Ain Shams Engineering Journal 14 (2023) 102140

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

The use of crowdsourcing data for analyzing pedestrian safety in urban areas

Mostafa Telima [a,*], Mohamed El Esawey [b], Karim El-Basyouny [c], Ahmed Osama [d]

[a] Center of Mobility Research, Public Works Department, Faculty of Engineering, Ain-Shams University, Cairo, Egypt
[b] Public Works Department, Faculty of Engineering, Ain-Shams University, Cairo, Egypt
[c] City of Edmonton's Urban Traffic Safety Research Chair, Civil and Environmental Engineering Department, Faculty of Engineering, University of Alberta, Edmonton, Canada
[d] Center of Mobility Research, Public Works Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt

ARTICLE INFO

Article history:
Received 25 October 2022
Revised 26 December 2022
Accepted 6 January 2023
Available online 11 January 2023

Keywords:
Streetguards
Crowdsourcing
Spatial analysis
Pedestrian-vehicle collisions
Hotspot locations

ABSTRACT

Pedestrians are the most affected vulnerable road users by traffic collisions. Due to incomplete and inconsistent collision statistics, assessing pedestrian safety remains a complex issue in developing countries. This study investigates the potential of using crowdsourced data to identify hotspot locations by observing pedestrian-vehicle interactions. Safety analysis was carried out using traffic incident data in Eastern Cairo, Egypt. Incident data included collisions, near misses, and infrastructure issues. Spatial autocorrelation analysis was undergone to determine whether incidents are clustered, dispersed, or randomly distributed. The results showed that incidents in the study area are generally dispersed. Nevertheless, local spatial autocorrelation showed that some locations on four major corridors were identified as hotspots with a 99% confidence level. The approach proposed in this study shall help transportation authorities in developing countries to identify and prioritize sites that require more safety attention.

1. Introduction

Traffic collisions are a considerable health concern worldwide and seriously threaten the social well-being and economic growth of developing countries such as Egypt. According to World Health Organization's (WHO) latest global report on road safety, approximately 1.35 million people die, and more than 50,000 people are injured or disabled every year due to road collisions. Collisions involving pedestrians represent a significant proportion, 27 %, of road fatalities. More than 90 % of road traffic deaths occur in low- and middle-income countries despite only forming about 45 % of the world's vehicles [1]. According to the NADA Foundation for Safer Egyptian Roads, an Egyptian nongovernmental organization (NGO) under the supervision of the Ministry of Social Solidarity, Egypt has one of the highest road-related fatality rates in the world, with 42 deaths per 100,000 individuals. The foundation indicated that traffic collisions are the leading cause of death for Egyptian youth aged 15 to 35 and that an average of four children die in road collisions daily. Egypt lost 3.2 % of its Gross Domestic Product (GDP) or $10.74 billion annually due to traffic fatalities and serious injuries [2]. Traffic collisions involving pedestrians are more likely to occur in urban areas with increased pedestrian activity and traffic volume. In 2003, 75 % of pedestrian fatalities in the United States occurred in urban areas [3].

Since 2014, the Egyptian government has focused on expanding the country's road network and widening main roads in many neighbourhoods throughout Greater Cairo Region (GCR) to mitigate traffic congestion. However, programs focused on improving pedestrian safety are scarce. Identifying hotspot locations involving pedestrians are usually the first step in a Pedestrian Safety Improvement Program (PSIP). Assessing pedestrian safety and identifying pedestrian-vehicle collision hotspots is essential for understanding the causes of these problematic locations and developing effective countermeasures.

Detection of hotspot locations requires reliable historical traffic and collision data. Unfortunately, in many developing countries, collision data suffer from several issues, including unavailability,

* Corresponding author.
E-mail addresses: Mostafa_Gamal_telima@eng.asu.edu.eg (M. Telima), m.elesawey@eng.asu.edu.eg (M. El Esawey), basyouny@ualberta.ca (K. El-Basyouny), ahmed.osama@alumni.ubc.ca, ahmed.osama@eng.asu.edu.eg (A. Osama).

Peer review under responsibility of Ain Shams University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.asej.2023.102140
2090-4479/

incompleteness, and inconsistency. More specifically, collisions involving pedestrians are usually underreported. According to previous studies, fatality underreporting ranges between 2 and 5 percent in developed countries and 25 to 50 percent in developing countries [4]. One of the most significant challenges in Egypt is the inconsistency of collision statistics issued by different authorities, which varies depending on their standards and requirements for reporting. As a result, crowdsourced data collected from volunteers through websites and mobile applications emerged as an alternative or supplementary source of information for traffic safety analysis. The main advantages of crowdsourcing include saving costs for collecting data and promoting sustainability. In addition, there is no need for roadside sensors or devices to be installed or maintained. On the other hand, crowdsourcing has several drawbacks, including redundancy in traffic incident reports and the possibility of false data being recorded.

In this study, pedestrian-vehicle collisions and near misses were collected from 'Streetguards'; a crowdsourcing tool developed to collect and map traffic incident data using a GIS platform. A near miss or conflict is defined as "an observational situation in which a vehicle [can also be a pedestrian or a bicyclist] and pedestrian [can also be a bicyclist or a vehicle] approach or encroach each other in space and time to such an extent that a collision is imminent if their movements remain unchanged" [5]. Conflicts occur more frequently and provide insight into the failure mechanism that leads to collisions. The collected data was used to undertake spatial pedestrian safety analysis using Geographical Information Systems (GIS). Identifying high-risk locations analysis was divided into two parts: firstly, a spatial analysis was carried out using the Kernel density estimation method, which could identify locations with a high spatial concentration of pedestrian-vehicle incidents.

Consequently, these locations were verified via spatial autocorrelation analysis to determine whether incidents are statistically clustered, dispersed, or randomly distributed.

The main objective of this study is to investigate the potential of using crowdsourced data to advance pedestrian safety assessments by identifying pedestrian hotspot locations. Findings from this research could be used to assist decision-makers and transportation planners in prioritizing the safety of vulnerable road users. This ensures that the rights of those who frequently do not receive attention in public spaces are properly addressed. It also allows pedestrians to enjoy a safer and more efficient mobility system and healthier access to public spaces.

2. Literature review

With the widespread availability of web and mobile applications, crowdsourcing has become one of the most widely used methods for enhancing traffic databases by collecting various levels of information that would remain unreported. Many previous studies tested using crowdsourcing to improve vulnerable users' safety. Rahman et al. [6] developed an android-based crowdsourced data collection application to locate and characterize conflicts for vulnerable modes. The crowdsourced data obtained from the application was compared to traditional fatality data for hotspot analysis. Medury et al. [7] analyzed pedestrian and bicycle collisions in and around three university campuses in California by comparing police-reported crash data with traffic safety information sourced from the campus communities. The crowdsourced traffic safety data comprised self-reported crashes and perceived hazardous locations via the web tool. SimRa is a crowdsourcing application for collecting data on bicycle routes and near-miss incidents to identify dangerous near-miss hotspots [8]. Saad et al. [9] utilized cycling crowdsourced data from STRAVA to develop safety performance functions (SPFs) for bicycle collisions at intersections. One popular crowdsourced cycling data source is BikeMaps.org, a global website that visualizes and collects bike collisions, theft incidents, and near misses [10]. Amin et al. [11] utilized Waze as a crowdsourcing data tool that can provide additional coverage of accidents with low false alarms to conventional traffic management systems. Timely reporting has also been found compared to the probe-base alternative. The redundancy of crowdsourced traffic incident reports was a major challenge for using unofficial data sources [11]. To reduce redundancy in Waze data, Amin et al. [12] proposed validating incident reports by fusing/clustering various types of information, such as spatiotemporal and semantic information, including incident type, road name, and direction. Brotoisworo et al. [13] examined the use of collision data from Waze and Twitter for populating the national collision database of the Philippines.

Crowdsourcing of traffic data is still in its early stages in Egypt, with only a few studies exploring its potential. In 2010, a cross-platform mobile application called "Bey2ollak" was launched in Egypt to allow users to share real-time traffic information in Cairo and Alexandria. Ezzat et al. [14] proposed extracting road intersections and turns from the trip data extracted from Bey2ollak. A newly developed global mapping tool named Streetguards.com allows users to report traffic collisions, near misses, infrastructure issues, and threatening incidents [15].

Hotspot identification typically relied on road collision data. There are currently several techniques for hotspot identification. These approaches vary in their level of complexity and data requirements. For example, simple techniques for identifying hotspot locations use collision frequencies [16] or collision rates [17,18]. Regression-to-the-mean (RTM) artifact is a well-known drawback of these simple methods. As such, more advanced statistical methods were developed and applied, such as the Empirical Bayes (EB) [19–22] and the Full Bayes (FB) [23,24]. Potential for safety improvement (PSI) is suggested as a safety ranking measure that takes into account the effects of traffic volume and other exposure indicators on safety [25,26]. Only a few studies have considered collision severity while developing risk criteria [27].

The study explores the power of open crowdsourcing data in identifying hotspot locations for pedestrian safety. Agencies and community organizations can use this data to identify and prioritize strategies for responding to potential pedestrian public health concerns.

3. Methodology

The contribution of the study is to identify locations with potential safety problems for pedestrians using crowdsourced data, through various spatial analyses, as an alternative to collision data. This research will significantly improve pedestrian safety in locations with no/poor collision data. Diagnosing those locations with safety issues will enable transportation engineers and policymakers to better understand the problem's root causes, and develop a set of countermeasures that can mitigate safety problems The flowchart in summarises the used methodology for the proposed study. Each part of this methodology will then be discussed in detail in the following sub-sections Fig. 1.

4. Data acquisition

Data was collected through a crowdsourcing tool 'Streetguards,' an open free website released on February 1st, 2020. Users can access the developed tool by visiting the website https://www.streetguards.com/. They can record traffic incidents such as collisions/near misses, infrastructure issues, and threatening events.

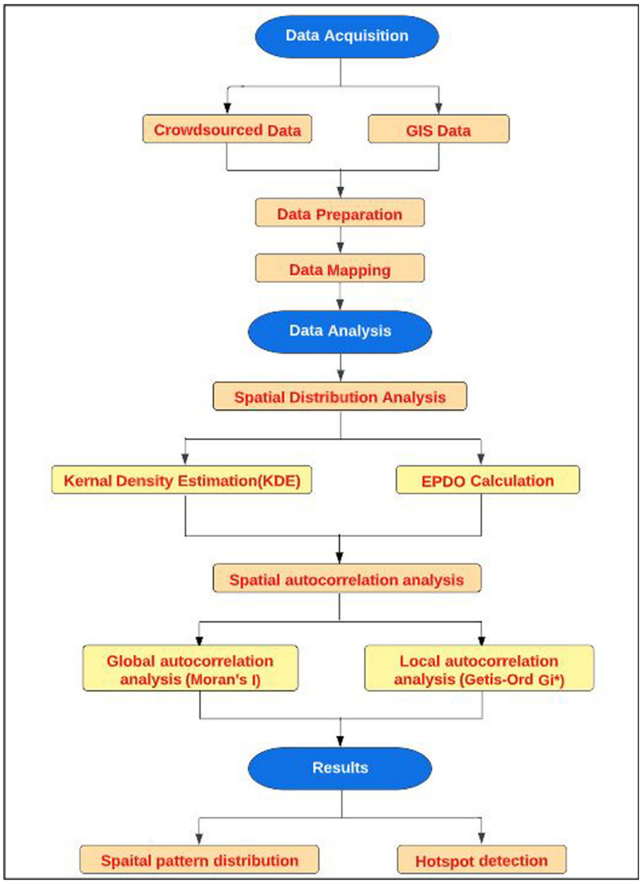

Fig. 1. Flowchart of methodology.

Information can be recorded in detail, including the incident location on a map, incident date, incident type, reporter contact details, number of injuries and fatalities, number of involved vehicles and pedestrians, text description of the incident, and photos of the incident, if available. The data used in this study was recorded on the website between February 2020 and May 2022. The majority of reports were in GCR, which is populated by nearly 19 million people, with the World Bank (WB) projecting the population to be 24 million by 2027.

This study focused on three urban districts in Eastern Cairo, Egypt: Al-Nozha, Masr El-Gedida, and East Nasr City. These districts have a total area of 1082.05 km^2 with around 1,042,900 residents (Cairo Governorate, 2020). These areas have mixed land use, including residential, commercial, and educational. These three areas are characterized by a high rate of car ownership due to the higher social classes of their residents. Private cars and taxis became the most viable mode of travel in these areas due to the limited frequency of buses and the worn-out then removed tramlines [28]. Recently, the three neighborhoods have undergone a significant transformation by widening the existing streets and constructing flyovers in most of the main squares. This allowed more space for traffic flow, easing traffic congestion within the neighborhoods and linking these districts to the main highways leading to the new communities.

4.1. Data collection and preparation

The data collected from Streetguards was exported from the back-end database, and records were filtered manually, case by case, through the following steps:

1. Removing all the duplicate points (i.e., having the same location, time, and description by the same person),
2. Deleting all recorded events that contained false information (such as false location),
3. Verifying that each incident was classified correctly by carefully checking the description of each reported event,
4. Reclassifying the records with incorrect incident categories while having accurate details and descriptions of an incident.
5. Dividing the incident categories into subcategories to cluster similar reports within each category. This was based on the underlying causes of the incidents as mentioned in the description.

The keywords used in the incident description were analyzed using a word cloud, which is a visual representation of word repetition. A word would appear more prominently in the image depending on how frequently the term appears in the analyzed text. shows the results of this analysis. The cloud of words reveals several problems related to pedestrian safety, including a lack of pedestrian crossings, traffic signal malfunction, and vehicle speeding. In other cases, medians, which are meant as refuge areas for pedestrians, were found too narrow and improperly designed to accommodate pedestrians. In other locations, access to the pedestrian overpass was not properly maintained. Finally, other locations suffered poor lighting causing the environment to be less safe and secure for pedestrians Fig. 2.

4.2. Data mapping

Pedestrian-vehicle incident data was exported into an ArcGIS platform for analysis. Incident data included pedestrian-vehicle collisions and near misses and road network data. Incidents were included in a point shape file projected on the road network map that includes road characteristics such as segment length, road classification, and the number of lanes, etc. Fig. 3 shows the locations of pedestrian-vehicle incidents throughout the three districts. A total of 226 pedestrian-vehicle collisions and 295 conflicts were recorded on the road network of the study area.

To emphasize the importance of the information provided by the reporters about the incidents, as well as how to report and classify each incident, a brief video was posted on the Streetguards' Facebook page for this purpose. Furthermore, posts on the homepage promoted the website and encouraged people to report and raise awareness of the importance of road safety measures.

One of the common challenges of crowdsourcing is data reliability. To verify the reported incidents, the research team conducted many site visits to diagnose the spots with potential

Fig. 2. Word cloud created from user incident descriptions.

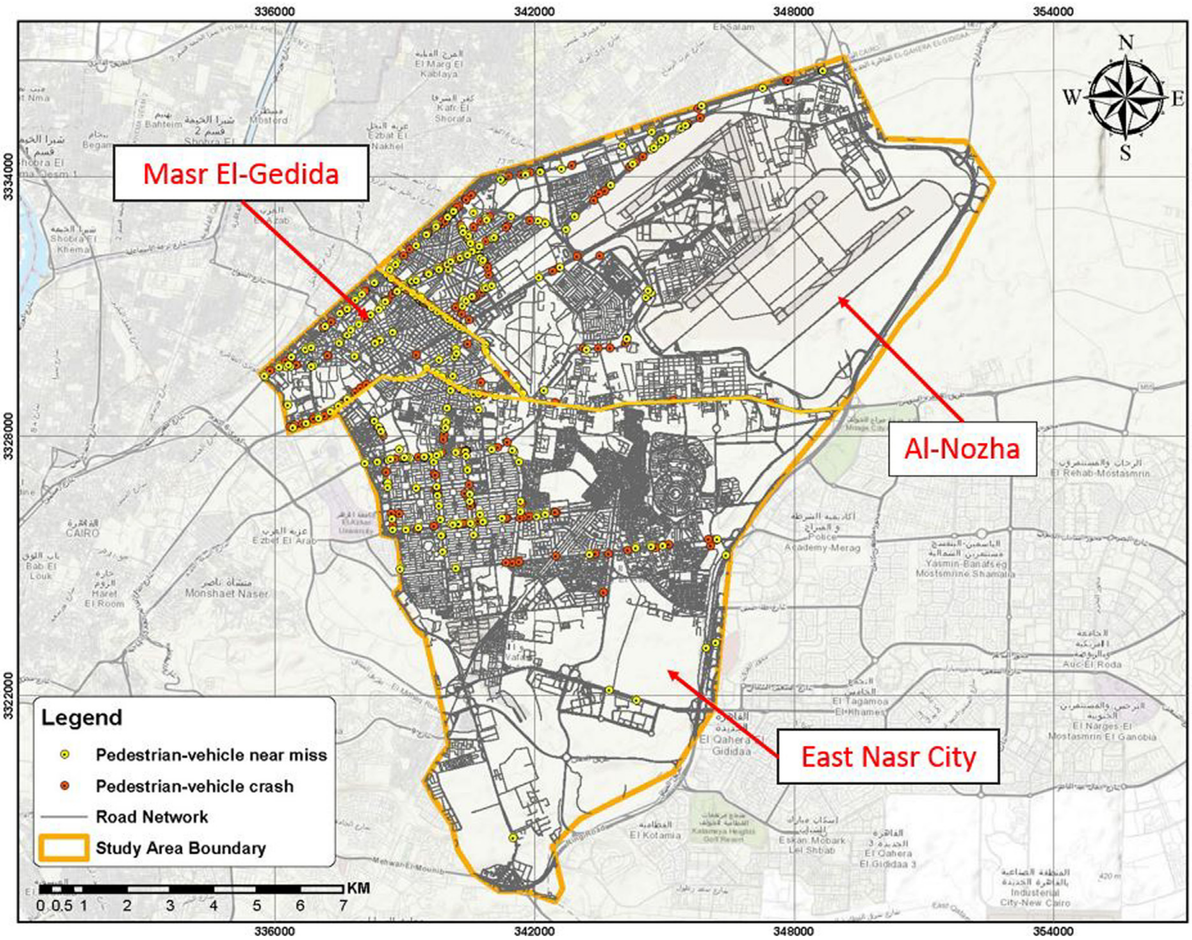

Fig. 3. Map of pedestrian-vehicle collisions and conflicts in three districts.

safety problems. Several issues were explored, such as the difficulty of crossing the road, street lighting, and driving behavior (i.e., speeding). To enhance and improve the dataset, researchers recorded collisions/near misses with attached photos periodically every week.

5. Analysis and results

Spatial distribution analysis of pedestrian incidents was undertaken to identify hotspot locations. Kernel Density Estimation (KDE) is one of the most widely used techniques to profile the concentration of events. Two types of KDE techniques currently exist planar KDE and network KDE. The first technique, planar KDE, estimates the density of point events over a specific area, whereas network KDE estimates the density over a distance. Both planer [29–32] and network [33–35] KDE were used in many previous studies for hotspot analysis. One major drawback of KDE techniques is that they cannot be used to test for statistical significance [30]. Alternatively, spatial autocorrelation is utilized to validate the statistical significance of spatial patterns of incident data. Moran's I Index and Geary's C Ratio measures simultaneously determine the spatial autocorrelation based on feature locations and feature values. Global statistics examine whether a dataset's overall spatial pattern is clustered, dispersed, or random. Local autocorrelation is adapted to measure spatial associations between each feature and neighboring features. Local Moran's I [36] and Getis-Ord Gi* statistics [37] are utilized as measures to analyze local spatial autocorrelation. In many recent studies [38–40], hotspot locations were identified using spatial autocorrelation analysis of collision data.

In this study, spatial safety analysis of pedestrian-vehicle incidents was carried out in a Geographical Information Systems (GIS) platform. Consequently, the identified hotspot locations were verified using spatial autocorrelation analysis to determine whether they were statistically clustered, dispersed, or randomly distributed. The analysis is described in more detail in the following sub-sections.

5.1. Spatial distribution analysis

The planer KDE technique was used to identify hotspot locations by creating a heatmap that profiles incidents and near misses in the study area. The purpose of KDE is to generate a smooth density surface of point events over space by counting the number of incidents at each location. The heatmap has the highest density value in the center, decreasing as one moves away from the center. The KDE function is given by Eq. (1) [41].

$$f(x,y) = \frac{1}{n \cdot h^2} \sum_{i=1} K\left(\frac{d_i}{h}\right) \tag{1}$$

Where f is the estimated density at location (x, y), h is the search radius (i.e. bandwidth or kernel size), n is the number of points (i.e. total number of incidents), d_i the distance between the event point i and the (x,y) location, and K is the kernel function.

The search radius, h, is given by:

$$h = 0.9\, min\left(SD, \sqrt{\frac{1}{In(2)}}D_m\right) * n^{-0.2} \tag{2}$$

Where SD is the standard distance, D_m is the weighted median distance from the weighted mean center, and n is the number of points (if no population field is used) or the sum of the population field values (if a population field is applied). The population field is a field denoting population values for each feature. The population field is the count or quantity spread across the landscape to create a continuous surface. No population field is used in this study so each incident will be counted once.

Fig. 4 shows the heatmap for the density of pedestrian-vehicle incidents across the study area. The areas highlighted in red have the highest density, which ranges between 24.9 and 37.2 incidents/Km^2 while the uncolored areas have the lowest incident density (i.e., less than 7.01). Most incident concentration was found in Masr El-Gedida and the northwest of East Nasr City and Al-Nozha districts, respectively. A high frequency of incidents occurred at the intersection of Gesr Al Suez and Fareed Semeika roads, known as Alf Maskan Square. This square experiences heavy pedestrian activity due to the existence of a metro station, microbus stops, and commercial shops. Furthermore, it was also noted that Abou Bakr El-Sedeek Road experiences high incident occurrence. The red areas in Fig. 4 show where incidents are more likely to occur. Incident severity should be considered when determining the actual hotspots. The seriousness (i.e., severity) of incidents was determined based on the description of the recorded events. The average economic cost of different incident severity levels was assumed based on their relative weights. The average cost of a collision in Egypt in 2008 was estimated as 1,994,002, 120,735.3, 84,200.2, and 55,472.7 Egyptian Pounds (EGP) for fatal, serious injury, slight injury, and Property Damage Only (PDO) collisions, respectively [42]. These costs were corrected as per the current economic indicators. Relative weights 35.95, 2.18, 1.52, and 1.00 were used as PDO equivalency factors. Equation (3) was used to calculate Equivalent Property Damage Only (EPDO) values for incidents at each location:

$$EPDO = 35.95 \times X_1 + 2.18 \times X_2 + 1.52 \times X_3 + X_4 \tag{3}$$

Where.

$X1$ = total number of fatal collisions.

$X2$ = total number of serious injury collisions.

$X3$ = total number of other injury collisions.

$X4$ = total number of property-damage-only collisions.

Fig. 5 illustrates the severity of incidents according to EPDO estimates in the study area. The severity of incidents is divided into five classes. Class 1 includes a range of 1.0 to 10.9, which indicates low severity. High severity locations fall under Class 5, with EPDO that ranges between 393 and 719.

The spatial distribution analysis of incident severity led to many important observations. Firstly, an area around one crosswalk, Abou Bakr El-Sedeek-Haroun El-Rashid, exhibited high incident severity. Abou Bakr El-Sedeek is a main arterial with high traffic volumes and high exposure of pedestrians due to a metro station and a school nearby. In addition, the pedestrian traffic signal was malfunctioning most of the time, leading to multiple interactions between pedestrians and vehicles. These interactions, in turn, caused many severe conflicts around the crosswalk.

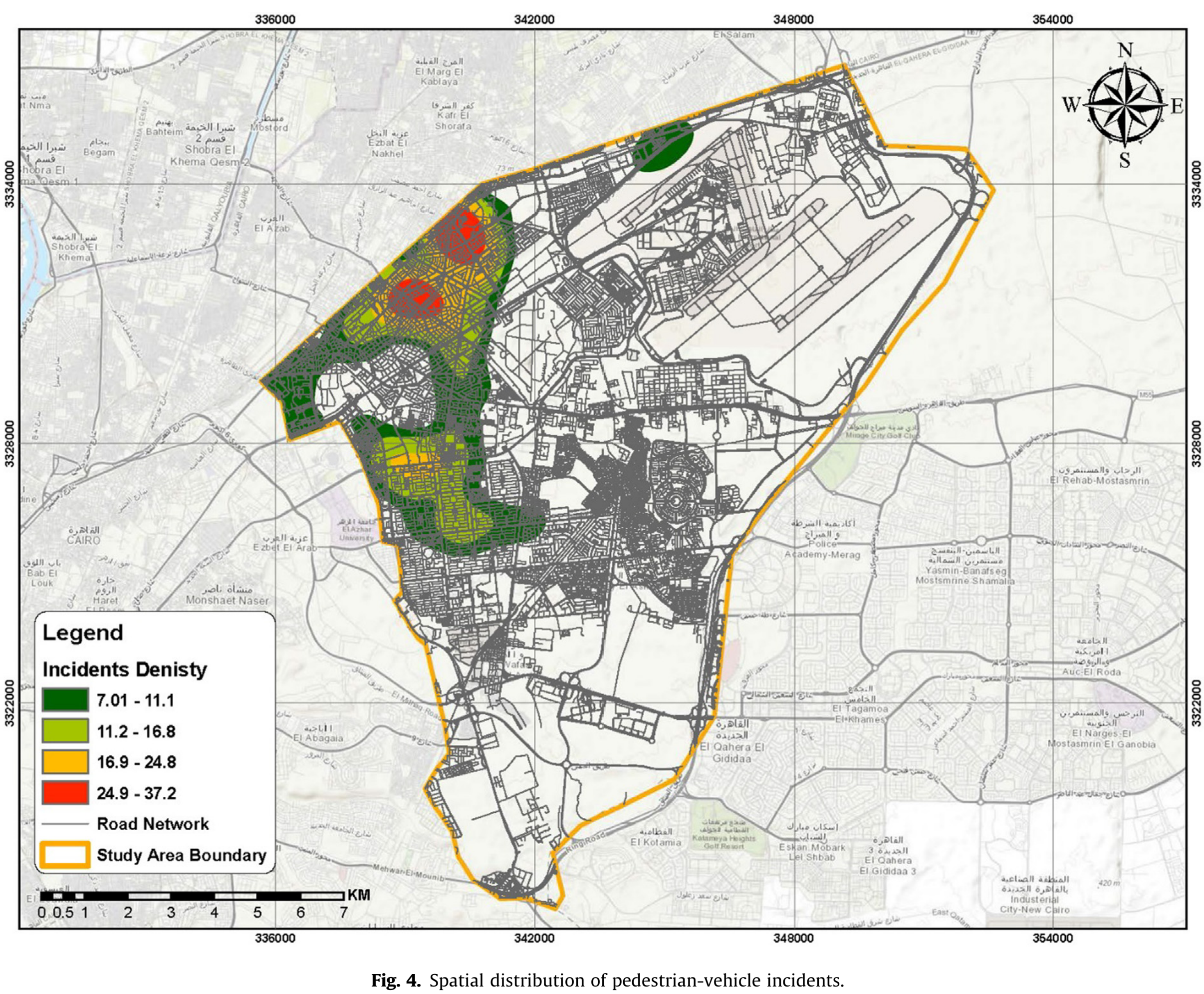

Fig. 4. Spatial distribution of pedestrian-vehicle incidents.

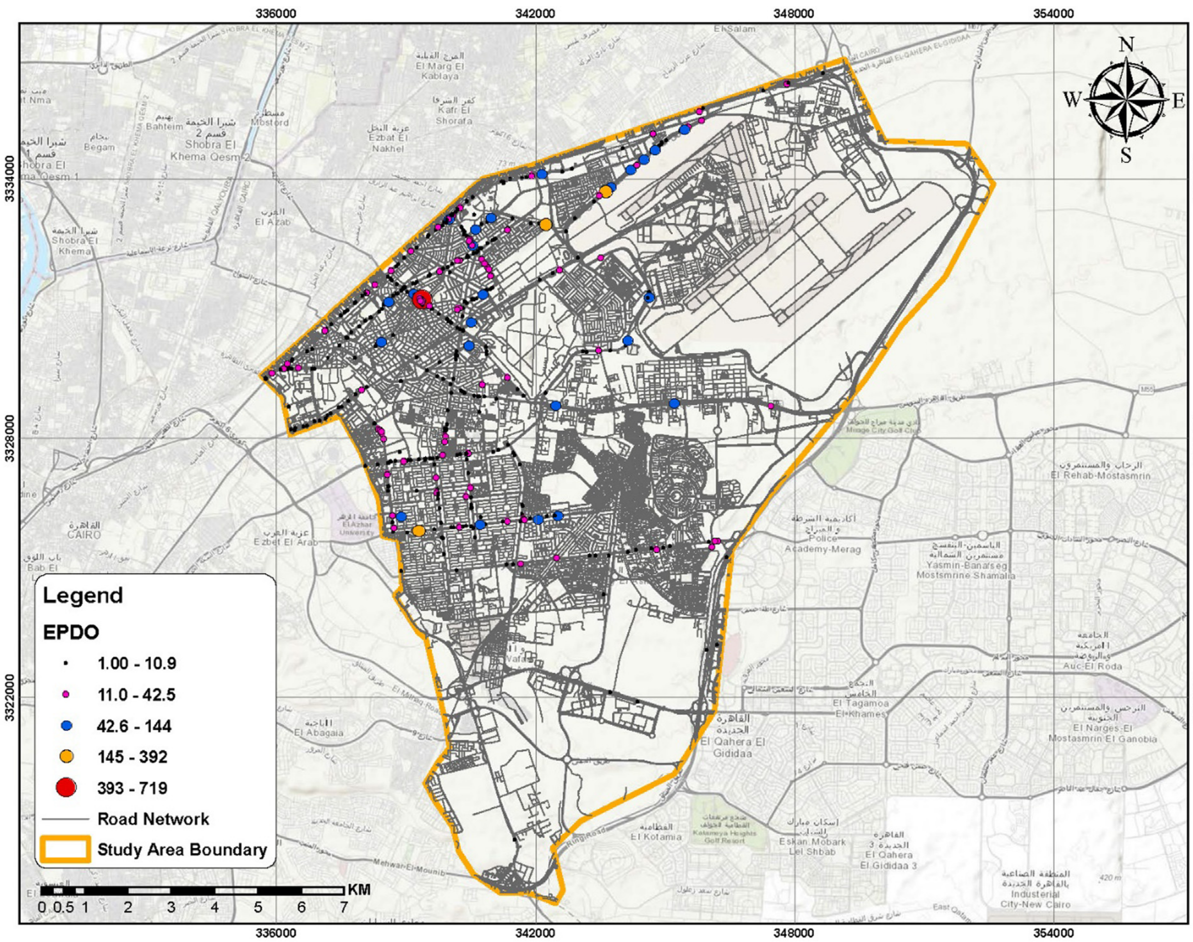

Fig. 5. Distribution of EPDO incidents.

Another location, Joseph Tito Road, a major corridor, was found to have clusters of moderate to severe incidents attributed to the high speed of vehicles and lack of adequate night lighting. Another corridor, Mostafa El-Nahass, was found to have an irregular distribution of incident severity. Frequent bus and microbus stops and illegal pedestrian crossings at the start and end of flyovers on this road were observed. Another location, Nasr Road, was noticed to have some moderate incidents due to vehicle speeding, and access to the pedestrian overpass was not properly maintained. Fig. 6 shows examples of high-risk locations for pedestrian crossing. The description of the problems in those locations (as input by the users) is consistent with the word cloud analysis. Moreover, a previous study found that speeding is the most significant factor that led to collisions in the same study area [15], followed by aggressive drivers' behaviour and illegal pedestrian crossings. These factors accounted for more than 95 % of traffic collision causes.

The visual examination of the incident severity map gave some valuable insights into the overall distribution of incidents in the study area. Nevertheless, it was still necessary to statistically examine the degree of clustering in the dataset based on more quantitative measures.

5.2. Spatial autocorrelation analysis

The method of identifying consistent groups of objects based on the quantities of their attributes is known as spatial autocorrelation analysis [43]. There are two common approaches to spatial autocorrelation analysis: global and local spatial autocorrelation [44]. These two types of spatial analysis are applied in this study.

5.2.1. Global autocorrelation analysis

Moran's I statistic is a measure of global autocorrelation which was used to examine the overall spatial pattern of pedestrian-vehicle incident data. The Moran's I measure defines whether incident data is clustered, dispersed, or random. Moran's I is calculated based on the simultaneous measurement of both incident locations and their EPDO values. The inverse distance between any two points defines their location proximity weight. This is expressed mathematically as:

$$I = \frac{n}{S_0} \cdot \frac{\sum_i^n \sum_{j=1}^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i^n (x_i - \bar{x})^2} \tag{4}$$

where, x_i, x_j denote the i th and j th EPDO values, respectively, $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i, w_{ij} = \frac{1}{d_{ij}}$ represents the spatial weight between location i and its adjacent location j, n which is the total number of features. The aggregate value of all spatial weights is given by:

$$S_o = \sum_{i=1}^n \sum_{j=1}^n w_{i,j} \tag{5}$$

Values of Moran's I range from −1 to 1, and a value greater than 0 indicates a positive correlation (i.e., clustering) in the data, while a value less than 0 then indicates a negative correlation (i.e., dispersion) meaning that the same area has a large difference in attributes. Values of the Global Moran's I statistic that are closer to zero

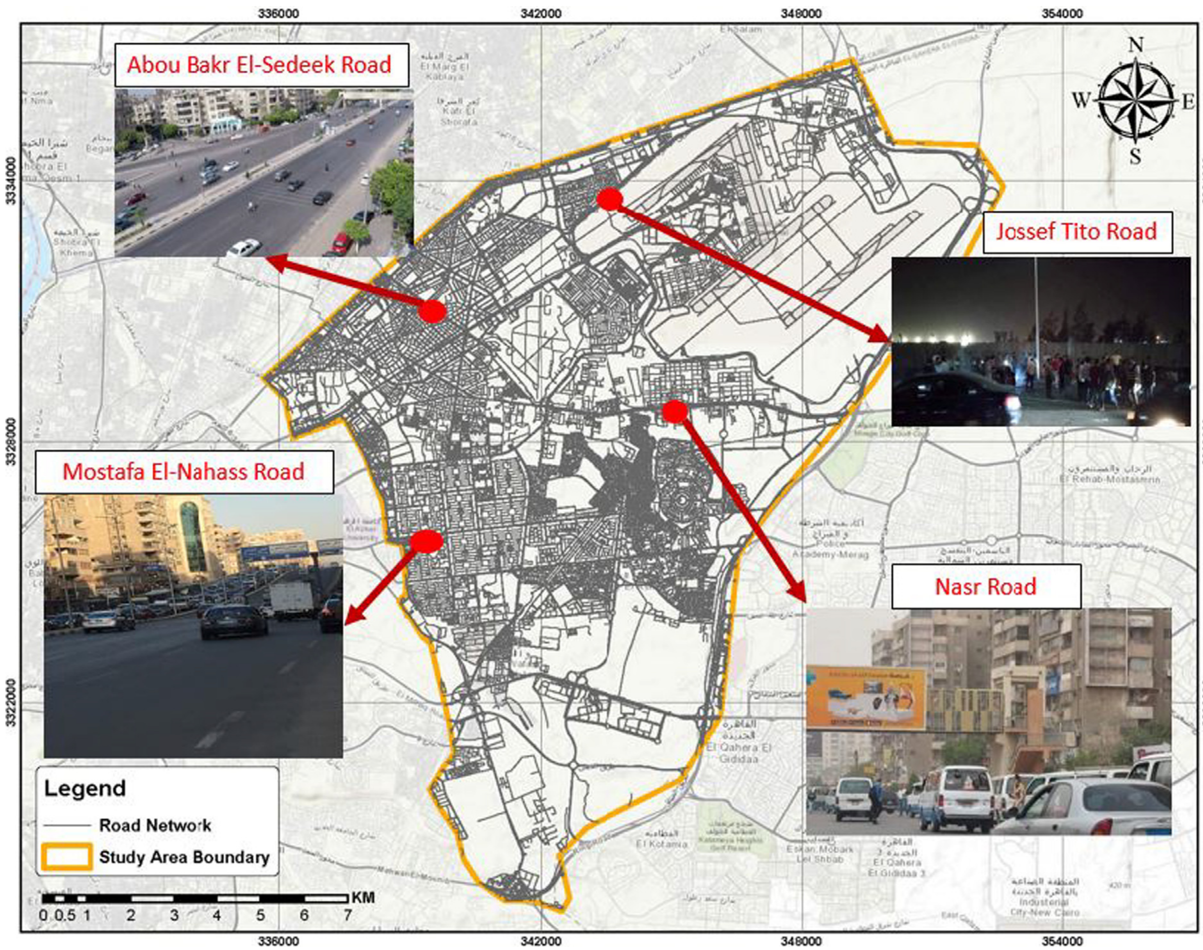

Fig. 6. Examples of high-risk locations for pedestrian crossing.

suggest a random distribution in the spatial pattern of the data. To evaluate the statistical significance for Moran's I, the z-score is calculated such that:

$$Z = \frac{I - E[I]}{\sqrt{V[I]}} \tag{6}$$

$$E[I] = -\frac{1}{n-1} \tag{7}$$

Where $E[I]$ denotes the expected Moran's and $V[I]$ is the variance of values in the dataset.

$$V[I] = E[I^2] - E[I]^2 \tag{8}$$

The null hypothesis for the Global Moran's I statistic states that the spatial pattern exhibits a random distribution. The alternative hypothesis has two possible outcomes:

- When the p-value is statistically significant and the z-score is positive. The spatial distribution of high and/or low values in the dataset is more spatially clustered than expected if the underlying spatial processes were random. In simple words, a positive z-score indicates that the values of the adjacent features are similar.
- When the p-value is statistically significant and the z-score is negative. The dataset's spatial distribution of high and/or low values is more spatially dispersed. A dispersed spatial pattern often reflects some type of competitive process. A feature with a high value repels other features with high values and vice versa. A negative z-score suggests that the values of the adjacent features are different [45].

The results of the global autocorrelation analysis are shown in Fig. 7. Moran's I was found to be −0.1072 with a p-value of 0.0193 and a z-score of −2.3397. These results indicate that incidents are dispersed in the study area with a 95 % confidence level. In other words, nearby locations have more dissimilar EPDO values.

5.2.2. Local autocorrelation analysis

To identify pedestrian-vehicle hotspot locations, the Getis-Ord Gi* statistic was used to analyze local spatial autocorrelation. This statistic is used to assess whether each attribute (i.e. EPDO value of incident) is similar/different from the others in their neighborhood [37]. The Getis-Ord Gi* statistic is calculated as [46]:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2}{n-1}}} \tag{9}$$

Where x_j is the EPDO value of j th element, $w_{i,j}$ is the spatial weight of i th element and j th element, and n is the total number of incidents.

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{10}$$

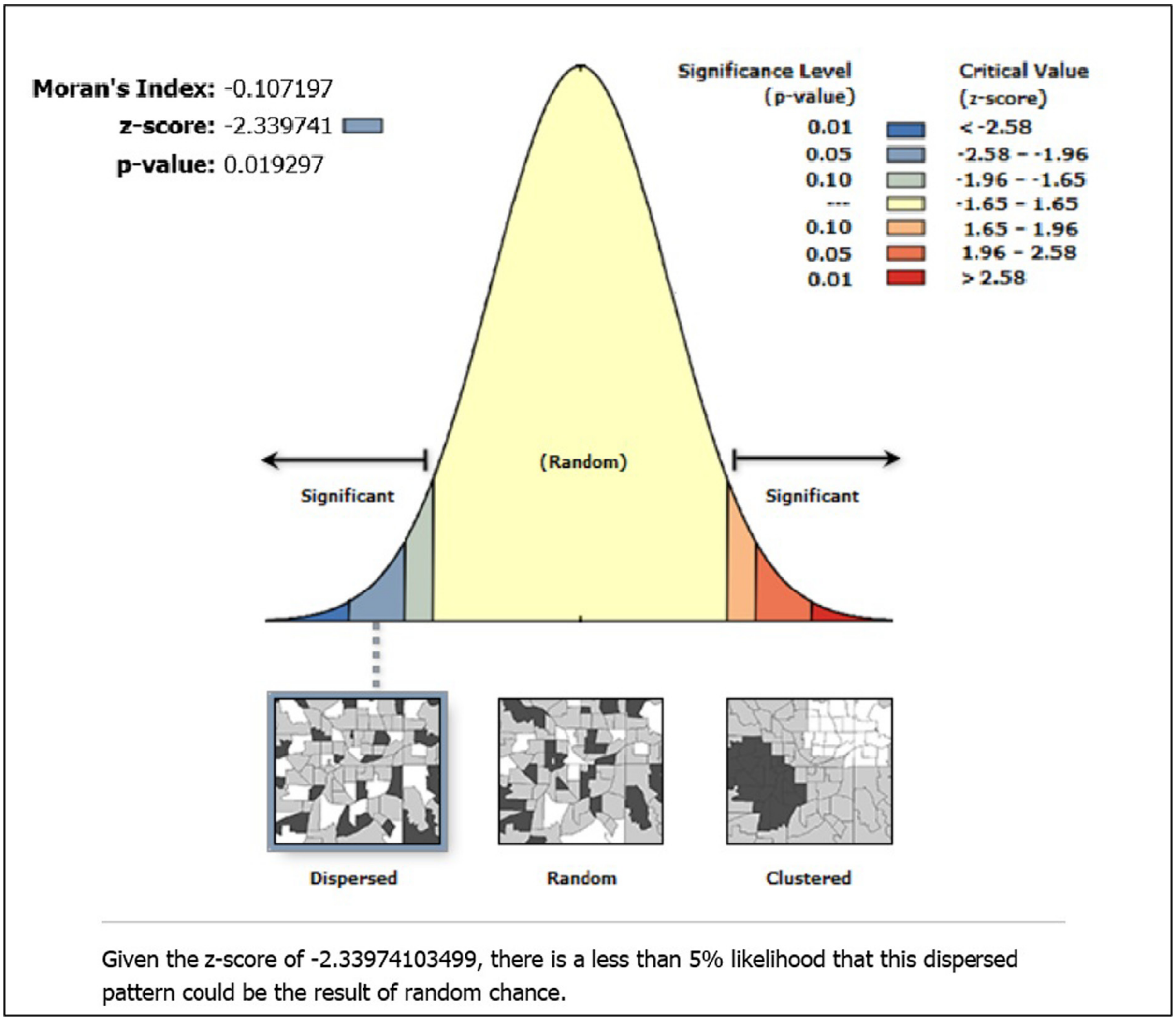

Fig. 7. Global Moran's I report.

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \quad (11)$$

High values of the Getis-Ord Gi* statistic indicate that the location is a hotspot. To examine the statistical significance of clustering values, the z-score is calculated. The higher the z-score (i.e., greater than zero), the more likely the site is a hotspot and vice versa.

The red zones in Fig. 8 are hotspots with a 99 % confidence level. The results confirmed the outcomes of the spatial distribution analysis using the KDE technique. Abou Bakr El-Sedeek-Haroun El-Rashid crosswalk was identified as a hotspot location with a 99 % confidence level. As well some locations on Joseph Tito road were also identified as hotspots.

6. Conclusions

Pedestrians are the most vulnerable type of road users. However, assessing pedestrian safety is a complex issue in developing countries due to limited, incomplete, and inconsistent collision data. This study focuses on analyzing pedestrian safety in urban areas in Cairo, Egypt. The main objective was to investigate the potential of using crowdsourced data for pedestrian risk assessment. More specifically, to identify potential hazardous locations for pedestrian safety using pedestrian-vehicle collisions and near misses. A crowdsourcing tool, Streetguards, was utilized to collect reported collisions and near misses in three urban neighborhoods in Eastern Cairo between February 2020 to May 2022. The collected data was filtered, mapped, and analyzed in a GIS environment. Spatial distribution analysis using the KDE technique was initially employed to determine the density of hazardous events/incidents in the study area. Consequently, global autocorrelation analysis was used to test whether these events are random, dispersed, or clustered. Moran's Index statistic demonstrated that the distribution of the collisions in the study area is dispersed with a 95 % confidence level. Local autocorrelation analysis was consequently used to identify pedestrian hotspot locations.

The employed techniques identified many locations as hotspots. Site visits to these locations confirmed the existence of potential safety problems for pedestrians for various reasons. For example, the Abou Bakr El-Sedeek-Haroun El-Rashid crosswalk was identified as a hotspot with a 99 % confidence level. The crosswalk is located on a major arterial with high traffic volumes where a metro station and a school exist. Pedestrian traffic signal experienced a continuous malfunction, which resulted in a lack of control on the crosswalk. Due to the high traffic and pedestrian volumes with a lack of control, many serious interactions and severe conflicts

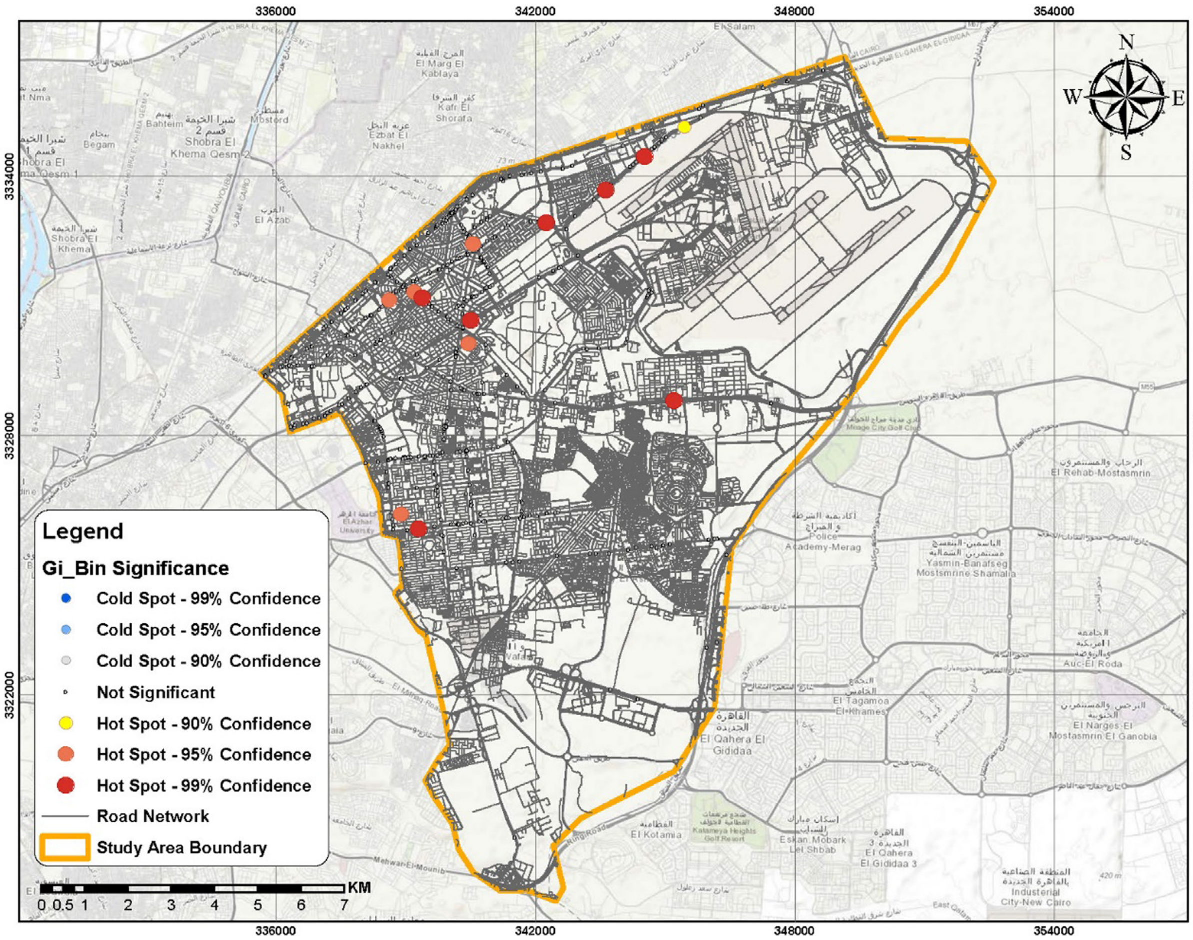

Fig. 8. Significance of pedestrian-vehicle incidents hotspots.

occurred between pedestrians and vehicles around this crosswalk. Another location, Joseph Tito Road, was also identified as a hotspot location with a 99 % confidence level. High speeds of vehicles and lack of adequate night lighting characterize this road.

The research reported in this study aimed to identify locations with potential safety problems for pedestrians using crowdsourced data as an alternative to collision data. The next stage of this research will involve collecting video data for the identified hotspot locations to undertake a more rigorous pedestrian safety diagnosis further. Computer vision techniques will be used to determine surrogate safety measures such as traffic conflicts and violations. This will give a better grasp of the magnitude of the safety problem at these locations. It will also enable analyzing the temporal variation of the safety level at these facilities over different periods of the day. This research will significantly improve pedestrian safety in locations with no/poor collision data. Identifying and diagnosing locations with safety issues will enable a better understanding of the problem's root causes, which will enable the development of a set of potential countermeasures that can efficiently mitigate safety problems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] World Health Organisation (WHO), Global status report on road safety 2018: summary. Geneva. 2018; 2018.
[2] World Bank 2020, Road Safety Opportunities & Challenges: Low- and Middel-Income Country Profiles, News.Ge, p. https://news.ge/anakliis-porti-aris-qveynis-momava, 20189.
[3] Safety H. Analysis of elderly pedestrian accidents and recommended countermeasures. J Safety Res 1996;27(2):128. doi: https://doi.org/10.1016/0022-4375(96)86964-7.
[4] Mahmud SM, Ferreira L, Hoque M, Tavassoli A. Reviewing traffic conflict techniques for potential application to developing countries. J. Eng Sci Technol 2018;13(6):1869–90.
[5] Hydén C. The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique. 70th ed. BULLETIN LUND INSTITUTE OF TECHNOLOGY, DEPARTMENT Trafikteknik Tekniska Hoegskdan i Lund; 1987.
[6] Rahman Z et al. Using crowd sourcing to locate and characterize conflicts for vulnerable modes. Accid Anal Prev 2019;128(June 2018):32–9. doi: https://doi.org/10.1016/j.aap.2019.03.014.
[7] Medury A, Grembek O, Loukaitou-Sideris A, Shafizadeh K. Investigating the underreporting of pedestrian and bicycle crashes in and around university campuses – a crowdsourcing approach. Accid Anal Prev 2019;130(July 2016):99–107. doi: https://doi.org/10.1016/j.aap.2017.08.014.
[8] Karakaya AS, Hasenburg J, Bermbach D. SimRa: Using crowdsourcing to identify near miss hotspots in bicycle traffic. Pervasive Mob Comput 2020;67. doi: https://doi.org/10.1016/j.pmcj.2020.101197.
[9] Saad M, Abdel-Aty M, Lee J, Cai Q. Bicycle safety analysis at intersections from crowdsourced data. Transp Res Rec 2019;2673(4):1–14. doi: https://doi.org/10.1177/0361198119836764.
[10] Nelson TA, Denouden T, Jestico B, Laberee K, Winters M. BikeMaps.org: a global tool for collision and near miss mapping. Front Public Heal 2015;3(MAR):1–8. doi: https://doi.org/10.3389/fpubh.2015.00053.
[11] Amin-Naseri M, Chakraborty P, Sharma A, Gilbert SB, Hong M. Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports

from Waze. Transp Res Rec 2018;2672(43):34–43. doi: https://doi.org/10.1177/0361198118790619.
[12] Amin-naseri M. Adopting and incorporating crowdsourced traffic data in advanced transportation management systems; 2018.
[13] Brotoisworo P, Sunio V. Exploring the Utilization of Crash Data from Waze and Twitter for Official Proceedings of the Eastern Asia Society for Transportation Studies , Vol. 12, 2019 Exploring the Utilization of Crash Data from Waze and Twitter for Official National Crash Datab," no. September 2019, 2020.
[14] Ezzat M, Attia R, Elgohary M, Khalifa ME. Extracting road turns and intersections from crowd source GPS tracks. In: International Conference on Communication and Signal Processing (ICCSP), 2017, pp. 0947–0951.
[15] Osama A, El Masry H, Farrag M, Telima M. Street Guards: A Tool for Crowdsourcing Road Safety and Security Data in Developing Countries. In: Transportation Research Board 101st Annual Meeting, 2022, no. TRBAM-22-01650.
[16] Deacon JA, Zegeer CV, Deen RC. Identification of hazardous rural highway locations. Transp Res Rec 1975;543:16–33.
[17] Barker, Judith, C. Baguley, A road safety good practice guide; 2001.
[18] Taylor MA, Bonsall PW, Understanding Traffic Systems: Data, Analysis and Presentation. 2nd ed. Aldershot: Ashgate; 2000.
[19] Li L, Zhang Y. Bayesian approach based on geographic information system to identify hazardous roadway segments for traffic crashes. Transp Res Rec 2007;2024:63–72. doi: https://doi.org/10.3141/2024-08.
[20] Hauer E. Observational Before/After Studies in Road Safety. Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Bingley, UK: Emerald publishing Ltd.; 1997.
[21] Hauer E. Identification of sites with promise. Transp Res Rec 1996;1542:54–60. doi: https://doi.org/10.3141/1542-09.
[22] Elvik R. State-of-the-art approaches to road accident black spot management and safety analysis of road networks; 2007.
[23] Huang H, Chin HC, Haque M. Empirical evaluation of alternative approaches in identifying crash hotspots. Transp Res Rec 2009;2103:32–41. doi: https://doi.org/10.3141/2103-05.
[24] Miaou SP, Song JJ. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accid Anal Prev 2005;37(4):699–720. doi: https://doi.org/10.1016/j.aap.2005.03.012.
[25] Hauer E, Kononov J, Allery B, Griffith MS. Screening the road network for sites with promise. Transp Res Rec 2002;1784:27–32. doi: https://doi.org/10.3141/1784-04.
[26] Persaud B, Lyon C, Nguyen T. Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. Transp Res Rec 1999;1665:7–12. doi: https://doi.org/10.3141/1665-02.
[27] Miranda-Moreno LF. Statistical models and methods for identifying hazardous locations for safety improvements. Ph.D.Thesis: Department of Civil Engineering, University of Waterloo; 2006.
[28] Herzog Y, Meuron J, de, Herz P, Rahbaran M, Zhou S. Helipolis City in a City; 2010.
[29] Pulugurtha SS, Krishnakumar VK, Nambisan SS. New methods to identify and rank high pedestrian crash zones: an illustration. Accid Anal Prev 2007;39(4):800–11. doi: https://doi.org/10.1016/j.aap.2006.12.001.
[30] Anderson TK. Kernel density estimation and K-means clustering to profile road accident hotspots. Accid Anal Prev 2009;41(3):359–64. doi: https://doi.org/10.1016/j.aap.2008.12.014.
[31] Erdogan S, Yilmaz I, Baybura T, Gullu M. Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. Accid Anal Prev 2008;40(1):174–81. doi: https://doi.org/10.1016/j.aap.2007.05.004.
[32] Khan G, Qin X, Noyce DA. Spatial analysis of weather crash patterns. J Transp Eng 2008;134(5):191–202. doi: https://doi.org/10.1061/(ASCE)0733-947X(2008)134:5(191).
[33] Young J, Park PY. Hotzone identification with GIS-based post-network screening analysis. J Transp Geogr 2014;34:106–20. doi: https://doi.org/10.1016/j.jtrangeo.2013.11.007.
[34] Xie Z, Yan J. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: An integrated approach. J Transp Geogr 2013;31:64–71. doi: https://doi.org/10.1016/j.jtrangeo.2013.05.009.
[35] Mohaymany AS, Shahri M, Mirbagheri B. GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. Geo-Spatial Inf Sci 2013;16(2):113–9. doi: https://doi.org/10.1080/10095020.2013.766396.
[36] Anselin L. Local Indicators of Spatial Association—LISA. Geogr Anal 1995;27(2):93–115. doi: https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.
[37] Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Geogr Anal 1992;24:189–206.
[38] Huang Y, Wang X, Patton D. Examining spatial relationships between crashes and the built environment: a geographically weighted regression approach. J Transp Geogr 2018;69(May):221–33. doi: https://doi.org/10.1016/j.jtrangeo.2018.04.027.
[39] Dai D. Identifying clusters and risk factors of injuries in pedestrian-vehicle crashes in a GIS environment. J Transp Geogr 2012;24:206–14. doi: https://doi.org/10.1016/j.jtrangeo.2012.02.005.
[40] Yu H, Liu P, Chen J, Wang H. Comparative analysis of the spatial analysis methods for hotspot identification. Accid Anal Prev 2014;66:80–8. doi: https://doi.org/10.1016/j.aap.2014.01.017.
[41] Jayan DK, Ganeshkumar B. Identification of accident hotspots: a GIS based implementation for Kannur district, Kerala. Int J Geomatics Geosci 2010;1(1):51–9.
[42] Ismail MA, Abdelmageed SM. Cost of Road Traffic Accidents in Egypt; 2010, pp. 1308–1314.
[43] Mohamed MG, Saunier N, Miranda-Moreno LF, Ukkusuri SV. A clustering regression approach: A comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US and Montreal, Canada. Saf Sci 2013;54:27–37. doi: https://doi.org/10.1016/j.ssci.2012.11.001.
[44] Waters N. A Review of Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications, vol. 102, no. 1; 2012.
[45] Siddiqui C, Abdel-Aty M, Choi K. Implications of pedestrian safety planning factors in areas with minority and low-income populations. Int J Sustain Transp 2014;8(5):360–81. doi: https://doi.org/10.1080/15568318.2012.702853.
[46] Songchitruksa P, Zeng X. Getis-ord spatial statistics to identify hotspots by using incident management data. Transp Res Rec 2010;2165:42–51. doi: https://doi.org/10.3141/2165-05.