

TRAFFIC ANALYSIS ZONE LEVEL ROAD TRAFFIC ACCIDENT PREDICTION MODELS BASED ON LAND USE CHARACTERISTICS

Kamatalam Mahaboob Peera¹, Ravinder Singh Shekhawat², CSRK Prasad³

^{1,2,3} Transportation Division, Department of Civil Engineering, National Institute of Technology, Warangal-506004, Telangana, India

Received 3 June 2019; accepted 7 September 2019

Abstract: Studies have shown that there is a link between different land use characteristics and road traffic accidents. The present study investigates this relationship between road traffic accidents and land use characteristics for Cyberabad area of Hyderabad city. Generalized linear regression techniques are used for developing the relationship between different land use characteristics with the road traffic accidents. Total accidents and KSI (Killed and Severe Injury) accidents per year for each traffic analysis zone (TAZ) are taken as dependent variables whereas employment density and land use proportions for each TAZ are taken as independent variables. The correlation matrix developed between the independent variables showed a high degree of correlation between the population density with the residential area. The results from the generalized linear regression analysis show that the employment density, residential area, and road length are positively correlated with road traffic accidents whereas mixed land use is negatively correlated with road traffic accidents. It also found that mixed land use was also negatively correlated and road length is positively correlated to the KSI.

Keywords: Land use, generalized linear model, traffic analysis zone, spatial analysis.

1. Introduction

According to Association for Safe International Road Travel (ASIRT, 2016), road traffic accidents are very common all over the world and annual global road traffic accident statistics states that nearly 1.35 million people die in road crashes each year, on average 3,287 deaths a day with an additional 20-50 million are injured or disabled. According to the World Health Organization (WHO, 2015), nearly 3400 people die on the world roads every day. Tens of millions of people are injured or disabled every year. In developing countries, the total number of road accidents occur are

more than that of developed nations because of insufficient infrastructure towards road safety. Road safety is a term which describes the safety of road user by implementing sufficient improvement in infrastructure, policies to reduce road accidents.

Over the past decades, various studies have been discussed between road accident studies with road geometric conditions, road user behavior, traffic parameters, and environmental conditions. Few studies have been attempted to establish the relationship between land use behavior and road traffic accidents. (Levine *et al.*, 1995) investigated the relationship between

¹ Corresponding author: kmpeera218@student.nitw.ac.in

land use characteristics and road traffic accidents over the zonal level and they found that residential population density, retail, manufacturing, and services industry

were positively correlated with the number of road traffic accidents. (Wedagama *et al.*, 2006) have discussed this relationship with an approach which is shown in Figure 1.

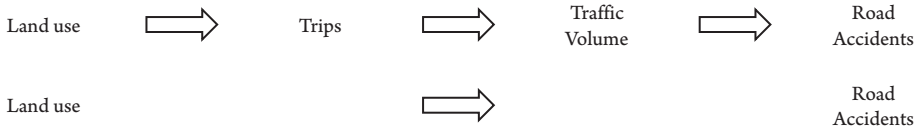


Fig. 1.
An approach between land use with road accidents
 Source: (Wedagama *et al.*, 2006)

While analyzing and developing the road accident prediction model, normally, land use parameters is not considered as direct input while traffic flow is considered as a direct input. However, a hypothesis can be considered as the land-use is the direct source of trip generates and trip attracts, which later leads to traffic, therefore, to model the occurrence of accidents in terms of land-use without consideration of traffic flows. (Pulugurtha *et al.*, 2013) developed log-link negative binomial crash estimation models and concluded that the single-family residential area is negatively correlated with the total numbers of crashes. Yang and Loo (2016) discussed the spatial distribution of road traffic collisions their relationships by using the Empirical Bayes estimation and developed different hazardous road locations patterns over the study area for different road segment classes with different land use characteristics. (Wang *et al.*, 2016) stated that the higher prioritization should assign to TAZs with the medium land use intensity for improving the pedestrian safety. (Kim *et al.*, 2010) showed that the business and commercial areas are strongly correlated with the injury and fatal crashes whereas demographic variables are highly associated with injury crashes, pedestrian and bike

crashes. (Aryaija *et al.*, 2007) investigated the influence of land use characteristics on child pedestrian road accidents. After developing the GLM models, they concluded that secondary retail and high-density residential land use are negatively associated with child pedestrian casualties whereas educational sites were positively associated with child pedestrian casualties. (Kim *et al.*, 2006) have identified the influence of land use, population, employment and economic activity on motor vehicle accidents. They have compared Linear, Poisson, and Negative Binomial model and concluded heavy employment, park, commercial, and school activities were positively and significantly correlated with accidents. (Ivan *et al.*, 2010) modeled for single and multi-vehicle crashes by using traffic density, land use as number of driveways, ambient light condition and time of day. (Wedagama *et al.*, 2007) estimated and forecasted that 30% of adult pedestrian's accident rate will increase with the increment of 1% of retail land use in urban area during weekdays. (Dissanayake *et al.*, 2009) concluded by developing six different GLM for showing a better relationship between child pedestrian casualties and land use and also found secondary retail and high-density residential area as main land use type. Kim

and Yamashita (2002) studied on motor vehicle crashes and land use by using GIS technique to plot crashes at different land use.

In summary, the main purpose of this study is to identify land-use types which influence the road traffic accidents of the study area. As per this approach to road accident investigation, a methodology had to be developed and suitable data sources identified.

2. Methodology

The accident data collected for minimum four consecutive years from police records, the required data from accident review report from FIR (First Information Report) of police record should be tabulated according to the date and day of occurrence, time of accident, location of accident, direction of accused vehicle, cause of the accident, type of accident (fatal, non-fatal and property damage only), number of people injured and dead, accused and victim's type of vehicle, age of accused and victim driver. Coordinate of each accident location was identified by using GPS (Global Positioning System) device and also Google map. These geo co-ordinates of each accident location with details were imported to QGIS software for further spatial analysis. Each accident locations were identified as a point on study area map with its actual position.

QGIS is an open source software for a geographical information system, which used for geo-referencing and digitizing process. Geo-referencing is the process of assigning real-world coordinates to each pixel of the raster while digitizing in GIS is the process of converting geographic data either from a hardcopy or a scanned image into vector data by tracing the features. During the

digitizing process, features from the traced map or image are captured as coordinates in point, line, or polygon format. Geocoding uses a description of a location, most typically a postal address or place name, to find geographic coordinates from spatial reference data such as building polygons, land parcels, street addresses, and postal codes and so on. The coordinate for any location captured from GPS device and also with online map sources e.g. Google map, Google earth, Bing map, Open street map etc. A shapefile is an ESRI (Environmental Systems Research Institute) vector data storage format for storing the location, shape, and attributes of geographic features. It is stored as a set of related files and contains one feature class. Road, location and land use details are stored in that shapefile.

Numbers of accidents per year are selected as the dependent variable. Independent variables are selected based on socio-economic factors of each TAZ area. The study area is divided into a number of traffic analysis zones for identification of variable such as area of each TAZ, road density, population, employment, and land use proportions from each TAZ. Accident prediction model for selected dependent and independent variables are formed. The Generalized Linear models (GLM) i.e. Negative binomial GLM and Poisson GLM are developed using SPSS Statistics Software is used for the formation of the accident prediction model.

3. Data

Study area selected for conducting this study is Cyberabad area, Hyderabad City, Telangana. Accident data for this study is collected from Cyberabad police (traffic) in the form of FIR segregated. Required data

was extracted from this accident data such as time and date, location, vehicle involved, number of people dead and injured, sex, average age, police station and the section of law under which accident recorded. TAZ map and Land Use data are collected from GHMC in the form of georeferenced AutoCAD file or shapefile. This TAZ file having polygon shapefile showing boundary and shape of each TAZ. Land Use data file having the polygonal shapefile showing shape area of the different classes of land use like Residential, Commercial, Public and Semi-Public, Industrial, Institutional, Offices and Recreational. Zonal socio-economic characteristics were found from

Comprehensive Transportation Study (CTS), 2011 documents. HMDA with association with LEA Associates South Asia Pvt., Ltd., New Delhi, India conducted Comprehensive Transportation Study (CTS) for Hyderabad Metropolitan Area (HMA) in 2011.

The collected accident data, during the year 2011 to 2015, of Cyberabad city from 44 police stations were extracted. Figure 2 represents total number of cases recorded as death and injury cases of 44 police stations for the study area. Total 18402 numbers of accident cases were reported during the year 2011-2015 under jurisdiction area of 44 police stations.

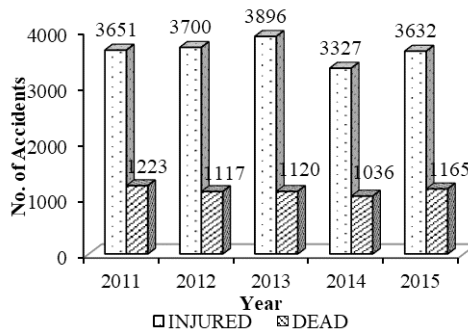


Fig. 2.
Types of Accident Recorded for the Year 2011-2015

3.1. Geo-Referencing

Police station and accident location coordinates are found by using Google map in the accuracy of 5-10 meter. These accident locations were imported to QGIS as delimited text layer and point shapefile generated. This shapefile included various accidents location with an attribute having the total number of accidents occur at that location, date and time of accident occur, location name, location type, the victim and

accused vehicle involved, the age of road user, number of dead and injured person.

Import the TAZ polygon file to QGIS and identify the TAZ location in which accident location exist by following command Processing > Toolbox> count points in a polygon, this will give the number of accident location per TAZ and the TAZ in which accident location exist. Figure 4 shows the TAZ locations in which accident location exist.

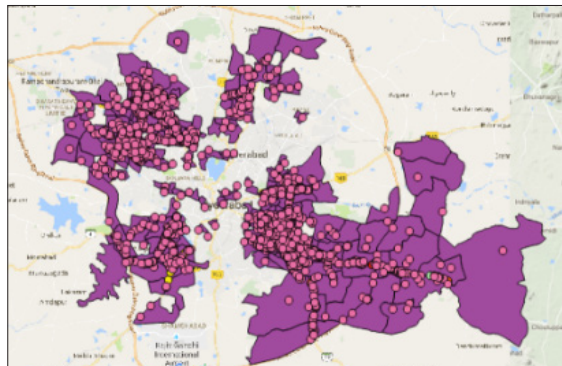


Fig. 3.
TAZ Locations of the Study Area in which Accident Location Exist in QGIS

The study area consists of 215 traffic analysis zones which are shown in Figure 3. TAZ presented as a polygon shapefile in QGIS, which included the name of zone, jurisdiction, district as an attribute. The socio-economic data which contains Population and employment for every TAZ were also added to attribute of TAZ shapefile.

3.2. Land use Data Extraction

Land use, which is polygon shapefile, consists of various attributes of different land use

classes viz., Residential, Commercial, Public Semi-Public, Mixed Land Use, Industrial, Institutional, Recreational and Offices with their shape length and shape area imported to QGIS. This shapefile gets intersected with TAZ shapefile in which accidents occurred using Spatial query tool and Intersection tool in QGIS which result in the formation of new shapefile which includes shape area of different land use classes for each TAZ. Figure 4 shows the land use shapefile required for study area TAZs in which accident occurred. Table 1 shows the land use proportion for each TAZ.

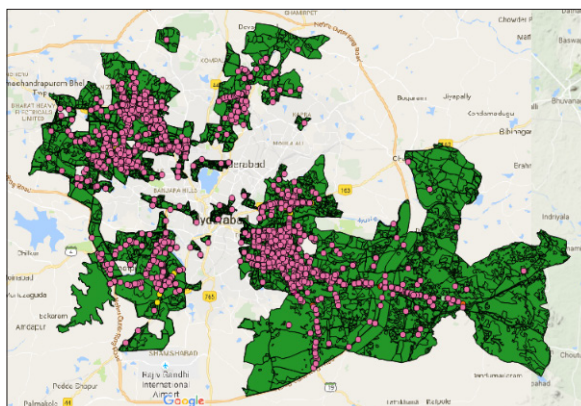


Fig. 4.
Land use Shape File Required for Study Area TAZs

Table 1*Land use Proportion for Each TAZ*

Land use variables	N	Minimum	Maximum	Mean	Std. Deviation
Residential	215	0.00	1.78	.4278	.29757
Commercial	215	0.00	.34	.0277	.05713
Public & Semi-public	215	0.00	.40	.0258	.06330
Industry	215	0.00	.90	.0365	.13388
Mixed_landuse	215	0.00	.90	.0628	.13873
Recreational	215	0.00	.36	.0098	.04084
Institutional	215	0.00	.92	.0397	.12708
Offices	215	0.00	.48	.0054	.04298
ln_Road length	215	.50	2.57	1.5549	.30432

4. Model Development

The validity and success of any mathematical model depend on the variables which used for its development. Variables are selected based on their correlation with other variables and logical sign with respect to dependent variables. The combination of various independent variables is used for the development of the model. These independent variables are 1) Population density (number of people per square km of TAZ area), 2) Employment density (number

of employees per square km of TAZ area), 3) Natural log of Road length for each TAZ area, 4) Land use proportions for each TAZ which are residential, commercial, public semi-public, mixed land-use, industry, institutional, recreational, offices. 'Number of overall accidents per year per TAZ' and 'Number of KSI accidents per year per TAZ' are used as a dependent variables. The relationship between all variables is found by the correlation matrix which is shown in table 2. Priority is given to the less correlation between the dependent variables.

Table 2*Correlation Matrix*

Variables	Population density	Employment density	Residential	Commercial	Public & Semi-public	Industry	Mixed land use	Recreational	Institutional	Offices	ln (Road length)
Population density	1.00										
Employment density	0.16	1.00									
Residential	0.63	0.05	1.00								
Commercial	0.29	0.40	0.11	1.00							
Public & Semi-public	0.02	0.00	-0.07	0.14	1.00						
Industry	-0.08	0.02	-0.28	0.00	-0.07	1.00					
Mixed land use	-0.17	0.04	-0.36	-0.05	-0.08	-0.09	1.00				
Recreational	-0.02	0.01	-0.10	0.03	-0.03	0.06	0.11	1.00			
Institutional	-0.10	-0.06	-0.19	-0.10	-0.03	-0.06	-0.08	0.06	1.00		
Offices	0.00	0.00	-0.03	0.01	-0.04	0.00	0.26	0.06	-0.02	1.00	
ln (Road length)	-0.44	-0.28	-0.23	-0.26	-0.02	-0.12	0.17	-0.12	-0.12	0.04	1.00

From the correlation matrix, the population density is highly correlated with the Residential area proportion, thus population density excluded from the independent variables used for the development of the models. Two different models are developed considering ALL Type of accidents and KSI type of accidents using generalized linear models in SPSS Statistics Software for selected dependent and independent variables. As per literature (Example: Miaou, 1994; Chengye 2013), it has been found that standard least square regression would not effectively fit to model the crash data as compared to GLM and it has been seen that Poisson with log-link and Negative binomial with log-link is the best way to model accident estimation. Poisson distribution is used to model count data if the mean and variance of the data are equal and if not so then the data is said to be over-dispersed or under-dispersed. If the variance of data is greater than mean then the data is said to be over-dispersed and if the mean of data is greater than variance then the data is said to be under-dispersed which is a very rare case. The negative binomial distribution model is used for modeling overdispersed data because they can take into account the effect of unnoticed heterogeneity due to the variable which is omitted or not considered among TAZs. The models are developed by

using Generalized Linear technique can be written as:

$$\ln(\mu) = \beta_0 + \sum_{i=1}^n \beta_i X_i \tag{1}$$

$$\ln(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n \tag{2}$$

Where μ is the dependent variable and X_1, X_2, X_3 and X_n are the explanatory variables with their respective coefficients $\beta_1, \beta_2, \beta_3$ and β_n . β_0 is considered as constant.

The first model was developed considering the Total number of accidents i.e. 7147 accidents for 215 TAZs using Negative binomial GLM and Second model was developed considering KSI type of accidents i.e. 2071 accidents for 169 TAZs using Poisson GLM. Independent variables with Wald Chi-square value more than 1 or level of significance less than 0.05 (95% confidence level) were considered to have statistically significant effect on accidents.

The statistical parameters with their respective coefficient, standard error, Wald Chi-square value, and significance value in the model to estimate the total number of accidents per year per TAZ are shown in Table 3. Similarly, to estimate, the KSI type of accidents per year per TAZ are shown in Table 4.

Table 3

Model to Estimate the Total Number of Accidents per Year per TAZ

Variables	Coefficients	Std. Error	Wald Chi-Square	Significance value (p-value)
(Intercept)	-1.209	.5893	4.206	.040
Employment density	.016	.0060	6.767	.009
Residential	.876	.3661	5.720	.017
Commercial	-.488	1.6941	.083	.773
Public Semi-public	.469	1.1173	.176	.675
Industry	1.005	.6979	2.074	.150
Mixed-land use	-2.296	.8510	7.283	.007
Recreational	2.272	1.8133	1.570	.210
Institutional	-1.151	.8425	1.866	.172
Offices	3.617	2.2062	2.688	.101
Ln(Road length)	1.701	.3077	30.571	.000
Summary Statistics				
Number of TAZ	215			
Number of accidents	7147			
Statistical Test	Deviance/Deg. of freedom	Negative binomial Model Fitted LR Test for NB		
	1.842	X ² = 46.927	P value = .000	Significant at p=.05

ALL Model can be written as:

$$Ln(ALL) = -1.209 + .016 * Employment Density + .876 * Residential proportion - 2.296 * Mixed land use proportion + 1.701 * Ln(Road length) \tag{3}$$

Table 4

Model to Estimate the KSI Accidents per Year per TAZ

Variables	Coefficients	Std. Error	Wald Chi-Square	Significance value (p-value)
(Intercept)	-.972	.3217	9.121	.003
Employment density	-.001	.0046	.077	.781
Residential	.196	.2087	.879	.348
Commercial	-.836	1.1637	.517	.472
Public Semi-public	1.206	.7562	2.544	.111
Industry	.426	.4068	1.095	.295
Mixed-land use	-1.689	.5434	9.654	.002
Recreational	1.269	1.1978	1.122	.290
Institutional	-1.216	.6339	3.679	.055
Offices	2.074	1.1564	3.216	.073
Ln(Road length)	1.150	.1629	49.808	.000
Summary Statistics				
Number of TAZ	169			
Number of accidents	2071			
Statistical Test	Deviance/Deg. of freedom	Negative binomial Model Fitted LR Test for NB		
	3.292	X ² = 74.757	P value = .000	Significant at p=.05

KSI Model can be written as:

$$Ln(KSI) = -.972 - 1.689 * Mixed land use proportion + 1.150 * Ln(Road length) \tag{4}$$

5. Results and Discussions

From table 3, it has been seen that Negative binomial (NB) GLM is the best-fitted model for predicting Total Accidents per year per TAZ with LR (Likelihoods ratio) test for NB having p-value = 0.000 which is less than p-value = 0.05 at 95% confidence level which shows the model is statistically significant. From model, it has seen that Employment density (in thousands per square kilometer), Residential area proportion, Mixed Land-use proportion and Road length (in kilometer) are the statistically significant at 95% confidence level with p-value less than 0.05. From the coefficients of the significant variable it has seen that the Total accidents per year per TAZ are positively associated with the Employment density (in thousands per square kilometer), Residential area proportion and Road length

whereas negatively associated with the Mixed Land-use proportion.

From table 4, it has been seen that Poisson GLM is the best-fitted model for predicting KSI type of accidents per year per TAZ with LR (Likelihoods ratio) test having p-value = 0.000 which is less than p-value = 0.05 at 95% confidence level which shows the model is statistically significant. From model, it has seen that Mixed Land-use proportion and Road length (in kilometer) are the statistically significant at 95% confidence level with p-value less than 0.05. From the coefficients of the significant variable, it has seen that KSI type of accidents per year per TAZ is positively associated with Road length whereas negatively associated with the Mixed Land-use proportion. Summary of the Model Results is shown in table 5.

Table 5
Summary of the Model Results

Model	Significant Variable
ALL	(+) Employment density, (+) Residential, (-) Mixed Land-use, (+) Road length
KSI	(-) Mixed Land-use, (+) Road length

Employment density of a TAZ means the total number of employed in that particular TAZ area. From model, it has seen that total accidents per year per TAZ are positively associated with the Employment density. If the number of employees increases in TAZ than it leads to increase number of work trips which results in more traffic and more accidents. As per an article of Indian express, about 3.5 lakh employees work in Cyberabad area and almost 80% of them use their personal vehicles, mostly four-wheelers. So, accidents will increase with increase in employment density makes sense.

The residential area is mainly responsible for the Trip generation because most of the trips

are initiated from home like home to work, home to shopping etc. and more residential area leads to increase in more population density so trip generation will be more and more trips leads to increase in more traffic and more accidents. So, if residential area increases accidents will also increase until unless proper traffic calming measure were not taken.

Mixed land-use is the combination of two or more land-use like residential-commercial etc. Both models (ALL and KSI) predicted that mixed land-use is negatively associated with the accidents. Mixed land-use leads to reduction in trip length and promote more

walk or non-motorized trips which result in less vehicular traffic and fewer accidents, so if mixed land-use with proper traffic calming measures increases accidents will decrease.

Road length is a factor that is directly related to road accidents so if there is no road in an area then there would be no road accidents. As the road length increases, exposure to the traffic also increases which leads to increase in the chance of more road accidents and both model (ALL and KSI) predicted that the road length is positively associated with the accidents.

6. Summary and Conclusions

This study aims at the investigational study of accident patterns in the study area of Cyberabad. The main cause of transportation problems is the rapid growth of a city and road safety is the main concern of rapid growth. There has been a curious raising rate of road crashes and casualties from year to year. The present study also intends to determine the effective independent variables for the cause of accidents. Observation of data have been made to select the variables and detailed analysis of accident data for 2011 to 2015 year was carried out. Variables value have been found using QGIS for each TAZ using spatial analysis.

ALL accidents model used employment density, residential area proportion, mixed land-use proportion and road length as independent variables for predicting accidents per year from each zone. While the KSI accidents model used mixed land-use proportion and road length as independent variables for predicting KSI accidents per year from each zone. Total accidents per year per TAZ are positively associated with the Employment density (in

thousands per square kilometer), Residential area proportion and Road length whereas negatively associated with the Mixed Land-use proportion. KSI type of accidents per year per TAZ is positively associated with Road length whereas negatively associated with the Mixed Land-use proportion. Based on outcomes from this study maintaining a balance between different land uses in a TAZ would improve road user safety and introducing traffic calming measures and pedestrian crossing would minimize the accidents in the future.

Acknowledgment

Authors express their thanks to the Hyderabad Metropolitan Development Authority officials and LEA Associates, Hyderabad for providing the necessary data. Authors express also their thanks ACP Traffic, Hyderabad Traffic Police for providing the accident data.

References

- Aryaija, J.; Dissanayake, D.; Bird, R. 2007. Land use factors and the risk for child pedestrians, *Journal of the Eastern Asia Society for Transportation Studies* 7: 2814-2829.
- ASIRT. 2016. *Road Safety Facts*. The Association for Safe International Road Travel. Available from Internet: <<https://www.asirt.org/safe-travel/road-safety-facts/>>. Accessed on 18th December 2018.
- Chengye, P.; Ranjitkar, P. 2013. Modelling motorway accidents using negative binomial regression, *Journal of the Eastern Asia Society for Transportation Studies* 10: 1946-1963.
- Dissanayake, D.; Aryaija, J.; Wedagama, D. P. 2009. Modelling the effects of land use and temporal factors on child pedestrian casualties, *Accident Analysis & Prevention* 41(S): 1016-1024.

- Ivan, J. N.; Wang, C.; Bernardo, N. R. 2000. Explaining two-lane highway crash rates using land use and hourly exposure, *Accident Analysis & Prevention* 32(6): 787-795.
- Kim, K.; Brunner, I. M.; Yamashita, E. Y. 2006. Influence of land use, population, employment, and economic activity on accidents, *Transportation research record* 1953(1): 56-64.
- Kim, K.; Pant, P.; Yamashita, E. 2010. Accidents and accessibility: Measuring influences of demographic and land use variables in Honolulu, Hawaii, *Transportation Research Record* 2147(1): 9-17.
- Kim, K.; Yamashita, E. 2002. Motor vehicle crashes and land use: empirical analysis from Hawaii, *Transportation Research Record: Journal of the Transportation Research Board* (1784): 73-79.
- Levine, N.; Kim, K. E.; Nitz, L. H. 1995. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators, *Accident Analysis & Prevention* 27(5): 675-685.
- Miaou, S. P. 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, *Accident Analysis & Prevention* 26(4): 471-482.
- Pulugurtha, S. S.; Duddu, V. R.; Kotagiri, Y. 2013. Traffic analysis zone level crash estimation models based on land use characteristics, *Accident Analysis & Prevention* 50: 678-687.
- Wang, X.; Yang, J.; Lee, C.; Ji, Z.; You, S. 2016. Macro-level safety analysis of pedestrian crashes in Shanghai, China, *Accident Analysis & Prevention* 96: 12-21.
- Wedagama, D. M. P.; Bird, R.; Dissanayake, D. 2007. The influence of urban land use on pedestrians accidents during congested and uncongested periods, In *Proceedings of the Eastern Asia Society for Transportation Studies The 7th International Conference of Eastern Asia Society for Transportation Studies*, 364-364.
- Wedagama, D. P.; Bird, R. N.; Metcalfe, A. V. 2006. The influence of urban land-use on non-motorised transport casualties, *Accident Analysis & Prevention* 38(6): 1049-1057.
- WHO. 2015. *Global status report on road safety 2015: supporting a decade of action*. World Health Organization. Violence, Injury Prevention, & World Health Organization.
- Yang, B. Z.; Loo, B. P. 2016. Land use and traffic collisions: A link-attribute analysis using Empirical Bayes method, *Accident Analysis & Prevention* 95: 236-249.