Research Article

Pedestrian Crash Exposure Analysis Using Alternative Geographically Weighted Regression Models

Seyed Ahmad Almasi ¹, Hamid Reza Behnood ¹ and Ramin Arvin ²

¹Department of Transportation Planning, Faculty of Engineering, Imam Khomeini International University, Qazvin, Iran

²Department of Civil & Environmental Engineering, University of Tennessee, Knoxville, USA

Correspondence should be addressed to Hamid Reza Behnood; hr.behnood@gmail.com

Received 30 November 2020; Revised 25 January 2021; Accepted 8 February 2021; Published 18 February 2021

Academic Editor: Mohamed Hussein

Copyright © 2021 Seyed Ahmad Almasi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to develop a sustainable, safe, and dynamic transportation system, proper attention must be paid to the safety of pedestrians. The purpose of this study is to analyze the surrogate measures related to pedestrian crash exposure in urban roads, including the use of sociodemographic characteristics, land use, and geometric characteristics of the network. This study develops pedestrian exposure models using geographical spatial models including geographically weighted regression (GWR), geographically weighted Poisson regression (GWPR), and geographically weighted Gaussian regression (GWGR). In general, the results of the GWPR model show that the presence of a bus station, population density, type of residential use, average number of lanes, number of traffic control cameras, and sidewalk width are negatively associated with increasing the number of crashes. In this study, in order to identify traffic analysis zones (TAZ) based on the observed and predicted crash data, spatial distance-based methods using GWPR outputs have been used. This study shows the dispersion and density of pedestrian crashes without possessing the volume of pedestrians. Comparison of the performance of GWPR and Poisson models shows a significant spatial heterogeneity in the analysis.

1. Introduction

Pedestrians are known as vulnerable road users, and the severity of pedestrian injuries in motor vehicle crashes is relatively high. Today, ensuring the safe movement of pedestrians is one of the most challenging concerns for transportation engineers. In general, in urban accidents, drivers and passengers have the largest share of the comprehensive cost of traffic accidents (94%) [1]. In order to develop a sustainable, safe, and dynamic transportation system, proper attention must be paid to the safety of pedestrians. The proportion of pedestrian casualties in the world has increased by an average of 11 percent to 14 percent over the past decade, so addressing pedestrian safety and raising awareness about safe pedestrianization is an important issue. This study was conducted with the aim of addressing the safety of pedestrians and identifying the extent of exposure to pedestrian crashes in urban areas and

identifying accident-prone areas. Obviously, the number of people walking on the streets (i.e., pedestrian trips) and the factors that cause the presence of more pedestrians on the streets is one of the best measures for pedestrian exposure [2–5].

However, continuous measurement of pedestrian travels is difficult considering all the effective variables because it requires the use of significant resources and many factors that play a role in creating pedestrian travels. The purpose of this study is to investigate the available criteria and select the most effective measures in order to predict the variable of pedestrian exposure (pedestrian trips) and identify areas prone to pedestrian crashes. In other words, the purpose of this study is to analyze the surrogate measures related to pedestrian exposure in urban roads, including the use of sociodemographic characteristics, land use, and geometric characteristics of the network. The three-step process in the study involves the development of exposure models using

geographical spatial models including geographically weighted regression (GWR), geographically weighted Poisson regression (GWPR), and geographically weighted Gaussian regression (GWGR). Exposure models in this study are compared with the study model developed by Lee et al. [5] which were performed using Tobit method and generalized linear models (GLMs) and predicted pedestrian travel. In their study, it was suggested that the effect of the geographical location of exposure variables for pedestrian crashes be investigated in future studies. In the current study, the effect of spatial exposure variables based on their geographical location has also been investigated. Then, in order to identify the best exposure model between GWR, GWPR, and GWGR models, the Akaike Information Criterion (AIC) index and P value were used after validating the models to predict pedestrian crashes. This method can be described as a pedestrian safety analysis on urban roads (microlevel) with macrolevel data. Also in this study, in order to identify traffic analysis zones (TAZs) based on the observed and predicted crash data, spatial distance-based methods using GWPR outputs have been used.

The city of Tehran in Iran and its urban areas have been considered in this study. Although the two-step process (i.e., first identifying pedestrian collision variables and second crash prediction and identifying high-risk areas) has a relatively larger modeling error than the one-step model (pedestrian crash prediction only), but still by analyzing the volume of pedestrians and crashes, their output can lead to a better understanding of safety [5]. This study addresses pedestrian safety in the study area by identifying the best model for dealing with pedestrian crashes on major urban roads (first- and second-degree arteries and collector streets) as well as creating a safety analysis process for regions where pedestrian crash data are not available. This process can help transportation authorities create safer paths for pedestrians by implementing appropriate safety interventions.

2. Literature Review

Pedestrian safety is a growing concern, and so extensive studies have been conducted to ensure pedestrian safety. Researchers have tried to identify the factors contributing to pedestrian fatalities as well as identifying the urban areas with the highest risk of crashes for pedestrians by developing spatial relationships. For this purpose, in this section, past studies on the development of exposure models and spatial analysis of crashes have been reviewed.

Lee and Abdel-Aty [6] conducted a comprehensive study of pedestrian crashes at intersections in Florida. This study followed the Keall [7] method to use pedestrian personal travel data to create a rational model of a pedestrian exposure with crashes. In the proposed exposure measure, different walking patterns were reflected by different age groups of pedestrians. Miranda-Moreno et al. [8] analyzed two important relationships between land development and pedestrianization: (a) between land use and pedestrian activities and (b) between risk exposure (pedestrian and vehicle activities) and pedestrian crash frequency. The authors concluded that the land use pattern affects the level of

pedestrian activities with a direct impact on pedestrian safety. Ukkusuri et al. [9] developed a pedestrian count crash frequency model for New York City using a negative binomial model and a stochastic parameter. This model found that the ratio of illiterate population, business areas, school areas, functional characteristics of the intersection, type of access control on the roads, and the number of lanes had a positive effect on pedestrian crashes.

In order to select the variables used in the proposed exposure model, several previous studies have been reviewed. Previous researchers have shown that pedestrian volume is a significant measure of exposure that has a positive effect on the occurrence of vehicle-pedestrian collisions [2–5, 10]. Another significant measure of exposure is the effects of land use patterns that have long been studied by researchers [5, 11–13]. Wier et al. [13] found that the number of pedestrian crashes is relatively higher in commercial and residential areas. There are many studies that describe the impact of demographic and socioeconomic characteristics on pedestrian safety (e.g., [12]).

Although previous researchers have attempted to explain pedestrian exposure, there are very few studies that specifically identify the exact causes of this criterion. Another issue is the reliability of pedestrian volume data. There have been many cases where pedestrian volume data were not available or were not sufficiently accurate to perform a safety analysis. A reliable process for identifying surrogate measures is needed to express the pedestrian exposure criteria in such cases. Apart from these issues, the use of the negative binomial (NB) or even zero-inflated negative binomial (ZINB) model in microlevel safety analysis has been questioned by many authors [5, 14–16]. However, some authors have confirmed it in macrolevel analysis [17].

In general, spatial prediction of crashes using localized parameters gives us more accurate predictions compared to methods in Highway Safety Manual [18] that use global parameters. In addition, traffic exposure criteria (such as AADT and length of segment) are considered as predictors of crash frequency and have been widely used by transportation professionals to predict the occurrence of crashes at a particular site. Therefore, in the field of safety performance functions (SPFs), understanding the different spatial relationships between the main factors of exposure and the frequency of crashes has significant potential for the development of localized SPFs that can potentially provide more accurate crash predictions at separate sites [19].

In the current study, in order to identify high-risk TAZs based on observed and predicted crash data, two methods have been adopted: (1) frequency-based methods and (2) distance-based models. The first group measures the severity of point events based on the density of an area. These methods include kernel density estimation (KDE). The second group measures the spatial dependence of point events based on the distance of points from each other. This group includes methods such as nearest neighbor distances, K -functions, and Moran I [20, 21]. Hadayeghi et al. [22] presented traditional crash prediction models for 463 TAZs in Toronto using traditional NB (global) and GWPR general regression models. The results showed that GWPR models

were able to partially deal with spatial dependence as well as spatial heterogeneity resulting from these factors and TAZs. Xu and Huang [23] modeled the total crash frequency as a function of road length density, population density, average household income, and percentage of road sections with different speed limits and showed that the GWPR model due to the instability of the crash location has acceptable accuracy compared to the NB model with random parameter. A parametric GWPR model was also developed to estimate some parameters globally and some locally [24]. Similarly, a study by Rhee et al. [25] investigated traffic accidents with spatial correlation and spatial relevance using advanced spatial modeling methods. The results showed that the statistical performance of GWR was superior in the correlation coefficient of localization.

3. Methods

In this study, in the first step, which is the identification of exposure variables, several statistical methods have been used to identify these variables and the crash frequency is examined based on different modeling methods. In the second step, crash prediction models are presented at the TAZ level using surrogate variables. The following is a brief description of the modeling techniques used in these two steps.

3.1. Models Used to Identify Exposure Variables

3.1.1. Generalized Linear Models. Generalized linear models (GLMs) are a general class of statistical models that include many common models with specific features. A typical GLM is as follows:

$$Y = \sum_{i=1}^m \beta_i \chi_i + \varepsilon_i. \quad (1)$$

In this equation, Y is the linear prediction and ε_i is the error parameter. In the generalized linear model, the assumptions of independent and normal distribution in Y are given. This distribution includes such cases as normal, Poisson, gamma, and binomial distributions [26]. The GLM is a flexible generalization of ordinary linear regression that allows the use of response variables that have error distribution models other than the normal distribution.

3.1.2. Tobit Model. In this study, in order to eliminate any negative prediction of pedestrian crash, Tobit model was used to identify the measure of exposure. The Tobit model is a statistical model used to describe the relationship between a censored dependent variable y_i and an independent variable (or vector) x_i . The Tobit model is as follows:

$$\begin{aligned} y_i^* &= \beta \chi_i + \varepsilon_i \quad i = 1, 2, \dots, N, \\ y_i &= \begin{cases} y_i^*, & \text{if } y_i^* > 0, \\ 0, & \text{if } y_i^* \leq 0. \end{cases} \end{aligned} \quad (2)$$

In this relation, y_i^* is a hidden variable that can only be seen if it is positive. Also, N is the number of observations, y_i is a dependent variable, x_i is a vector of explanatory variables, β is a vector of estimable parameters, and ε_i is a normal and independent distribution. The error parameter also has a mean of zero and a variance of σ^2 [27].

3.1.3. Variable Importance for Exposure Model Using Random Forest. Important explanatory variables can be determined using a random forest exposure model. The first step in this process is to place a random forest of data. During the fitting of this process, an out-of-bag error (a method for measuring random forest prediction error) is recorded for each data point and averaged in the forest. In order to measure the importance of the j^{th} attribute after training, the values of the j^{th} attribute can be changed among the training data and the out-of-bag error is re-estimated in this turbulent dataset [28].

3.2. TAZ Level Crash Predictive Models

3.2.1. Network KDE. As mentioned earlier, many recent studies have used the KDE network method developed by Okabe et al. [29] to examine the spatial correlation of point events in a road network. In this study, the KDE network method has been used to estimate the density of road sections in the Tehran road network. This method is based on the study method of Okabe and Sugihara [30].

In this study, the KDE network was performed on 1000-meter sections, similar to those proposed by Xie and Yan [20] and Nie et al. [31]. Also, according to studies [32, 33] in order to achieve more accurate KDE results, three values of 100, 200, and 500 meters have been considered for bandwidth measure. See Okabe and Sugihara [30] for more details on the computational process.

3.2.2. Geographically Weighted Regression Models. Geographically weight regression (GWR) is an exploratory method that has been adopted in the relevant literature mainly to deal with spatial variables. Past studies on the relationship between urban form and pedestrian behavior have mainly used global regression models. However, since the present study includes urban areas with the main arterial functional class and collectors, the behavioral characteristics of pedestrians may be different in each one. Therefore, it is likely that the relationship between urban form and walking varies across the study area.

In this study, a Gaussian GWR model is used to evaluate the relationship between exposure variables (land use and street characteristics around houses as independent variables). To determine which model is appropriate, a comparison was made between Gaussian GWR and geographically weighted regression and geographically weighted Poisson regression. A model with lower AIC values is a more appropriate model [34–36].

A GWPR model is also used in this study. In a GWPR model, the frequency of crashes is predicted by a set of

explanatory variables in which the parameters are allowed to change in space. This model can be written as follows [37]:

$$\ln(\lambda_i) = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)\ln(E_{vi}) + \sum_{k=1}^k \beta_k(u_i, v_i)x_{ij}. \quad (3)$$

In this relation, (u_i, v_i) specifies the coordinates of region i . It should be noted that, in GWPR, $\beta_k(u_i, v_i)$ is a function of the coordinates of the center for region i .

In this study, GWR4.0 software was used to identify high-risk points in which the chances of determining the walking exposure variable increase because changes in independent variables are given in the first step of modeling. The results of GWR, GWPR, and GWGR in ArcGIS, version 10.2, were mapped to visualize spatial relationships. Also, even if there is a discontinuity in the study area, the optimal bandwidth has been selected based on several experiments to ensure that the blank spaces are outside the crash points of the study area.

3.3. Measures of Goodness of Fit. To evaluate and compare the performance of the models, three statistics were used to measure the accuracy of the estimates. First, we used AIC, which indicates that the lower the AIC, the better the model [38]. The AIC is measured as follows:

$$\begin{aligned} \text{AIC} &= D + 2K, \\ \text{AICc} &= \text{AIC} + \frac{2k(k+1)}{n-k-1}, \end{aligned} \quad (4)$$

where D represents the model deviation and k is the number of parameters. In GWPR, due to the nonparametric framework of the model, the number of parameters is meaningless. Therefore, an effective number of parameters must be considered, which can be written as follows [39]:

$$\begin{aligned} K &= \text{trace}(S), \\ \text{MAD} &= \frac{\sum_{i=1}^n |y_{\text{pred}} - y_{\text{obs}}|}{n}, \\ \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^n (y_{\text{pred}} - y_{\text{obs}})^2}{n}}, \end{aligned} \quad (5)$$

where S is the hat matrix. In addition to AIC, we also used mean absolute error (MAE) and root mean square error (RMSE) to compare model performance. Lower MAE and RMSE values indicate better model performance. Finally, Moran's I statistics model was used to validate the models. Statistically, Moran's I statistics is a measure of spatial correlation. In this study, the Moran test was used to examine whether the residuals of city-wide crash predictions were spatially related to neighboring TAZs. Negative (positive) value of Moran's I statistics indicates a negative (positive) spatial correlation at the overall level.

3.4. Data Preparation. The main source of data for this study is Tehran Municipality. The data used in this study are shown in Table 1. In this study, the analyzed zones have been considered for model development based on the variables of exposure in TAZs, which are 560 zones for Tehran. Of course, it should be considered that, in order to match the analyzed zones, the whole city can be divided into equal units, but due to the lack of homogeneous distribution of pedestrian crashes, many of the identical zones will have zero observed crashes.

The TAZ characteristics selected for the crash analysis include all the variables in Table 1. All items selected as exposure variables are items that affect the frequency of crashes. Crash data variable (CR) shows the total number of pedestrian crashes in Tehran. The density of speed cameras (TCC) indicates a risk factor at their installation site, as these devices are typically installed in locations where drivers need more focus and are at greater risk of road crashes [32]. Bus station (BS) can be a risk factor as a large number of pedestrians get on and off in one place and some of them tend to cross the street [5]. The presence of schools (SC) is one of the most important places to attract pedestrians, so the presence of schools in TAZs during the hours of the day is a risk factor for pedestrian crashes [5]. The presence of a pedestrian bridge (PB) based on studies has improved pedestrian safety in conflicting with vehicles [5]. Also, the presence of intersections in any zone increases the risk of pedestrian collisions. It is obvious that population density (TP0) in each zone and the density of vulnerable users (TP1 and TP2) increase the risk of pedestrian collisions [5, 32, 33]. This study was conducted in two steps including [1] identifying the variables of pedestrian exposure and [2] investigating the spatial-geographical relationship between the variables and the spatial crash prediction at the TAZ level. GIS and SPSS software were used to extract and process data for the first step and GWR4 for the second step. The integration of the database with all the information collected in TAZs is done with the help of standard tools in GIS that allow spatial search, layer addition, and spatial operations based on topological relationship.

Table 1 shows the pedestrian crash dataset of Tehran, which includes 1231 observed cases. Descriptive variables were also classified into three categories: "demographic and socioeconomic," "land use," and "traffic and geometric". Out of 25 variables collected, 15 variables are listed in Table 1 based on the results of the first step in the study. Road network in Tehran, including arterial roads and collectors, has been used for analysis. Data were collected from various sources. Figure 1 shows the study area and the status of existing crashes. Figure 2 shows the KDE crash density function based on crash point and crash density per kilometer in three bandwidths of 100, 200, and 500 meters.

The correlation between the descriptive variables used in the Tobit model (selected model based on the results of the first step of the study) was investigated before the modeling process. Pairs of variables with correlation coefficients higher than 0.6 are not included in the models simultaneously [5]. In the modeling process, first the explanatory variables with the lowest correlation values were included in

TABLE 1: Description of variables used in the study.

Variable	Description
Crash data (CR)	Pedestrian crash data for the years 2017 to 2019 in Tehran
Bus stops (BS)	Location of bus stops in the existing situation and detailed plan of Tehran
Schools (SC)	Location of all schools in Tehran urban zones including both existing and planned schools
Pedestrian bridges (PB)	Location of pedestrian overpasses
Intersections (TS)	Location of all controlled and noncontrolled intersections
Total population (TP0)	Based on the last census in 2016
Children population (TP1)	Based on the last census in 2016
Elder people population (TP2)	Based on the last census in 2016
Motorcycles (TM)	Based on the data recorded in the 2016 census and registered in police databases
Vehicles (TC)	Based on the data recorded in the 2016 census and registered in police databases
Residential land use (RE)	Data gathered by Tehran Municipality's staff in 2017
Business land use (BU)	Data gathered by Tehran Municipality's staff in 2017
Recreational land use (RCE)	Data gathered by Tehran Municipality's staff in 2017
Average road width (AW)	Based on GIS layers of the Tehran Municipality
Average number of lanes (AL)	Based on GIS layers of the Tehran Municipality
Average length of median refuges (TR)	Based on GIS layers of the Tehran municipality
Average sidewalk width (PP)	Average sidewalk width in each traffic zone based on aerial photos
Average speed (SA)	Based on the data adopted from the Traffic Control Center of Tehran Municipality
Average road slope (AS)	Based on GIS layers of the Tehran Municipality
Speed cameras (TCC)	Based on the data adopted from the Traffic Control Center of Tehran Municipality

FIGURE 1: The study area and the status of existing crashes.

the model and the variables with relatively lower correlation values were preferred in the model (Tables 2 and 3).

4. Results and Discussion

A total of six exposure models have been developed in this study (Table 4). Because two different modeling methods are used (GLM vs. Tobit) to compare the best model, it is not appropriate to compare the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). Therefore, to compare the models, the mean absolute deviation (MAD) and the root mean square error (RMSE) for each model have been used [5]. Table 4 shows that the Tobit

model using all variables performs best with the lowest MAD and RMSE values. The Tobit model also shows any predicted negative pedestrian crashes using the exposure variables equal to zero because the lower limit is set at zero.

Between 2017 and 2019, 1228 pedestrian crashes were reported in Tehran, in which 44 people died and 1184 were injured. A total of 4979 schools, 4831 bus stops, 927 pedestrian bridges, 801 lighted intersections, and 5386 traffic control cameras located in 560 different TAZs were included in this analysis (Table 1), with an average number of variables ranging from 0 to 80 in all TAZs. This initial benchmark shows that the difference between the scatterings of exposure variables at the level of TAZs is significant. The

FIGURE 2: KDE crash density functions in different bandwidths. (a) KDE crash density function for bandwidth = 100 m. (b) KDE crash density function for bandwidth = 200 m. (c) KDE crash density function for bandwidth = 500 m. (d) KDE crash density function per km for bandwidth = 100 m. (e) KDE crash density function per km for bandwidth = 200 m. (f) KDE crash density function per km for bandwidth = 500 m.

TABLE 2: Pearson correlation coefficients.

Correlation		SC	BS	PB	TS	TP0
SC	Pearson	1				
	Sig.					
BS	Pearson	0.54**	1			
	Sig.	0.00				
PB	Pearson	0.39**	0.60**	1		
	Sig.	0.00	0.00			
TS	Pearson	0.36**	0.41**	0.21**	1	
	Sig.	0.00	0.00	0.00		
TP0	Pearson	0.63**	0.60**	0.47**	0.37**	1
	Sig.	0.00	0.00	0.00	0.00	

*Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

results of GWR, GWGR, and GWPR models are shown in Tables 5–7, respectively. Table 8 compares the GWR, GWGR, and GWPR models. According to the results shown in these tables, the GWPR model has a higher accuracy in predicting crashes based on exposure variables. In Figure 3, based on crash predictive models and using ArcGIS software, a pedestrian crash map of Tehran has been produced. Table 9 shows the ANOVA values for the GWPR model.

Tables 5–9 show the results of the GWR, GWGR, and GWPR models with adaptive bisquare kernel for predicting crashes. In the kernel density function model, the lowest AICC value was obtained based on the adaptive bisquare

kernel. Notably, we found that the results were largely consistent with the adaptive bisquare kernel. Also, the Lagrange coefficient values for the GWPR model and the global model were 0.12 and 0.15, respectively, which is less than the critical Lagrange value (3.48). The study by Hezaveh et al. [1] also confirms the adaptive bisquare kernel in relation to the Gaussian adaptive for urban TAZs.

The comparison of AIC, AICC, deviation, MAE, and RMSE presented in Table 8 shows that the GWPR model is more appropriate than the global model. The value of Moran I (0.031) indicates that, in the GWPR model, the residues are not related to each other. In addition, VIF values

TABLE 4: Comparison of the exposure models developed in the study.

Model type (exposure)	MAD	RMSE
Exposure model (GLM) using all variables	26.91	40.12
Exposure model (GLM) using PCA variables	37.21	49.26
Exposure model (Tobit) using all variables	22.69	37.99
Exposure model (GLM) using random forest variables	31.06	42.40
Exposure model (Tobit) using RF variables	33.76	42.43
Exposure model (Tobit) using PCA variables	30.61	43.24

TABLE 5: Summary results of the GWR model for predicting crashes.

Variables	Intercept 2.053	C-export	Residual -0.044	R^2 (local) 0.39	Std. error	Std. resid -0.024
SC		0.248			0.183	
BS		0.228			0.196	
PB		0.118			0.120	
TS		0.178			0.085	
TP0		0.725			0.210	
TM		-0.423			0.248	
TC		0.354			0.155	
RE		-0.423			0.122	
BU		0.572			0.135	
RCE		-0.028			0.123	
AW		0.39			0.190	
TR		0.121			0.088	
TCC		-0.055			0.012	
TP1		0.514			0.145	
TP2		0.325			0.162	
AL		0.254			0.065	
PP		-0.416			0.321	
SA		0.213			0.174	
AS		0.085			0.162	

TABLE 6: Summary results of the GWGR model for predicting crash.

Variables	Estimate	Stand. error	Z (Est/SE)	Mean	Std.	Min	Max	Median	Lower quartile	Upper quartile	Local
Intercept	2.173	0.101	21.462	1.046	1.810	3.093	-2.52	2.009	1.922	2.284	Yes
SC	0.209	0.150	1.394	0.214	0.228	0.531	-0.36	0.285	0.150	0.381	Yes
BB	-0.084	0.170	-0.494	0.553	0.513	1.809	-0.41	0.332	0.128	0.862	Yes
PB	0.012	0.164	0.071	0.227	0.337	0.898	-0.22	0.324	0.181	0.501	Yes
TS	0.202	0.121	1.676	0.173	0.179	0.555	-0.42	0.164	0.107	0.275	Yes
TP0	-0.548	0.519	-1.057	1.081	0.344	3.677	-1.79	-0.007	-0.342	0.817	Yes
TP1	-0.137	0.411	-0.333	1.311	-0.13	4.750	-2.03	-0.518	-0.907	0.385	Yes
TP2	0.154	0.322	0.477	0.742	0.301	1.933	-1.31	0.188	-0.219	0.796	Yes
TM	0.273	0.187	1.465	0.519	0.246	1.123	-2.04	0.330	0.098	0.536	Yes
TC	0.091	0.419	0.218	1.292	-1.32	0.605	-4.83	-1.137	-1.925	-0.244	Yes
RE	-0.429	0.143	-2.999	0.343	-0.56	-0.03	-1.45	-0.545	-0.835	-0.235	Yes
BU	0.229	0.116	1.970	0.520	-0.08	0.923	-1.23	-0.038	-0.509	0.350	Yes
REC	0.556	0.133	4.172	0.253	0.769	1.429	0.139	0.718	0.636	0.886	Yes
AW	0.086	0.121	0.716	0.158	0.005	0.497	-0.39	0.008	-0.125	0.064	Yes
AL	-0.348	0.117	-2.970	2.477	-1.16	-0.00	-11.6	-0.234	-0.576	-0.140	Yes
TR	0.641	0.191	3.347	0.373	0.386	1.473	-0.33	0.324	0.124	0.621	Yes
AS	0.232	0.112	2.068	0.208	0.170	0.715	-0.46	0.201	0.050	0.266	Yes
TCC	0.795	0.175	4.548	0.456	0.813	2.017	0.114	0.665	0.470	1.085	Yes
PP	-0.423	0.125	-3.25	-0.25	0.145	1.025	-0.21	0.361	0.251	0.189	Yes

(mean = 1.8; maximum = 3.9) indicate that the local multicollinearity issue is not critical in this study.

In the study by Lee et al. [5], it is proposed to investigate the local effect of pedestrian crash exposure variables. The

statistical results of the second step models in the current study show that, in the GWPR model, all variables had a local effect. Figure 3 shows the spatial effect of the estimated parameter on crashes. This figure shows only those

TABLE 7: Summary results of the GWPR model for predicting crash.

Variables	Estimate	Stand. error	Z (Est/SE)	Mean	Std.	Min	Max	Median	Lower quartile	Upper quartile	Local
Intercept	-13.165	0.033	-399.863	-14.48	3.208	-31.32	-12.16	-13.15	-14.101	-12.813	Yes
SC	0.104	0.036	2.883	0.111	0.126	-0.160	0.436	0.110	0.008	0.213	Yes
BS	-0.355	0.040	-8.943	-0.028	0.294	-0.824	0.732	-0.073	-0.226	0.139	Yes
PB	0.042	0.038	1.100	0.062	0.168	-0.358	0.433	0.058	-0.068	0.183	Yes
TS	0.218	0.032	6.771	0.077	0.157	-0.401	0.284	0.117	-0.006	0.198	Yes
TP0	-0.312	0.122	-2.559	0.681	1.509	-2.625	5.228	0.402	-0.415	1.370	Yes
TP1	-0.401	0.109	-3.672	-0.467	1.147	-3.593	2.591	-0.618	-1.044	0.201	Yes
TP2	-0.055	0.074	-0.744	0.286	0.416	-0.822	1.068	0.312	-0.012	0.631	Yes
TM	0.254	0.044	5.724	-0.004	0.473	-2.010	1.211	0.026	-0.241	0.290	Yes
TC	0.343	0.105	3.257	-1.014	0.917	-3.034	1.017	-0.992	-1.831	-0.254	Yes
RE	-0.230	0.041	-5.627	-0.224	0.189	-0.709	0.200	-0.261	-0.365	-0.043	Yes
BU	-0.012	0.026	-0.464	-0.220	0.359	-1.242	0.727	-0.101	-0.508	0.049	Yes
REC	0.201	0.026	7.724	0.212	0.189	-0.117	0.764	0.148	0.055	0.328	Yes
AW	0.139	0.034	4.017	-0.019	0.172	-0.631	0.412	0.005	-0.103	0.097	Yes
AL	-0.153	0.039	-3.920	-3.683	7.218	-40.32	0.502	-0.146	-2.609	-0.056	Yes
TR	0.327	0.044	7.352	0.181	0.314	-0.555	1.083	0.157	-0.064	0.367	Yes
AS	0.211	0.028	7.588	0.124	0.323	-0.527	1.141	0.084	-0.078	0.245	Yes
TCC	-0.251	0.037	-6.715	-0.005	0.252	-0.709	0.621	-0.029	-0.149	0.136	Yes
PP	-13.165	0.033	-399.863	-14.48	3.208	-31.32	-12.16	-13.15	-14.101	-12.813	Yes

TABLE 8: Comparison and fitted goodness model.

Models	Global model result		Regression model
	GWGR	GWPR	GWR
Unbiased sigma estimate	2.3961	2.4921	2.1748
-2 log-likelihood	2569.6562	2989.7522	2347.2654
AIC	2587.4516	2767.4426	2507.5604
AICc	2589.0339	2799.0559	2534.7239
Adjusted R square	0.3214	0.44	0.41
MAD	0.851	0.482	0.952
RMSE	1.052	0.591	1.34
Moran's index	0.042	0.031	0.061
P value	<0.001	<0.001	<0.001

coefficients that have a significant effect and the small coefficients are shown in white. It is noteworthy that the estimated coefficients in common fixed models are in the range of similar values in spatial models [1] and this shows that the estimated parameters in global models (i.e., fixed models) are characteristics of the average of the factors affecting the dependent variable.

VIF values do not change between 1.12 and 3.9 (the critical value of VIF between 5 and 10 means a complete spatial correlation between the independent variables). This could be due to excessive scatter in the exposure variables. According to studies [1, 23, 40], the location of VIF in this limit can also justify the geographical distribution of Poisson in crash prediction, but this issue can be explored in the future. Traffic parameters such as intersections and speed cameras have a significant impact on pedestrian crashes and in TAZs where the speed camera density is higher, and fewer crashes will occur due to the variable estimation coefficient (negative).

The model predicts that crashes decrease as population density increases. The sign of population density is negatively associated with crash frequency which is consistent with previous studies [41]. This could be due to the fact that,

in residential areas without commercial and recreational land uses, due to the low speed of vehicles and the presence of speed bumps, as well as distracting effects, the crash density has decreased. In general, the results of the GWPR model show that the presence of a bus station, population density, type of residential land use, average number of lanes, number of traffic control cameras, and sidewalk width have a negative effect on increasing the number of crashes. In the GWR model, the number of motorcycles, residential land use, recreational land use, the average number of lanes, and the number of speed cameras in TAZs had a negative relation with increasing pedestrian crashes. Finally, in the GWGR model, the number of bus stops, population density, residential land use density, average number of lanes, and the number of speed cameras were negatively associated with increasing crashes. It should be noted that, in the three mentioned models, a significant relationship between dependent variable and independent variables has been obtained, which has been confirmed by previous studies (e.g., [13, 40, 42]).

One of the explanations for the negative sign of the bus station in urban areas can be the reduction of the volume of motor vehicles around the residential area, which reduces

FIGURE 3: Crash predictive maps for GWR, GWGR, and GWPR models.

traffic congestion and ultimately exposes the motor traffic of other residents [1, 5]. On the other hand, poor design of a multimode network can negatively affect the safety of nonmotorized users and public transportation. The difference between the signs of the estimated coefficients in several different models requires more details in future studies. We

may expect older people to suffer more severe injuries due to vulnerability [1]. Conversely, older people travel less than other groups [43–45]. As a result, in this study, the percentage of elderly people compared to other age groups was negatively associated with the increase in crash density in the GWPR model. The negative sign of motorcycle users on the

TABLE 9: ANOVA table for GWPR model.

Source	SS	DF	MS	F	P value
Global residuals	3111.824	542			0.001
GWPR improvement	943.826	83.0214	11.286		0.001
GWPR residuals	2168.097	458.2651	4.730	2.3861	0.001

increase of pedestrian crashes can be due to the increase in motorcycle travel; hence, the number of pedestrian trips has decreased. In this study, the variables of number of schools, number of intersections, and pedestrian bridges have been shown with a positive sign, in which the impact of intersections on pedestrian crashes is not significant but has a positive effect on causing crash, which is confirmed by the study by Lee et al. [5]. Due to the presence of parents and children on the school routes, there is more walking activity in TAZs with more schools. Households with less than two vehicles (0 or 1 vehicle) are another important source of pedestrian activity. Car ownership is directly related to household income levels, which reflects the socioeconomic impact on pedestrian activity [5]. It is obvious that family members without transportation meet their transportation needs through public transportation or walking. On the other hand, the amount of car ownership in TAZs also has a significant impact on increasing pedestrian crash exposure. Of course, the crash may occur with vehicles that are outside the TAZ and collide with a pedestrian while traveling in different TAZs. In this study, the TC variable is the vehicle ownership variable in the study area, considering the two factors, and the model results show a positive effect on pedestrian exposure. As mentioned earlier, in this study, the average width of the sidewalk was negatively associated with pedestrian exposure and did not cause crashes, which is consistent with previous studies (e.g., [5, 45, 46]). These findings are also consistent with studies of human factors that show that some groups (e.g., low income, low education, and young urban road users) are more prone to abnormal behaviors [1, 47, 48].

According to the modeling results, the effect of slope and average width of the route on crashes is positive and the reason could be that pedestrians in wider passages have to travel longer to cross the street so they are more exposed in the passage of vehicles. Also, the medium slope has a positive effect on pedestrian crashes compared to zero-slope roads due to more difficult control of vehicles in adverse weather conditions.

5. Conclusion

In this study, a systematic approach has been developed that uses pedestrian surrogate measures based on exposure information. In this study, in the first step, which is the identification of exposure variables, several statistical methods have been used to identify these variables and the frequency of crashes is investigated based on different modeling methods. In the second step, crash prediction models were presented at the zone level using surrogate variables. In this study, three models GWR, GWGR, and GWPR have been used to spatially predict the crash frequency based on exposure variables, and the results of the study showed that the GWPR model makes more accurate predictions. In addition, identifying effective criteria such as

the presence of school, car and motorcycle ownership, bus station, sidewalk width, pedestrian bridges, type of intersection control and the presence of midroad refuge, population density, type of land use, width of roads, average number of routes, average road slopes, and number of speed cameras in dealing with pedestrians is important in this study. In this study, it is emphasized that while providing safety measures for pedestrians, cases such as improving traffic calming should be done in areas with high density of schools as well as schools in the area of intersections and increase the width of sidewalks in areas with more bus stations, because in areas where the bus is the main mode of transportation, there is a tendency to walk and consequently pedestrians are exposed to crashes.

The proposed two-step method in this study involves two consecutive modeling processes. The first model identifies the exposure variables in pedestrian crashes and the second model estimates the number of pedestrian crashes using three spatial models GWR, GWGR, and GWPR. However, this trend is limited because the result can be affected by the errors accumulated in the first stage due to the existence of an uncontrollable confounding variable as well as information biases. It is possible to solve the problem by adopting a simultaneous modeling approach. This study has shown the dispersion and density of pedestrian crashes without possessing the volume of pedestrians and thus by taking safety measures in places prone to pedestrian crashes, social costs, and casualties can be decreased. In this study, Poisson regression was used to evaluate the relationship between sociological variables and crashes at the zone level. Comparison of the performance of GWPR and Poisson models shows a significant spatial heterogeneity in the analysis. The increase in residential density in urban areas has been associated with a decrease in speed and therefore has led to a reduction in crash frequency. On the other hand, increasing travel time and consequently increasing traffic exposure affect the social costs of crashes. Identifying traffic-prone zones can be a useful element in developing policies to support mitigation measures related to pedestrian exposure to traffic. We expect that, in future studies, negative geographic binomial distribution models and the experimental Bayesian geographic model will be used to identify pedestrian exposure variables.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors of this article would like to thank the Deputy of Transportation and Traffic of Tehran Municipality, Legal Medicine Organization, Statistics Center of Iran, and Deputy of Architecture and Urban Planning of Tehran Municipality in order to collect the required data and cooperate in field visits and geometric information of the road network.

References

- [1] A. M. Hezaveh, R. Arvin, and C. R. Cherry, "A geographically weighted regression to estimate the comprehensive cost of traffic crashes at a zonal level," *Accident Analysis & Prevention*, vol. 131, pp. 15–24, 2019.
- [2] D. G. Davis and J. P. Braaksma, "Adjusting for luggage-laden pedestrians in airport terminals," *Transportation Research Part A: General*, vol. 22, no. 5, pp. 375–388, 1988.
- [3] X. Qin and J. N. Ivan, "Estimating pedestrian exposure prediction model in rural areas," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1773, no. 1, pp. 89–96, 2001.
- [4] W. W. Y. Lam, S. Yao, and B. P. Y. Loo, "Pedestrian exposure measures: a time-space framework," *Travel Behaviour and Society*, vol. 1, no. 1, pp. 22–30, 2014.
- [5] J. Lee, M. Abdel-Aty, and I. Shah, "Evaluation of surrogate measures for pedestrian trips at intersections and crash modeling," *Accident Analysis & Prevention*, vol. 130, pp. 91–98, 2019.
- [6] C. Lee and M. Abdel-Aty, "Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida," *Accident Analysis & Prevention*, vol. 37, no. 4, pp. 775–786, 2005.
- [7] M. D. Keall, "Pedestrian exposure to risk of road accident in New Zealand," *Accident Analysis & Prevention*, vol. 27, no. 5, pp. 729–740, 1995.
- [8] L. F. Miranda-Moreno, P. Morency, and A. M. El-Geneidy, "The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1624–1634, 2011.
- [9] S. V. Ukkusuri, S. Hasan, and H. M. A. Aziz, "Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2237, no. 1, pp. 98–106, 2011.
- [10] W. W. Y. Lam, B. P. Y. Loo, and S. Yao, "Towards exposure-based time-space pedestrian crash analysis in facing the challenges of ageing societies in Asia," *Asian Geographer*, vol. 30, no. 2, pp. 105–125, 2013.
- [11] R. Cervero, "Mixed land-uses and commuting: evidence from the American housing survey," *Transportation Research Part A: Policy and Practice*, vol. 30, no. 5, pp. 361–377, 1996.
- [12] D. J. Graham and D. A. Stephens, "Decomposing the impact of deprivation on child pedestrian casualties in England," *Accident Analysis & Prevention*, vol. 40, no. 4, pp. 1351–1364, 2008.
- [13] M. Wier, J. Weintraub, E. H. Humphreys, E. Seto, and R. Bhatia, "An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning," *Accident Analysis & Prevention*, vol. 41, no. 1, pp. 137–145, 2009.
- [14] D. Lord, S. P. Washington, and J. N. Ivan, "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory," *Accident Analysis & Prevention*, vol. 37, no. 1, pp. 35–46, 2005.
- [15] D. Lord, S. Washington, and J. N. Ivan, "Further notes on the application of zero-inflated models in highway safety," *Accident Analysis & Prevention*, vol. 39, no. 1, pp. 53–57, 2007.
- [16] Y. J. Kweon, "Development of crash prediction models with individual vehicular data," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1353–1363, 2011.
- [17] Q. Cai, J. Lee, N. Eluru, and M. Abdel-Aty, "Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models," *Accident Analysis & Prevention*, vol. 93, pp. 14–22, 2016.
- [18] American Association of State Highway Transportation Officials (AASHTO), *Highway Safety Manual*, AASHTO, Washington, DC, USA, 2010.
- [19] J. Liu, A. J. Khattak, and B. Wali, "Do safety performance functions used for predicting crash frequency vary across space? Applying geographically weighted regressions to account for spatial heterogeneity," *Accident Analysis & Prevention*, vol. 109, pp. 132–142, 2017.
- [20] Z. Xie and J. Yan, "Kernel density estimation of traffic accidents in a network space," *Computers, Environment and Urban Systems*, vol. 32, no. 5, pp. 396–406, 2008.
- [21] T. Steenberghen, K. Aerts, and I. Thomas, "Spatial clustering of events on a network," *Journal of Transport Geography*, vol. 18, no. 3, pp. 411–418, 2010.
- [22] A. Hadayeghi, A. S. Shalaby, and B. N. Persaud, "Development of planning level transportation safety tools using geographically weighted poisson regression," *Accident Analysis & Prevention*, vol. 42, no. 2, pp. 676–688, 2010.
- [23] P. Xu and H. Huang, "Modeling crash spatial heterogeneity: random parameter versus geographically weighting," *Accident Analysis & Prevention*, vol. 75, pp. 16–25, 2015.
- [24] R. Amoh-Gyimah, M. Saberi, and M. Sarvi, "The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models," *Analytic Methods in Accident Research*, vol. 13, pp. 28–51, 2017.
- [25] K.-A. Rhee, J.-K. Kim, Y.-I. Lee, and G. F. Ulfarsson, "Spatial regression analysis of traffic crashes in Seoul," *Accident Analysis & Prevention*, vol. 91, pp. 190–199, 2016.
- [26] U. Olsson, *Generalized Linear Models: An Applied Approach*, Lightning Source, La Vergne, TN, USA, 2002.
- [27] S. P. Washington, M. G. Karlaftis, and F. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*, CRC Press, Boca Raton, FL, USA, 2010.
- [28] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [29] A. Okabe, T. Satoh, and K. Sugihara, "A kernel density estimation method for networks, its computational method and a GIS-based tool," *International Journal of Geographical Information Science*, vol. 23, no. 1, pp. 7–32, 2009.
- [30] A. Okabe and K. Sugihara, *Spatial Analysis along Networks: Statistical and Computational Methods*, John Wiley & Sons, Hoboken, NJ, USA, 2012.
- [31] K. Nie, Z. Wang, Q. Du, F. Ren, and Q. Tian, "A network-constrained integrated method for detecting spatial cluster and risk location of traffic crash: a case study from Wuhan, China," *Sustainability*, vol. 7, no. 3, pp. 2662–2677, 2015.
- [32] M. J. T. L. Gomes, F. Cunto, and A. R. Da Silva, "Geographically weighted negative binomial regression applied to zonal level safety performance models," *Accident Analysis & Prevention*, vol. 106, pp. 254–261, 2017.
- [33] D. Chimba, A. Musinguzi, and E. Kidando, "Associating pedestrian crashes with demographic and socioeconomic

- factors,” *Case Studies on Transport Policy*, vol. 6, no. 1, pp. 11–16, 2018.
- [34] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons, Hoboken, NJ, USA, 2003.
 - [35] J. Tu and Z. Xia, “Examining spatially varying relationships between land use and water quality using geographically weighted regression I: model design and evaluation,” *Science of the Total Environment*, vol. 407, no. 1, pp. 358–378, 2008.
 - [36] Q. Wang, J. Ni, and J. Tenhunen, “Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems,” *Global Ecology and Biogeography*, vol. 14, no. 4, pp. 379–393, 2005.
 - [37] A. Ramin, M. Kamrani, A. Khattak, and J. Rios-Torres, *Safety Impacts of Automated Vehicles in Mixed Traffic*, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2018.
 - [38] H. Bozdogan, “Model selection and Akaike’s information criterion (AIC): the general theory and its analytical extensions,” *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.
 - [39] T. Nakaya, A. S. Fotheringham, C. Brunsdon, and M. Charlton, “Geographically weighted Poisson regression for disease association mapping,” *Statistics in Medicine*, vol. 24, no. 17, pp. 2695–2717, 2005.
 - [40] D. Lord and F. Mannering, “The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives,” *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.
 - [41] W. E. Marshall and N. N. Ferenchak, “Assessing equity and urban/rural road safety disparities in the US,” *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, vol. 10, no. 4, pp. 422–441, 2017.
 - [42] N. Dong, H. Huang, J. Lee, M. Gao, and M. Abdel-Aty, “Macroscopic hotspots identification: a Bayesian spatio-temporal interaction approach,” *Accident Analysis & Prevention*, vol. 92, pp. 256–264, 2016.
 - [43] A. F. Williams and O. Carsten, “Driver age and crash involvement,” *American Journal of Public Health*, vol. 79, no. 3, pp. 326–327, 1989.
 - [44] D. L. Massie, K. L. Campbell, and A. F. Williams, “Traffic accident involvement rates by driver age and gender,” *Accident Analysis & Prevention*, vol. 27, no. 1, pp. 73–87, 1995.
 - [45] KRTPO, *2008 East Tennessee Household Travel Survey Final Report*, Knoxville Regional Transportation Planning Organization, Austin, TX, USA, 2008.
 - [46] A. J. Khattak and D. Rodriguez, “Travel behavior in neo-traditional neighborhood developments: a case study in USA,” *Transportation Research Part A: Policy and Practice*, vol. 39, no. 6, pp. 481–500, 2005.
 - [47] J. Davey, D. Wishart, J. Freeman, and B. Watson, “An application of the driver behaviour questionnaire in an Australian organisational fleet setting,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 10, no. 1, pp. 11–21, 2007.
 - [48] T. Nordfjærn, A. M. Hezaveh, and A. R. Mamdoohi, “An analysis of reported driver behaviour in samples of domestic and expatriate Iranians,” *Journal of Risk Research*, vol. 18, no. 5, pp. 566–580, 2015.