

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2026

Volume 42, Pages 162-175

ICBASET 2026: International Conference on Basic Sciences, Engineering and Technology

DBSCAN-Based Spatial Hotspot Detection of Motorcycle and Three-Wheeled Vehicle Crashes

Ghaidaa A. Majeed
Mustansiriyah University

Abeer K. Jameel
Mustansiriyah University

Abstract: Because of the lack of physical protection and the often-complex interaction between motorcycles and 3-wheelers and the other components of urban transportation systems, motorcycles and 3-wheelers are among the most vulnerable road users. Therefore, it is important to determine spatial crash hotspots to enable evidence-based road safety management and targeted interventions. The purpose of this study is to detect crash hotspot areas of motorcycles and three-wheel vehicles for the Rusafa district in Baghdad using the density-based spatial clustering of applications with noise (DBSCAN) algorithm and the Haversine distance measure. There were 802 crash records that were georeferenced and analyzed to identify areas that have been more prone to crashes. We used a Ball Tree nearest-neighbor search algorithm to implement DBSCAN and preprocessed the geographic coordinates before we applied DBSCAN, converting them to radians. We did a sensitivity analysis and selected the optimal clustering parameters ($\text{eps} = 0.5 \text{ km}$, $\text{min_samples} = 5$). Three spatial indicators (Zone_Label, Zone_Count, Hotspot_Flag) were then calculated to describe the identified hotspot regions. The framework recommended 37 spatial clusters, with 656 (81.8%) of these crashes being identified as a hotspot and 146 crashes (18.2%) as spatial noise. The greatest number of crashes recorded occurred in the largest hotspot (Zone 8) with 116 crashes, or 14.46% of all crashes recorded. Results also indicate that the most common crash locations are not necessarily the most severe, suggesting that, in addition to severity, crash frequency should be taken into account when assessing road safety. The framework proposed here may assist transportation agencies in determining persistent crash locations, prioritizing engineering and enforcement measures, and combining with machine learning methods to forecast crash severity.

Keywords: DBSCAN, Spatial hotspot detection, Haversine distance, GIS, Road safety

Introduction

Motorcycles and three-wheelers are indispensable elements of urban transport networks, especially in developing countries, where they offer low-cost, flexible, and efficient transportation. They are maneuverable and can be used for both passenger and commercial traffic in busy streets. Unlike enclosed vehicles, however, these modes provide little physical protection to riders and passengers and put them at significant risk of serious injury or death in the event of a traffic accident. Motorcycle and three-wheeled vehicle crashes have thus emerged as a significant public health and transportation safety problem, particularly in rapidly urbanizing cities with mixed traffic, high traffic volumes, and complex roadway environments (WHO, 2023).

Traffic accidents do not happen uniformly throughout the urban road network. Rather, they usually focus in one or more geographic areas where roadway geometry, traffic exposure, land use, operational characteristics, and human behavior combine to raise the risk of crashes. These high-risk locations, also known as crash hotspots, are key to evidence-based road safety management. Hotspot identification helps the transportation agency prioritize engineering improvements, optimize enforcement efforts, allocate limited safety resources more

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

© 2026 Published by ISRES Publishing: www.isres.org

effectively, and target limited education resources to reduce the number and severity of crashes (National Research Council, 2010; Shokat & Jameel, 2025).

The typical approaches to identifying hot spots are based on crash frequency rankings, administrative boundaries, roadway segments, or statistical aggregation techniques. While these methods offer valuable insights, crash patterns seldom align with administrative or roadway boundaries, making them less useful for capturing localized spatial patterns. In addition, standard clustering algorithms may fail to capture crash patterns around complex intersections, commercial areas, or urban activity centers, where crash clusters may form in irregular shapes (Li & Wang, 2022; Jameel & Jameel).

New developments in Geographic Information Systems (GIS), spatial data analytics, and machine learning have enabled more sophisticated strategies for crash hotspot detection. The growing popularity of this type of clustering algorithm is due to its ability to detect spatial clusters in the data that are not created by aggregating data in arbitrary spatial units. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm has proven to be one of the most popular methods for traffic safety analysis. One advantage of DBSCAN over standard clustering methods is that it automatically determines the number of clusters (no prior knowledge of it), automatically finds clusters of any shape (no requirement for them to be convex), and classifies isolated data points as noise. The advantages of DBSCAN render it particularly suitable for traffic crash analysis, as crash locations often are not arranged in a compact geometric pattern, but in an irregular road corridor (Ali & Alhusaynat, 2026; Mandala et al., 2026).

Choosing a distance measure is also crucial if geographic coordinates are used to analyze relationships. Latitude and longitude are not Cartesian coordinates, but rather angular measurements on the surface of the Earth. For that reason, spherical distance measures must be used in place of Euclidean distance measures. The Haversine distance is a perfect choice because it provides the shortest distance between two points on Earth's surface while accounting for its curvature. The Haversine metric is combined with DBSCAN to enhance the reliability of hotspot detection by augmenting the neighborhood relationships between crash locations with a geographical sense (Sharmila & Sabarish, 2021; Pusuluri, & Dangeti, 2024).

Although density-based clustering has been increasingly applied in transportation safety studies, several challenges remain. The performance of DBSCAN is strongly affected by both the neighborhood radius (eps) and the minimum number of samples (min_samples) required to form a cluster. These parameters required to form a cluster strongly affect DBSCAN's performance. Furthermore, most previous studies have examined all traffic crashes, with relatively few specifically examining motorcycles and three-wheeled vehicle crashes using coordinate-level spatial analysis. Further, there is little research that investigates the sensitivity of DBSCAN parameters or that converts clustering results into more practical hotspot indicators to inform further crash severity modeling and transportation planning (Lai et al., 2019; Zhang et al., 2024).

With urbanization and rising traffic volumes, Iraq's capital, Baghdad, has seen an increase in the number of motorcycles and 3-wheelers. The city is home to several signalized intersections, commercial corridors, mixed land-use zones, and freewheeling road networks in district Rusafa, one of the city's busiest urban areas, which contributes to high crash rates. While enhancing road safety in this context is important, there is a lack of in-depth coordinate-based hotspot analyses of vulnerable road users.

In view of this, the aim of this study is to use the DBSCAN algorithm, with Haversine distance as the metric, to detect crash hotspots for motorcycles and three-wheeled vehicles in the city of Rusafa in Baghdad. A set of 802 georeferenced crash records was analyzed to identify dense crash clusters and isolated crash locations, and to assess the impact of varying neighborhood radii through a sensitivity analysis. To quantify the detected clusters and support integration with crash severity prediction and transportation safety assessment models in the future, three spatial indicators were developed: Zone_Label, Zone_Count, and Hotspot_Flag.

This study helps in the following four ways:

- It introduces a hotspot detection system based on a coordinate-based approach and specifically tailored to motorcycle and three-wheeled vehicle crashes, based on the DBSCAN cluster algorithm and Haversine distance measure.
- It tests the sensitivity of DBSCAN neighborhood parameters to determine the balance between the accuracy of the hotspot detection and spatial resolution.
- It develops spatial hotspot indicators that can be integrated into future machine learning and crash severity prediction models.

- It offers a first-hand account of how transportation planners and road safety agencies can identify high-risk areas in urban settings in Baghdad and prioritize interventions that will enhance safety.

The results provide valuable insights into the evolving field of GIS-based traffic safety analysis and highlight the potential of density-based spatial clustering for data-driven decision-making on enhancing road safety for vulnerable road users in urban areas.

Methods

Research Framework

The methodology used in this study was density-based spatial clustering (DBSCAN), which was used to detect crash hotspots for motorcycles and 3-wheelers in Baghdad. The proposed framework is a sequential six-step process of crash data collection and preprocessing, Geographic coordinate preparation, Spatial distance computation (with the Haversine metric), Hotspot detection (with the DBSCAN algorithm), Sensitivity analysis for parameter selection, spatial visualization, and descriptive analysis of the identified hotspots. The overall research framework is presented in Figure 1.

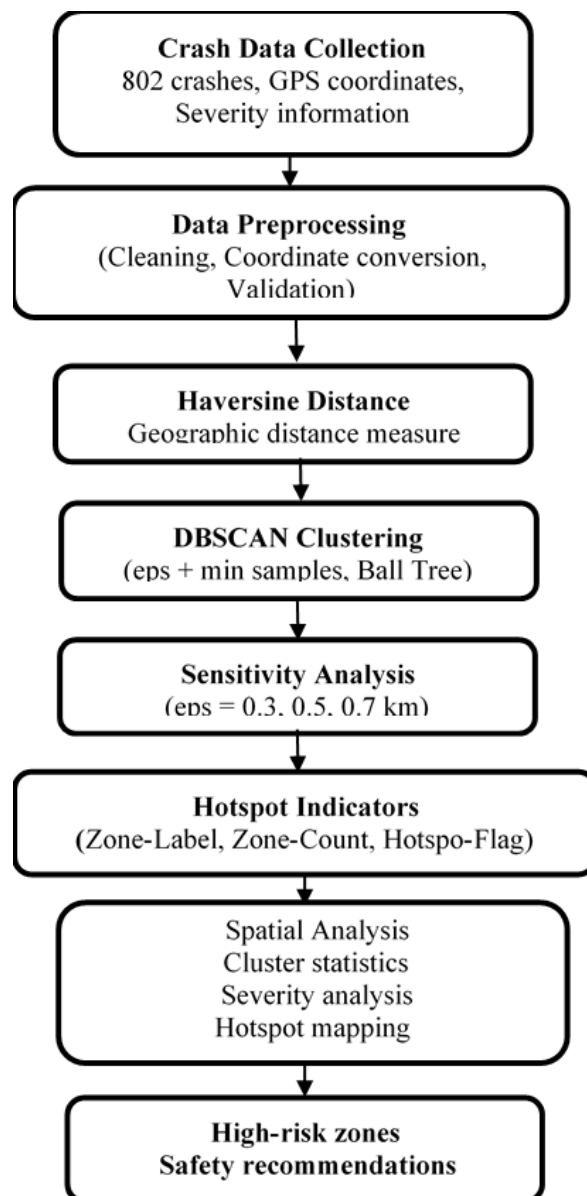


Figure 1. Proposed research framework for DBSCAN-based spatial hotspot detection of motorcycle and three-wheeled vehicle crashes.

Study Area and Crash Database

The study was conducted in Baghdad, Rusafa, one of the most densely populated urban areas with a dense network of transportation corridors, including arterial roads, signalized intersections, commercial corridors, residential neighborhoods, and mixed land-use developments. The district has a high number of motorcyclists and high traffic demand, which makes it ideal for studying spatial patterns of crash concentrations.

The 802 motorcycle and three-wheeled vehicle crashes in the crash database were officially recorded traffic crashes with geographic information. Geographic coordinates (latitude and longitude) were included with the severity information, which was classified into 3 levels for each crash record: Fatal crashes, Injury crashes, and Property damage only (PDO) crashes. Crash severity was used to describe the characteristics of the detected hotspot zones but did not affect the clustering process, whereas geographic variables were used as the main variables for detecting hotspot zones.

Data Preparation

All crash data were first extensively processed to enhance data quality and facilitate accurate spatial analysis prior to clustering. The preprocessing consisted of the following steps:

- Geographic information system (GIS) data; and
- Removing duplicated observations;
- Correcting incomplete or incorrect geographic coordinates;
- Converting Degrees–Minutes–Seconds (DMS) to decimal degrees;
- Converting degrees to radians for Haversine distance calculation.

Angles recorded in DMS were converted into decimal degrees (Eleiche, 2010):

$$DD = D + M / 60 + S / 3600 \quad (1)$$

where:

DD is the coordinate in decimal degrees, D = degree, M = minute, S = second.

Since Scikit-learn uses angular coordinates to implement the haversine metric, all latitudes and longitudes were converted from decimal degrees to radians prior to clustering.

Geographic Distance Measurement

Geographic coordinates are used to determine points on the Earth's surface, which is curved, unlike planar coordinate systems (Mahmoud & Akkari, 2016). As such, the Euclidean distance is not suitable for measuring the real distance between crash sites. Therefore, this study used the Haversine distance: the shortest great-circle distance between two points (A and B) on the Earth's surface. The Haversine metric ensures neighborhood relationships reflect real geographic distances, rather than planar approximations.

The Haversine distance is the distance between two points on the globe.

$$d = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \quad (2)$$

Where R is the Earth's radius (6371.0088 km), d is the geographic distance between two crash locations, ϕ_1 and ϕ_2 are the latitudes of the two points in radians, $\Delta\phi$ is the difference between the two latitudes, and $\Delta\lambda$ is the difference between the two longitudes.

DBSCAN Spatial Clustering

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was employed with the Python library Scikit-learn to identify crash hotspots. DBSCAN is different from partition-based clustering methods like K-means in that it doesn't rely on a fixed number of clusters. Unlike partition-based clustering

algorithms like K-means, DBSCAN doesn't assume a fixed number of clusters (Zhang et al., 2018; Saputra & Sihabudin, 2023; Kamh et al., 2024). The characteristics of DBSCAN are particularly interesting for traffic crash analysis because hotspot areas often exhibit irregular spatial patterns along road corridors, at intersections, and in urban activity centers.

DBSCAN divides observations into 3 categories:

- Core points are those that have a minimum of min-samples crashes within a search radius.
- Border points that are in the neighborhood of a core point but have fewer observations in the neighborhood.
- Noise points are those that don't meet the density requirement and aren't clustered together.

There are two parameters that control the algorithm:

- eps: radius of the maximum neighborhood; The term “eps” refers to the maximum distance between points in the clusters in DBSCAN. eps must be in radians and not kilometers when using Haversine as the distance metric (Mahmoud & Akkari, 2016; Prasetya et al., 2020; Sharmila & Sabarish, 2021).
- min_samples: minimum number of crashes in neighbors to form a dense area.

Based on preliminary experiments and sensitivity analysis, the following values were selected: eps = 0.5 km and min_samples = 5. The neighborhood radius was converted to radians for calculation by scikit-learn, which uses radians to compute the Haversine distance. In neighborhood search, a Ball Tree nearest-neighbor search structure was used to speed up the computation.

Sensitivity Analysis

The neighborhood radius parameter is sensitive to DBSCAN's performance; hence, a sensitivity analysis was conducted to determine the value of EPS. The term “eps” refers to the maximum distance between points in the clusters in DBSCAN. eps must be in radians and not kilometers when using Haversine as the distance metric (Mahmoud & Akkari, 2016; Prasetya et al., 2020; Sharmila & Sabarish, 2021).

The following performance indicators were used to assess each radius:

- The number of clusters identified;
- Clustered crash records;
- Noise observations;
- Percent of crashes that are clustered;
- Spatial coverage;
- Maintaining local spatial resolution.

The chosen parameter should be maximized to detect hotspots and prevent over-segmentation or merging of hotspot areas.

Spatial Hotspot Indicators

Three spatial indicators were derived for each crash record after the clustering process.

- Zone_Label: The identifier of the DBSCAN cluster to which a crash was linked was used as the identifier for that crash. Noise observations received Zone_Label = -1

Zone labels are simply categories and do not specify the intensity of the hotspot.

- Zone_Count: It is the total number of crashes in each cluster zone. This variable represents the magnitude of hotspots and can be used to compare hotspot regions.
- Hotspot_Flag: A binary variable was created to differentiate between clustered crashes and isolated crashes:

$$\text{Hotspot_Flag} = \begin{cases} 1, & \text{cluster member} \\ 0, & \text{noise point} \end{cases} \quad (3)$$

This indicator aids in future statistical analyses and may be included in future crash prediction or machine learning models.

Descriptive Spatial Analysis

Characteristics of the identified clusters were quantified with descriptive statistics.

The analysis included: The sum of the number of clusters; The ratio of hotspot to noise proportions;

The number of crashes for each hotspot; largest hotspot; Average hotspot size; and

Distribution of crash severity by hotspot areas.

The average cluster size was computed as:

$$\text{Mean Cluster Size} = \frac{Nc}{K} \quad (4)$$

Where Nc is the total number of clustered crash points, and K is the total number of detected clusters. Noise points were excluded from the calculation of mean cluster size.

Spatial visualization was performed using Python GIS libraries, with OpenStreetMap as the base map. The visualization consisted of: a crash severity map; a hotspot map and a non-hotspot map; the spatial distribution of clusters from the DBSCAN algorithm; hotspot frequency distribution, the composition of crash severity by hotspot; and a detailed visualization of the largest hotspot. These visualizations enable intuitive interpretation of spatial clustering results and help transportation agencies determine locations for priority safety interventions.

Results and Discussion

Sensitivity Analysis of DBSCAN Parameters

The density-based clustering algorithm also relies heavily on the choice of neighborhood radius (eps), as it affects hotspot identification results. Three neighborhood radii (0.3, 0.5, and 0.7 km) were studied for each neighborhood size of 5 crash points to assess the robustness of the clustering results. The clustering performance for the three values that were considered is summarized in Table 1.

Table 1. Sensitivity analysis of the DBSCAN neighborhood parameter eps for hotspot detection

eps (km)	eps (radians)	Hot spot points	Total Points	Percentage (%)	Note
0.3	0.00004708	452	802	57 %	Smaller radius; detects fewer hotspot points
0.5	0.00007846	656	802	82 %	Balanced value; selected for final modeling
0.7	0.00010985	760	802	95 %	Larger radius; broader coverage, lower spatial resolution

As predicted, the neighborhood radius showed a progressive increase in the number of crash records included in the clusters and a progressive decrease in the number of observations deemed spatial noise. If the radius was reduced to 0.3km, only 452 crashes (57%) were assigned to hotspot clusters, suggesting that only a small number of crashes were classified as highly concentrated. This setup provided good local spatial resolution but caused several meaningful hotspot areas to be spread apart and many local crashes to be treated as single observations.

The values of the radius 1.0 km and 1.5 km, however, resulted in a high proportion of “clustered” crashes (96% and 95%) and a marked reduction in noise points. This larger neighborhood, however, geographically grouped clusters of crashes together into larger areas, making it difficult for the algorithm to detect and separate individual hazardous areas. The usefulness of hotspot analysis for engineering interventions might be reduced if too many clusters are merged, as different roadway environments would end up being represented as a single hotspot.

The intermediate radius of 0.5 km, yielded a balanced solution, with 656 crashes (81.8%) being clustered together, and with good spatial separation between the adjacent hotspots. This parameter was selected for further analysis due to the good hotspot coverage and spatial resolution. The results validate the impact of neighborhood definition in the sensitivity of DBSCAN and highlight the importance of parameter tuning before interpreting the spatial hotspot patterns.

Statistical Summary of the Derived Spatial Variables

The identified crash hotspots are summarized in Table 2.

Spatial Indicators	Value
Total Crash Records	802
Number of DBSCAN-clustered	37
Clustered Crash Points	656
Noise Points	146
Cluster Crashes%	81.8%
Noise Points%	18.2%
Large Cluster	Zone 8
Crashes in Large Cluster	116
Large Cluster%	14.46%
Mean Cluster Size*	17.73

Using the same parameters, DBSCAN resulted in 37 spatial clusters and 146 noise observations, which accounted for 81.8% and 18.2% of the total number of crash records, respectively. The hotspot clusters had an average of 17.7 crashes, with some having more and others having fewer. There were 116 crashes in the largest hotspot (Zone 8), which made up 14.46% of all motorcycle/3-wheeled vehicle crashes.

The clustered crashes indicate that the locations of motorcycle crashes are not random in the transportation network but are concentrated at certain locations. Clustering suggests that the variables related to roadway, traffic exposure, and land use near the roadway are significant factors in the occurrence of a crash. Spatial heterogeneity is further suggested by the distribution of crashes with a relatively low number of crashes occurring at a disproportionately high share of locations.

Spatial Distribution of Crash Frequencies

The distribution of the crash frequencies along the detected DBSCAN clusters is shown in Figure 2.

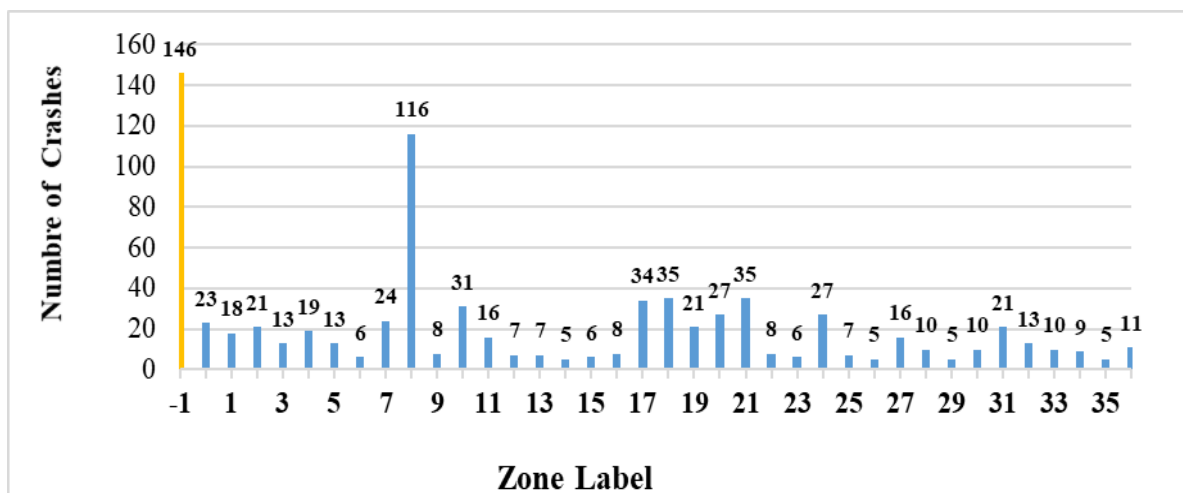


Figure 2. Distribution of crashes across the identified spatial zones

These results show significant variability in hotspot sizes. The highest crash frequency was in Zone 8, followed by Zones 18, 21, 17, and 10. However, a few clusters had fewer than 10 crashes, and 146 observations fell outside the identified hotspots. The dominance of Zone 8 implies that the area is not only subject to isolated crashes, but also to a persistent occurrence. These concentrations tend to occur where there are high traffic demand areas, traffic conflict areas, high pedestrian volumes, and intricate intersection geometries. The small groups also show that the risk of a crash is not concentrated in a single hazardous corridor but is spread across a variety of smaller regions. This spatial heterogeneity aligns with urban transportation systems that exhibit significantly different roadway functions, land uses, and traffic operations over relatively short distances. Therefore, hotspot identification data are more useful than crash frequencies, as they can be used to direct safety interventions to specific locations.

Hotspot Classification

The DBSCAN classification identified the crash locations as hotspot or non-hotspot areas. Figure 2 presents the proportion of crash records assigned to each category. About 82% of the crashes occurred in hotspot areas, and 18% occurred in isolated areas. There is evidence of a high level of geographic dependence in motorcycle crashes because of the high percentage of clustered crashes. Non-hotspot locations, however, should not be taken as a given to be safe locations. Instead, these observations are of crashes that did not meet the density threshold criteria. Even if no specific hotspots are identified, individual severe crashes can occur outside of these areas, as many factors contribute to their severity. To further examine the geographic distribution of hotspot locations, the DBSCAN results were visualized using spatial mapping techniques. Figure 3 illustrates the locations of hotspot and non-hotspot crashes within the study area.

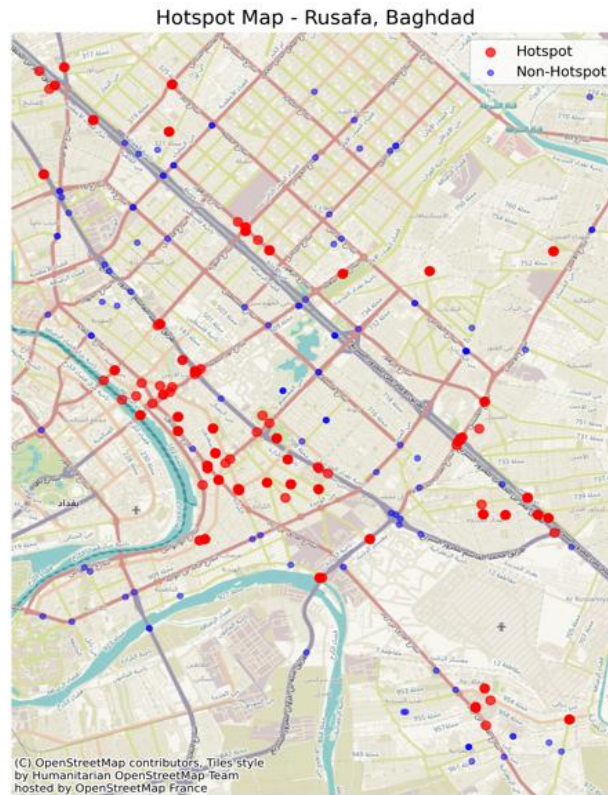


Figure 3. Hotspot and non-hotspot crash locations in Rusafa, Baghdad

The hotspot locations (red points) in Figure 3 are clearly concentrated in urban corridors and intersections in Rusafa, Baghdad with non-hotspot locations (blue points) being more spatially dispersed. The distribution indicates that there is a strong relationship between the distribution pattern of motorcycle and three-wheeled vehicle accidents and traffic interaction and urban activity. Identified hotspot areas are useful to transportation planners and decision makers as they will be able to prioritize safety interventions in areas where crashes are concentrated over a period of time.

Distribution of Crash Severity Inside and Outside Hotspots

The distribution of fatal, injury, and PDO crashes was analyzed between hotspot and non-hotspot areas to investigate the relationship between the severity of the crashes and the presence of the hotspots. The distribution of crash severities by hotspot and non-hotspot locations is presented in Figure 4. In absolute numbers, hotspot areas showed more fatal, injury, and PDO crashes than non-hotspot areas. This observation is not surprising, as most reported crash data comes from hotspot areas.

The findings indicate that hotspots have higher crash rates across all severity levels. This analysis, however, is descriptive and based on crash frequencies rather than severity proportions. Thus, hotspot designation should not be assumed to be associated with more severe crashes. Instead, the location of hotspots mainly corresponds to higher absolute crash frequencies. Further zone-wise analyses are needed to pinpoint zones with a disproportionately high incidence of severe crash outcomes.

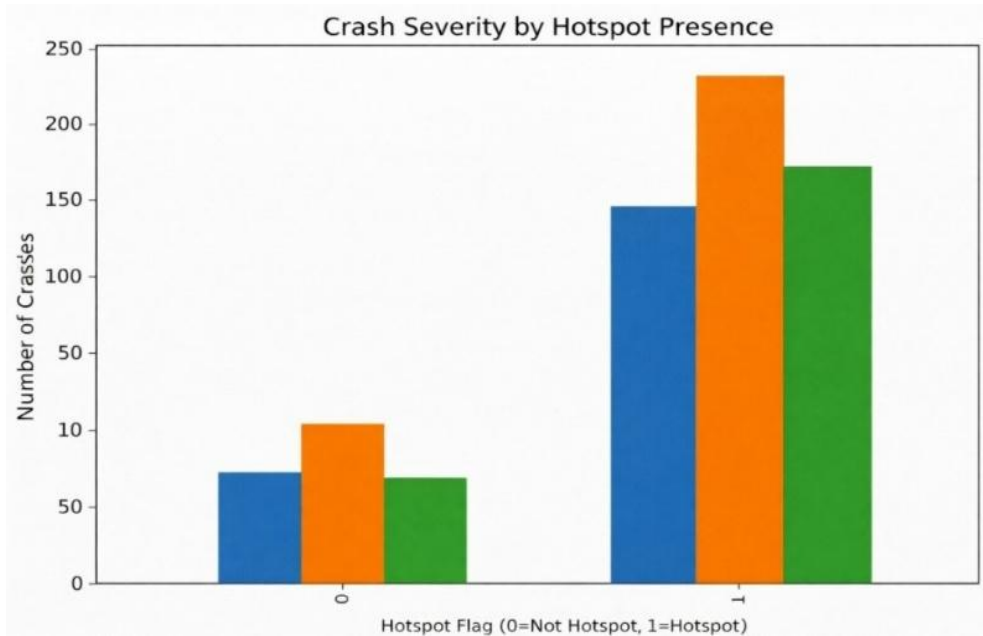


Figure 4. Distribution of crash severity according to hotspot presence.

Spatial Distribution of Crash Severity

The distribution of severity of the crashes in the identified DBSCAN zones is shown in Figure 5. Zone 8 experienced the most crashes (116), and therefore has the greatest concentration of crashes in the study area. Relatively high crash frequencies were also found in other zones (Zones 17, 18, and 21).

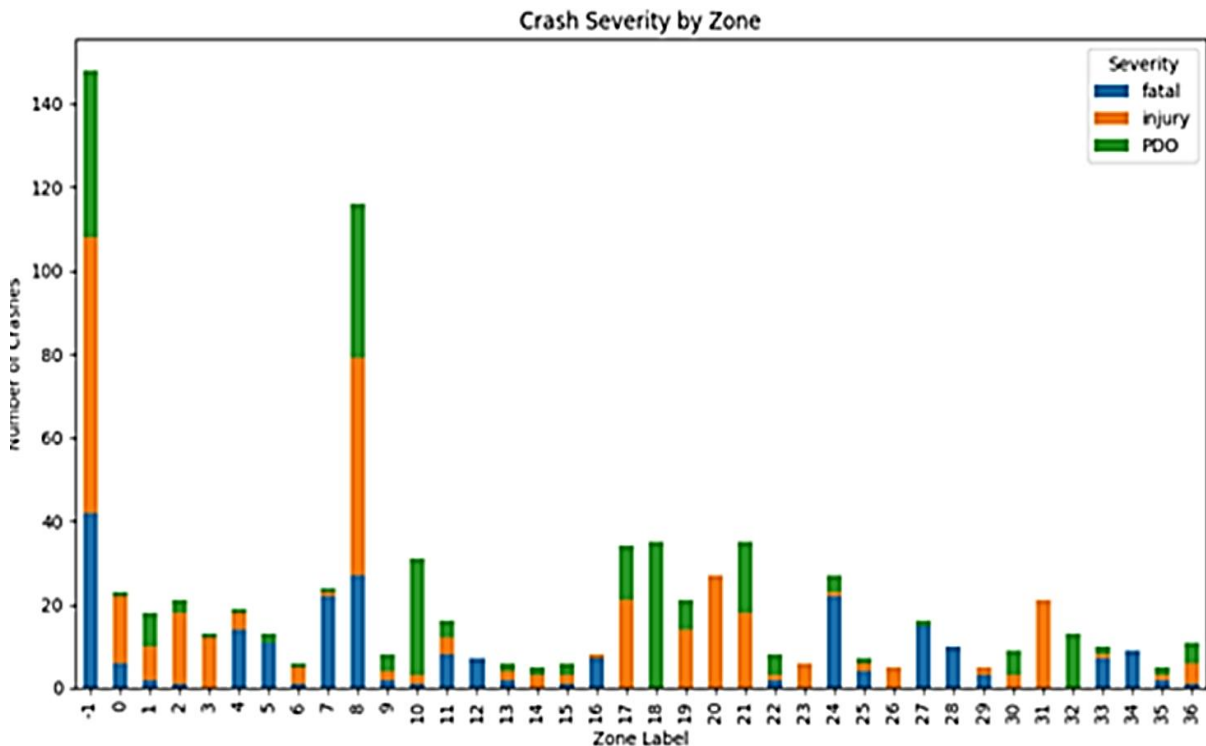


Figure 5. Distribution of crash severity across the identified DBSCAN zones.

The spatial distribution of fatal crashes was different, however. There were a few zones with relatively high numbers of fatal crashes and relatively low overall crash counts compared to Zone 8, including Zones 24, 27, and 28. This finding suggests that there is no correlation between the severity of crashes and their frequency. The findings show that frequency and severity of crashes are two complementary aspects of road safety. Although DBSCAN clustering works well to find clusters of crashes, further analyses based on the severity of

crashes are necessary to find areas with disproportionately severe crash outcomes. Therefore, both indicators should be taken into account if prioritizing road safety interventions in urban areas.

Spatial Distribution of Motorcycle Crashes in Baghdad

The overall spatiotemporal distribution of the severity of crashes and hotspot classification in Baghdad is shown in Figure 6. The severity map presents the geographic distribution of fatal, injury, and PDO crashes, and the hotspot map differentiates between crashes of multiple clusters and non-crashed locations. The maps depict a similar pattern of high rates of crashes in the middle, high-traffic-volume portions of the study area. The hotspot map also reflects that a significant percentage of crashes did occur in spatially clustered locations, while crashes at non-hotspot points were more spread out.

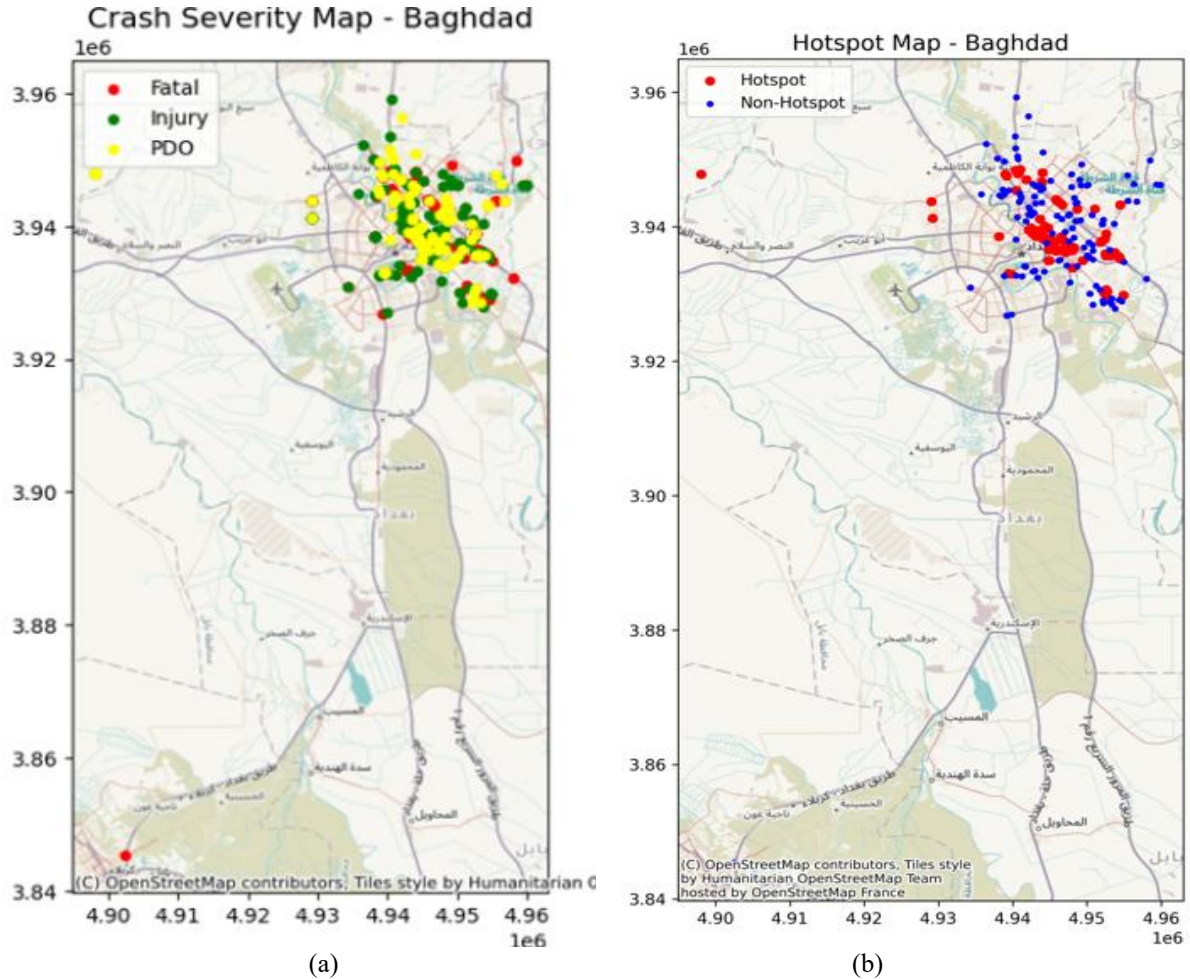


Figure 6. Spatial distribution of crash severity and hotspot classification in Baghdad: (a) crash severity map and (b) hotspot map.

Spatial Distribution of DBSCAN Clusters

Figure 7 shows the spatial distribution of the DBSCAN clusters found in the study area. There were 37 clusters detected with heterogeneous spatial patterns throughout Rusafa. There were relatively few crashes in several of the clusters but the highest number of crash events occurred in Zone 8. The detailed visualization of Zone 8 shows the concentration of crashes around major intersections and in urban corridors, and the algorithm's ability to detect concentrated areas of crashes within the urban road network.

Results indicate that the majority of the crashes occurred in a small area, including around major intersections and traffic corridors. The spatial distribution of crashes in the area of Zone 8 (Rusafa-Baghdad) shows some major urban features like Al-Firdaws Square, Al-Saadoon Street, Kahramana Square, and Al-Nidhal Street. The location of crashes is slightly jittered to bring out clear visualizations.

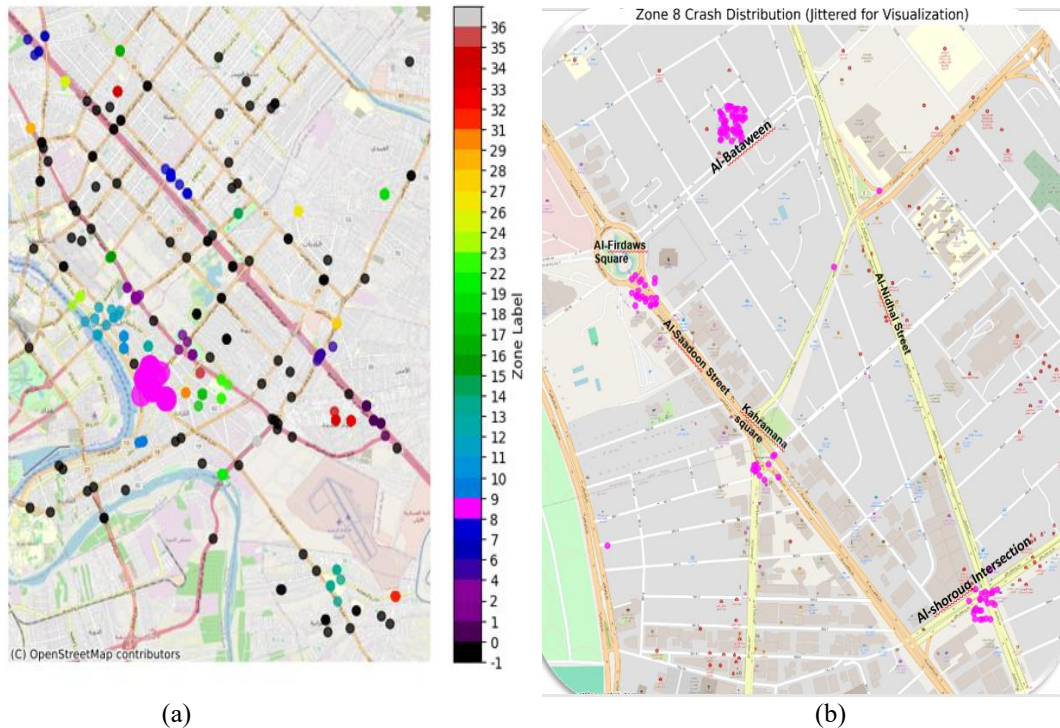


Figure 7. DBSCAN clustering results: (a) spatial distribution of all detected clusters in Rusafa, Baghdad, and (b) detailed visualization of Zone 8, the largest identified crash cluster.

Discussion

This study introduced a spatial hotspot detection methodology that could be implemented in GIS to identify spatial clusters of motorcycle and three-wheeled motorcycle crashes in the Rusafa district of Baghdad, using the DBSCAN (density-based spatial clustering of application with noise) clustering algorithm and the Haversine distance metric. The proposed framework effectively identified repeat crash locations by analyzing 802 geo-referenced crash records and separating isolated locations that were not deemed crash hotspots based on the density requirement. The sensitivity analysis showed that the clustering performance was strongly influenced by the neighborhood radius, and that $\text{eps} = 0.5 \text{ km}$ was the optimal compromise between the number of hotspots covered and the spatial resolution. With these optimal parameters, 37 spatial clusters were found; 81.8% of these crashes were in hotspot areas, and 18.2% were classified as spatial noise. In this largest hotspot (Zone 8) there were 116 crashes (14.46% of all crashes recorded).

The results demonstrate that motorcycle and 3-wheeled vehicle crashes are not randomly distributed over the road network, but clustered in a few urban areas. Furthermore, the analysis will show that hotspot frequency and severity of crashes are complementary and not interchangeable measures. The highest number of crashes was found in the largest hotspot, but some smaller hotspots had relatively high percentages of fatal crashes, suggesting that hotspots with the highest number of crashes are not hotspots with the highest percentage of fatal crashes. The overall conclusion is that both frequency- and severity-based analyses are important for setting road safety priorities.

In terms of method, the study shows the effectiveness of using the Haversine distance metric with DBSCAN for the geographic analysis of crash data. The proposed spatial indicators, Zone_Label, Zone_Count, and Hotspot_Flag, are additional explanatory variables that can be added to future statistical and machine learning models for crash severity prediction, transportation risk assessment and intelligent transportation systems. Beyond visualization, these indicators bring spatial information to hotspot detection to enhance the value of the information to predictive safety analysis.

At a practical level, the identified hotspot locations could be used by transportation planners and road safety authorities to gain insight into what aspects of the roadway and traffic network require improvement, traffic management, enforcement and public education. Interventions can be targeted to persistent hotspots to increase the safety of vulnerable road users and maximize the use of limited transportation safety resources.

While all of these contributions have been made, there are a few things that need to be recognized. All crash locations are analyzed without accounting for traffic exposure, roadway geometry, traffic control characteristics, weather, land-use attributes or temporal variations. These factors should be integrated into spatiotemporal and machine learning models in future studies to gain a fuller understanding of motorcycle crash risk. Furthermore, comparing Hotspot Detection Method with other methods, like Kernel Density Estimation (KDE), Getis–Ord G_i^* , Local Moran's I , OPTICS, and HDBSCAN, would better support the robustness and generalizability of the Hotspot Detection Method.

In general, the proposed framework shows that density-based spatial clustering is an effective and practical method for detecting hotspot areas of motorcycle and three-wheeled vehicle crashes in urban areas. The methodology helps the transportation safety analyst conduct a safety analysis based on GIS and it also serves as a basis for data-driven decision making to increase road safety for vulnerable road users.

Conclusion

This research aimed to develop a spatial hotspot detection methodology to identify areas of motorcycle and three-wheel vehicle (TWV) crash concentration in the Rusafa district of Baghdad city, utilizing the DBSCAN clustering algorithm, the Haversine distance measure, and GIS. The proposed framework identified persistent crash hotspots and excluded crash locations that were not dense enough to warrant inclusion in the persistent hotspot list by analyzing 802 geo-referenced crash records. The sensitivity analysis showed that the clustering results were highly sensitive to the neighborhood radius; $\text{eps} = 0.5$ km was found to be the most suitable value to balance hotspot coverage and spatial resolution. With these optimal parameters, 37 spatial clusters were identified, with 81.8% of crashes falling into hotspot areas and 18.2% considered spatial noise. A total of 116 crashes occurred in the largest hotspot (Zone 8), accounting for 14.46% of all reported crashes.

The results show that the crashes involving motorcycles and three-wheel vehicles are not randomly distributed on the road network but are clustered in relatively small numbers of locations within the city. Furthermore, this analysis shows that hotspot frequency and crash severity must be regarded as complementary, not synonymous. The highest frequency of crashes was in the largest hot spot; however, several smaller hot spots had relatively high percentages of fatal crashes, meaning that areas with the highest crash rates are not necessarily the highest risk of injury. This discovery is significant in terms of the need to combine frequency and severity-based analysis for setting road safety priorities.

From the methodological point of view, the study shows the effectiveness of this methodology for analyzing geographic crash data based on a hybrid approach of DBSCAN and Haversine distance. The proposed spatial indicators: Zone_Label, Zone_Count, and Hotspot_Flag, are additional explanatory variables that can be utilized by future statistical and machine learning models to predict crash severity, transportation risk assessments, and intelligent transportation systems. The indicators are useful beyond visualization to provide spatial information as a tool to predictive safety analytics.

As a practical example, the identified hotspot locations can serve as evidence for transportation planners and road safety authorities to prioritize engineering improvements, traffic management, enforcement, and public awareness programs. Strategic deployment of interventions at hotspot locations can enhance the safety of vulnerable road users and maximize the use of limited transportation safety resources.

Several limitations should, however, be noted in the face of these contributions. This analysis only examined the spatial distribution of crash locations, and the other data elements such as traffic exposure, roadway geometry, traffic control characteristics, weather, land use, and temporal changes were not included. These factors should be incorporated into spatiotemporal and machine learning models in future studies to gain a more comprehensive understanding of motorcycle crash risk. Furthermore, comparative assessments with other hotspot detection methods (e.g., Kernel Density Estimation (KDE), Getis–Ord G_i^* , Local Moran's I , OPTICS, and HDBSCAN) would help to support the robustness and generalizability of hotspot identification techniques.

In general, the proposed framework is an effective and practical method for identifying persistent crash hot spots involving motorcycles or three-wheeled vehicles in urban areas. The methodology is part of GIS transportation safety analysis and serves as a foundation for data-driven decision-making for improving road safety for vulnerable road users.

Scientific Ethics Declaration

* The authors declare that the scientific, ethical, and legal responsibility for this article published in EPSTEM Journal belongs to the authors.

Conflict of Interest

* The authors declare that they have no conflicts of interest

Funding

* Authors have not received financial support.

Acknowledgments or Notes

* This article was presented as an oral presentation at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Istanbul/Türkiye on September 17-18, 2026.

* The authors would like to thank Mustansiriyah University, Baghdad, Iraq, and the College of Engineering, Highway and Transportation Department, for their support in the present work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This study has not been supported financially by any public or private funding agencies.

References

- Ali, W., & Alhusaynat, A. D. (2026). An intelligent model for traffic accident hotspot detection. *Journal of Education for Pure Science*, 16(1).
- Eleiche, M. (2010, June). A minutes-based metric system for geographic coordinates in mobile GIS. In *2010 18th International Conference on Geoinformatics* (pp. 1–5). IEEE.
- Jameel, A. H., & Jameel, A. (2024). Risk assessment of the Latifiya expressway segment using a proactive approach. *Journal of Engineering and Sustainable Development*, 28(5), 601–610.
- Kamh, H., Alyami, S. H., Khattak, A., Alyami, M., & Almujiab, H. (2024). Exploring road traffic accident hotspots using clustering algorithms and GIS-based spatial analysis. *IEEE Access*, 13, 60944–60954.
- Lai, W., Zhou, M., Hu, F., Bian, K., & Song, Q. (2019). A new DBSCAN parameters determination method based on improved MVO. *IEEE Access*, 7, 104085–104095.
- Li, J., & Wang, X. (2022). Hotspot identification on urban arterials at the meso level. *Accident Analysis & Prevention*, 169, 106632.
- Mahmoud, H., & Akkari, N. (2016, March). Shortest path calculation: A comparative study for location-based recommender system. In *2016 World Symposium on Computer Applications & Research (WSCAR)* (pp. 1–5). IEEE.
- Mandala, S. K., Gurrupu, N., Praneeth, K. S., Vikas, M., Sanskar, K., & Rohith, T. (2026, March). Accident hotspot detection and spatial analysis. In *2026 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)* (Vol. 4, pp. 1–6). IEEE.
- National Research Council U.S. (2010). *Transportation research board, task force on development of the highway safety manual, & transportation officials, joint task force on the highway safety manual*. . Highway Safety Manual (Vol. 1). Washington, DC: AASHTO.
- Pusuluri, V. L., & Dangeti, M. R. (2024). Applications of QGIS and machine learning for road crash spot identification. *Earth Science Informatics*, 17(3), 2331–2346.
- Saputra, R., & Sihabudin, A. (2023). Hotspot identification through pick-up and drop-off analysis of ride-hailing transport service. *International Journal of Advanced Computer Science and Applications*, 14(11).
- Sharmila, S., & Sabarish, B. A. (2021, February). Analysis of distance measures in spatial trajectory data clustering. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1085, No. 1, p. 012021). IOP Publishing. Retrieved from <https://iopscience.iop.org/article/10.1088/1757-899X/1085/1/012021/meta>
- Shokat, D. M., & Jameel, A. K. (2025, September). Detecting high-risk road segments using kriging and inverse distance weighted methods. In *Proceedings of the Institution of Civil Engineers–Transport* (Vol. 178, No. 6, pp. 430–442). Emerald Publishing Limited.

- World Health Organization. (2023). *Global status report on road safety 2023*. Geneva, Switzerland: World Health Organization. Retrieved from <https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/road-safety/global-status-report-on-road-safety-2023>
- Zhang, J., Li, Z., Pu, Z., & Xu, C. (2018). Comparing prediction performance for crash injury severity among various machine learning and statistical methods. *IEEE Access*, 6, 60079–60087.
- Zhang, L., Wan, W., Lv, B., Li, Q., Sun, K., & Wan, Q. (2024, October). Identification of urban traffic accident hotspots using Getis-Ord Gi and DBSCAN algorithms. In *2024 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE)* (pp. 93–99). IEEE.

Author(s) Information

Ghaidaa A. Majeed

Mustansiriyah University, College of Engineering, Highway and Transportation Engineering Department, Baghdad, Iraq
ORCID iD: <https://orcid.org/0009-0004-3014-9102>

Abeer K. Jameel

Mustansiriyah University, College of Engineering, Highway and Transportation Engineering Department, Baghdad, Iraq
ORCID iD: <https://orcid.org/0000-0001-9999-9974>

*Contact e-mail: abeer_khudhur@uomustansiriyah.edu.iq

*Corresponding author(s) contact email

To cite this article:

Majeed, G. A., & Jameel, A. K. (2026). DBSCAN-based spatial hotspot detection of motorcycle and three-wheeled vehicle crashes. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM)*, 42, 162–175.