Research Article

Spatial and Temporal Distribution Analysis of Traffic Accidents Using GIS-Based Data in Harbin

Meina Wang [ID]**, Jing Yi, Xirui Chen, Wenhui Zhang** [ID]**, and Tiangang Qiang** [ID]

School of Traffic and Transportation, Northeast Forestry University, Harbin 150040, China

Correspondence should be addressed to Wenhui Zhang; rayear@163.com and Tiangang Qiang; qiangtg2021@163.com

Road traffic safety is a social issue of widespread concern. It is important for traffic managers to understand the distribution patterns of road traffic accidents. To this end, this study examines the spatial and temporal patterns of road traffic accidents from both accident frequency and accident severity perspectives. Road traffic accident data from 2016 to 2018 in Harbin, China, were used for the analysis. First, the spatial localization of accidents was completed using geocoding, and the localized accident data were classified by season. Then, density analysis was performed both with and without considering road network density. The results of the density analysis showed that when road network density was considered, accidents were mainly distributed in urban centers, while accidents were more dispersed when road network density was not considered. Third, a cluster analysis considering accident severity found that low-severity accident clusters occurred mostly in urban centers. High-severity accident clusters were mostly present in suburban areas. Finally, the results of these two methods are shown by using the comap technique. Areas of the city with a high frequency and severity of crashes in each season were identified. This study will help traffic management to have a more visual and intuitive understanding of the urban traffic safety situation and to take targeted measures to improve it accordingly.

1. Introduction

The casualties and property damage caused by road traffic accidents are serious. According to the World Health Organization [1], approximately 1.3 million deaths and injuries occur as a result of traffic accidents each year worldwide. Road traffic-related deaths are the 8th leading contributor to human deaths worldwide. In 2016, 63,093 people died in traffic accidents in China, 1.7 times more than in the United States. The death rate per 10,000 vehicles reached 2.72, 2.1 times that in the United States [2]. To decrease the number of traffic accidents, it is necessary to determine where and when accidents occur frequently. Previous studies have shown that accidents occur with certain spatial and temporal patterns. Areas with a high frequency and severity of crashes vary over time and occur in different areas within cities. Therefore, the frequency and severity of accidents are combined to understand the spatial and temporal distribution patterns of accidents. This information helps traffic managers take

targeted preventive measures to reduce fatalities and injuries.

Most of the previous traffic accident data analyses used mathematical statistical methods [3–5]. These methods mainly use accident frequency to determine the location of accident hot spots [6]. Their operation and expression are relatively simple, but these methods have many limitations, such as the lack of visualization and inability to connect space and time. In contrast, spatial statistics methods can be used to visualize the spatial distribution characteristics of traffic accidents through GIS technology. Compared with traditional mathematical statistics, spatial statistics fully utilize the advantages of GIS in spatial data processing. On the one hand, the distribution of traffic accidents can be visualized through GIS visualization technology [7–9]. On the other hand, by using a variety of spatial analysis tools in GIS, scholars can explore the spatial distribution characteristics of traffic accidents and the spatial relationship between different traffic accidents from a variety of

perspectives [10–12]. The most common spatial statistical methods in GIS are density analysis, which accomplishes spatial visualization of accidents through kernel density and point density methods [13–16], cluster analysis, which can identify the spatial distribution of traffic accidents as aggregation, diffusion, or random distributions by nearest neighbor distance, and Ripley's K function method [17–19], which can identify traffic accident hotspot areas by hotspot analysis [20–22] and spatial autocorrelation analysis [23–25].

In general, existing studies have reported a series of results in the analysis of traffic accident data, but there are still some shortcomings. First, in previous studies on accident density analysis, only the accident density was considered, and the influence of road network density on the accident density was not considered. Second, the measure of traffic safety level includes not only the frequency of traffic accidents but also the consequences of the severity of the accidents themselves. Although previous studies have examined the spatial distribution characteristics of regions with higher accident severity, few have linked accident frequency with the spatial distribution characteristics of accident severity.

Therefore, in view of the shortcomings of previous studies, accidents were first spatially located by geocoding methods in this paper. Second, accidents were divided into four categories according to the season in which they occurred. Then, using density analysis and cluster analysis, areas with a high frequency and severity of traffic accidents were identified. Finally, the results of the two analysis methods were combined to determine accident-prone areas of different severities.

This paper aims to investigate the spatial and temporal patterns of accidents from two perspectives: accident frequency and accident severity. The remainder of the article is arranged as follows. Section 2 describes the traffic accident data. Section 3 shows the main methods used in this study. The results of density analysis and cluster analysis are presented in Section 4. Finally, Section 5 presents the conclusions.

2. Data Processing

This study focused on the city of Harbin, located in northeastern China. Harbin covers an area of approximately 53,100 km^2 and had a population of 9.55 million in 2017. This study focused on only the main urban area of the city. Each data point contains basic accident information, such as time of occurrence, location of occurrence, accident casualties, road type, and weather data. In this research, some of this information (shown in Table 1) was selected for the study.

In GIS, the location of a traffic accident is generally marked by latitude and longitude coordinates. However, latitude and longitude were not included in the raw data. Therefore, the longitude and latitude coordinates needed to be determined from the description of the traffic accident location. Then, the spatial location of the accident was

finalized. This process is called geocoding [26]. Tian et al. [27] evaluated the quality of four mainstream geocoding services in China (Baidu, Gaode, Sogou, and Tencent). The service quality of Tencent's geocoding API was considered relatively high, with higher data quality and more complete address data than the other services. Therefore, Tencent was selected to complete the conversion of the accident coordinates and then import the converted data into GIS. When the Tencent API returns the coded result, it also returns the reliability of the results. Reliability values range from 1 (low reliability) to 10 (high reliability). A result is considered credible when it has a reliability score of 7 and above, so these results were retained in this study. After the geocoding process, 5850 accidents were identified for further study.

Figure 1 illustrates the overall spatial distribution of traffic accidents from 2016 to 2018. The figure suggests that the accidents occurred mostly in the central urban area.

3. Methodology

3.1. Comap Method. The comap technique can help us to recognize the location of traffic accidents over time. It has been widely used in temporal-spatial integration [28, 29]. In this paper, traffic accident data from 3 years were divided into four subsets by season in accordance with Harbin's climatic conditions. Then, density analysis and cluster analysis were applied to calculate the intensity of each subset. Finally, the results were arranged sequentially in a graphic to show the spatial distribution of traffic accidents over time. According to the suggestion of related literature [30], class boundaries should overlap. As shown in Table 2, accidents were divided into four subsets by season. There is some overlap between subsets to avoid temporal boundaries.

3.2. Density Analysis Method. This study used point density and line density to identify spatial patterns of traffic accidents. The former is obtained by calculating the number of accidents per unit area. The latter is obtained by calculating the length of the section per unit area.
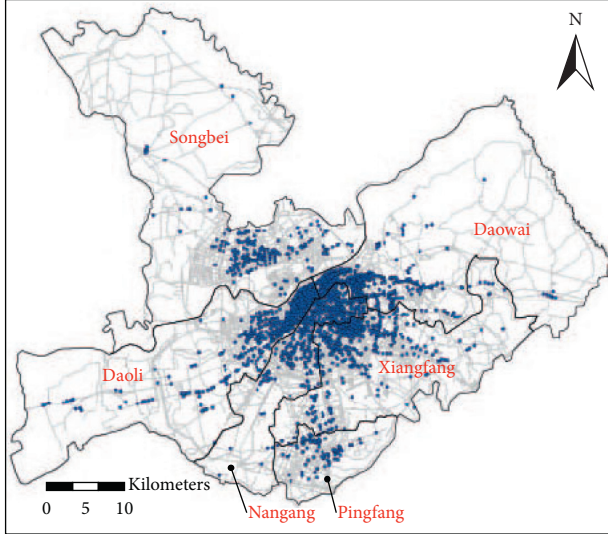
The calculation of the density analysis was performed in GIS with the neighborhood method. For example, the study area was divided into several small square cells with side length d when calculating the density of the accident points. Each cell ultimately corresponds to a pixel in the output map. The accident density in the region where cell k is located is D_k^{accident}, and the radius of the neighborhood is set to r. $N_k(r)$ is the number of accidents within a neighborhood centered at the center of cell k and with radius r. The point density is calculated as follows:

$$D_k^{\text{accident}} = \frac{N_k(r)}{\pi r^2}. \tag{1}$$

The road network density in the area where cell k is located is D_k^{road}. Similarly, $L_k(r)$ is the length of the road within the same neighborhood. The specific formula is as follows:

TABLE 1: List of information selected from raw traffic data.

Field name	Field description	Data format	Example
Accident time	The specific time of the accident	Float	2019-06-10 15:35
Administrative district	The district where the accident occurred	String	Daoli District
Accident location	Detailed description of the accident site	String	Intersection of Dongzhi Road and Hongqi Street
Accident evaluation	The final casualties caused by the accident	String	The accident caused 1 death, 2 injuries, and so on

FIGURE 1: Study zone and locations of all accidents in Harbin (2016–2018).

TABLE 2: Start and end dates of the four seasons.

Season	Spring	Summer	Fall	Winter
Start date	Mar 5	Jun 5	Sep 5	Nov 5
End date	Jun 10	Sep 10	Nov 10	Mar 10

$$D_k^{\text{road}} = \frac{L_k(r)}{\pi r^2}. \tag{2}$$

In addition, d and r need to be determined in the actual calculation. They are usually obtained in GIS from the minimum of the output image height and width, which are 1/30th and 1/250th of the minimum. The study area is between 126.15 and 127.15 east longitude and 46.09 and 45.52 north latitude. The minimum output image width was obtained after longitude and latitude were converted to the actual distance. A cell length of 230 m and a neighborhood radius of 1900 m were selected for the density analysis.

3.3. Cluster Analysis Method.
Cluster analysis is a more rigorous data analysis process. Spatial clustering analysis divides collections of physical or abstract objects in spatial data into similar classes. In turn, the spatial patterns of similar classes of data are obtained [31]. In this paper, outlier analysis was performed in GIS to explore the spatial pattern

of accident severity. First, this approach is in contrast to traditional cluster analysis approaches, such as hierarchical or divisional cluster-based methods. These methods can determine only whether a sample belongs to a certain category [32]. However, outlier analysis identifies samples that do not belong to any category. This provides a more comprehensive analysis of the spatial pattern of accidents. Second, the calculation of the outlier analysis is based on the attributes of the individual accident sample points. Original data attributes are largely preserved. This facilitates an in-depth study of the accidents.

Outlier analysis is performed by calculating the local Moran's I of an accident. It measures the correlation between the attributes of each incident point and the values of other neighboring points. Outlier analysis is calculated as follows:

$$I_i = \frac{x_i - \overline{X}}{S_i^2} \sum_{j=1, j\neq i}^{n} \omega_{i,j}\left(x_j - \overline{X}\right), \tag{3}$$

where I_i is the local Moran's I statistic of data point i, n is the total number of accidents, x_i and x_j are the attributes of data points i and j (which correspond to the accident severity in this paper), $\overline{X}$ is the global mean of the attribute, $\omega_{i,j}$ is the spatial weight between i and j, and S_i^2 is a second-order sample matrix of the attributes of data points.

Formally, S_i^2 can be expressed as

$$S_i^2 = \frac{\sum_{j=1, j\neq i}^{n}\left(x_j - \overline{X}\right)^2}{n-1}, \tag{4}$$

where the z_{I_i} score for the data point is calculated as

$$z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}}, \tag{5}$$

where $E[I_i]$ and $V[I_i]$ can be expressed as

$$E[I_i] = -\frac{\sum_{j=1, j\neq i}^{n} \omega_{ij}}{n-1},$$

$$V[I_i] = E[I_i^2] - E[I_i]^2. \tag{6}$$

There were five types of statistical results of the outlier analyses: high-high clustering (H-H), high-low clustering (H-L), low-high clustering (L-H), low-low clustering (L-L), and nonsignificant. In general, a 95% confidence level was used to signify statistical significance. In other words, a result was considered statistically significant when the p value was less than 0.05. The corresponding z-score should have ranged between -1.96 and $+1.96$ according to the normal distribution. If a result was statistically significant and $I > 0$,

then the data point had the same level of high or low attributes as the adjacent points. The attribute values of the point were compared with the average attribute values of all the data points to determine whether the results indicated H-H or L-L clustering. If $I < 0$, then the properties of the data point differ significantly from those of adjacent points, and the point is an outlier.

4. Analysis Results

4.1. Density Analysis

4.1.1. Distribution of Accident Point Density. The frequency of accidents per unit area is an indicator. This indicator is used to measure the level of traffic safety on urban roads. In this section, point density analysis was used to calculate this value. Moreover, the temporal and spatial distribution of traffic accidents was generated. This approach helps to determine whether an accident hot spot is subject to temporal fluctuations in the accidents. In addition, the maximum value normalization method was used to normalize the density values. This method facilitates the classification and comparison of accident density intervals.

According to relevant literature [28], the accident density results were classified into 3 levels. Density values of 0.5 and 0.8 were used as the two cutoff points. Density values between 0 and 0.5 indicate a low-density area of accidents, values between 0.5 and 0.8 indicate an intermediate density area of accidents, and values between 0.8 and 1 indicate a high-density area of accident. The density calculation results are shown in Figure 2.

As shown in Figure 2, the number of accidents was similar in spring and fall. The locations of intermediate to high densities of accidents were similar in these two seasons. In addition, the number of accidents was higher in summer than in the other seasons. The areas with intermediate to high densities of accidents were also more widely distributed in summer. The high-density accident area statistics are shown in Table 3. The high-density accident areas were found in the same administrative divisions in spring, summer, and autumn. In winter, these areas were located only in the Daoli District. The high-density area was the largest in summer, at approximately $5 \, \text{km}^2$. The accident rates in high-density areas were highest in the fall.

Statistics of areas with an intermediate density of accidents are shown in Table 4. The areas with high densities of accidents were found in the same administrative divisions in spring, summer, and autumn, and these areas were mainly located in the Daoli, Daowai, Xiangfang, and Nangang Districts. The areas with a high density of accidents in winter in addition to the other three seasons include the Pingfang District. The area with an intermediate density of accidents is largest in summer, at approximately $35.22 \, \text{km}^2$. The accident rates in areas with an intermediate density of accidents were highest in winter.

In conclusion, areas with intermediate to high densities of accidents were concentrated near large shopping malls, schools, and hospitals (yellow circles), especially in the vicinity of the First Hospital and Ha Station. These areas were

larger in the summer and winter and smaller in the spring and fall. In addition, the accident density in some areas (yellow circles) was found to have not changed significantly in space and time through the comap technique. However, the density in some areas fluctuated in space and time. For example, the green box in Figure 2(d) was not identified in any of the other seasons. Notably, the southern area that had a high density of accidents in the winter (highlighted by the box in Figure 2(b)) was not identified as a high-density area in any of the other seasons.

4.1.2. Distribution of Accident Point Density considering Road Network Density. The density of accident points per unit area was determined in the previous section. However, that analysis did not consider the density of the road network. The spatial-temporal pattern of accidents was not fully reflected. Therefore, a new spatial-temporal accident pattern was obtained by calculation. This calculation was performed by dividing the point density by the road network density ($D_k^{\text{accident}}/D_k^{\text{road}}$) to obtain the accident frequency per unit road length. The division of the density value is consistent with the above. The new pattern, which considers the road network density, is shown in Figure 3.

A significant change is shown in Figure 3. The spatial and temporal patterns of accidents differed significantly by season. In addition, new areas with intermediate to high densities of accidents were identified. The statistics of new high-density areas of accidents are shown in Table 5. A comparison of the accident density with the administrative divisions indicated that, in spring and fall, accidents occurred in the Daowai and Xiangfang Districts, while in summer and winter, they occurred in the Daoli District. A comparison of accident density and road networks indicated that, in spring and autumn, accidents were concentrated on Nanzhi Road, Pioneer Road, Gongbin Road, and Hongqi Street. In summer and winter, accidents were concentrated on Pioneer Road and South Straight Road. The area with a high density of accidents in autumn was the largest, at approximately $14.11 \, \text{km}^2$. Accident rates in the high-density areas were largest in spring.

The statistics of the updated intermediate density areas of accidents are shown in Table 6. A comparison with administrative divisions indicated that, in spring, autumn, and winter, an intermediate density of accidents occurred in the Daoli, Daowai, and Xiangfang Districts. In summer, the area with an intermediate density of accidents also contains Nangang District. A comparison with the road network indicated that the areas with an intermediate density of accidents in each season were mostly concentrated on Gongbin Road, Xinyang Road, Dongzhizhi Road, Nanzhizhi Road, Pioneer Road, and so on. In summer and autumn, these areas also included Tianheng Street and so on. In winter, they included Haping Road, Heping Road, Hexing Road, and so on. The area with an intermediate density of accidents was largest in summer, at approximately $41.39 \, \text{km}^2$. The accident rate in areas with an intermediate density of accidents was similar among seasons.
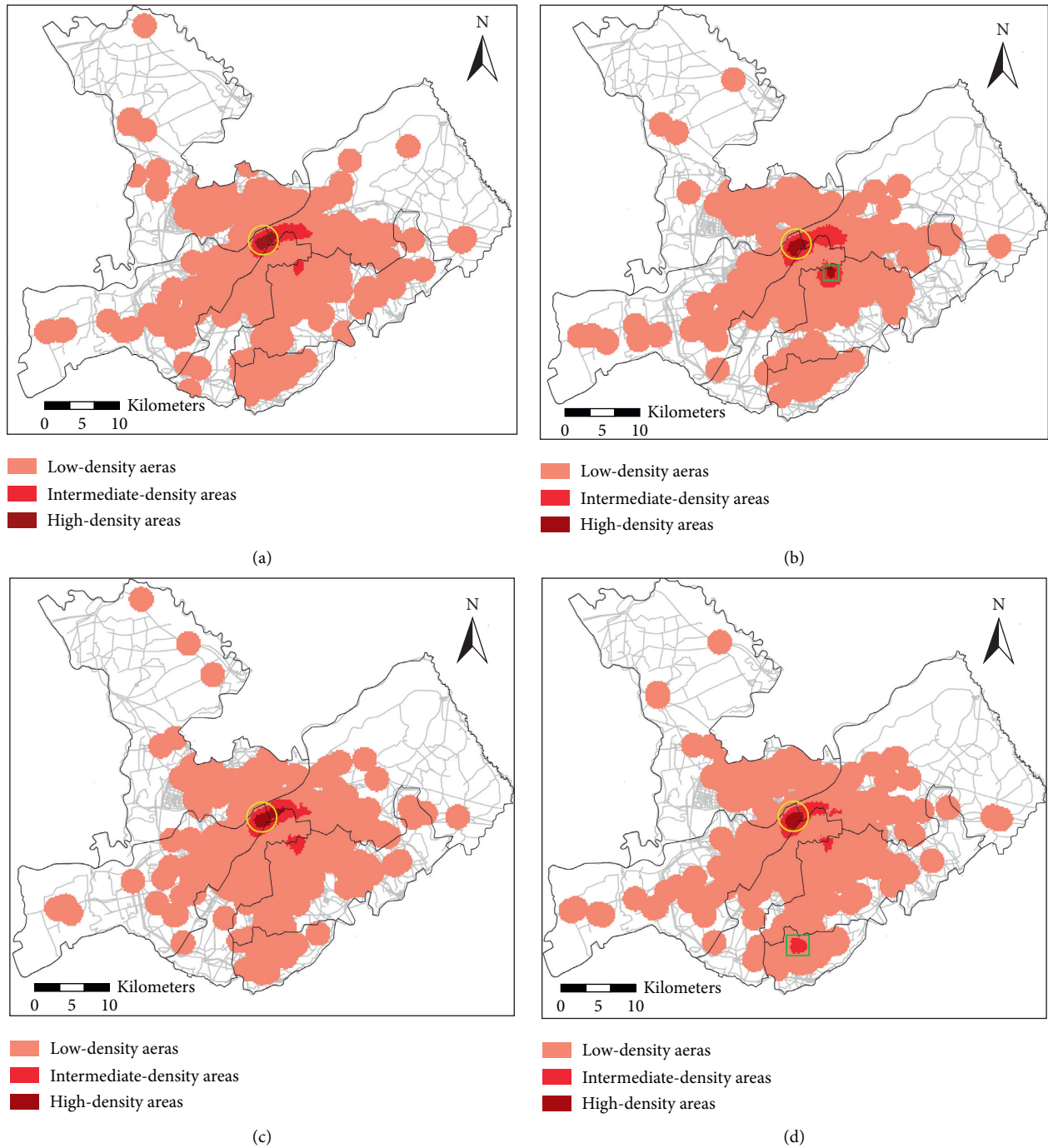
(a)

(b)

(c)

(d)

FIGURE 2: Accident point density distribution. (a) Spring. (b) Summer. (c) Fall. (d) Winter.

TABLE 3: Statistical description of areas with a high density of accidents.

Description	Spring	Summer	Fall	Winter
District	Daoli and Xiangfang	Daoli and Daowai	Daoli and Daowai	Daoli
Area (km^2)	1.10	6.02	2.3	1.8
Accident rate (%)	0.073	0.099	0.102	0.087

In summary, the size of areas with a high density of accidents in spring and winter was smaller, while that in summer and autumn were larger and more widely distributed. In addition, a comap was generated to demonstrate temporal-spatial patterns of accidents in each season. Some areas (marked by boxes 1, 2, and 3 in Figure 4)

TABLE 4: Description of the accidents in area with an intermediate density of accidents.

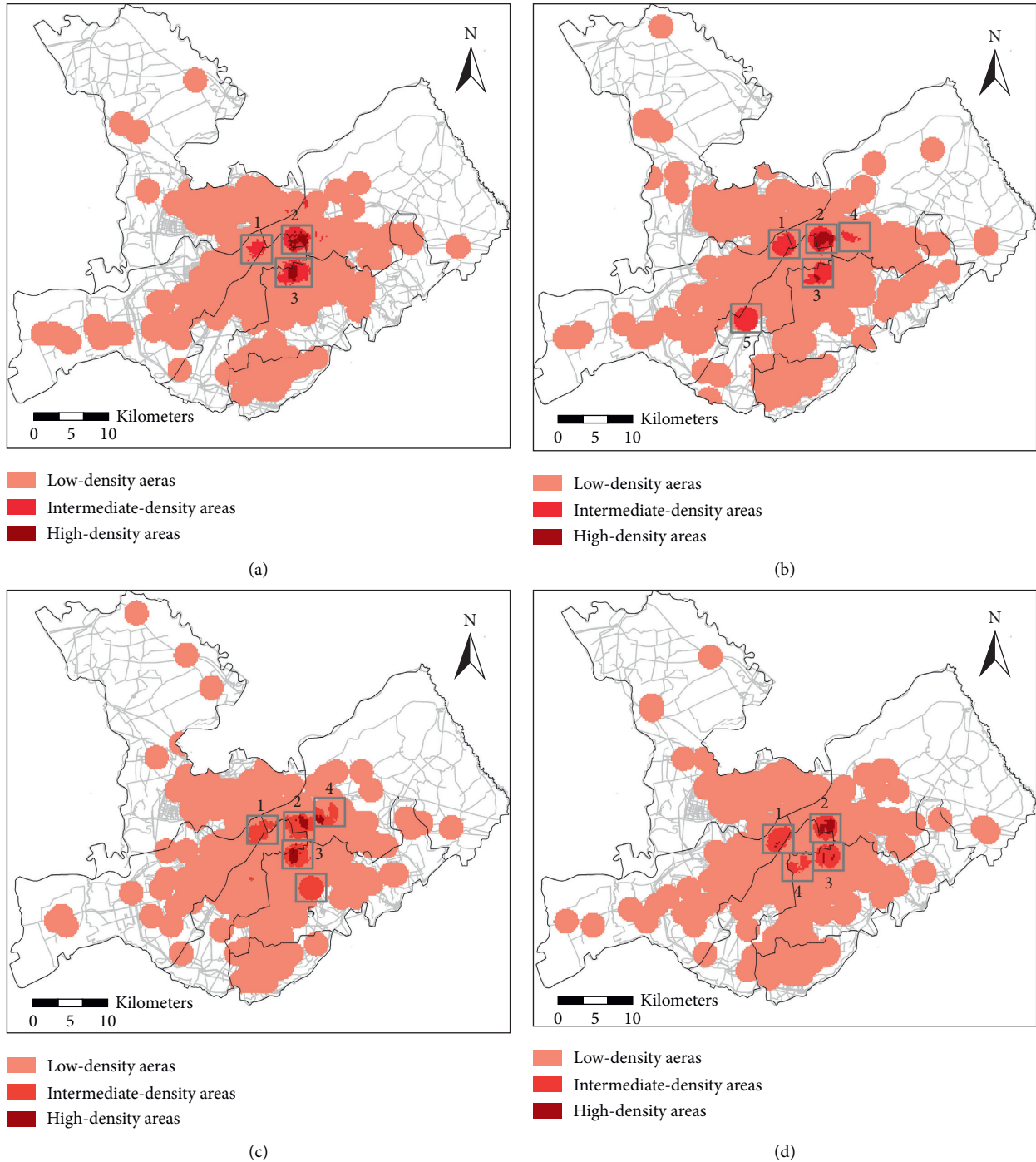| Description	Spring	Summer	Fall	Winter
District	Daoli, Daowai, Xiangfang, and Nangang	Daoli, Daowai, Xiangfang, and Nangang	Daoli, Daowai, Xiangfang, and Nangang	Daoli, Daowai, Xiangfang, Nangang, and Pingfang
Area (km^2)	25.03	35.22	27.16	30.64
Accident rate (%)	0.315	0.364	0.329	0.221

(a)

(b)

(c)

(d)

FIGURE 3: Density distribution of accident points combined with road networks. (a) Spring. (b) Summer. (c) Fall. (d) Winter.

TABLE 5: Statistics of areas with a high density of accidents after considering road network density.

Description	Spring	Summer	Fall	Winter
District	Daowai and Xiangfang	Daowai	Daowai and Xiangfang	Daowai
Area (km^2)	6.37	5.65	14.11	4.40
Accident rate (%)	0.075	0.041	0.050	0.029

TABLE 6: Statistics of areas with an intermediate density of accidents considering road network density.

Description	Spring	Summer	Fall	Winter
District	Daoli, Daowai, and Xiangfang	Daoli, Daowai, Nangang, and Xiangfang	Daoli, Daowai, and Xiangfang	Daoli, Daowai, and Xiangfang
Area (km^2)	25.13	41.39	39.18	34.39
Accident rate (%)	0.210	0.252	0.261	0.227

showed essentially no fluctuations over time. In some areas, there were changes in both spatial and temporal patterns. Examples of these areas include box 5 in Figures 4(b) and 4(c) and box 4 in Figure 4(d). These areas have a low density of accidents in all the other seasons.

Compared with the results shown in Figure 2, both analyses demonstrated relatively similar areas of intermediate and high densities of accidents, although significant changes were observed in some areas. Some findings were obtained after considering the road network density. First, the color of some areas was lighter in Figure 3 than in Figure 2. For example, the areas in the yellow circle in Figure 2 and the green box in Figure 2(d) became areas with low or intermediate densities of accidents. This indicates that the density results in these areas were due to an overly dense road network. Second, new areas with intermediate to high densities of accidents were identified in each season. The frequency of accidents per unit road length is higher in boxes 4 and 5 in Figure 3(b) and box 4 in Figure 3(b). Finally, there were no significant changes in the area (box 3 in Figure 3(b)). Whether considering the frequency of accidents per unit area or per unit road length, intermediate-to high-density areas of accidents were identified in this area.

4.2. Cluster Analysis.

The previous section presents the results obtained when the frequency of accidents was analyzed without considering the severity of accidents. Traffic managers measure the severity of accidents considering not only the frequency of accidents but also the property damage and casualties caused by the accidents. In fact, if an area occasionally has a particularly serious accident, it deserves more attention than areas where minor accidents occur frequently. Therefore, according to the relevant literature and the data, the accident severity was classified into 3 levels in this section (as shown in Table 7).

There are certain patterns in the distribution of traffic accidents of differing severity in time and space. In this analysis, accident severity was used as a factor to evaluate the results of the spatial clustering of accidents. The clustering results are shown in Figure 4.

The position of the box in Figure 4 corresponds to that in Figure 3. The dark red points (H-H) represent the high-severity accident class. The dark blue points (L-L) represent the low-severity accident class. The light red points (H-L) indicate a few high-severity accident points contained within the spatial extent occupied by many low-severity accident points. The light blue points (L-H) indicate a few low-severity accident points within the spatial area occupied by many high-severity accident points. The gray points indicate that the incident points had no obvious clustering features.

As shown in Figure 4, different types of clustering features were identified in each season. Some areas of Pingfang District showed a tendency for high-severity accidents. Likewise, this pattern was also observed in the southwestern part of the study area in spring. Most of these areas were concentrated in peripheral and suburban areas away from urban centers. In contrast, most accidents are dispersed within urban centers. To a certain degree, traffic accidents can cause casualties and property damage, although most accidents have a low severity. In addition, the traffic accidents in the northeastern area of Daoli District in winter were in the L-L class, indicating that this area was a hot spot for low-severity accidents.

4.3. Combined Density Analysis and Clustering Analysis.

This approach allows us to clearly understand the results of combined clustering and density analyses. By combining these analyses, accident severity in areas with intermediate and high densities of accidents was identified. More importantly, these areas had a ranking that should be noted. Traffic managers could thus target individual areas for better management and regulation. For example, in spring, there were three areas with intermediate to high densities of accidents (surrounded by boxes in Figure 4). For example, M_m is the number of accidents in Region 1 after the cluster analysis, with $m = 1, 2, 3, 4, 5$, representing accidents in the H-H, H-L, L-H, and L-L classes and nonsignificant accidents, respectively. The proportion P_m, which indicates the likelihood that this region will eventually exhibit a certain clustering feature, is calculated as
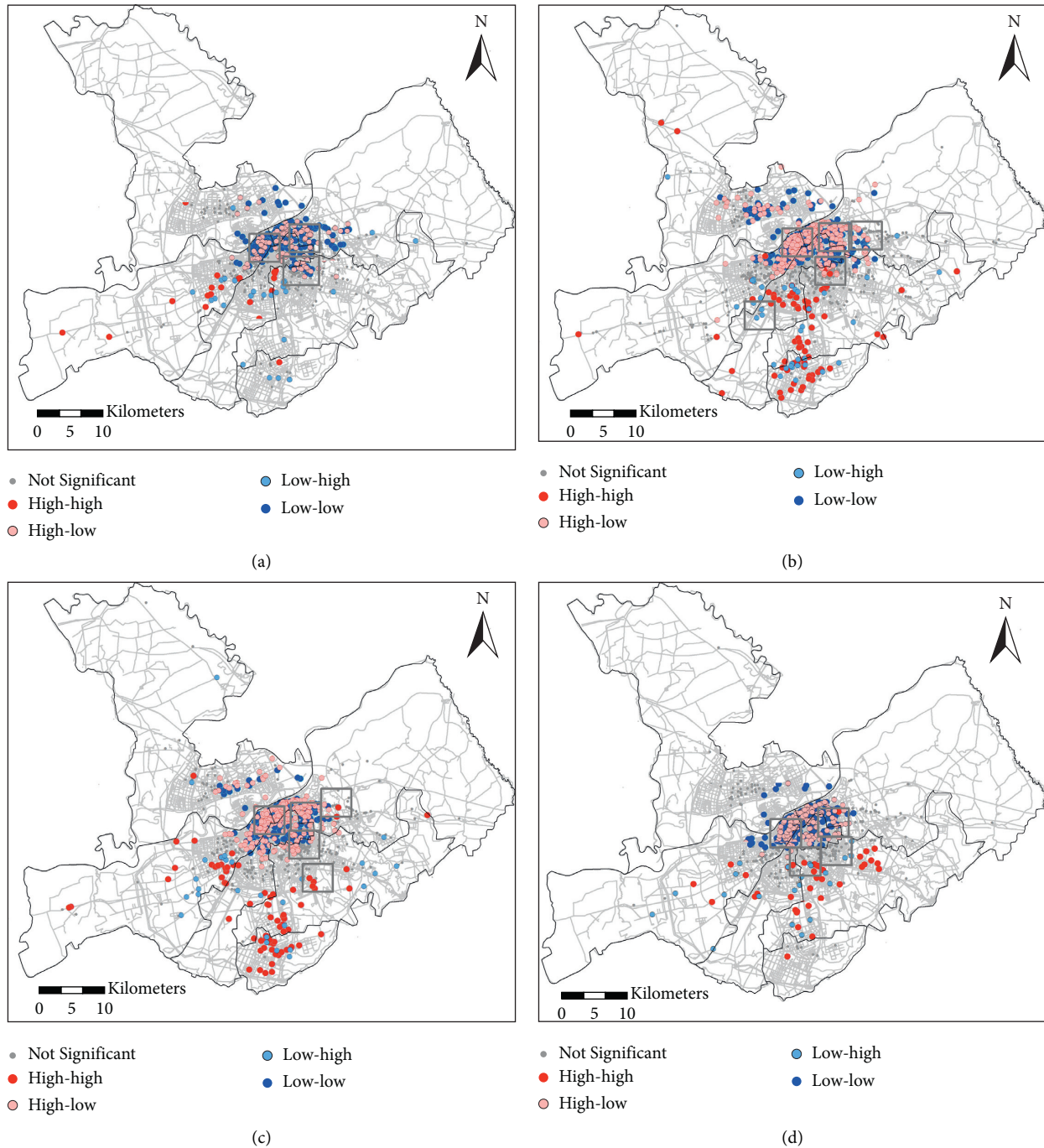
(a)

(b)

(c)

(d)

FIGURE 4: Results of the analysis of accident severity anomalies. (a) Spring. (b) Summer. (c) Fall. (d) Winter.

TABLE 7: Accident severity level descriptive statistics.

Grade	Description	Frequency	Proportion (%)
1	Property damage	2965	50.68%
2	Injury	2302	39.35%
3	Fatality	583	9.97%

$$P_m = \frac{M_m}{\sum_{m=1}^{5} M_m}. \qquad (7)$$

Similarly, the proportion of clustering results for each season was calculated using equation (7). The calculation results are shown in Table 8. The clustering results for some regions were nonsignificant and are not presented in this table. As shown in Table 8, accidents in the L-L class account for a relatively large proportion of accidents in Region 1 (Figure 4(a)), in Regions 1 and 4 (Figure 4(b)), and in Regions 1 and 2 (Figure 4(d)). This result indicated that although the frequency of accidents per unit length of roadway is greater in these areas, the severity of accidents is

TABLE 8: Spatial clustering statistics for areas with intermediate and high densities of accidents.

Season	Region	Percentage of clustering results (%)				
		HH	HL	LH	LL	Not significant
Spring	Region 1	—	3.85	—	80.77	15.38
	Region 1	—	26.53	—	56.12	17.35
Summer	Region 4	0.82	—	—	83.33	15.85
	Region 5	—	—	76.92	—	23.08
	Region 1	—	52.53	—	30.74	17.78
Autumn	Region 2	—	64.00	—	11.00	17.00
	Region 5	68.33	5.36	—	—	26.31
Winter	Region 1	—	22.03	—	63.63	14.34
	Region 2	—	27.86	—	69.29	12.86

generally lower. Conversely, accidents in the G-G class account for a relatively large proportion of accidents in Region 5 (Figure 4(c)). This result indicated that the frequency of accidents per unit road length is higher in this area and the severity of accidents is higher.

5. Conclusion

(1) In this paper, density analysis was combined with comap technology to study the spatial and temporal patterns of traffic accidents by season from the perspective of accident frequency. The results show that the accident density distribution is more diffuse in summer and winter when the road network density is not considered. After considering road network density, the accident density distribution is more diffuse in summer and autumn.

(2) Accident severity is divided into three levels: accidents that cause property damage, accidents that lead to injury, and fatal accidents. Based on these three levels, cluster analysis was used to explore the spatial and temporal patterns of accidents. The following conclusions were drawn. Traffic accidents in urban centers mostly show characteristics of the H-L and L-L classes. Traffic accidents in the central part of the city are generally low-severity accidents. Conversely, the results show that accidents with characteristics of the H-H class were mainly found in the outer urban areas and suburban areas. This indicates that most traffic accidents in these areas are high-severity accidents.

(3) Density analysis is a regional analysis. This method reflects a coarse picture of the spatial-temporal pattern of accidents. Clustering analysis can be accurate at the level of the accident points. Urban areas prone to accidents of different severities were identified by combining two methods. The results show that the accident severity was lower in Region 1 in spring, summer, and winter, although the frequency of accidents was higher. In Region 5 in the fall, not only was the frequency of accidents greater but also the severity of accidents was generally higher.

(4) Due to data limitations, further analysis of causal factors of accidents leading to spatial patterns is needed. The next step will be to conduct an in-depth study of the causes of various types of traffic accidents, taking into account road characteristics and infrastructure, with a view to providing a more detailed basis for traffic safety management.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Meina Wang and Wenhui Zhang contributed to conceptualization, methodology, and original draft preparation; Meina Wang contributed to software; Jing Yi and Tiangang Qiang contributed to validation; Wenhui Zhang and Meina Wang contributed to formal analysis; Meina Wang, Xirui Chen, and Jing Yi contributed data curation; Xirui Chen, Meina Wang, and Tiangang Qiang contributed to review and editing. All the authors have read and agreed to the published version of the manuscript.

Acknowledgments

References

[1] World Health Organization, *Global Status Report on Road Safety*, World Health Organization, Geneva, Switzerland, 2018.

[2] National Bureau of Statistics of the People's Republic of China, *Traffic Accident Statistics*, National Bureau of Statistics of the People's Republic of China, Beijing, China, 2016.

[3] C. Liu and A. Sharma, "Exploring spatio-temporal effects in traffic crash trend analysis," *Analytic Methods in Accident Research*, vol. 16, pp. 104–116, 2017.

[4] L. A. Ahmed, "Using logistic regression in determining the effective variables in traffic accidents," *Applied Mathematical Sciences*, vol. 11, no. 42, pp. 2047–2058, 2017.

[5] I. Bargegol, M. Najafi, and M. Abolfazlzadeh, "Statistical analysis of the railway accidents causes in Iran," *International Journal of Engineering*, vol. 30, no. 12, pp. 1822–1830, 2017.

[6] H. Sun, Q. Wang, P. Zhang et al., "Spatialtemporal characteristics of tunnel traffic accidents in China from 2001 to present," *Advances in Civil Engineering*, vol. 2019, Article ID 4536414, 12 pages, 2019.

[7] Y. Li, M. Abdel-Aty, J. Yuan, Z. Cheng, and J. Lu, "Analyzing traffic violation behavior at urban intersections: a spatio-temporal kernel density estimation approach using automated enforcement system data," *Accident Analysis & Prevention*, vol. 141105509 pages, 2020.

[8] N. Haji Mirza Aghasi, "Introducing GIS as legitimate instrument to deal with road accident data: a case study of Iran, Tehran," *Spatial information research*, vol. 25, no. 1, pp. 151–159, 2017.

[9] X. Zhu, H. Cong, Y. Zhi et al., "Accident-prone location analysis based on GIS spatial clustering," *Urban Transport of China*, vol. 16, no. 3, pp. 21–27+86, 2018.

[10] F. Jiang, K. K. R. Yuen, and E. W. M. Lee, "Analysis of motorcycle accidents using association rule mining-based framework with parameter optimization and GIS technology," *Journal of Safety Research*, vol. 75, pp. 292–309, 2020.

[11] L. Hu, X. Wu, J. Huang, Y. Peng, and W. Liu, "Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China," *Safety Science*, vol. 127104710 pages, 2020.

[12] M. Özcan and M. Küçükönder, "Investigation of spatio-temporal changes in the incidence of traffic accidents in kahramanmaraş, Turkey, using GIS-based density analysis," *Journal of the Indian Society of Remote Sensing*, vol. 48, no. 7, pp. 1045–1056, 2020.

[13] Y. Wang and L. Wang, "An identification method of traffic accident black point based on street-network spatial-temporal kernel density estimation," *Scientia Geographica Sinica*, vol. 39, no. 8, pp. 1238–1245, 2019.

[14] Á. Briz-Redón, F. Martínez-Ruiz, and F. Montes, "Spatial analysis of traffic accidents near and between road intersections in a directed linear network," *Accident Analysis & Prevention*, vol. 132105252 pages, 2019.

[15] Y. Kang, N. Cho, and S. Son, "Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density estimation," *PLoS One*, vol. 13, no. 5, Article ID e0196845, 2018.

[16] H. Harirforoush and L. Bellalite, "A new integrated GIS-based analysis to detect hotspots: a case study of the city of Sherbrooke," *Accident Analysis & Prevention*, vol. 130, pp. 62–74, 2019.

[17] X. Chen, L. Huang, D. Dai, M. Zhu, and K. Jin, "Hotspots of road traffic crashes in a redeveloping area of Shanghai," *International Journal of Injury Control and Safety Promotion*, vol. 25, no. 3, pp. 293–302, 2018.

[18] G. A. Shafabakhsh, A. Famili, and M. S. Bahadori, "GIS-based spatial analysis of urban traffic accidents: case study in Mashhad, Iran," *Journal of Traffic and Transportation Engineering*, vol. 4, no. 3, pp. 290–299, 2017.

[19] A. K. Al-Aamri, G. Hornby, L.-C. Zhang, A. A. Al-Maniri, and S. S. Padmadas, "Mapping road traffic crash hotspots using GIS-based methods: a case study of Muscat Governorate in the Sultanate of Oman," *Spatial Statistics*, vol. 42100458 pages, 2021.

[20] F. Ouni and M. Belloumi, "Pattern of road traffic crash hot zones versus probable hot zones in Tunisia: a geospatial analysis," *Accident Analysis & Prevention*, vol. 128, pp. 185–196, 2019.

[21] M. B. Ulak, E. E. Ozguven, L. Spainhour, and O. A. Vanli, "Spatial investigation of aging-involved crashes: a GIS-based case study in Northwest Florida," *Journal of Transport Geography*, vol. 58, pp. 71–91, 2017.

[22] H. E. Colak, T. Memisoglu, Y. S. Erbas et al., "Hot spot analysis based on network spatial weights to determine spatial statistics of traffic accidents in Rize, Turkey," *Arabian Journal of Geosciences*, vol. 11, no. 7, pp. 1–11, 2018.

[23] A. Soltani and S. Askari, "Exploring spatial autocorrelation of traffic crashes based on severity," *Injury*, vol. 48, no. 3, pp. 637–647, 2017.

[24] E. Vaz, S. Tehranchi, and M. Cusimano, "Spatial assessment of road traffic injuries in the greater toronto area (GTA): spatial analysis framework," *Journal of Spatial and Organizational Dynamics*, vol. 5, no. 1, pp. 37–55, 2017.

[25] S. S. Radiah Shariff, H. Abdul Maad, N. Narsuha Abdul Halim, and Z. Derasit, "Determining hotspots of road accidents using spatial analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 1, pp. 146–151, 2018.

[26] B. Wilson, N. Wilson, and S. Martin, "Using GIS to advance social economics research: geocoding, aggregation, and spatial thinking," *Forum for Social Economics*, vol. 2019, pp. 1–25, 2019.

[27] Q. Tian, Y. Gong, Y. Gong et al., "A comparative evaluation of online geocoding services in China," *Geomatics and Information Science of Wuhan University*, vol. 41, no. 10, pp. 1351–1358, 2016.

[28] K. G. Le, P. Liu, and L.-T. Lin, "Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam," *Geo-Spatial Information Science*, vol. 23, no. 2, pp. 153–164, 2020.

[29] A. Asgary, A. Ghaffari, and J. Levy, "Spatial and temporal analyses of structural fire incidents and their causes: a case of Toronto, Canada," *Fire Safety Journal*, vol. 45, no. 1, pp. 44–57, 2010.

[30] C. Plug, J. Xia, and C. Caulfield, "Spatial and temporal visualisation techniques for crash analysis," *Accident Analysis & Prevention*, vol. 43, no. 6, pp. 1937–1946, 2011.

[31] R. Jia, A. Khadka, and I. Kim, "Traffic crash analysis with point-of-interest spatial clustering," *Accident Analysis & Prevention*, vol. 121, pp. 223–230, 2018.

[32] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *D-J Series*, vol. 2, no. 2, pp. 226–235, 2019.