average sidewalk width, land use in vicinity of intersections and traffic control type at intersections have a significant relationship with the dependent variable and multicollinearity issue doesn't exist between significant independent variables. The backward elimination method has been used for model calibrations. In this method, first all variables were entered into the equation and then sequentially removed in order to improve significant of variables. The significance level was considered to be 0.1, indicating a moderate level in the selection of independent variables. Calibration process has been summarized in table 4.

In order to choose the most appropriate model, some controls such as the significance of parameters, a logical sign of parameters, R^2 and predicted R^2 have been done and model 5 has been chosen as the most suitable model.

Table 2. Chi-squared test for fitting Poisson distribution to the pedestrian crash data

Num.	Dependent variable (PVC)	Frequency	Relative frequency	Probability of F(y) with Poisson distribution	$(f(py)-f(y))^2$	$\frac{(f(py)-f(y))^2}{f(y)}$
1	1	8	0.11429	0.0024	0.0125	5.1174
2	2	9	0.12857	0.0442	0.0071	0.1610
3	4	13	0.18571	0.1332	0.0028	0.0207
4	5	8	0.11429	0.1602	0.0021	0.0132
5	6	5	0.07143	0.1606	0.0080	0.0495
6	7	5	0.07143	0.1380	0.0044	0.0321
7	8	6	0.08571	0.1037	0.0003	0.0031
8	10	7	0.10000	0.0417	0.0034	0.0816
9	13	5	0.07143	0.0053	0.0044	0.8279
10	15	2	0.02857	0.0009	0.0008	0.8407
11	16	1	0.01429	0.0003	0.0002	0.5683
12	17	1	0.01429	0.0001	0.0002	1.6579
	Total	70	100			9.3735

Table 3. Chi-squared test for fitting negative binomial distribution to the pedestrian crash data

Num.	Dependent variable (PVC)	Frequency	Relative frequency	Probability of F(y) with negative binomial distribution	$(f(py)-f(y))^2$	$\frac{(f(py)-f(y))^2}{f(y)}$
1	0	8	0.11429	0.0169	0.0095	0.5590
2	3	9	0.12857	0.0181	0.0122	0.6721
3	4	13	0.18571	0.0226	0.0266	1.1759
4	5	8	0.11429	0.0169	0.0095	0.5590
5	6	5	0.07143	0.0130	0.0034	0.2625
6	7	5	0.07143	0.0130	0.0034	0.2625
7	8	6	0.08571	0.0144	0.0051	0.3533
8	10	7	0.10000	0.0157	0.0071	0.4525
9	13	5	0.07143	0.0130	0.0034	0.2625
10	15	2	0.02857	0.0078	0.0004	0.0547
11	16	1	0.01429	0.0053	0.0001	0.0153
12	17	1	0.01429	0.0053	0.0001	0.0153
	Total	70	100			4.6445

Table 4. Negative binomial models for predicting pedestrian crashes at intersections in Tehran's 11th district

Model No. / Variable	1	2	3	4	5	6
LC	0.741**	0.621**	0.398*	0.443*	0.451**	0.425*
LR	0.284	0.311				
LE	0.095*	0.109*	0.07	0.023		
MB	0.154**	0.167**	0.143**	0.126**	0.134**	0.131**
N_L	-0.084***	-0.077***	-0.065**	-0.06**	-0.055**	-0.042**
S	0.092	0.113	0.097	0.102		
D_{RC}	0.408**	0.315*	0.228	0.125*	0.257*	0.291**
D_{cont}	0.234	0.163	0.217			
EW	-0.275**	-0.333**	-0.214*	-0.181*	-0.161*	
Ln(VT)	0.55**	0.551**	0.574**	0.597**	0.607**	0.551**
Ln(VP)	0.38**	0.273**	0.251*	0.269**	0.261*	0.263*
E	-1.829					
C	-6.098**	-5.372**	-5.385**	-5.793**	-5.911**	-5.893**
R^2	0.787	0.751	0.731	0.717	0.693	0.659
Predicted R^2 (Eq. 12)	**0.611**	**0.563**	**0.542**	**0.511**	**0.538**	**0.505**

Note: ***, **, * denote Significance at 1%, 5%, 10% level.

4.4 Model Validation

Validating is the most important step in modeling. In order to model validation, T-Test was used for a paired sample. For this purpose, some of the intersections that didn't contribute in modeling process were considered. According to Model 5, the number of pedestrian crashes occurred at intersections were predicted. T-Test was used for a paired sample to compare the difference between the number of predicted and observed crashes. This test was performed using SPSS software. The results presented in Tables 5 and 6. In order to validation, 5% confidence level was considered. Therefore, according to the P-value reported in Table 6, it can be concluded that the model has the required validity. Scaled Deviance and Scaled Pearson Chi-Square of model 5 were 1.234 and 1.089 which both of them are larger than one which indicates overdispersion and confirms using negative binomial model.

Table 5. Summarization of predicted and observed crashes

	Mean	Standard deviation	Standard error
Predicted	4.433	2.529	0.653
Observed	4.267	2.631	0.679

5. Discussion

Based on the results, it was found that Model No. 5 is a suitable model for predicting pedestrian crashes in Region 11. Details of this model are shown in Table 7.

In the final proposed model, since the correlation of Ln(VT) is larger than VT with the independent variable, and it's impossible to use both of them simultaneously (due to the Multicollinearity), the Ln(VT) was used. The Ln(VT) and Ln(PT) is significant with a positive sign, which indicates that as the volume of pedestrians and vehicles increase, the probability of crashes at intersections will increase, in other words, the safety level of intersections will decrease. One of the factors affecting pedestrian crashes is

Table 7. Final proposed model for prediction of pedestrian crashes

Variables	Coefficient	Standard error	Z	P > |z|	ARF
LC	0.451	0.217	2.080	0.037	56.99
Ln (VT)	0.607	0.196	3.096	0.002	83.49
Ln (VP)	0.261	0.134	1.948	0.051	29.82
EW	-0.161	0.099	-1.613	0.104	-14.87
MB	0.134	0.064	2.112	0.035	14.34
NL	-0.055	0.019	-2.788	0.005	-5.35
D_{RC}	0.258	0.141	1.825	0.068	29.43
Constant	-5.911	1.864	-3.172	0.002	--
	0.000	Prob > χ^2		Likelihood ratio χ^2 = 127.36	

surrounding land use that has a positive sign, which indicates that pedestrian crashes are more likely to occur in commercial land uses like CBD. Number of metro and bus stations in vicinity of the intersection have been significant with a positive sign which indicates that by increasing number of metro and bus stations in vicinity of the intersections, the volume of pedestrians will increase and causes a reduction in pedestrian precision and ignore the safety principles when crossing the street.

Table 6. t-student results

t-test	DOF	p-value	Result
0.617	14	0.512	validate

One of the important variables in increasing the safety of pedestrians is to increase the effective width of sidewalks. The presence of hawkers, motorcycle parks in sidewalks and occupation of sidewalks by shops reduce the width of sidewalks and compel pedestrians to use the carriageway and increase the probability of crashes. In the proposed model, this variable has a negative sign that confirms the significance of the variable.

In our case study, due to the old texture of streets, especially in the southern regions, high density of shopping centers and the lack of suitable systems for goods transportation, Odd and Even and traffic restriction zones has reduced the car volumes and in the absence of proper supervision, motorcycle volumes with commercial applications have been increased which leads to increasing the probability of pedestrian crashes. Therefore, the number of entry and outlet lanes in the model is consistent with the above description.

At the intersections of the main streets, due to increase in the volume of vehicles and vehicle speeds, the risk of pedestrian crashes increases as the sign of this variable in the proposed model reflects this claim.

5.1 Prioritization results based on crash prediction model

Prioritization of intersections in Region 11 is illustrated in Figure 6. Based on the model, intersections 37, 29 and 21 have the most potential for improvement, respectively, with a potential improvement of 6.93, 4.31, and 3.68. According to Figure 6, it can be concluded that the intersection of 37 (Valiasr- Emam Khomeini) with features such as being in a commercial area and an old texture with narrow sidewalks, high car and pedestrian volumes, proximity to the bus stations and the lack of facilities to improve the safety of pedestrians has the greatest potential for

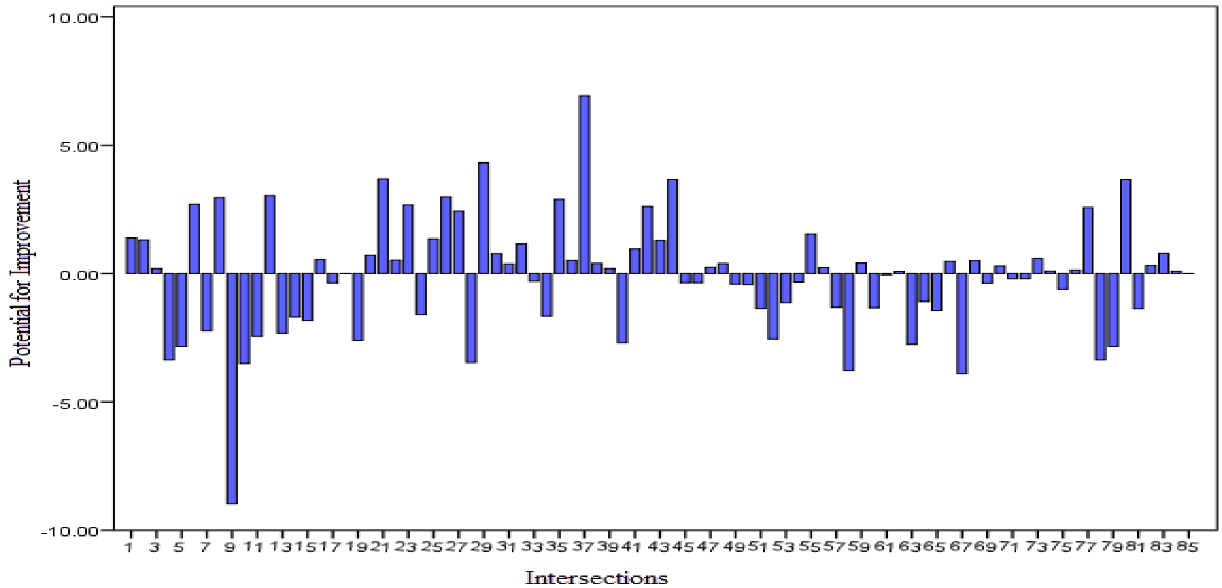

Figure 6. The potential for improvement of intersections based on prediction of pedestrian crashes method

improvement. Figure 6 also shows that the intersection of 9 (Enghelab-Valiasr) with features such as being in a commercial area with relatively wider sidewalks, high car and pedestrian volumes, proximity to the bus and metro stations has the least potential for improvement, which can be attributed to the development of underground walkways for pedestrians in this place.

6. Conclusions

In this research, the factors affecting pedestrian crashes have been investigated and a negative binomial model was used to estimate the frequency of pedestrian crashes based on effective factors. The most important results of this research are the following:

- Increasing pedestrian volumes, especially in the central areas such as CBD with worn-out texture and narrow sidewalks lead to using street and increase the probability of pedestrian crash. In the proposed model, the natural logarithm of the average pedestrian volume is significant with a positive sign, which indicates as the pedestrian volumes increase, pedestrian crash frequencies will increase at intersections. Based on crash reduction factor, one-unit increase in natural logarithm of average pedestrian volume will increase the crash frequencies by 29.82% while the other variables remain constant.

- In central areas with mostly commercial land uses, along with the application of traffic restriction plans (restriction and Odd-Even zones) and the lack of proper transportation systems lead to increasing in the number of motorcycles. In the proposed model, the natural logarithm of average vehicle volume is significant with a positive sign, which indicates that as vehicle volumes increased, the pedestrian crash frequencies increase at intersections. Based on the crash reduction factor, one-unit increase in the natural logarithm of average vehicle volume will increase the crash frequencies by 83.49% while the other variables remain constant.

- One of the factors affecting pedestrian crashes is surrounding land use which in the proposed model has been significant with a positive sign which indicates that pedestrian crash is more likely to occur in commercial land uses. according to the proposed model, pedestrian crash probability in areas with commercial land use is more likely than other land uses. Based on the crash reduction factor, one-unit increase in the density of central business distinct will increase the crash frequencies by 56.99% while the other variables remain constant.

- In the proposed model, the number of metro and bus stations variable has a positive sign. The existence of metro and bus stations in vicinity of the intersection leads to increase in the volume of pedestrians, reduce pedestrian precision and ignore the safety principles when crossing the street. One-unit increase in the number of public transport stations will increase the frequency of pedestrian crashes at intersections by 14.34% while the other variables remain constant.

- One of the important parameters in increasing the safety of pedestrians is to increase the effective width of sidewalks. The presence of hawkers, park of motorcycles on sidewalks and capture of walkways by shops, compel pedestrians to use the street and increase the probability of crashes. In the proposed model, this variable was significant with a negative sign, indicating that by one-unit increase in the effective width of sidewalks, the probability of pedestrian crashes at intersections will reduced by 14.87% while the other variables remain constant.
- According to the proposed model, the probability of pedestrian crashes increases due to high vehicle volumes and high vehicle speeds at the intersections of main streets and the sign of this variable in the proposed model reflects this claim.

- The results of the prioritization based on the prediction model show that the intersection of Imam Khomeini-ValiAsr with the potential for improvement of 6.93 has the highest potential for improvement and the Enghelab-Valiasr intersection with an improvement potential of -8.97 has the least potential for improvement.

- Some factors affect the safety of pedestrians such as sidewalks in most commercial areas in the city center with a large pedestrian volumes and presence of hawkers, occupation of sidewalks by shops, forcing pedestrians to use the street and thus increase the probability of crashes.

7. References

- AASHTO (2010) "Task force on development of the highway safety manual, and transportation officials", Transportation Research Board, Highway safety manual.

- Ayati, A. and Abbasi, A. (2011) "Application of zero-inflated regression models in modeling crashes on urban highways", IQBQ, Vol. 11, No. 4, pp. 1-15.

- Ayati, E., and Ghasemi Noqabi, M. (2010) "Research note: prediction of the frequency of damage-only crashes at four-leg signalized intersections", Journal of Transportation, Vol. 6, No. 3, pp. 287-300. (In Persian)

- Ayati, E., Zakeri Sardroodi, J. A. and Sadeghi, A. A., (2010) "Modeling the frequency of crash occurrence at rail-road crossings", Journal of

Transportation Engineering, Vol. 2 No. 2. (In Persian)

- Bennet, S. A. and Yiannakoulias, N. (2015) "Motor-vehicle collisions involving child pedestrians at intersection and mid-block locations", Accident Analysis and Prevention, Vol. 78, pp. 94-103.

- BITRE (The Bureau of Infrastructure, Transport and Regional Economics), (2013) "Road Deaths Australia", Statistical Summary, Canberra ACT.

- Chang, L. Y. (2005) "Analysis of freeway crash frequencies: negative binomial regression versus artificial neural network", Safety science, Vol. 43, No. 8, pp. 541-557.

- Demetriades, D., Murray, J., Martin, M., Velmahos, G., Salim, A., Alo, K. and Rhee, P. (2004) "Pedestrians injured by automobiles: relationship of age to injury type and severity", Journal of the American College of Surgeons, Vol. 199, No. 3, pp. 382-387.

- Elyasi, M., Saffarzadeh, M., Boroujerdian, A., Semnarshad, M., and Mazaheri, M. (2017) "Prioritization of Suburban Crash Factors Based on Analytical Network Process", International Journal of Transportation Engineering, Vol. 5, No. 2, pp. 197-209.

- European Commission, (2016) "Annual Crash Report", Directorate General for Transport.

- Hauer, E., Council, F. M., and Mohammedshah, Y. (2004) "Safety models for urban four-lane undivided road segments", Transportation Research Record, Vol. 1897, No. 1, pp. 96–105.

- Highway Capacity Manual, (2010) "HCM 2010", Transportation Research Board, National Research Council, Washington DC.

- Hosseinpour, M., Yahaya, A. S. and Sadullah, A. F. (2014) "Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads", Crash Analysis and Prevention, Vol. 62, pp. 209-222.

- Jermprapai, K. and Srinivasan, S. (2014) "Planning-level model for assessing pedestrian safety", Transportation Research Record, Vol. 2464, No. 1, pp. 109-117.

- Jonsson, T., Ivan, J.., and Zhang, C. (2007) "Crash prediction models for intersections on rural multilane highways: differences by collision type". Transportation research record, Vol. 2019, No. 1, pp. 91–98.

- Kim, J. Y. (2016) "The influences of intersection roadway characteristics on pedestrian-vehicle collisions", (Doctoral dissertation, UCLA).

- Kim, D. G., Washington, S. and Oh, J. (2006) "Modeling crash types: new insights into the effects of covariates on crashes at rural intersections", Journal of Transportation Engineering, Vol. 132, No. 4, pp. 282–292.

- LaScala, E. A., Gruenewald, P. J. and Johnson, F. W. (2004) "An ecological study of the locations of schools and child pedestrian injury collisions", Crash Analysis and Prevention, Vol. 36, No. 4, pp. 569-576.

- Leden, L., Gårder, P. and Johansson, C. (2006) "Safe pedestrian crossings for children and elderly", Crash Analysis and Prevention, Vol. 38, No. 2, pp. 289-294.

- Li, D., Ranjitkar, P., Zhao, Y., Yi, H. and Rashidi, S. (2017) "Analyzing pedestrian crash injury severity under different weather conditions", Traffic Injury Prevention, Vol. 18, No. 4, pp. 427-430.

- Li, Y., Nan, S., Gong, X. and Ma, R. (2019) "A geometric design method for intersections with pre-signal systems using a phase swap sorting strategy", PloS one, Vol. 14, No. 5.

- Litman, T. and Steele, R. (2017) "Land use impacts on transport. Canada: Victoria", Transport Policy Institute.

- Miranda-Moreno, L. F., Morency, P. and El-Geneidy, A. M. (2011) "The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections", Accident Analysis and Prevention, Vol. 43, No. 5, pp. 1624-1634.

- Mirbaha, B., Saffarzadeh, M. and Noruzoliaee, M. (2013) "A model for predicting schoolchildren accidents in the vicinity of rural roads based on geometric design and traffic conditions", International Journal of Transportation Engineering, Vol. 1, No. 1, pp. 25-33.

- Mohammadi, M. A., Samaranayake, V. A., and Bham, G. H. (2014) "Crash frequency modeling using negative binomial models: An application of generalized estimating equation to longitudinal data", Analytic Methods in Crash Research, Vol. 2, pp. 52-69.

- Mooney, S. J., DiMaggio, C. J., Lovasi, G. S., Neckerman, K. M., Bader, M. D., Teitler, J. O. and Rundle, A. G. (2016) "Use of Google Street View to assess environmental contributions to pedestrian injury", American Journal of Public Health, Vol. 106, No. 3, pp. 462-469.

- Moudon, A. V., Lin, L., Jiao, J., Hurvitz, P., and Reeves, P. (2011) "The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington", Crash Analysis and Prevention, Vol. 43, No. 1, pp. 11-24.

- Mothafer, G. I., Yamamoto, T. and Shankar, V. N. (2017) "A negative binomial crash sum model for time invariant heterogeneity in panel crash data: Some insights", Analytic Methods In Crash Research, Vol. 14, pp. 1-9.

- Murphy, B., Levinson, D. M. and Owen, A. (2017) "Evaluating the safety in numbers effect for pedestrians at urban intersections", Crash Analysis and Prevention, Vol. 106, pp.181-190.

- National Highway Traffic Safety Administration (2017) "Traffic Safety Fact", U.S. Department of Transportation NSA, Vol. 230, pp. 8-10.

- Noland, R. B., Quddus, M. A. (2005) "Congestion and safety: a spatial analysis of London", Transportation Research Part A: Policy and Practice, Vol. 39, No. 7, pp. 737–754.

- Pulugurtha, S. S. and Sambhara, V. R. (2011) "Pedestrian crash estimation models for signalized intersections", Accident Analysis and Prevention, Vol. 43, No. 1, pp. 439-446.

- Qureshi, M., Virkler, M. R., Sanford Bernhardt, K. L., Spring, G., Avalokita, S., Yathapu, N. and

Gibbons, K. (2003) "Highway rail crossing project selection" (No. RDT 03-017,).

- Semnarshad, M., Elyasi, M., Saffarzadeh, M. and Saffarzadeh, A. (2018) "Identification and prioritization of crash-prone segments using international roughness index", International Journal of Transportation Engineering, Vol. 6, No. 1, pp. 35-48.

- Siddiqui, C., Abdel-Aty, M. and Choi, K. (2012) "Macroscopic spatial analysis of pedestrian and bicycle crashes", Crash Analysis and Prevention, Vol. 45, pp. 382-391.

- Tehran Municipality (2017) "Abstract data of transportation and traffic organization of Tehran municipality", Deputy of Transportation and Traffic of Tehran Municipality. (In Persian)

- Ticali, D. and Corriere, F. (2018) "Turbo roundabouts: geometric design parameters and performance analysis", GSTF Journal on Computing (JoC), Vol. 2, No. 1.

-

- Vaziri, E., Khashayeepur, M., Abedini, M., Ayati, A., Chaychi Motlagh, M. and Naghdizadeh, M. (2014) "Modeling the forecast number of traffic accidents in Tehran", Tehran Traffic Control Company.

- Washington, S. P., Karlaftis, M. G. and Mannering, F. (2010) "Statistical and econometric methods for transportation data analysis", Chapman and Hall/CRC.

- Xie, K., Ozbay, K., Kurkcu, A., and Yang, H. (2017) "Analysis of traffic crashes involving pedestrians using big data: Investigation of contributing factors and identification of hotspots", Risk Analysis, Vol. 37, No. 8, pp. 1459-1476.

- Xie, S. Q., Dong, N., Wong, S. C., Huang, H., and Xu, P. (2018) "Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure", Accident Analysis and Prevention, Vol. 121, pp. 285-294.