ICIC Express Letters
Part B: Applications
Volume **11**, Number **10**, October **2020**

ICIC International ©2020 ISSN 2185-2766
pp. **961–967**

ANALYSIS OF CHILDREN TRAFFIC ACCIDENTS AROUND SCHOOL AREA USING GIS AND DATA MINING TECHNIQUES

Hee Seon Han and Gun Ho Lee

Department of Industrial and Information Systems Engineering
Soongsil University
369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea
ghlee@ssu.ac.kr

Received March 2020; accepted June 2020

Abstract. *The purpose of this study is to discover patterns and knowledge of traffic accidents in children to prevent traffic accidents. We analyze traffic accident data to reduce children traffic accidents. The past accident history data is visualized, and data mining is performed to find factors of frequent traffic accidents and the relationship between the factors. We identify frequent patterns of accidents, according to factors such as the danger zone of children's accidents, dangerous time periods, types of dangerous vehicles, climate, and roads, and analyze the relationship of these factors. The patterns of the accidents are visualized so that the safety manager can easily recognize and use them for preventive management. We develop the rules that make a warning signal if there is a high probability of a traffic accident depending on various situations. The necessary information is also visualized on the map based on Geographical Information System (GIS) by region, road, and time for preventive safety management.*
Keywords: Car accident, Accident analysis, Geographical information system, Data mining

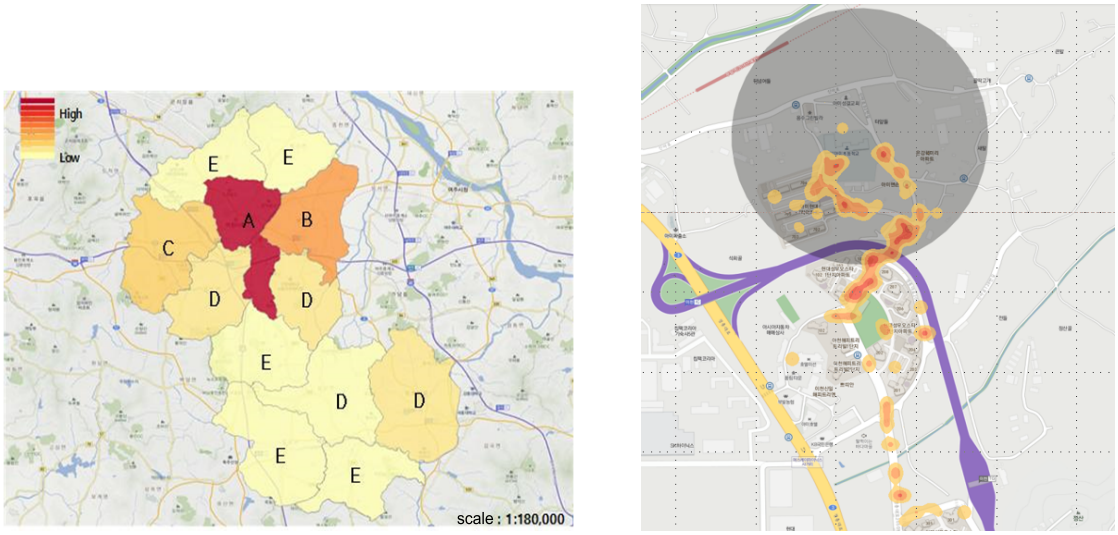
1. **Introduction.** Traffic accidents are caused by various factors depending on the state of the road, traffic lights, the existence of the crosswalk, observation of the traffic laws of the driver, walking habits of the pedestrian, and compliance with the traffic rules of pedestrians. Safety precautionary activities may reduce or prevent accidents. Because the patterns of traffic accidents vary from region to region, it is necessary to establish management plans according to regional characteristics. If a certain type of accident frequently occurs in a particular area, it can be a cause of repeated accidents such as the driver's fault, the structure of the road, and the state of the walkway. In this paper, we perform as follows. Traffic accident data are analyzed by various factors such as gender of an injured party, region, weather, vehicle type, observation of traffic rules, time, and season. We analyze the relationship between the factors and interactions with each other to find out the factors that highly affect traffic accidents. The purpose of this study is to discover patterns and knowledge of traffic accidents in children to prevent the traffic accidents. We present the patterns of traffic accidents of children and the guidelines for carrying out accident prevention activities according to the situations of weather, seasons, region, etc. In order to understand the traffic accident area, which is not known only from accident data, we analyze the data collected from the past traffic accident history and the questionnaire survey to derive a traffic accident area with the high potential risk of accidents. We present the hot spot of children's traffic accidents on the heat map for safety managers to assess the risk of traffic accidents on the walkway in the children's school zone and also to indicate areas where traffic accidents are likely to occur in the future.

DOI: 10.24507/icicelb.11.10.961

2. **Related Study.** Data mining techniques such as fuzzy K-Nearest Neighbor (K-NN), ID3, and function tree, a priori rule mining, Naïve Bayes, and k-means clustering are used to analyze road traffic accidents, to predict traffic incident duration, and to identify factors which influence the injury severity [1-5]. Classification technique and a priori algorithm are used to investigate the influence of other factors, the road condition, weather condition, road feature, and road intersection type and controls strongly affect the nature of accidents that lead to fatal and series injury severity [6]. The algorithms such as Self Organizing Map (SOM), K-modes, Support Vector Machine (SVM), Naïve Bayes, and decision tree are used to achieve better accuracy by using the clustering technique [7]. The result shows that accidents between the driver, passenger, or pedestrian are more involved. Contemporary Geographical Information System (GIS) technology is a popular tool for visualization of accident data and analysis of hot spots in the road. Many traffic agencies use GIS for accident analysis. Hirasawa and Asano [8] developed a traffic accident analysis system that uses a GIS. The system links great volumes of traffic accident data accumulated for more than ten years with digital map data indicating accident locations. The user extracts accident data by road section and by specific condition and analyzes accident frequency, accident rate, and fatality rate. Erdogan et al. [9] developed a system, transforming these textual data in a tabular form, and then this tabular data was georeferenced onto the highways. Then, the hot spots in the highways, in Afyonkarahisar administrative border were explored and determined with two different methods of kernel density analysis and repeatability analysis. Nanzala et al. [10] improved road traffic data storage and accessibility through the use of WebGIS by creating a road traffic accident data management system which was developed to encompass a database managed by PostgreSQL and a web server. Famili et al. [11] studied in GIS spatial analysis using multiple geostatistical methods to identify pedestrian crash hot spots and to identify clusters of pedestrians using kernel density estimation and ordinary kriging analysis and make a comparison between these two methods.

3. **Data Acquisition.** We get car accident data, safety facility data and the road data such as road width, the surface of the road, the speed limit zone through the open Application Programming Interface (API) of the database of Traffic Accident Information Distribution System (TAIDS) in South Korea and elementary school location data from the local education office, population data from National Statistical Office. After acquiring car accident data, we encode and process javascript object notation data and change it into objects of the R language. The acquisition period of data is from October 2015 to October 2017, and the total number of accidents is 14,624. The types of car accidents are classified as 'car-car' accident, 'car-people' accident, and 'car itself' accident. Data attributes include the occurrence time of the accident, the area where it occurred, the types of accident, violation of traffic regulations, road, accident vehicles, and longitude and latitude for the location of accident, pedestrian sex, age, types of vehicle behavior, driver behavior, pedestrian behavior, safety accident, area, day of the week, and means of travel, etc.

We visualize the pattern of car accident data obtained from TAIDS DB on the map of the web-based survey system. The accidents that occurred in the same place are accumulated for analysis and statistics. We cluster car accident spots on the map and create a 1st heat-up map to select the survey area, as shown in Figure 1(a). After visualizing the pattern of car accident areas and selecting the areas for analysis, we collect further data about car accidents around the selected areas to analyze the potential risk and to find the spots where traffic accidents are likely to occur in the future.



(a) The priority for preventive management (b) Hot spot of past car accidents

FIGURE 1. (color online) Visualization on the map

We collect 1,866 cases of direct or indirect experiences of car accidents in 56 elementary school students in Y city in accordance with the children's behavioral characteristics and accident risk perception area through a survey. We obtain further information on the environment of the accident area, route, and behavior patterns of pedestrians, etc.

The respondent displays the locations and routes where an accident is likely to occur on the map of the web-based survey system. They mark the spot and the route of a pedestrian that the respondents feel dangerous on the map with the reasons why he/she selects it as a dangerous area.

The attributes of the data are filtered, integrated, or reprocessed. For example, the date of birth and age are integrated since the two characters are similar in one field, and the unit of length measurement of the road is converted into the same unit. Address and location data are converted into coordinate information by geocoding, and the different coordinate system is converted into the World Geodetic System (WGS84). The data are represented by different representations, but in fact the same concept for each field is converted into an identical expression.

4. **Analysis of Traffic Accidents.** Car accident factors are classified as follows. Factors of an accident caused by the illegal act of drivers include signal violation, illegal U-turn, right (left) turn at his/her own risk, speed violation, illegal parking on the road, illegal driving, etc. Road factors of a car accident include crosswalks, intersection, alleyway, downhill, tunnel, bridge, etc. Infrastructure factors of a road include signaling systems, pedestrian crossing points, traffic lights (sign) post, speed limit zone panel, and road widths. Behavioral characteristics of children as pedestrians include inadequate awareness of risk for running himself/herself or running cars, lack awareness of speed of running the car, lack of recognition of traffic accidents with automobiles, etc. We analyze the location, the situation, the location of the area that is perceived to be dangerous, and the risk information.

In this study, QGIS (Quantum GIS) is used for data refinement, coordinate data conversion (latitude/longitude), and spatial unit mapping. QGIS is a free and open-source, cross-platform, desktop GIS application software that provides data viewing, editing, and analysis. We also use PostgreSQL for scalability and standardization for data storage and spatial operations, refined data relations, data fusion, and filtering. We use the R language to analyze and correlate the effects of traffic accidents and to visualize them. In order to visualize the participatory data on the map, the coordinate information of the

collected points is plotted on the map in point and line form. Spatial data are mapped to 20m × 20m grid data and processed in grid units. After aggregating in grid units, we visualize colors or contrasts to match the frequency of each grid. It is generated as point coordinates at the lattice center by the number of points aggregated per lattice and visualizes the heat map. The greater number of points in the grid, the darker the red color appears in Figure 1(b).

Existing studies have produced hot spot, which is a frequent accident spot among traffic accident points and presented the results of the analysis. Although the characteristics and circumstances of roads in each region are not the same, there is a limit to expanding in the regional area of this study. However, a general pattern in which traffic accidents occur frequently can be applied to accident risk areas to prevent car accidents. A common pattern in which traffic accidents occur frequently can be applied to detecting similar hazardous areas and areas of potential accidents.

4.1. **Frequent pattern mining by Exploratory Data Analysis (EDA).** This study performs exploratory data analysis to understand the general patterns of children in traffic accidents. In the monthly trend, the frequency of pedestrian traffic accidents is high in the 5th and 6th grades of elementary students during the beginning seasons of school, i.e., in March, September, and May, doing a lot of outdoor activities. The group of female students frequently experiences a car accident against pedestrians during the beginning school season in March and September. It is shown that the lower grade of the female group is more influenced by the elementary school curriculum schedule relatively. Infants and toddlers are unaffected by the onset of school because car accident against pedestrian occurs evenly between April and September. The survey shows that car accidents against pedestrians mainly occur on a single path, crossroads, or near crossroads. As a factor of the accident against pedestrians, according to the driver's behavior, there are vigilant, neglect to keep their eyes on the road, radio manipulation, delay of detection due to other vehicles, delay of detection due to abnormal climate, drunk driving hurdle, misjudgment of the surrounding situation, etc. Most accidents occur on a clear day, and the occurrence of accidents is rather low in foggy, snowy, or cloudy days. In order to prevent car accidents, pedestrians and drivers need to be aware of the distribution of accident types caused by car driving. Straight-line driving is capable of speeding, while left-right turning, U-turn, driving, or departing are available at low speed. 65.35% of car accidents against children occurs when the car moves straight forward. The percentage of traffic accident casualties of pedestrians in our city is relatively low compared to that of the total population and is lower than that of adults. Car accidents against children pedestrians occur at the time going to school and coming home from school after 2 pm since they are active during those times. Car accidents occur more during the time coming home from school than during the time going to school. One of the reasons for the small number of accidents during the time going to school is that the guardian, such as a parent or grandparent, is accompanying children.

In the case of adults, car accidents are more intense during the morning rush hour than those during the evening rush hour. The accidents against adults occur mainly from 6 pm after business hours until 11 pm in the night time.

The accidents against children occur on a single path, crossroads, and near crossroads rather than on pedestrian crossing during the time coming home from school. Car accidents against pedestrians in car-people accidents are mainly caused by passenger cars. The number of children accidents caused by passenger cars is less than that of an adult's accidents. The ratio of accidents against children caused by passenger cars is 3.13 times the ratio of accidents by other vehicle types, which is 1.96 times higher than the ratio of accidents of adults caused by passenger cars. The car accidents caused by passenger cars are frequently caused by the lack of compliance with pedestrian traffic regulations. The

Table 1. Support (%) by traffic accident characteristics

Main category	Accident characteristics (subcategory)								
	Support (%)								
Vehicle	Sedan	Van	Truck	Agricultural machinery	Construction	Special car	Cart	Motor-Cycle	Bicycle
	67.80	4.28	14.25	0.10	0.68	0.31	4.70	7.57	0.31
Location	Within intersection	Near interaction		Single path	Crosswalk vicinity	On bridge		Unknown	
	20.79	22.43		50.12	0.18	0.12		6.37	
Weather	Snow	Sunny			Rain	Fog		Cloudy	
	1.2	82.13			11.54	0.7		4.4	
Driver violation	At intersection	Pedestrian priority		Signal violation	Centerline invasion	Failure to comply with safe driving		Illegal turn	etc.
	0.37	12.98		5.74	1.70	76.28		0.64	2.29
Movement of car	U-turn	Go straight		Forward movement	Backward movement	Stopping, parking		Left, right turn	etc.
	0.59	65.35		1.49	8.53	0.85		18.71	4.48

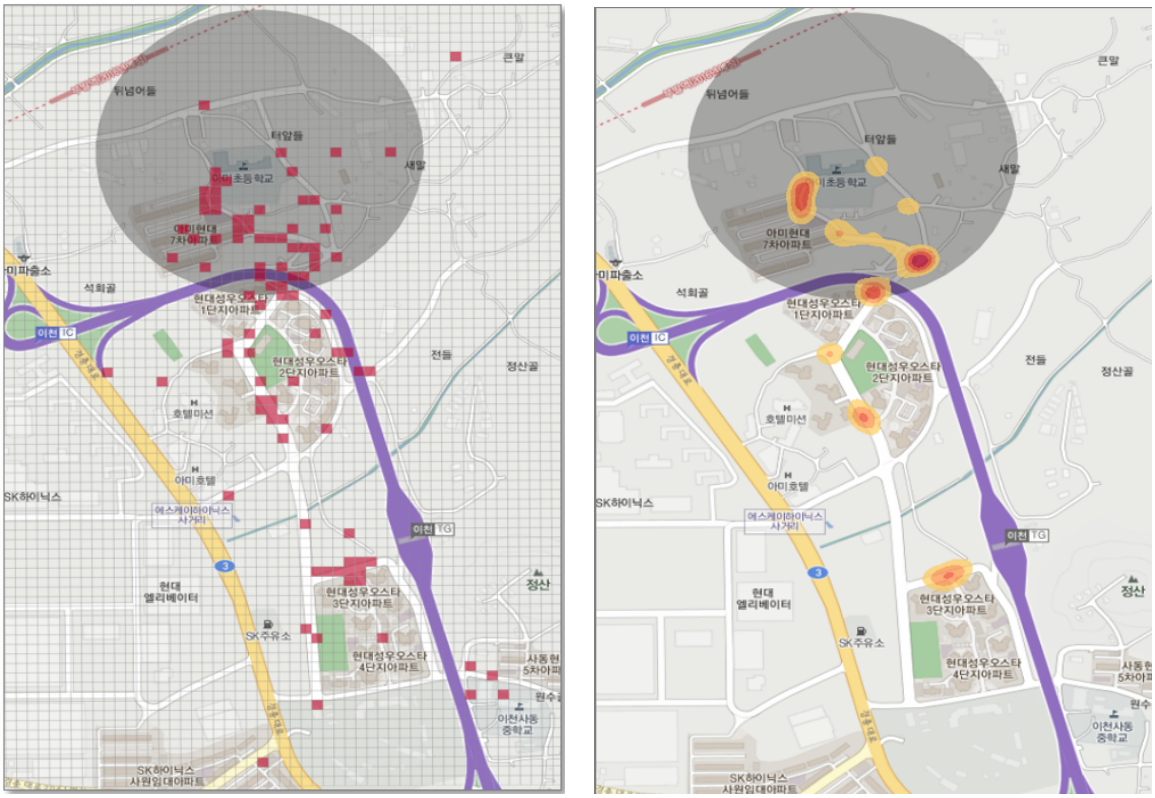
car accidents caused by children's violations of traffic regulations are higher than those by adult's violations of traffic regulations. In infants and young children, the accidents of male infants and toddlers are higher than those of female's but female children in the upper grades of elementary school are vulnerable to car accidents. The car accident frequency in the group of 1 to 3 years old is less than half of the frequency in the children's group of kindergarten and elementary school. In-car accidents against children under 12 years old, there is a pattern of rapid increase from 4 years old or older. The frequency of car accidents of a male in the adult is higher than that of a female. Car accidents of the lower grades of elementary students occur mainly between 2:00 pm and 3:00 pm and those of the upper grades frequently occur between 4:00 pm and 6:00 pm. The survey shows that 48.39% of children traffic accidents involve in a road crossing. On the other hand, 41.33% of adult car accidents occur on a road crossing. The car accident against children frequently occurs while crossing the roadway. It is necessary to take measures for the installation of facilities such as fences in order to prevent accidents caused by illegally crossing.

4.2. **Rules for traffic accident prevention management.** The rules in Table 2 are the combination of the factors of the accident, and the accuracy of the prediction of the rules is high. The rules can be used as a guideline for the preventive management of car accidents in the area by showing on maps.

Table 2. Example of rules developed

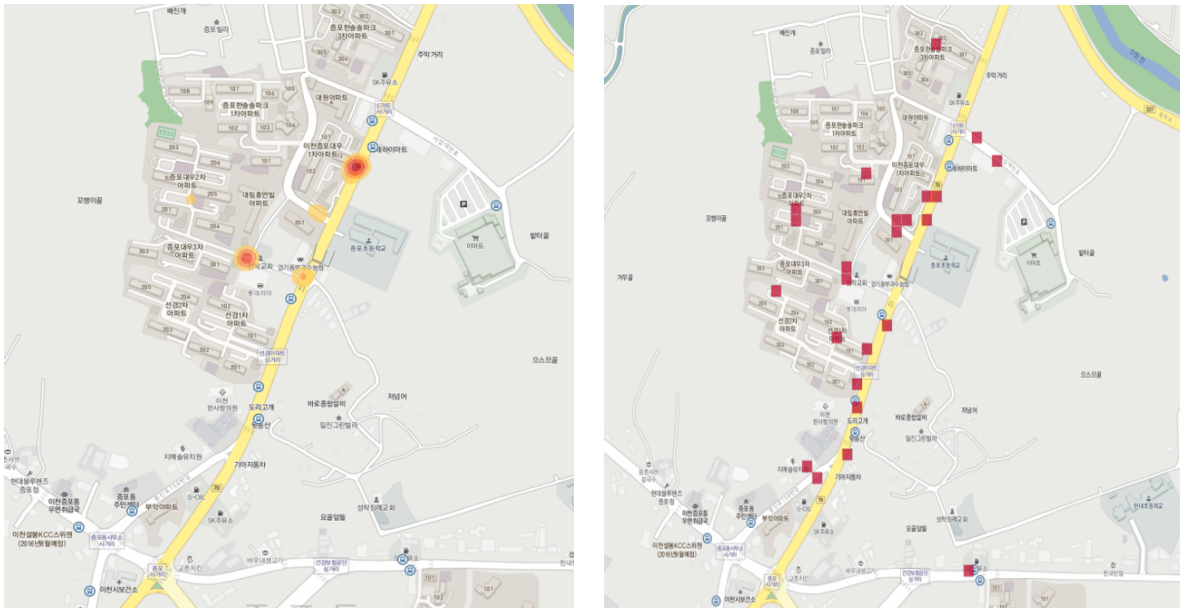
IF infants and elementary school students come near the crossroad in area A on a sunny day, send alert signals.
IF infants and elementary school students come near a crosswalk in area B on a sunny day, send alert signals.
IF elementary school students walk on the single path in area C on a sunny day, send alert signals.
IF a car goes to the crosswalk in the D area during the time from 17:00 to 17:59 hours, send a serious warning signal to the driver and the elementary school students who are coming to the crosswalk or near the crosswalk.

4.3. **Visualization of priority for preventive management of accidents using GIS.** We define the criteria to classify the areas that the children come and go as high-accident zone, quasi-accident zone, safe zone with a decision tree. Using the GIS, the map



(a) The crosswalk required (b) Non-crosswalk

FIGURE 2. (color online) Hot spot of areas with frequent illegal crossings



(a) Hot spots on non-crosswalk (b) Hot spots on crosswalk

FIGURE 3. (color online) Hot spots of frequent accident prone area

shows the priority for preventive management of the children's activity area according to color. Figure 2 shows the accident spots of the crosswalk required and non-crosswalk in the past. The darker red color indicates the point at which the accident occurred frequently. In Figure 3, we predict areas where accidents are likely to happen.

5. **Conclusion.** This study analyzed children traffic accidents using data mining and GIS techniques and consequently found the patterns of children traffic accidents. We have found frequent patterns of the times and spots where children traffic accidents occur frequently. We developed rules for children's traffic safety management through analysis of actual accident data and the data obtained by the survey. These rules can be used to prevent traffic accidents in children. When identifying vulnerable points to the car accident and establishing safety measures for children safety, it can be used as basic data for customized traffic risk points in the area. Based on GIS, we identify potential accident spots by the elementary school on maps through mashup analysis. It provides important information on traffic accidents so that it can develop effective traffic safety management plans tailored to the region while taking consideration of the centralized management areas that are not included within the 300m radius of the school's major route.

Acknowledgment. This work is supported by Soongsil University. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers.

REFERENCES

[1] T. Sayed, W. Abdelwahab and F. Navin, Identifying accident-prone locations using fuzzy pattern recognition, *J. of Transportation Engineering*, vol.121, no.4, pp.352-358, 1995.

[2] L. Guan, W. Liu, X. Yin and L. Zhang, Traffic incident duration prediction based on artificial neural network, *Proc. of 2010 International Conference on Intelligent Computation Technology and Automation (ICICTA)*, vol.3, pp.1076-1079, 2010.

[3] Kashani, A. Shariat-Mohaymany and A. Ranjbari, A data mining approach to identify key factors of traffic injury severity, *PROMET-Traffic & Transportation*, vol.23, no.1, pp.11-17, 2011.

[4] A. Komolafe and A. Olutayo, Using data mining technique to predict cause of accident and accident prone locations on highways, *American J. of Database Theory and Application*, vol.1, no.3, pp.26-38, 2012.

[5] L. Li, S. Shrestha and G. Hu, Analysis of road traffic fatal accidents using data mining techniques, *Proc. of 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, London, UK, pp.363-370, 2017.

[6] B. Atnafu and G. Kaur, Survey paper on analyze and predict the nature of road traffic accidents using data mining techniques in Maharashtra, India, *International J. of Engineering Trends and Technology*, vol.53, no.1, pp.23-31, 2017.

[7] G. Velivela, K. P. Purna, Y. Challa, G. V. Krishna and S. Krishna, Traffic accidents analysis with respect to road user using data mining techniques, *International Journal of Emerging Trends & Technology in Computer Science*, vol.6, no.3, pp.15-20, 2017.

[8] M. Hirasawa and M. Asano, Development of traffic accident analysis system using GIS, *Proc. of the Eastern Asia Society for Transportation Studies*, vol.4, pp.1193-1199, 2003.

[9] S. Erdogan, I. Yilmaz, T. Baybura and M. Gullu, Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar, *Accident Analysis and Prevention*, vol.40, pp.174-181, 2008.

[10] J. Nanzala, K. David and M. Kenneth, Development of a WebGIS supported road traffic accident data management system: A case study of Nairobi county, *Proc. of Sustainable Research and Innovation (SRI) Conference*, pp.218-227, 2015.

[11] A. Famili, W. A. Sarasua, J. H. Ogle, A. Shams and N. Asadi, GIS based spatial analysis of pedestrian crashes: A case study of South Carolina, *Proc. of International Conference on Transportation and Development 2018: Connected and Autonomous Vehicles and Transportation Safety*, Pittsburg, PA, pp.368-376, 2018.