

Article

Road Traffic Accident Hotspot Detection: A GIS-Based Machine Learning Approach Using HDBSCAN and Spatial Clustering Techniques

Subham Roy ^{1,*} , Alireza Mohammadi ²  and Ranjan Roy ¹ ¹ Department of Geography and Applied Geography, University of North Bengal, Siliguri 734013, West Bengal, India² Department of Geography and Urban Planning, Faculty of Social Sciences, University of Mohaghegh Ardabili, Ardabil 5619911367, Iran

* Correspondence: rs_subhamroy@nbu.ac.in or subhammn2@gmail.com

Abstract

Road Traffic Accidents (RTAs) represent a significant public safety issue in rapidly urbanising nations, resulting in considerable fatalities, injuries, and economic losses. This research investigates the spatio-temporal distribution and hotspot dynamics of RTAs in Siliguri City, India, a principal transnational transport corridor connecting northeastern India with adjacent countries. A geocoded dataset comprising RTA incidents from 2021 to 2023 was analysed using integrated GIS-based machine learning and statistical methods. Temporal clusters were identified through Kulldorff's purely temporal scan statistics, while Kernel Density Estimation (KDE) quantified accident density during morning peak, midday/off-peak, evening peak, and lean/night-time intervals. Spatial clustering was further assessed using LISA-Moran's I, purely spatial scan statistics, and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). Emerging Hotspot Analysis (EHA) was employed to detect evolving hotspot patterns over time. The findings indicate that major accident hotspots are concentrated at key intersections and transport corridors, such as Hill Cart Road, Darjeeling More, Sevoke Road, Eastern Bypass, and Burdwan Road. Moran's I (0.157; $p = 0.007$) demonstrates significant but moderate spatial autocorrelation, and spatial scan statistics identified three principal high-risk zones. HDBSCAN classified 81.90% of incidents within clustered areas. Lean/night-time periods exhibited the highest accident densities, reaching 14.21 accidents/km² at critical intersections. These results underscore the utility of integrating GIS and machine learning techniques for urban traffic safety planning and hotspot-focused intervention strategies.

Keywords: accidents; fatalities; injury; hotspot detection; traffic safety

Academic Editor: Gianluca Grippelli

Received: 15 April 2026

Revised: 20 May 2026

Accepted: 26 May 2026

Published: 30 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Road traffic accidents (RTAs) continue to increase as populations grow, especially in emerging nations where increasing urbanisation and improved transport infrastructure has led to greater reliance on automobiles [1,2]. This rise in automobile usage leads to increased traffic deaths and injuries [3,4]. Traffic accidents significantly disrupt traffic flow by raising safety concerns and generating delays [1]. Environmental and terrain-related factors such as heavy rainfall [5], seasonal fog, reduced visibility [6], drainage-related waterlogging, and complex road geometry [7,8] may significantly influence the spatial distribution and severity of road traffic accidents. Similarly, hilly transport corridors [9] and high-volume

transit routes [4] often experience increased accident vulnerability due to traffic congestion and heterogeneous vehicle movement. According to research, RTAs do not occur randomly; they are impacted by factors such as road geometry, traffic volume, physical conditions, and weather conditions [7]. Therefore, to effectively develop accident prevention strategies and reduce the frequency of accidents, it is critical to identify high-risk sites and evaluate the timing of accidents [10].

Road traffic accidents result in about 1.35 million fatalities annually, making them the primary cause of death among children and young adults aged 5 to 29 [11]. Despite possessing just roughly 60% of the world's automobiles, low- and middle-income nations are responsible for 92% of worldwide road deaths [12,13]. More than half of these fatalities involve vulnerable road users such as cyclists, pedestrians, and motorcyclists [14]. Furthermore, RTAs have killed more people than malaria, and by the end of this decade, they are predicted to overtake HIV / AIDS as the top cause of mortality [15]. In 2021, India reported over 400,000 road traffic accidents (RTAs), resulting in more than 150,000 deaths, with a road traffic fatality rate of 14.6 per 100,000 persons [16,17]. Road traffic accidents are a major concern in Indian cities. High population density, rising vehicle numbers, and poor infrastructure in major cities like Delhi, Mumbai, Bengaluru, and Chennai all contribute to greater accident rates and deaths, putting public safety at risk [18].

Locations with substantially higher accident rates than surrounding areas are referred to as “hotspots” or “black spots” [19]. These hotspots are the geographic locations where high concentrations of RTA crashes occur compared to the total distribution of accidents in the area [2]. A hotspot is a road segment characterised by a high risk of vehicle accidents. Elvik [20] defined a *hotspot* as any road segment that has more incidents than other parts, frequently owing to unique hazard characteristics peculiar to that site. Therefore, identifying high-risk locations and accident clusters is a critical step in developing traffic safety strategies for urban areas [12]. This process involves identifying areas with frequent accidents and analysing the spatial risk patterns linked to these sites [21,22]. Geographic Information Systems (GIS) play an essential part in this study because they combine accident data with geographical information, allowing for the construction of comprehensive maps and heat maps [2,23]. These visual representations identify the distribution of RTA hotspots and geographical clustering patterns [2,21,22].

Previous studies have used a variety of methodologies to discover RTA patterns, including temporal and geographical analysis. Purely temporal approaches examine RTAs across time to identify trends and anomalies. These techniques detect trends, cycles, and anomalies in accident data, through time-series techniques such as moving averages [24], Autoregressive Integrated Moving Average (ARIMA) models [25], the K-means algorithm combined with hierarchical clustering methods, and seasonal decomposition [26]. However, solely temporal techniques have disadvantages, such as missing spatial patterns, ignoring spatial interdependence, and overlooking regional changes. They also introduce aggregation bias and provide a limited understanding of underlying causes by focusing just on when accidents occur rather than where they occur [27]. Other studies have utilised Getis-Ord G_i^* (Hotspot Analysis) Kernel Density Estimation (KDE), and Moran's I [7,11,28] to analyse spatial patterns and identify hotspots as well as purely spatial clusters of RTAs. In addition, Local Indicators of Spatial Association (LISA) and Ripley's K-function have also been employed to examine spatial clustering and identify localized patterns in road traffic accidents [29,30]. These studies have found these techniques effective in identifying and ranking RTA locations.

Other studies have utilised various spatiotemporal analysis algorithms to detect clustering patterns of RTAs in urban areas [7,28–30]. For instance, the Space-Time Kullback–Leibler Scan Statistics method [31], the Space-Time Cube, along with emerging hotspot

analyses [32–34], Dynamic Time Warping (DTW) [35] and Spatiotemporal Point Pattern Analysis [2] have been applied to simultaneously examine both spatial and temporal clusters of RTAs in urban areas. Spatiotemporal analysis methods applied to RTAs can be sensitive to scale, computationally intensive, and prone to interpretation challenges, but they offer comprehensive insights, improved cluster detection, and adaptable, accurate results through effective visualisation [21].

Despite the growing literature on road traffic accident (RTA) hotspot detection, many previous studies rely on single-method approaches that often fail to capture the complex and heterogeneous clustering behaviour of accidents in rapidly urbanising cities. The present study contributes to the literature by presenting a unified GIS-based spatio-temporal framework integrating temporal scan statistics, Kernel Density Estimation (KDE), Moran's I, LISA, purely spatial scan statistics, HDBSCAN machine-learning clustering, and Emerging Hotspot Analysis (EHA). First, the study incorporates time-segmented accident density analysis and total accident density mapping using aerial intersection-level visualisation. It then comparatively applies three clustering approaches—LISA-Moran's I spatial autocorrelation, purely spatial scan statistics, and HDBSCAN—to more effectively examine accident concentration patterns from multiple analytical perspectives. Unlike conventional clustering methods, HDBSCAN enables the detection of irregular and varying-density clusters while also identifying noise and probabilistic cluster memberships [36]. Furthermore, the study employs EHA to identify evolving spatio-temporal hotspot dynamics. The focus on Siliguri City, a strategically important transnational transport corridor in northeastern India, further provides insights into an underexplored geography of road traffic risk.

Although previous studies have successfully applied individual spatial, temporal, and spatio-temporal techniques for RTA hotspot detection, each method possesses certain limitations. Traditional hotspot and KDE approaches are effective for visualising accident concentration but are often sensitive to bandwidth selection and may overlook local spatial dependencies. Similarly, Moran's I and LISA efficiently detect spatial autocorrelation and local clustering patterns; however, they are limited in identifying irregular or varying-density clusters, which HDBSCAN has successfully addressed. Scan statistics provide statistically robust cluster detection but are constrained by predefined scanning windows and cluster shapes. Moreover, many earlier studies rely on a single analytical framework, limiting comprehensive interpretation of accident dynamics across space and time. Therefore, a more integrated and comparative analytical framework is required to better capture the heterogeneous and evolving nature of urban traffic accident patterns.

Therefore, the present study aims to evaluate the temporal and spatial patterns of RTAs and identify hotspot zones in Siliguri City, India, using a spatio-temporal dataset covering the period from 2021 to 2023. The study first employed Kulldorff's purely temporal scan statistics to detect temporal clusters, followed by KDE to examine accident density across four different time intervals: morning peak hours (8:01 a.m.–11:00 a.m.), midday/off-peak hours (11:01 a.m.–4:00 p.m.), evening peak hours (4:01 p.m.–8:00 p.m.), and lean hours (8:01 p.m.–8:00 a.m.). To improve spatial interpretation of accident concentration, aerial intersection-level visualisation using Google Earth imagery was also incorporated. Furthermore, three clustering approaches—LISA-Moran's I spatial autocorrelation, purely spatial scan statistics using SaTScan, v10.1 and HDBSCAN machine-learning clustering—were comparatively applied to identify and analyse RTA cluster patterns across the city. Finally, EHA was used to detect evolving spatio-temporal hotspot dynamics of RTAs over time.

2. Database and Methodology

2.1. Study Area

Siliguri City, located in the Darjeeling district of West Bengal, is a critical geostrategic corridor for India, frequently referred to as the “Chicken Neck of India.” The Siliguri Corridor represents a critical component of India’s overall infrastructure, which has a narrow territorial extent connecting the northeastern states with the rest of the country [37,38] (Figure 1). The region’s importance is not limited to its geographical location; it serves as a critical economic and transportation artery for the northeastern states of Arunachal Pradesh, Meghalaya, Assam, Nagaland, Mizoram, Tripura, Manipur, and Sikkim [39,40].

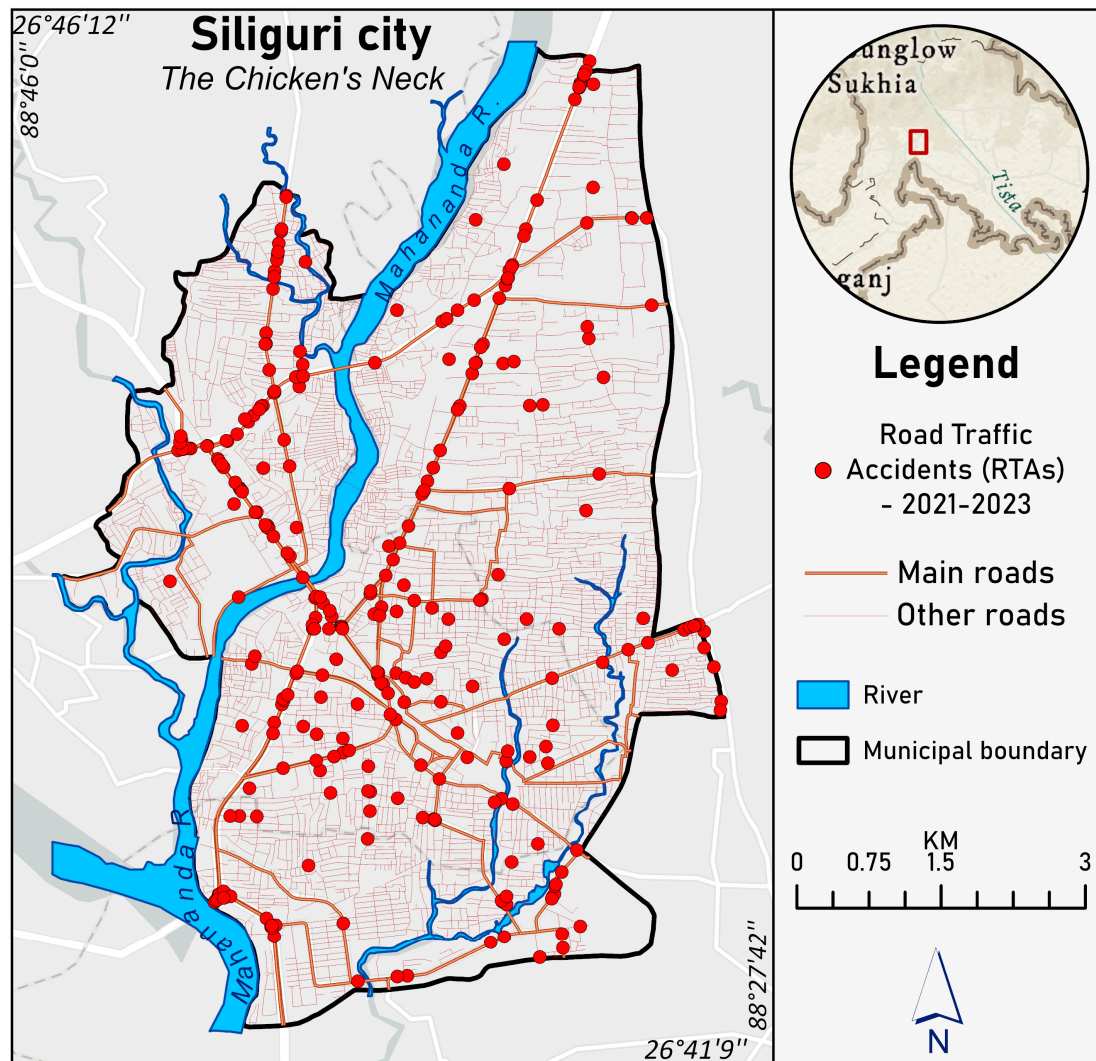


Figure 1. Location Map of the Study Area. The map also shows the spatial distribution of registered Road Traffic Accidents (RTAs) during 2021—2023 along major and other roads within Siliguri City.

Moreover, the city’s strategic relevance is amplified because it serves as a critical junction for multiple national highways, including NH10, AH2, and NH31, connecting the region to major Indian cities and states such as West Bengal, Assam, and Sikkim [41]. Siliguri’s location near key international borders significantly boosts its importance as a transportation hub [42,43]. The city is connected to Bhutan, Bangladesh, and Nepal and plays a crucial role in facilitating cross-border trade and movement. This proximity to international borders also positions Siliguri as a key centre for economic, political, and defence activities, making it an integral part of India’s national security strategy [41,44].

Together, these transport corridors facilitate the movement of goods, services, and people, strengthening Siliguri's reputation as a hub for logistics, trade, and tourism in the region.

Given its strategic location and vital role in linking Indian states with neighbouring countries, Siliguri is ideal for studying road traffic patterns. The high traffic volume—including commercial vehicles, international transport, and daily commuters—creates a dynamic setting for traffic analysis and hotspot mapping. This research focuses on exploring the spatial-temporal and clustering patterns of road traffic accidents in this unique geopolitical landscape, which serves as a critical point for both national and regional connectivity.

2.2. Database Generation

The dataset for the road traffic accident (RTA) study was collected from the Siliguri Metropolitan Police, comprising 315 RTA incidents reported over 36 months from 2021 to 2023. The database preparation included incidents that contain complete spatial and temporal information, making them suitable for GIS-based spatio-temporal analysis. Therefore, the analysis is based on officially registered incidents rather than a sampled subset of accident records. This raw data was converted into a shapefile format suitable for spatial analysis within GIS platforms. Each record includes coordinates of longitude (X) and latitude (Y) representing the location of the accident, along with additional details of the incident, such as date and time. Moreover, these incident data were collected and geocoded as point data within a GIS environment. The road network was collected from OpenStreetMap (OSM) and integrated with the dataset to improve spatial representation. For the present study, the road network was categorised into two types, namely, primary roads and other roads. Furthermore, administrative boundary data were obtained from the Siliguri Municipal Corporation. The dataset was further cleaned and organised for subsequent spatial analysis, and finally, a full-fledged GIS-based spatial accident database was prepared. This dataset was used in the present study to explore Siliguri City RTA patterns and trends in depth. The overall methodology of the present study is presented in Supplementary Table S1 and Figure 2.

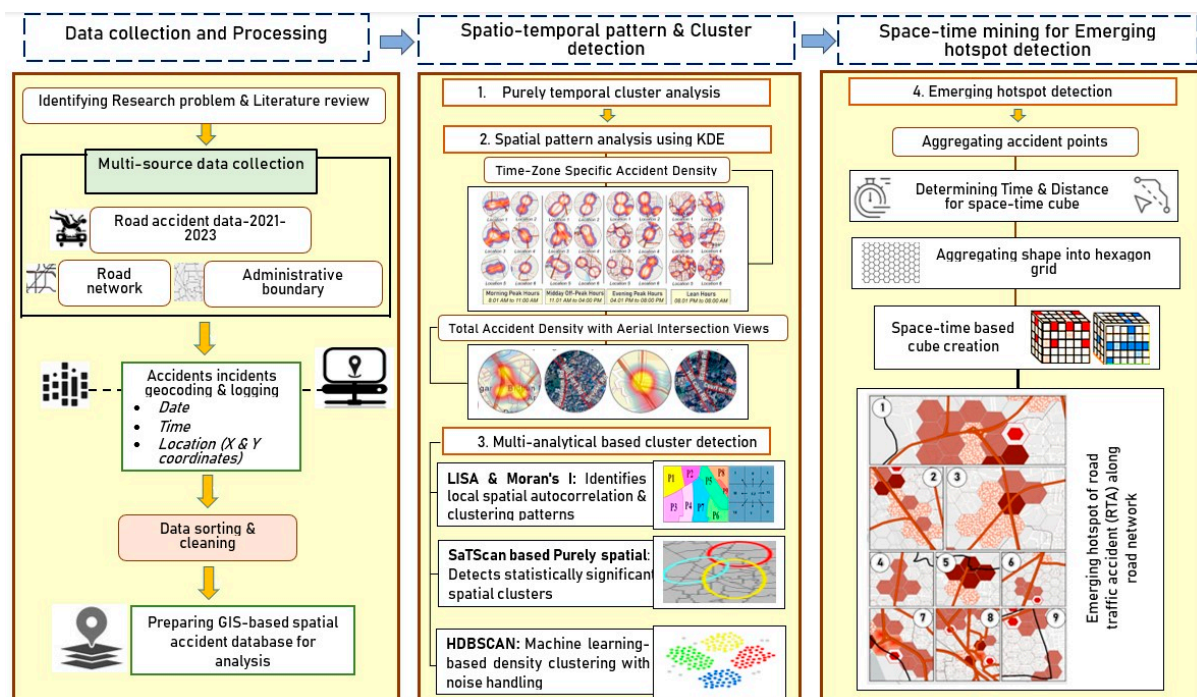


Figure 2. Methodological Flowchart for the present study.

2.3. Purely Temporal Cluster Analysis

To assess the temporal patterns of RTA over three years or 36 months, this study utilised SaTScan v.10.1.2 software package for temporal cluster analysis basis. The method is focused exclusively on identifying clusters within specified time intervals and does not consider any geospatial factor [37]. A retrospective, purely temporal cluster analysis focused on time-based clusters within the specified period and excluded spatial variations [45–47]. The Poisson discrete Scan statistics were used with a time aggregation period of 1 month. A monthly aggregation interval was selected to capture short-term temporal fluctuations in accident occurrences while avoiding excessive day-to-day variability and sparse event distribution, which may affect cluster stability [48]. The Poisson-based temporal scan statistic is particularly suitable for count-based event data such as RTAs, where incidents are assumed to occur independently across time. For the present study, the cluster size was set to a minimum of 1 generic, and the maximum was 50% of the study period. The 50% upper limit is widely recommended in scan statistic applications to avoid excessively large temporal clusters while still allowing the detection of meaningful high-risk periods. High-rate clusters required at least 2 cases and were otherwise unrestricted. No adjustments were made for weekly trends or known relative risks, as the primary objective was to identify overall temporal concentration patterns rather than model-specific covariate-driven variations. Temporal graphs were generated, focusing on the most likely clusters. The analysis involved 999 Monte Carlo replications with a significance cutoff of 0.05.

2.4. Spatial Pattern Analysis Using Kernel Density Estimation (KDE)

KDE is an effective technique for analysing the spatial distribution of road traffic accidents [2,4]. The KDE technique calculates the intensity of events within a specified bandwidth, which results in a smoothed surface representing spatial patterns [7]. A kernel function is used to assign weights to areas surrounding each point event based on their distance from the event [11]. This technique creates a continuous surface displaying the intensity of RTA occurrences across the study area. The raster surface estimates RTA occurrence intensity for a 30 m cell size at each raster cell. The KDE analysis was performed using ArcGIS Pro 3.3.0 using the Spatial Analyst tool with a quadratic kernel function, which is defined as follows in Equation (1).

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

In this equation, $f(x, y)$ refers to the projected density of RTA at the location (x, y) . The parameter n signifies the total number of RTA incidents being evaluated. The search radius or the bandwidth, is denoted by h . The distance between the location $(x \& y)$ and the i th incident is represented as d_i while K represents the kernel function utilised in calculating the density.

2.5. Detecting Spatial Clusters and Autocorrelations

2.5.1. Spatial Patterns Identification Through LISA and Moran's I

In this study, two methods have been employed to identify spatial patterns of RTAs, i.e., Global Moran's I and Local Indicators of Spatial Association (LISA). Moran's I is a global measure of spatial autocorrelation, helping to determine whether accidents are clustered, dispersed, or randomly distributed across the study area [37]. The global Moran's I evaluates spatial autocorrelation by examining both locations and feature attributes. Our study used the Global Moran's I as the initial method to assess spatial autocorrelation.

Additionally, we calculated the Z-score and p -value to determine the statistical significance of the Moran's I index. The formula for Moran's I is expressed in Equation (2):

$$I = \frac{N}{W} \cdot \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (2)$$

where I denotes Moran's I statistic, N represents the total number of locations, and x_i and x_j are the RTA counts at locations i th and j th, respectively. The mean RTA count is denoted by $\bar{x}$, while w_{ij} indicates the spatial weight between points i and j . W represents the sum of all spatial weights. Moderately positive values of Moran's I indicate clustering of comparable values, negative values point to dispersion, and near zero values suggest randomness.

For a more localised traffic accident analysis, LISA is used to detect the specific locations where RTA cluster occur or deviate significantly from the overall pattern. LISA identifies high-high clusters (H-H/hotspots) and reveals areas where high RTA rates are concentrated, in contrast to nearby high rates. In contrast, Low-Low Clusters (LL/Cold Spots) indicate areas with low RTA rates surrounded by similarly lower values. High-low (H-L) or Low-High (L-H) outliers also show locations where accident rates significantly differ from their neighbouring locations, such as a high-accident zone next to a low-accident area, indicating potential anomalies. LISA is represented by the following Equation (3):

$$I_i = \frac{(x_i - \bar{x})}{C^2} \sum_j sw_{ij} (x_j - \bar{x}) \quad (3)$$

In this equation, I_i is the local Moran's I for location i th, x_i and x_j are the accident counts for locations i and j , and C^2 is the dataset's variance. sw_{ij} represents the spatial weights between sites.

2.5.2. Purely Spatial Analysis Using Scan Statistics Approach

Purely spatial scan statistics is one of the approaches used to detect and evaluate case clusters in Siliguri City. This is accomplished by comparing the observed number of cases across different regions to the number expected in each region. The basic principle for this kind of method involves the movement of a circular window through the entire spatial domain to create a large number of potential clusters with each move. Key parameters, including relative risk (RR), p -value, and log-likelihood ratio (LLR), are calculated for each suspected cluster and analysed via descriptive tables and visual maps. In this study, the LLR values were prioritised to better understand cluster significance and spatial extent [45,46,49].

RR compares the risk within a cluster to the risk outside of it, emphasising areas with considerably higher or lower risks. RR is the predicted risk inside the cluster divided by the estimated risk beyond the cluster, and it can be mathematically determined using the following Equation (4):

$$RR = \frac{c/E[c]}{(C-c)/(E[C]-E[c])} = \frac{c/E[c]}{(C-c)/(C-E[c])} \quad (4)$$

where c denotes no. of observed cases (accidents in this study) that present within the cluster, and C indicates the total no. of cases in the dataset. However, as the analysis is conditioned on the total number of cases observed, $E[C]=C$ [47].

Subsequently, LLR assesses the likelihood that a cluster is genuinely significant by comparing the number of observed cases to the number of expected cases. Higher LLR values indicate stronger evidence of a true cluster. Here, I depict all of the spatial units (wards in this example) in research area S . Zone I is made up of neighbouring wards that

vary in size and form. c_i and n_i , on the other hand, represent the observed and predicted numbers of cases (or populations) in zone i th. Furthermore, C and N (equation) denote the total number of instances and predicted cases in S , respectively (Equation (5)).

$$C = \sum_i c_i \text{ and } N = \sum_i n_i \quad (5)$$

For accidents, a Poisson model is usually employed. The LLR of zone i is thus provided by Equation (6) [47,50]:

$$LLR(i) = \left\{ \left(\frac{c_i}{n_i} \right)^{c_i} \left(\frac{C - c_i}{N - n_i} \right)^{C - c_i} \right\} I(c_i > n_i) \quad (6)$$

If there is interest in scanning for ‘negative clusters’ with a lower rate than expected, the indicator function is replaced by $I(c_i < n_i)$ and if the interest is in clusters of both higher and lower rates, the indicator function is removed. It is equivalent but numerically easier to work with the logarithm, and the test statistic is given by Equation (4) that is the most likely cluster is the scanning window $i \in I$, which maximises the LLR [46]. It is expressed in Equation (7).

$$T = \max_i \log(LLR(i)) = \max_i LLR(i) \quad (7)$$

A retrospective purely spatial cluster analysis was performed to detect areas of high RTA rates while excluding any temporal variation [45,46]. A circular spatial Poisson discrete scan statistics was used with the maximum cluster size set as 50% of the at-risk population. Clusters were required to include at least 2 cases, and no other controls or modifications were made for published relative risks. The analysis was based on 999 Monte Carlo replications, using an alpha level of 0.05. Results with p -values were determined using all processors, and the overall processing time was estimated as 0 s, and no summary was sent by email.

2.5.3. HDBSCAN a Machine Learning-Based Clustering Algorithm

HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) is a self-adjusted advanced density-based clustering algorithm that extends DBSCAN by introducing a hierarchical approach to handle clusters of varying densities and noise [36,51]. HDBSCAN operates using a variety of epsilon values, combining the results to get the best stable grouping. This allows HDBSCAN to identify clusters with varying density levels and increases its responsiveness to parameter settings [36,52]. It is particularly effective for spatial analysis tasks like identifying road traffic accident (RTA) hotspots, where incident densities vary across regions and [53,54]. HDBSCAN builds the hierarchy between two core points, p and q , known as the *mutual reachability distance*, as provided by the parameter $\min_pts$. The mutual reachability distance is the minimal ϵ radius at which p and q are considered density-reachable [55]. For the HDBSCAN analysis, the Euclidean distance metric was applied using projected coordinates under the WGS-84 Datum UTM Projection, Zone 45N coordinate system. The minimum cluster size was 7 and minimum sample parameters were selected iteratively based on the spatial distribution characteristics and density variation in the 315 geocoded RTA points to effectively identify stable urban accident clusters while minimising noise classification.

2.6. Space-Time Cube Based Emerging Hotspot

The Emerging Hot Spot Analysis (EHA) tool explores space-time patterns by integrating two fundamental statistical methods, i.e., the Getis-Ord G_i^* statistic and the Mann–Kendall trend test, which analyses temporal trends [34]. For preparation of the

space-time cube, the tool requires several input parameters such as the aggregated RTA points, the determination of time step and distance interval for cube creation, aggregation shape type and a polygon analysis mask. EHA detects temporal patterns in spatial clustering by building a space-time cube out of numerous raster layers representing distinct periods layered with a common location ID and timestep ID. Each layer, or 'slice,' on the z-axis indicates a distinct time period, whereas the x-y plane depicts the geographical region throughout that era) [56].

For our investigation, we used three-month or quarterly time steps, a fixed distance band conceptualization, and stratified the analysis by RTAs with a spatial connection distance of 200 m. Although EHA has 9 different outcome categories, the output identified five major hotspot categories related to RTAs, which include new, consecutive, intensifying, persistent and sporadic hot spots. These categories were interpreted based on the necessity of understanding the spatiotemporal patterns of RTAs in the city.

3. Results and Discussion

3.1. Purely Temporal Analysis

The Retrospective Purely Temporal analysis, which used the Discrete Poisson model to search for clusters with high rates, discovered a cluster between February and August 2022. This cluster comprised all location IDs and had 80 cases, with an expected count of 59.78. The annual incidence rate for this cluster was 813.3 per 100,000 individuals, with an observed-to-expected ratio of 1.34. The relative risk was computed as 1.45, and the log-likelihood ratio was 3.906564. Despite these findings, the cluster was not statistically significant (p -value = 0.246, Monte Carlo rank = 51 out of 204). Figure 3 shows that the temporal cluster was identified between February and August 2022. While the number of RTAs increased throughout this time period, the absence of statistical significance indicates that the increase was likely due to random fluctuation rather than a meaningful surge. Consequently, the overall distribution of cases throughout time remained consistent, with no substantial periods of increased risk noted.

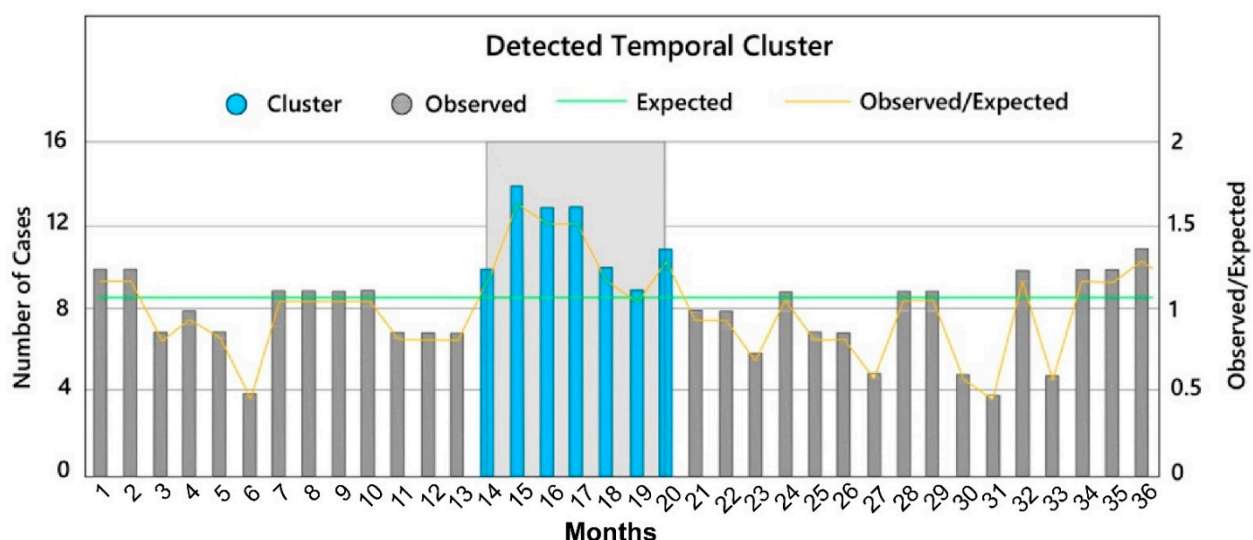


Figure 3. Detected Temporal clusters of Road traffic accident (RTA) over a 36-month period (2021–2023). It represents significant cluster between February 2022 to August 2022.

3.2. Spatial Pattern Analysis

3.2.1. Time-Segmented Accident Density Analysis

The investigation of RTA density in Siliguri City over different time periods indicates notable trends and high-risk zones (HRZ) during four key periods (Figure 4 and Table A1).

During the peak morning hours of 8:01 a.m. to 11:00 a.m., the accident density ranged between 1.61 and 3.85 accidents per sq. km in significant HRZ locations. The intersection of AH2 Noukaghat Road and Burdwan Road (spot 3) is the most hazardous location, accounting for 12.20% of all accidents in the region. Other high-risk intersections are Hill Cart Road near Darjeeling More (A1) and Sevoke Road near Salugara More (A2), which experience heavy traffic and complex road layouts that pose substantial safety risks. These significant sites account for 51.22% of accidents during this period. Moving into the Midday Off-Peak Hours (11:01 a.m.–4:00 p.m.), KDE values show an increase, with accident density peaking at 8.05 accidents/km² near the intersection of Hill Cart Road and Burdwan Road by Sevoke (Location 3). This location alone contributed to 8.33% of the day's accidents. Commercial areas like NH10 near Champasari More (A7) and the Sevoke Road check post (A9) also see a rise in accidents due to the combination of increased vehicle and pedestrian activity. Similar to the morning peak, central locations account for a substantial 51.19% of total accidents, highlighting the importance of monitoring these areas during the day.

During the evening rush hours (4:01 p.m.–8:00 p.m.), accident frequency increases in heavily congested areas, with densities ranging from 2.10 to 6.46 incidents per square kilometre. A key high-risk zone stretches along Hill Cart Road and AH2 near Darjeeling More up to Siliguri Junction (Location 3), recording 6.46 accidents/km² which contributed to 16.13% of all incidents. Other areas like Champasari Road (A11) and Satyen Bose Road (A9) also experience higher accident rates due to rush hour congestion. Nearly half of all accidents during this time occur in these critical areas. Interestingly, the lean hours (8:01 p.m.–8:00 a.m.) show the highest overall accident density, with KDE estimates reaching up to 14.21 accidents/km², particularly at the intersection of Venus More and Court More (Location 5), which accounts for 13.53% of accidents. Additional hotspots include the stretch from Mallaguri to Siliguri Junction on AH2 (Location 2) and Hill Cart Road near Airview (Location 6). Accidents during this time are likely influenced by factors like poor visibility, faster vehicle speeds, and driver fatigue [57–59]. A notable 67.67% of accidents during lean hours happen in significant locations, revealing the safety risks associated with nighttime driving in these critical junctions.

3.2.2. Spatial Accident Density with Aerial Intersection View

The spatial density map overlaid on the aerial view provides crucial insight into RTA patterns at major city intersections. Figure 5 depicts the number of accidents by road segment and colour correlates to the intensity/extent of accident occurrences. Yellow indicates a high frequency of accidents along busy roads, whereas blue or purple indicates areas where there have been minimal RTA incidents. Most strikingly, severe zones are aggregated at intersections and on heavily travelled urban roadways where repeat accidents may be more likely to occur. The spatial accident density map identifies dense road traffic accident hotspots overlaid on aerial imagery, showcasing key takeaways from the ten most prominent locations.

Every junction is where some of the most critical routes converge, making them very susceptible to frequent accidents. The condition becomes critical at the vital traffic hub of Darjeeling More (Location 1) and Champasari More (Location 2), where vehicles converge from the hilly regions via NH110, Bagdogra, Siliguri, and other sides through Hill cart road, thereby creating a high probability for accidents due to heavy vehicular volume mixed with public transport and local.

Similarly, along Sevoke Road, i.e., Location 3 (Salugara More) and Location 4 (Check Post) have been identified as the central points of maximum traffic density throughout the day, leading to a high possibility of resulting in accidents due to heavy traffic. In the case of Airview-Sevoke More (Location 5) and Venus More (Location 6), these are high-density

zones comprising a mix of both commercial and residential traffic along the intersections of Burdwan Road, Hill Cart Road, Sevoke Road, and Bidhan Road, boosting accident rates here to an all-time high. Similarly, Nougaghat More (Location 7) and Tinbatti More (Location 8) also present high accident density due to complex road geometry, leading to the recurrence of accidents. A pertinent factor is the complexity of intersections along Ghogomali road and Eastern Bypass near Ashghar More (Location 9) and IOC More (Location 10), which have been identified as key locations where a high level of traffic passing through at relatively high speed combine to form accident-prone environments.

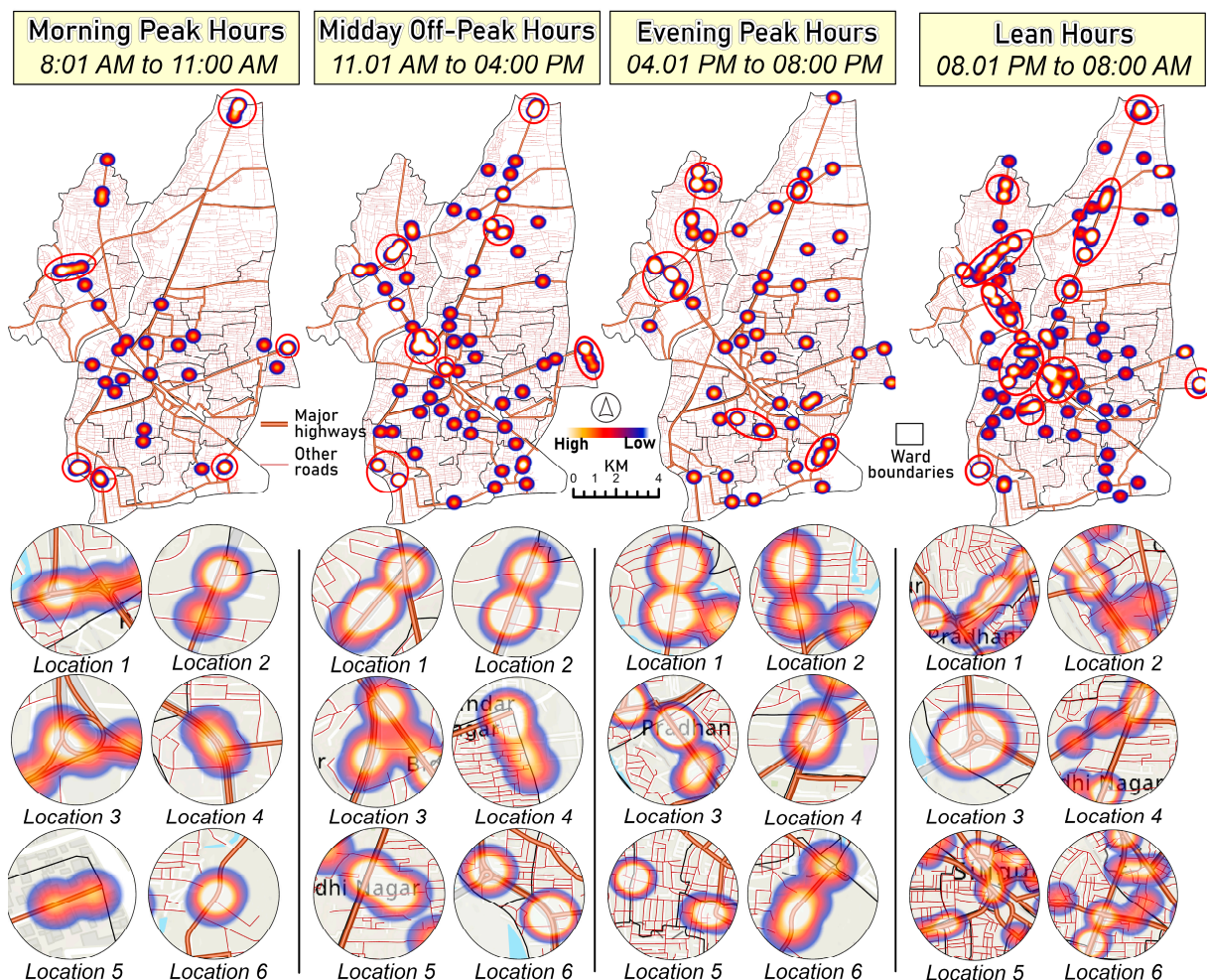


Figure 4. Spatial pattern of Major Road Traffic Accident (RTA) density zones across four-time intervals. The four time zones considered are: Morning Peak Hours, Midday/Off-Peak Hours, Evening Peak Hours, and Lean Hours. The red highlighted areas indicate zones with high RTA density. As a result, accident rates vary greatly depending on the time of day and location. High-density traffic zones, such as junctions and significant highways, see higher accident rates during peak hours (morning and evening) owing to congestion and complicated traffic patterns. Commercially active intersections such as Airview, Venus More, and Sevoke More also report more incidents around midday, which coincides with greater pedestrian and vehicle traffic for commercial purposes. Interestingly, the overall number of accidents greatly increases during lean hours despite reduced traffic. Other studies have revealed similar results: accidents were more frequent at night due to limited visibility, high speeds, and drivers' tiredness [60,61]. At this time, high-risk zones are frequently located in places with complex crossings and low lighting. However, these factors were not directly measured within the present dataset and should therefore be interpreted as reasonable contributing explanations rather than confirmed causal mechanisms. Additional studies incorporating behavioural, traffic-flow, and environmental variables would be necessary to establish direct causal relationships.

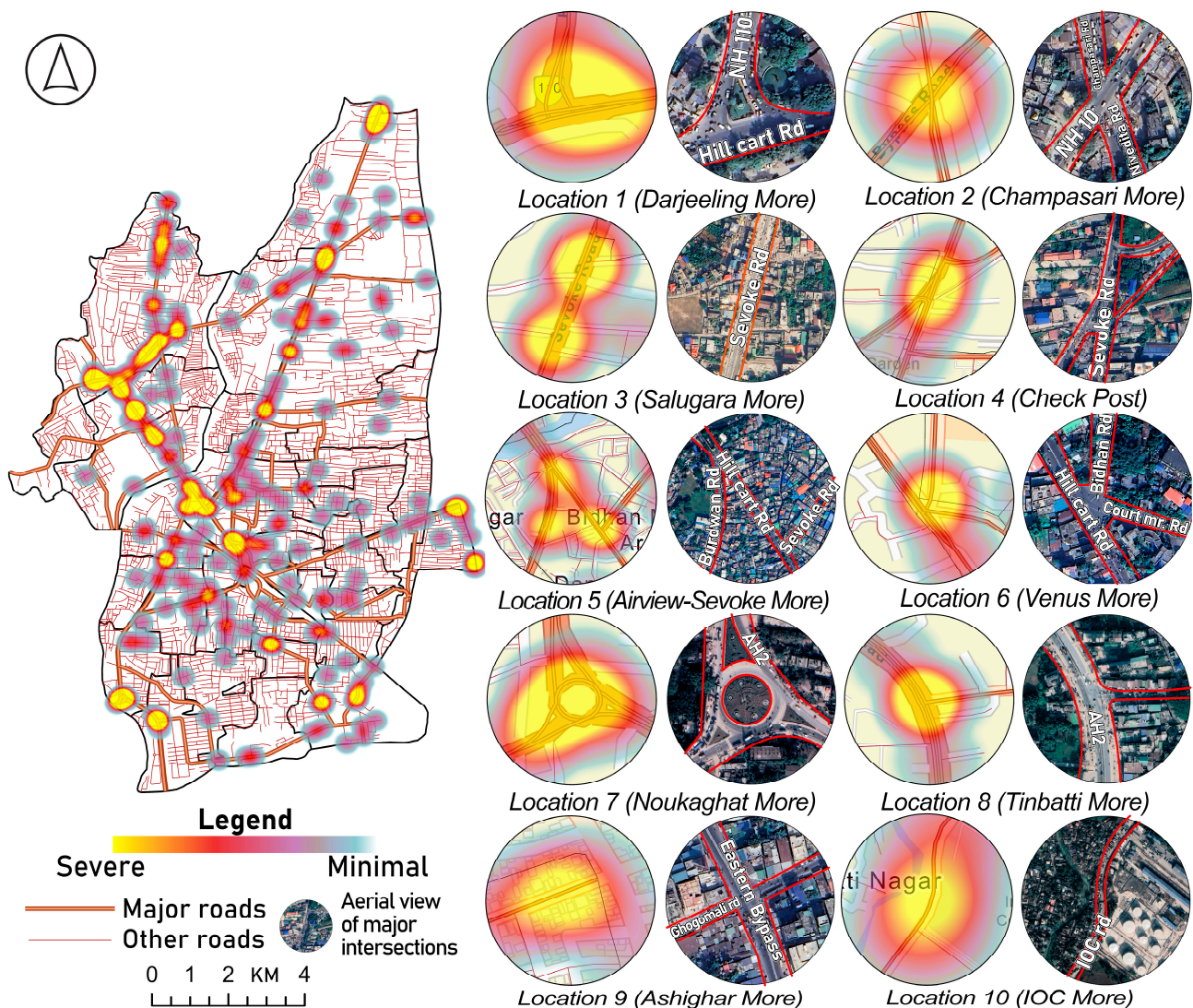


Figure 5. Total RTA density with an Aerial View of Major Roads and Traffic Intersection Points.

3.3. Multi-Analytical Clustering Detection of RTA

3.3.1. Spatial Patterns of RTA Clusters Using Moran's I and LISA

The LISA and Moran's I analysis reveals distinct spatial clustering patterns of road traffic accidents (RTAs), identifying both high-risk clusters and areas where accident patterns deviate from their surroundings [7,11]. The spatial map, combined with the Moran's I scatter plot and histogram, offers critical insights into these clusters, providing a basis for targeted safety interventions (Figure 6).

This map depicts four types of clusters: high-high clusters (red) represent anchor points where high RTA incidents are surrounded by other high-density areas. The high-low outliers (light red) are areas of high accident densities with lower-density surroundings, and the opposite is true for low-high cases (light blue). In contrast, low-low clusters (dark blue) represent low-density zones with similar low incidents of surrounding RTAs, while black dots indicate no obvious clustering.

The analysis reveals that high-risk clusters are distributed throughout the city, with significant high-risk clusters found along Hill Cart Road and AH2 near Darjeeling More up to Siliguri Junction. The concentration of traffic and intersections contribute to frequent accidents in this segment. Similarly, other major locations include Sevoke Road at Salugara and Check Post Intersection, Noukaghat More, Eastern Bypass near Ashighar More, and Burdwan Road. These areas are characterised by high crash rates, which are attributed

to problematic traffic flow, difficult road layouts, and poor safety measures. The analysis also finds high-low clusters along Champasari Road, S.F. Road, and IOC Road, which have relatively more accidents than the surrounding areas. These transition areas indicate possibly sudden changes in traffic or different road conditions that might be causing the RTA incidents to increase. These high-low clusters highlight the need for targeted traffic control measures to manage traffic transitions and prevent accidents.

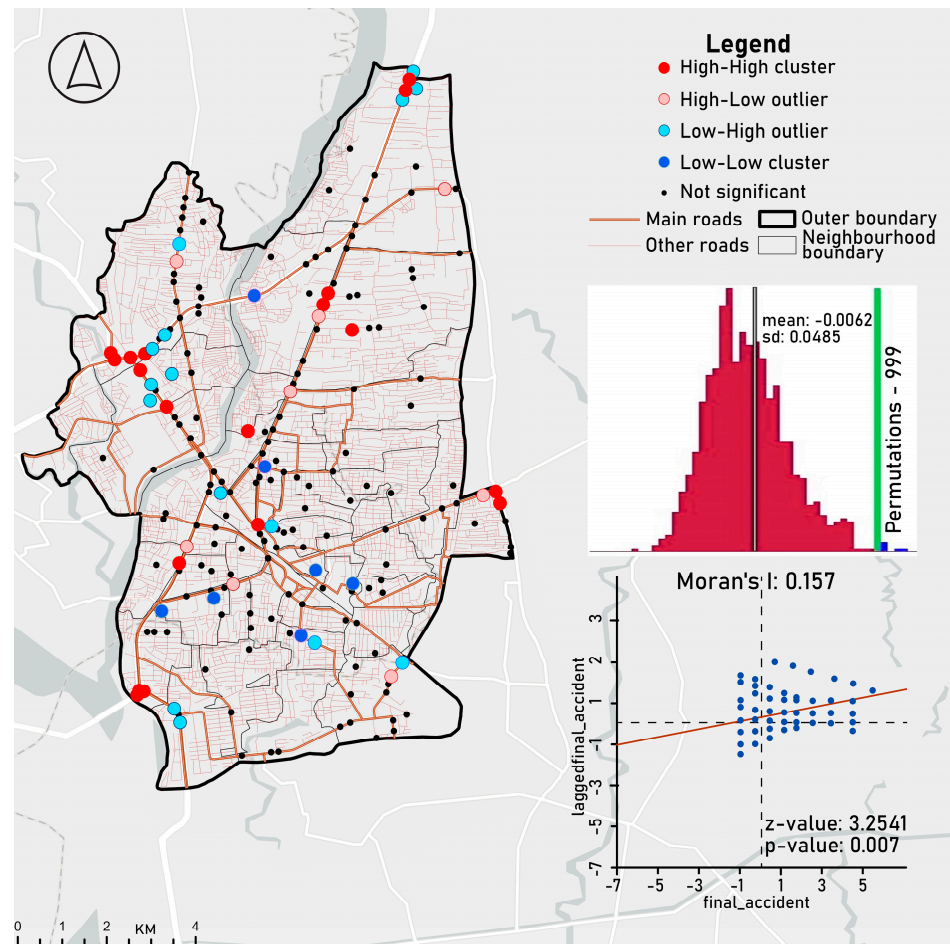


Figure 6. LISA analysis showing local spatial clustering and outliers of total RTAs, with the statistical permutation distribution and Moran's I scatter plot indicating local autocorrelation.

The Moran's I statistic of 0.157 indicates a moderate positive spatial autocorrelation, meaning that high and low accident densities tend to cluster rather than being randomly distributed. This is further supported by a z-value of 3.2541, with $p = 0.007$, confirming that these clusters are statistically significant. The histogram of Moran's I with 999 permutations further reinforces that accident hotspots are concentrated in specific regions, and these patterns are not random.

3.3.2. Identifying Purely Spatial Clusters of RTA with Scan Statistics

Scan statistics provided the most compelling spatial analysis result and proved highly valuable to our research. It identified three main clusters in the study area where RTA incidents are at much higher risk [21]. It offers key insights regarding accident concentration, this method can accurately identify spatial locations where the number of observed cases exceeds the expected number (Figure 7).

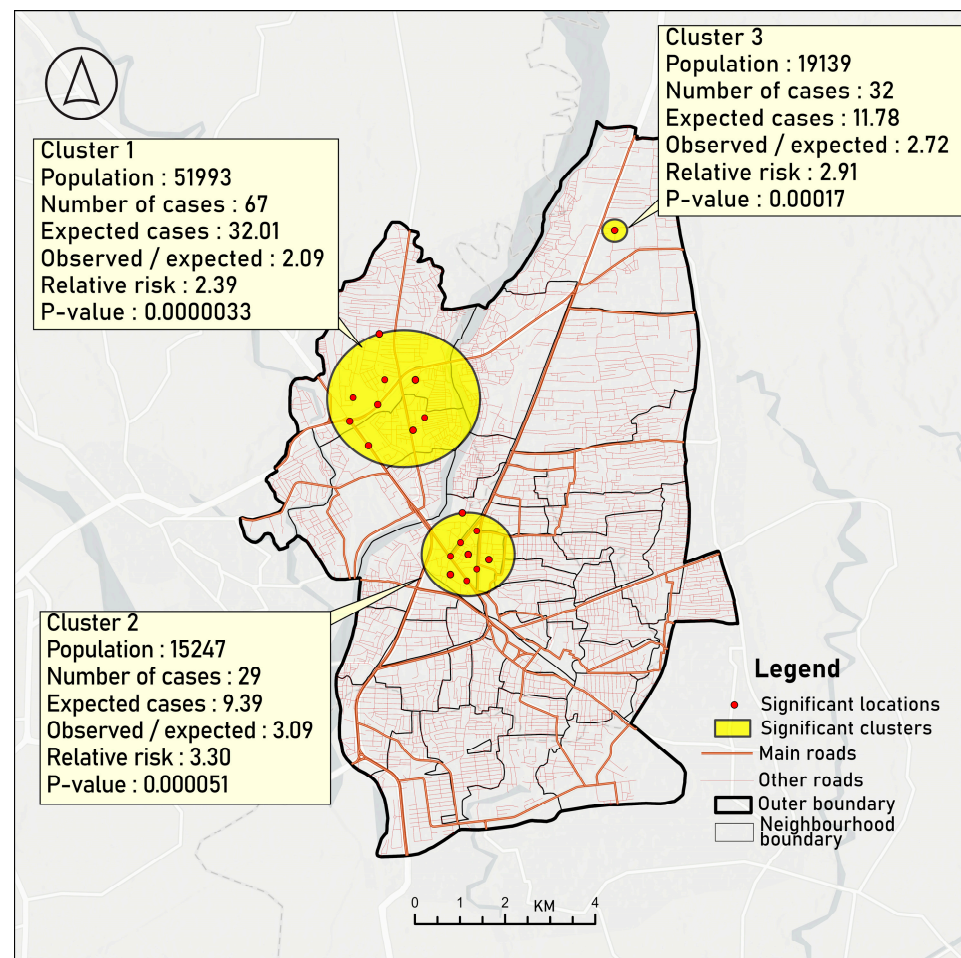


Figure 7. High-risk Purely Spatial Clusters of RTAs, with each circle representing significant locations and clusters.

The analysis reveals that Cluster 1 is the largest high-risk zone and a matter of urgent concern. It comprises ward number 45, 2, and 46, with a total population of around 51,993 and an observed count of 67 cases compared to the expected 32.01 cases. The RR of this cluster is 2.39, showing that accidents in the area are more than twice as likely to occur relative to what would be expected. The *p*-value of this cluster is significant at 0.000033, while the log-likelihood ratio (LLR) was a substantial 16.75, confirming the statistical significance of this zone. The area covers Hill Cart Road, AH2, Champasari Road, and Nivedita Road, which are highly congested main traffic arteries, leading to more frequent chances of both small and big accidents. The spatial analysis of this area further highlights the need for intervention, such as value-added traffic management measures, expanded infrastructure, and public safety provisions to reduce risks.

Similarly, Cluster 2 (Wards 11, 12, 10, and 6), located in the city centre, shows a similar RTA pattern. This cluster, with a population of 15,247, has an RR of 3.30, the highest significant relative risk in the study area, with 29 observed cases (compared to an estimated 9.39 instances). Statistically, this cluster is highly significant, with an LLR value of 13.74 and a *p*-value of 0.000051. These clusters include major routes such as Burdwan Road, Hill Cart Road, Sevoke Road, and Bidhan Road, which have been facing issues with high traffic volumes and busy commercial areas. This key area's high population density and traffic congestion make it an ideal target for safety improvements.

Lastly, Cluster 3, located near Sevoke Road in Ward 42, presents a population of 19,139 with 32 observed cases, compared to an expected 11.78 cases. This cluster has a relative risk

of 2.91 and an LLR of 12.44 (p -value = 0.00017), reveals another significant high-risk zone. The elevated accident rate along Sevoke Road, one of the city's busiest commercial and transit routes, suggests that traffic management improvements are also critical in this area. These clusters collectively reveal the geographic concentration of road traffic accidents in the study area, each distinguished by high relative risks and strong statistical significance.

3.3.3. Density-Based Clustering of Accident Hotspots Using HDBSCAN

The HDBSCAN analysis provides comprehensive insights into the RTA clusters in the city (Figure 8). Figure 8A indicates the actual clusters found, and Figure 8B reveals the high probability-based crash locations, giving important insights into hotspots for accidents. The panel map below includes a detailed view with probabilities over 90% for important cluster sites to consider in road safety interventions. Figure 8A shows the spatial distribution of the eight identified HDBSCAN clusters marked in different colours. These clusters are situated along important traffic corridors within the city, like NH10, Sevoke Road, AH2, and Bidhan Road, among others. Also, in the southern part, accidents happen in clusters across areas like Nougakhat More, S.F. Road, and DBC Road, which underline also underline several accident-prone zones scattered across the urban landscape.

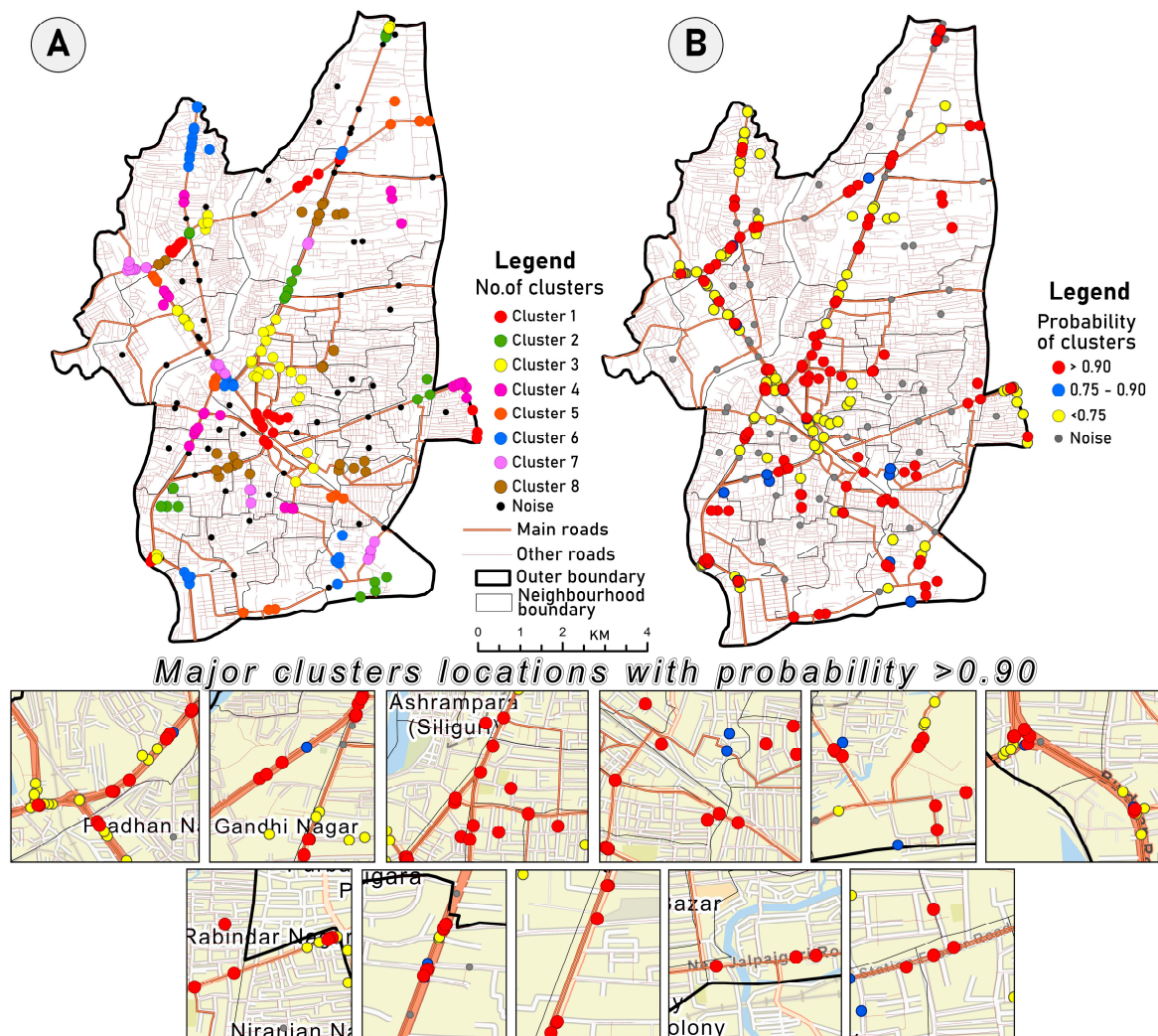


Figure 8. HDBSCAN algorithm-based clustering detection with noise. (A) Number of clusters detected, including noise, and (B) Cluster classification by probability values. The figure also highlights major RTA cluster locations with probability values > 0.90.

Conversely, Figure 8B highlights the probability of these clusters and visually reveals the certainty of where accidents are most likely to occur. The clusters with a probability greater than 0.90 are shown in red, thus indicating the high-risk areas of RTA concentration. These zones are mainly located in high-traffic areas, such as NH10 near Darjeeling More to Checkpost More, Sevoke Road, and Eastern Bypass. Areas with moderate cluster probabilities (0.75 to 0.90) are marked in yellow; similarly, blue markers also show clusters with a probability lower than 0.75 on the map. The maps displayed in the bottom panel give an overview of the most significant cluster points distributed along NH10, Sevoke Road, Eastern Bypass, Nougakhat More, Bhanumati Road, along with DBC Road and S.F. Road, all having a probability of more than 0.90. In addition, the HDBSCAN analysis-based silhouette score of 0.583 indicates a strong and cohesive clustering structure. Also, the Davies-Bouldin Index of 0.533 and a Calinski-Harabasz Index value of 439.29 strengthen the durability of this clustering model. The analysis identified 81.90% of accidents in the clusters, with an additional 18.10% classified as noise (outside of any cluster), reflecting accident patterns deemed too irregular or sparse to form coherent groupings.

Interestingly, Figure 9 reveals the membership probabilities distribution (where each point stands for how strongly it is associated with its assigned clusters). The *x*-axis shows the membership probabilities (0 to 1), and the *y*-axis shows the number of points count in each probability range. The critical observations that we can infer from the model, i.e., the majority of the points having a probability higher than 0.95, indicate that these points are strongly affiliated with their respective clusters. This suggests a high confidence in the clustering, meaning that the identified accident-prone zones are well-defined and reliable. In other words, points with a membership probability above 0.95 are highly likely to represent their clusters accurately, reflecting high confidence in the clustering.

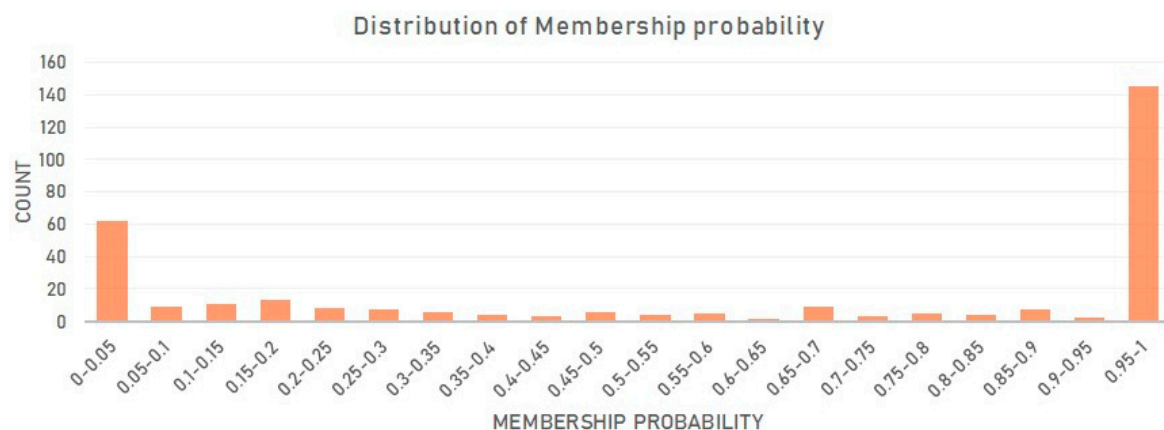


Figure 9. HDBSCAN —based distribution of membership probabilities for points within clusters.

In comparison, the number of points with lower membership probabilities (0 to 0.05) is much less frequent, especially for small memberships (<30). These points are not well differentiated within any cluster, and some might be borderline or noise, which can easily fall into lower-confidence clusters. However, the large number of points with a membership probability greater than 0.95 indicates that HDBSCAN has found strong and highly distinct clusters where most points strongly belong to just one cluster. This improves the overall reliability of the clusters in accurately identifying the areas associated with the most serious accidents.

3.3.4. Comparison of Clustering Techniques

The comparison of the three spatial clustering methods—LISA & Moran's I, SaTScan, and HDBSCAN—highlights differences in their ability to detect and characterise road

traffic accident (RTA) clusters within the city (Table 1). Each method successfully identified clusters, but the nature of the clusters and their interpretation varied. LISA & Moran's I detected clusters based on spatial autocorrelation, identifying 19 High-High (hotspot) clusters, 7 Low-Low (coldspot) clusters, and 23 outliers. In contrast, SaTScan detected three significant clusters, all statistically robust, with a total of 7 clusters observed overall. HDBSCAN, using a density-based approach, identified 8 clusters, with 81.90% of the points forming clusters and 18.10% classified as noise, reflecting its ability to manage scattered data points.

Table 1. Summary of Spatial Clustering Techniques.

Measure	LISA & Moran's I	Scan Statistic Based Purely Spatial	HDBSCAN
Cluster Detected	Yes	Yes	Yes
Cluster Types	High-High (Hotspot), Low-Low (Coldspot), Outliers	Significant cluster	Density-based cluster with noise
Number of Clusters	- High-High—19 - Low-Low—7 - Outliers—23	- Significant cluster—3 - Total cluster—7	Total cluster—8
Statistical measures	- Moran's I value—0.157 - z-score—3.254 - p-value—0.007	All significant cluster p-value—<0.00	- Silhouette Score—0.583 - Davies-Bouldin Index—0.533 - Calinski-Harabasz Index—439.29
Cluster detection success rate	N/A	N/A	- Clusters detected—81.90% - Noise—18.10%
Cluster probability rate	N/A	N/A	>0.90—46.98% 0.75–0.90—5.08% <0.75—29.84% - Noise—18.10%
Cluster Shape	Irregular	Circular	Irregular
Strength	Detects local spatial autocorrelation	Identifies statistically significant high-risk clusters	Flexible detection of irregular and varying-density clusters with probabilistic confidence

In terms of statistical performance, the RTA data showed strong spatial autocorrelation (LISA & Moran's I = 0.157, z-score = 3.254, p-value = 0.007). Similarly, the SaTScan-based solely cluster method produced comparable strong findings, with all clusters having a p-value less than 0.00, indicating the statistical significance of the observed clusters. On the other hand, HDBSCAN produced performance indicators such as a Silhouette Score of 0.583, a Davies-Bouldin Index of 0.533, and a Calinski-Harabasz Index of 439.29, indicating well-defined clusters with sufficient separation and cohesiveness.

One of HDBSCAN's unique qualities is its ability to account for noise, with 18.10% of data points classified as noise, as opposed to other algorithms that do not specifically recognise such points. Furthermore, HDBSCAN gives cluster probability data, revealing that 46.98% of clusters have a probability greater than 0.90, indicating high cluster confidence. This contrasts with purely spatial clustering methods, and Moran's I, which do not give probability-based confidence estimates. Similarly, the form of the clusters also differed across approaches. LISA and HDBSCAN discovered irregular-shaped clusters that

better represent the complicated, real-world distribution of RTAs. Purely spatial clustering, on the other hand, identified circular clusters, which are beneficial for distinguishing well-defined regions but may overlook more detailed spatial patterns found in irregularly shaped clusters. Therefore, compared to conventional spatial clustering approaches, HDBSCAN provides a more flexible and realistic representation of urban accident patterns by detecting clusters with varying densities and irregular spatial forms without requiring predefined cluster numbers. This capability is particularly useful in complex urban traffic environments where accident distributions are heterogeneous and spatially uneven. The probabilistic cluster membership and noise identification further enhance the reliability and interpretability of hotspot detection, making HDBSCAN a robust complementary approach for GIS-based RTA analysis.

3.4. Space-Time Based Emerging Hotspot Analysis

The EHA, which integrates the Getis-Ord G_i^* statistic with the Mann–Kendall temporal trend test, indicates statistically significant increasing spatio-temporal trends in RTAs across several transport corridors of the city ($p = 0.007$). Furthermore, hotspot categories identified through EHA represent statistically significant clustering patterns at confidence levels exceeding 90%, indicating that the detected hotspot dynamics are unlikely to be random (Figure 10 and Supplementary Table S2). The figure depicts the distribution of new, consecutive, intensifying, persistent, and sporadic hotspots along significant routes and an intersection. These hotspot categories provide a comprehensive picture of how accident-prone locations have changed across time and space. New hotspots have been found along numerous vital routes, including Hill Cart Road, Sevoke Road, S.F. Road, AH2, and Ashighar Road, suggesting regions with a recent spike in accidents. These newly detected hotspots require immediate intervention because they indicate areas where accidents are beginning to concentrate.

Consecutive hotspots, the most widespread form, account for 53.5% of all hotspots, as indicated in the table. These hotspots are seen at nearly every major intersection and route in the city, indicating persistent accident trends over time. The prevalence of consecutive hotspots suggests that these places are persistently risky and require ongoing road safety upgrades. Similarly, the intensifying hotspot, which is only located at the Venus more and Noukaghat intersections where AH2 and Noukaghat Road converge. This intensifying hotspot, which accounts for 4.0% of the total, indicates that accidents are not only continuing but also becoming more severe and frequent at this area. This intensification may indicate changes in traffic flow, road conditions, or infrastructure that must be addressed immediately.

Persistent hotspots, which account for 10.1% of all hotspots, are concentrated at major intersections such as Venus More, where Bidhan Road, Hill Cart Road, and Court More Road intersect, as well as Salugara More on Sevoke Road and the Champasari intersection, where Champasari Road, Nivedita Road, and NH10 converge. These persistent hotspots represent locations with a long history of accidents, showing a chronic problem that has not been appropriately addressed over time. Finally, sporadic hotspots, which account for 22.2% of the total, are distributed across all main consecutive hotspot sites. These random patterns indicate that, while accidents in these locations are less common, they nevertheless occur intermittently and add to the total accident risk.

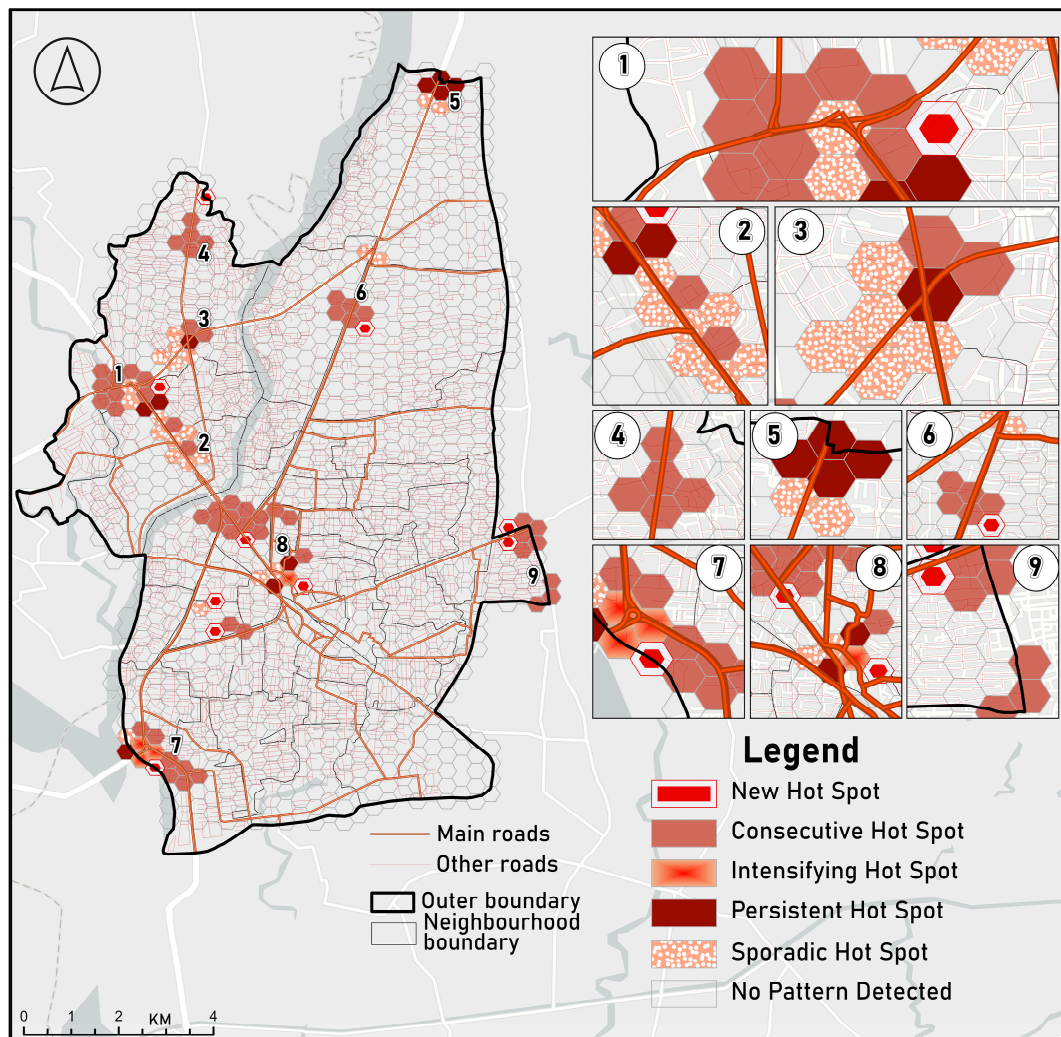


Figure 10. Identification of emerging RTA hotspots using space-time mining, with key hotspot locations highlighted on the map.

4. Policy Implications

The study's findings have significant implications for urban planning, traffic management, and the formulation of road safety legislation. High-risk zones are successfully identified using spatial density, clustering approaches, and hotspot analysis, offering a clear direction for focused safety improvements. Different hotspot categories identified through EHA require distinct intervention priorities. Newly emerging hotspots along Hill Cart Road, Sevoke Road, S.F. Road, AH2, and Ashighar Road require rapid monitoring and early preventive interventions before accident concentration intensifies further. Persistent and consecutive hotspots located at major commercial intersections and transport corridors, including Venus More, Salugara More, and Champasari intersections, demand long-term infrastructural improvements, intelligent traffic management systems, pedestrian safety measures, and continuous enforcement monitoring. Similarly, intensifying hotspots such as Venus More and Noukaghat intersections require immediate, targeted intervention due to the increasing frequency and severity of accidents. In contrast, sporadic hotspots may benefit from adaptive traffic surveillance and periodic enforcement during high-risk time periods, particularly during nighttime and lean-hour conditions.

Moreover, implementing time-specific strategies is essential, particularly considering the elevated accident rates observed during off-peak hours. Strengthening law enforcement during nighttime, coupled with installing improved street lighting and stricter speed

regulation measures, could significantly reduce risks associated with poor visibility and driver fatigue. While previous studies have associated nighttime accidents with visibility limitations, speeding behaviour, and driver fatigue, the present analysis does not directly evaluate these variables. Therefore, these interpretations should be considered indicative possibilities requiring further investigation rather than definitive causal explanations. Moreover, infrastructure development is also a vital strategy to reduce accidents. Measures like optimised traffic flow design and better pedestrian safety initiatives in high-risk areas like the Eastern bypass, Venus More, Hillcart road and Salugara More are necessary. The study shows the importance of data-driven approaches for decision-making in urban planning. Through the use of methods such as LISA, SaTScan, EHA, and HDBSCAN, this study presents a thorough framework for city planners to strengthen their decision-making while prioritising road security interventions. Similarly, public awareness campaigns focusing on driver behaviour could be launched during peak and off-peak times targeting specific areas which affect human behaviour, like speeding or fatigue, which are known to substantially contribute towards the occurrence of accidents.

The study also stresses the importance of incorporating road safety into more comprehensive urban planning. Given that commercial and residential areas have such high accident density, a broader approach that incorporates road infrastructure, land use, and traffic patterns is clearly warranted. New developments present an opportunity for urban planners to make road safety a priority, designing commercial zones and residential areas with access from major highways in ways that can keep gridlock down while decreasing the number of accidents. Working together, traffic engineers and city planners can design roads meant to meet the safety needs of all road users: drivers, cyclists, and pedestrians.

5. Limitations

Despite providing important insights into the spatial and spatio-temporal dynamics of RTAs, the study has certain limitations. The analysis was based on officially recorded and geocoded police-reported RTA incidents; however, detailed information regarding accident severity, crash type, involved road users, and potential reporting bias was not consistently available within the dataset. Similarly, traffic volume-based exposure normalisation and multivariate modelling were beyond the scope of the present study due to the unavailability of detailed traffic flow, exposure, and road-user datasets for the study area. Furthermore, although the accident locations were geocoded carefully within a GIS environment, minor positional inaccuracies may still exist due to address-based reporting and spatial referencing limitations. The study is also limited by its spatial scale and temporal coverage, as the analysis was conducted within the administrative extent of Siliguri City using a three-year dataset (2021–2023), which may not fully capture long-term accident dynamics and broader regional traffic interactions. Moreover, investigations that incorporate behavioural variables, traffic-flow characteristics, lighting conditions, and environmental factors may provide a clearer understanding of the higher concentration of RTAs during nighttime. Future studies incorporating traffic volume, road characteristics, weather conditions, and severity-based accident data may provide a more comprehensive understanding of urban traffic accident dynamics.

6. Conclusions

This study examined the spatial and spatio-temporal patterns of road traffic accidents (RTAs) in Siliguri City, India, using an integrated GIS-based analytical framework combining temporal analysis, density mapping, spatial autocorrelation, scan statistics, machine-learning clustering, and EHA. The findings reveal that major RTA hotspots are concentrated along critical transport corridors and intersection zones, particularly around

Hill Cart Road, Sevoke Road, Eastern Bypass, AH2, and other high-traffic commercial intersections. The analysis further indicates significant clustering patterns and evolving hotspot dynamics across different time periods, with higher accident density observed during nighttime and lean-hour conditions. Among the applied techniques, HDBSCAN demonstrated effectiveness in identifying irregular and varying-density accident clusters along with probabilistic cluster memberships and noise points, highlighting its suitability for complex urban traffic environments and its potential applicability in future GIS-based traffic safety studies.

The study highlights the importance of location-specific and time-sensitive traffic safety interventions in high-risk transport corridors and commercial intersections. The identified hotspot patterns may support urban planners, transport authorities, and traffic management agencies in prioritising infrastructural improvements, traffic monitoring, and targeted road safety strategies in accident-prone zones. Furthermore, the integrated analytical framework adopted in this study demonstrates potential applicability for RTA hotspot assessment in other rapidly urbanising cities experiencing complex traffic dynamics and increasing transportation pressure.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/geographies6020055/s1>, Table S1: Overview of Applied Analytical Techniques for RTA Density, Clustering and Hotspot Detection; Table S2: RTA based-Emerging Hotspots.

Author Contributions: S.R.—Writing—Original Draft, Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualisation, Review and Editing, Funding Acquisition, Supervision. A.M. and R.R.—Review and Editing, Conceptualization, Methodology, Software. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy.

Acknowledgments: Firstly, the authors would like to express cordial thanks to Department of Geography and Applied Geography, University of North Bengal for providing opportunity in conducting the research work. The authors are also deeply thankful to the Siliguri Metropolitan Police for providing the necessary data and support essential for the completion of this study. Finally, the authors extend their heartfelt appreciation to the six anonymous reviewers and the editor for their rigorous, insightful, and constructive comments, which significantly helped to improve the quality and overall standard of the manuscript. Any remaining errors, or omissions are solely my responsibility.

Conflicts of Interest: The authors declare that they have no competing interests.

Appendix A

Table A1. Road Traffic Accident (RTA) Density, Percentage, and High-Risk Zones based on Major Location and Time Zone.

Time-Zone	MLI	KDE Estimate (Accidents/km ²)	% of RTA Out of Total	Total % of RTA in MLs	HRZ	HRZ Code	Remarks
Morning Peak hours (8:01 a.m.–11:00 a.m.)	Location 1	2.79–2.38	9.76	51.22	Segment along Hill Cart Road near Darjeeling More	A1	MIHT
	Location 2	2.11–2.31	9.76		Segment along Sevoke road near Salugara more	A2	BCA
	Location 3	3.52–3.85	12.20		Intersection of AH2, Noukaghat road and Burdwan road near Noukaghat more	A3	MIHT
	Location 4	2.71–3.61	7.32		Segment along AH2 near Tinbatti more	A4	MIHT
	Location 5	1.99–2.02	7.32		Intersection of Eastern Bypass and Ghogomali Road near Ashighar More	A5	MIHT & BCA
	Location 6	1.61–1.6.3	4.88		Segment along IOC road near NJP Station	A6	HTZ
Mid-day off peak hours (11:01 a.m.–04:00 p.m.)	Location 1	3.71–6.50	7.14	51.19	Segment along NH10 near Champasari More	A7	MIHT & BCA
	Location 2	4.15–4.50	5.95		A2	A2	BCA
	Location 3	6.56–8.05	8.33		Intersection of Hill cart road and Burdwan road near Airview and Sevoke more	A8	MIHT
	Location 4	2.87–3.65	4.76		A5	A5	MIHT & BCA
	Location 5	3.86–4.53	4.76		Segment along Sevoke road near check post more	A9	HTZ & BCA
	Location 6	3.19–4.12	5.95		A3 & A4	A3 & A4	MIHT
	Other location	2.81–6.33	14.29		A1, A6, Intersection of Check post more—(A9), Venus more—(A10)	A1, A6, A9, A10	MIHT & BCA
Evening peak hours (04:01 p.m.–08:00 p.m.)	Location 1	3.07–4.03	8.06	50.00	Segment along Champasari road	A11	HTZ & BCA
	Location 2	2.41–3.38	6.45		Segment along Champasari road and NH10 near Champasari more	A12	MIHT, HTZ & BCA
	Location 3	3.23–6.46	16.13		Segment along Hill cart road & AH2 near Darjeeling more upto Siliguri Junction	A13	MIHT, HTZ & BCA
	Location 4	2.17–2.69	4.84		A8	A8	MIHT & BCA
	Location 5	2.54–3.36	8.06		Segment along Satyen Bose road near Babupara	A9	HTZ
	Location 6	2.10–2.94	6.45		A6	A6	HTZ

Table A1. Cont.

Time-Zone	MLI	KDE Estimate (Accidents/km ²)	% of RTA Out of Total	Total % of RTA in MLs	HRZ	HRZ Code	Remarks
Lean hours (08:01 p.m.–08:00 a.m.)	Location 1	5.64–14.08	11.28	67.67	A1, A7	A1, A7	MIHT & BCA
	Location 2	7.04–12.67	6.02		Segment along AH2 from Mallaguri upto Siliguri Junction	A14	
	Location 3	9.26–10.20	5.26		A3	A3	MIHT
	Location 4	4.64–8.19	8.27		A9	A9	MIHT & BCA
	Location 5	9.64–14.21	13.53		Intersection of Venus more and Court more	A10	MIHT & BCA
	Location 6	9.51–12.55	10.53		Intersection of Hill cart road, Burdwan Road and AH2 near Airview and Jhankar more (A15)	A8, A15	MIHT & BCA
	Other location	4.17–6.59	12.78		Segment along S.F. Road (A16), A2, A4, A5, A11	A2, A4, A5, A11, A16	MIHT & BCA

Note: MLI—Major location id, KDE—Kernal density estimation, HRZ—High risky zone, MIHT—Major intersection with heavy traffic, BCA—Busy commercial area, HTZ—Heavy Traffic Zone.

References

- Soltani, A.; Askari, S. Exploring spatial autocorrelation of traffic crashes based on severity. *Injury* **2017**, *48*, 637–647. [\[CrossRef\]](#) [\[PubMed\]](#)
- Le, K.G.; Liu, P.; Lin, L.T. Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam. *Geo-Spat. Inf. Sci.* **2020**, *23*, 153–164. [\[CrossRef\]](#)
- Tola, A.M.; Demissie, T.A.; Saathoff, F.; Gebissa, A. Severity, spatial pattern and statistical analysis of road traffic crash hot spots in Ethiopia. *Appl. Sci.* **2021**, *11*, 8828. [\[CrossRef\]](#)
- Afolayan, A.; Easa, S.M.; Abiola, O.S.; Alayaki, F.M.; Folorunso, O. GIS-based spatial analysis of accident hotspots: A Nigerian case study. *Infrastructures* **2022**, *7*, 103. [\[CrossRef\]](#)
- Keay, K.; Simmonds, I. Road accidents and rainfall in a large Australian city. *Accid. Anal. Prev.* **2006**, *38*, 445–454. [\[CrossRef\]](#)
- Petrova, E.G.; Shiryayeva, A.V. Road accidents in Moscow: Weather impact. *Adv. Environ. Sci.* **2019**, *11*, 19–30.
- Amiri, A.M.; Nadimi, N.; Khalifeh, V.; Shams, M. GIS-based crash hotspot identification: A comparison among mapping clusters and spatial analysis techniques. *Int. J. Inj. Control Saf. Promot.* **2021**, *28*, 325–338. [\[CrossRef\]](#)
- Gupta, P.; Shekhar, M.S.; Singh, G.P.; Gupta, D.S.; Singh, A.; Kumar, A.; Kumar, R.; Tomar, D.S. High-resolution analysis and prediction of heavy precipitation-induced GLOF events in North Sikkim Himalayas using the WRF model. *Phys. Chem. Earth Parts A/B/C* **2025**, *139*, 103968. [\[CrossRef\]](#)
- Joshi, A.K.; Joshi, C.; Singh, M.; Singh, V. Road traffic accidents in hilly regions of northern India: What has to be done? *World J. Emerg. Med.* **2014**, *5*, 112. [\[CrossRef\]](#)
- Harirforoush, H.; Bellalite, L. A new integrated GIS-based analysis to detect hotspots: A case study of the city of Sherbrooke. *Accid. Anal. Prev.* **2019**, *130*, 62–74. [\[CrossRef\]](#)
- Le, K.G.; Liu, P.; Lin, L.T. Traffic accident hotspot identification by integrating kernel density estimation and spatial autocorrelation analysis: A case study. *Int. J. Crashworthiness* **2022**, *27*, 543–553. [\[CrossRef\]](#)
- WHO. *Global Status Report on Road Safety 2015*; World Health Organization: Geneva, Switzerland, 2015.
- Tavakkoli, M.; Torkashvand-Khah, Z.; Fink, G.; Takian, A.; Kuenzli, N.; de Savigny, D.; Cobos Muñoz, D. Evidence from the decade of action for road safety: A systematic review of the effectiveness of interventions in low and middle-income countries. *Public Health Rev.* **2022**, *43*, 1604499. [\[CrossRef\]](#)
- WHO. Road Traffic Injuries. 2023. Available online: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries> (accessed on 6 December 2023).
- Ahmad, A.U.; Hossain, K.T.; Hossain, M.A. Identification of urban traffic accident hotspot zones using GIS: A case study of Dhaka Metropolitan Area. *J. Geogr. Stud.* **2020**, *3*, 36–42. [\[CrossRef\]](#)
- MRTH. Road Accidents in India. 2021. Available online: https://morth.nic.in/sites/default/files/RA_2021_Compressed.pdf (accessed on 6 December 2023).
- WHO. Road Traffic Mortality Rate per 100,000 Population in 2021. 2024. Available online: <https://data.who.int/indicators/i/B9D9E6A/D6176E2?m49=356> (accessed on 6 December 2023).

18. Mahata, D.; Narzary, P.K.; Govil, D. Spatio-temporal analysis of road traffic accidents in Indian large cities. *Clin. Epidemiol. Glob. Health* **2019**, *7*, 586–591. [[CrossRef](#)]
19. Dereli, M.A.; Erdogan, S. A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. *Transp. Res. Part A Policy Pract.* **2017**, *103*, 106–117. [[CrossRef](#)]
20. Elvik, R. A survey of operational definitions of hazardous road locations in some European countries. *Accid. Anal. Prev.* **2008**, *40*, 1830–1835. [[CrossRef](#)]
21. Mohammadi, A.; Kiani, B.; Mahmoudzadeh, H.; Bergquist, R. Pedestrian Road Traffic Accidents in Metropolitan Areas: GIS-Based Prediction Modelling of Cases in Mashhad, Iran. *Sustainability* **2023**, *15*, 10576. [[CrossRef](#)]
22. Soroori, E.; Kiani, B.; Ghasemi, S.; Mohammadi, A.; Shabanikiya, H.; Bergquist, R.; Kiani, F.; Tabatabaei-Jafari, H. Spatial Association Between Urban Neighbourhood Characteristics and Child Pedestrian–Motor Vehicle Collisions. *Appl. Spat. Anal. Policy* **2023**, *16*, 1443–1462. [[CrossRef](#)]
23. Mohammadi, R.; Taleai, M.; Otto, P.; Sester, M. Analyzing urban crash incidents: An advanced endogenous approach using spatiotemporal weights matrix. *Trans. GIS* **2024**, *28*, 368–410. [[CrossRef](#)]
24. Quddus, M.A. Time series count data models: An empirical application to traffic accidents. *Accid. Anal. Prev.* **2008**, *40*, 1732–1741. [[CrossRef](#)] [[PubMed](#)]
25. Getahun, K.A. Time series modeling of road traffic accidents in Amhara Region. *J. Big Data* **2021**, *8*, 102. [[CrossRef](#)]
26. Gu, J.; Jiang, Z.; Fan, W.D.; Wu, J.; Chen, J. Real-time passenger flow anomaly detection considering typical time series clustered characteristics at metro stations. *J. Transp. Eng. Part A Syst.* **2020**, *146*, 04020015. [[CrossRef](#)]
27. Mannering, F.L.; Bhat, C.R. Analytic methods in accident research: Methodological frontier and future directions. *Anal. Methods Accid. Res.* **2014**, *1*, 1–22. [[CrossRef](#)]
28. Alam, M.S.; Tabassum, N.J. Spatial pattern identification and crash severity analysis of road traffic crash hot spots in Ohio. *Heliyon* **2023**, *9*, e16303. [[CrossRef](#)] [[PubMed](#)]
29. Kan, Z.; Kwan, M.; Tang, L. Ripley's K-function for network-constrained flow data. *Geogr. Anal.* **2022**, *54*, 769–788. [[CrossRef](#)]
30. Zheng, M.; Xie, X.; Jiang, Y.; Shen, Q.; Geng, X.; Zhao, L.; Jia, F. Optimizing Kernel Density Estimation Bandwidth for Road Traffic Accident Hazard Identification: A Case Study of the City of London. *Sustainability* **2024**, *16*, 6969. [[CrossRef](#)]
31. Miao, C.; Chen, X.; Zhang, C. Assessing network-based traffic crash risk using prospective space-time scan statistic method. *J. Transp. Geogr.* **2024**, *119*, 103958. [[CrossRef](#)]
32. Cheng, Z.; Zu, Z.; Lu, J. Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections. *Sustainability* **2018**, *11*, 160. [[CrossRef](#)]
33. Mohammed, S.; Alkhereibi, A.H.; Abulibdeh, A.; Jawarneh, R.N.; Balakrishnan, P. GIS-based spatiotemporal analysis for road traffic crashes; in support of sustainable transportation Planning. *Transp. Res. Interdiscip. Perspect.* **2023**, *20*, 100836. [[CrossRef](#)]
34. Mesic, A.; Damsere-Derry, J.; Feldacker, C.; Mooney, S.J.; Gyedu, A.; Mock, C.; Kitali, A.; Wagenaar, B.H.; Wuaku, D.H.; Afram, M.O.; et al. Identifying emerging hot spots of road traffic injury severity using spatiotemporal methods: Longitudinal analyses on major roads in Ghana from 2005 to 2020. *BMC Public Health* **2024**, *24*, 1609. [[CrossRef](#)]
35. Zheng, Z.; Zhou, M.; Chen, Y.; Huo, M.; Sun, L.; Zhao, S.; Chen, D. A fused method of machine learning and dynamic time warping for road anomalies detection. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 827–839. [[CrossRef](#)]
36. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [[CrossRef](#)]
37. Roy, S.; Chowdhury, I.R. Intoxication in the city: Investigating spatial patterns and determinants of drugs and alcohol-related illegal activities in India's geostrategic corridor. *Appl. Geogr.* **2024**, *171*, 103386. [[CrossRef](#)]
38. Roy, S.; Chowdhury, I.R. Brighter Nights, Safer Cities? Exploring spatial link between VIIRS nightlight and urban crime risk. *Remote Sens. Appl. Soc. Environ.* **2025**, *37*, 101489. [[CrossRef](#)]
39. Kathuria, S.; Mathur, P. A policy framework to build on northeast India's strengths. In *Playing to Strengths*; World Bank Group: Washington, DC, USA, 2019; p. 1.
40. Roy, S. Claiming the night: Fear, exclusion, and the right to urban nocturnality. *Hum. Geogr.* **2025**, 19427786251377462. [[CrossRef](#)]
41. Ghosh, A. The importance of being Siliguri: Border effect and the 'untimely' city in North Bengal. In *Logistical Asia: The Labour of Making a World Region*; Palgrave Macmillan: Singapore, 2018; pp. 135–154.
42. Roy, S.; Majumder, S.; Bose, A.; Roy Chowdhury, I. Does geographical heterogeneity influence urban quality of life? A case of a densely populated Indian city. *Pap. Appl. Geogr.* **2023**, *9*, 395–424. [[CrossRef](#)]
43. Roy, S.; Singha, N. Analysis of ambient air quality based on exceedance factor and air quality index for Siliguri City, West Bengal. *Curr. World Environ.* **2020**, *15*, 235. [[CrossRef](#)]
44. Bhattacharyya, D.B.; Mitra, S. Making Siliguri a walkable city. *Procedia-Soc. Behav. Sci.* **2013**, *96*, 2737–2744. [[CrossRef](#)]
45. Costa, M.A.; Kulldorff, M. Applications of spatial scan statistics: A review. In *Scan Statistics*; Birkhäuser: Boston, MA, USA, 2009; pp. 129–152.
46. Kiani, B.; Raouf Rahmati, A.; Bergquist, R.; Hashtarkhani, S.; Firouraghi, N.; Bagheri, N.; Moghaddas, E.; Mohammadi, A. Spatio-temporal epidemiology of the tuberculosis incidence rate in Iran 2008 to 2018. *BMC Public Health* **2021**, *21*, 1093. [[CrossRef](#)]

47. Kulldorff, M. SaTScan User Guide for Version 10.1. 2022. Available online: <https://www.satscan.org/techdoc.html> (accessed on 21 November 2023).
48. Kulldorff, M. A spatial scan statistic. *Commun. Stat.-Theory Methods* **1997**, *26*, 1481–1496. [[CrossRef](#)]
49. Kulldorff, M.; Mostashari, F.; Duczmal, L.; Katherine Yih, W.; Kleinman, K.; Platt, R. Multivariate scan statistics for disease surveillance. *Stat. Med.* **2007**, *26*, 1824–1833. [[CrossRef](#)]
50. Maro, J.C.; Nguyen, M.D.; Dashevsky, I.; Baker, M.A.; Kulldorff, M. Statistical power for postlicensure medical product safety data mining. *eGEMs* **2017**, *5*, 6. [[CrossRef](#)]
51. Stewart, G.; Al-Khassaweneh, M. An implementation of the HDBSCAN* clustering algorithm. *Appl. Sci.* **2022**, *12*, 2405. [[CrossRef](#)]
52. Campello, R.J.; Moulavi, D.; Sander, J. Density-based clustering based on hierarchical density estimates. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Berlin, Heidelberg, 14–17 April 2013; pp. 160–172.
53. Jain, R.; Bhat, A. Determining Statistically Significant Road Accident Spatial Hotspots using Machine Learning Approaches. In *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*; IEEE: Piscataway, NJ, USA, 2022; pp. 214–221.
54. Wang, D.; Huang, Y.; Cai, Z. A two-phase clustering approach for traffic accident black spots identification: Integrated GIS-based processing and HDBSCAN model. *Int. J. Inj. Control Saf. Promot.* **2023**, *30*, 270–281. [[CrossRef](#)] [[PubMed](#)]
55. Cesario, E.; Lindia, P.; Vinci, A. Detecting multi-density urban hotspots in a smart city: Approaches, challenges and applications. *Big Data Cogn. Comput.* **2023**, *7*, 29. [[CrossRef](#)]
56. Sadler, R.C.; Melde, C.; Zeoli, A.; Wolfe, S.; O'Brien, M. Characterizing spatio-temporal differences in homicides and non-fatal shootings in Milwaukee, Wisconsin, 2006–2015. *Appl. Spat. Anal. Policy* **2022**, *15*, 117–142. [[CrossRef](#)]
57. Lucidi, F.; Mallia, L.; Violani, C.; Giustiniani, G.; Persia, L. The contributions of sleep-related risk factors to diurnal car accidents. *Accid. Anal. Prev.* **2013**, *51*, 135–140. [[CrossRef](#)] [[PubMed](#)]
58. Liu, J.; Li, J.; Wang, K.; Zhao, J.; Cong, H.; He, P. Exploring factors affecting the severity of night-time vehicle accidents under low illumination conditions. *Adv. Mech. Eng.* **2019**, *11*, 1687814019840940. [[CrossRef](#)]
59. Eboli, L.; Forciniti, C.; Mazzulla, G. Factors influencing accident severity: An analysis by road accident type. *Transp. Res. Procedia* **2020**, *47*, 449–456. [[CrossRef](#)]
60. Ackaah, W.; Apuseyine, B.A.; Afukaar, F.K. Road traffic crashes at night-time: Characteristics and risk factors. *Int. J. Inj. Control Saf. Promot.* **2020**, *27*, 392–399. [[CrossRef](#)]
61. Gu, Z.; Peng, B. Investigation into the built environment impacts on pedestrian crash frequencies during morning, noon/afternoon, night, and during peak hours: A case study in Miami County, Florida. *J. Transp. Saf. Secur.* **2021**, *13*, 915–935. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.