International Journal of Engineering and Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 13 Issue 12 December 2024 || PP 45-55

A GIS Approach for Spatial Analysis of Traffic Accidents in the city of Hyderabad

Nagulamalyala Saikrishna[1], Prof. D. Rajashekar Reddy[2]
[1]*(M.Tech. Scholar, Department of Civil Engineering, University College of Engineering, Osmania University, Hyderabad, Telangana, India)*
[2]*(Professor, Department of Civil Engineering, University College of Engineering, Osmania University, Hyderabad, Telangana, India)*

ABSTRACT: *Hyderabad is a rapidly growing city and one of Asia's fastest growing cities which also resulting in led to challenges such as, Traffic congestion and increasing in traffic accident rates continuously every year. Identifying hotspots is a crucial step in reducing accident rates and improving the effectiveness of traffic safety management. The main objective of this paper is to identify and ranking of hot spots for different category of accidents, Deaths, Injuries, Property damages only and Overall accidents from the year 2021 to 2023 with the application of Geographical Information System oriented Spatial and Statistical analysis techniques, Kernel Density Estimation (KDE) offers a visual representation of the density or "hot spots" in a given dataset. and Getis-ord Gi* will provide the statistical significant hot spots obtained from KDE results. The results provide valuable evidence for various agencies and can be effectively utilized to prioritize specific locations, implement improved planning, and develop management strategies to reduce traffic accidents.*

KEYWORDS - *Kernel Density Estimation(KDE), Getis-ord gi*, Hot spots, Hyderabad(HYD)*

Date of Submission: 15-12-2024 Date of Acceptance: 25-12-2024

I. INTRODUCTION

Hyderabad, the largest city in the Indian state of Telangana, is a prominent hub for the technology industry. Due to the large development of the road infrastructure the rates of the traffic accidents in the Hyderabad increasing continuously from past three years, 2021 to 2023 by 31.6%.

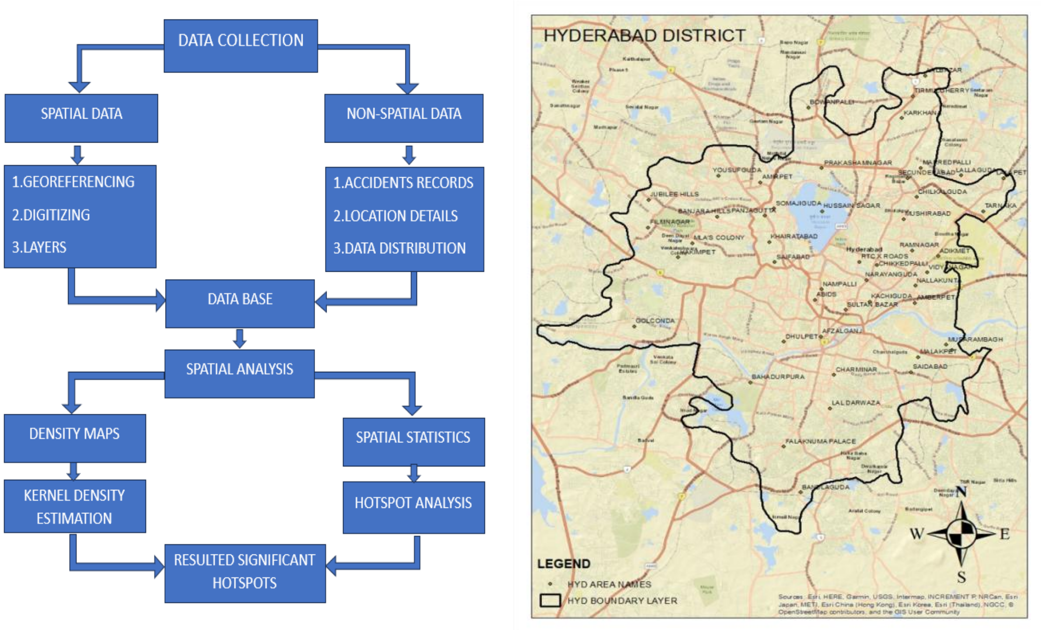
These traffic accidents are a major cause of human deaths and illness worldwide [1]. Injuries from accidents lead to substantial economic losses for individuals, their families, and society as a whole [2]. To reduce the traffic accidents, identification of hot spots is the foremost exercise. This paper highlights the application of spatial analysis to traffic accidents for identifying hot spot locations.

Kernel Density Estimation (KDE) is a non-parametric method used to estimate an unknown probability density function. Hotspot analysis with Gi* identifies statistically significant spatial clustering of high values, referred to as hotspots. The Gi* statistic provides a z-score and p-value to assess the statistical significance of these clusters. While GIS offers limited spatial information and does not provide a deeper understanding of accident patterns for traffic management, KDE offers a color-coded visual representation of accident density from high to low. Gi* is then applied to evaluate the statistical significance of the density values obtained from KDE [3].

In this paper GIS based spatial statistics have been used to identify the hot spot for different category of traffic accidents such as Deaths, Injuries, Property damages only and Overall accidents from the year 2021 to 2023 in the city of Hyderabad. The main objective is to analyse the hot spots for different category of accidents to obtain the information required to the traffic management to take appropriate decisions and proper measurements to enhance the safety in traffic management system and to reduce the accidents.

II. METHODOLOGY

The present study chosen three spatial statistic methods to obtain category wise accident hot spots such as Deaths, Injuries, Property Damage only and Overall accidents from the year 2021 to 2023 in the city of Hyderabad with the application of Geographical Information System(GIS). KDE (Kernel Density Estimation) creates a smooth, continuous surface that represents the density of points, such as traffic accidents, over a spatial area. It highlights regions with higher concentrations, making it easier to identify hotspots and visualize patterns for analysis or decision-making., Incremental Spatial Auto-correlation is used to obtain a threshold distance, using that particular distance the Hot spot analysis, Gi is analyzed. The detail research methodology is explained in the following sub sections. Fig 1 shows the procedure sequence of the research methodology. Study area chosen for the study is the Hyderabad, capital city of the state of Telengana. It is located in the central of the state.



(a) (b)

Fig. 1 Research Methodology(a) and Study area(b).

2.1 Non Spatial Data

While non-spatial data lacks location details like coordinates, it can still be analyzed for patterns, trends, and relationships (e.g., accident severity over time, vehicle types involved, or common causes). Adding spatial components, such as GPS coordinates or location details, can help transform it into spatial data for advanced mapping and GIS analysis. The following data were collected,

a. Date and Time of the accident occurrence.
b. Number of persons Injured or Dead.
c. Area of the accident occurrence.
d. Type of vehicle involved.
e. Demographic data(Age and Gender).

2.2 Spatial Data

2.2.1 Georeferencing

Georeferencing in GIS is the process of aligning a raster image (like a scanned map) to a known coordinate system using reference data. It involves adding control points by matching locations on the raster with corresponding points on the reference layer. The software calculates transformations (e.g., affine, spline) to align the image. After adjusting for accuracy and minimizing residual errors, the georeferencing can be saved to the raster or rectified as a new dataset. This process ensures the raster aligns accurately with other spatial data for mapping or analysis.

2.2.2 Digitizing

Digitizing in ArcGIS is the process of converting features from raster images or scanned maps into vector data (points, lines, or polygons). Using the Editor toolbar, you trace features manually by adding vertices to create new shapefiles or feature classes. This allows you to represent geographic elements like roads, boundaries, or buildings digitally. Digitizing can be done in two ways: manual digitizing, where features are traced by hand. The resulting vector data can then be analyzed, edited, or integrated with other GIS layers for spatial analysis.

2.3 Kernel Density Estimation

Kernel Density Estimation (KDE) is a non-parametric method that uses kernel functions as weights to smooth the influence of each data point across a defined neighborhood.

It is widely used for detecting traffic accident hotspots due to its simplicity and the ease of visualizing results. The entire study area is divided into a grid of predicted cells, and KDE is performed by placing a smooth, symmetric surface over each crash location. The method calculates density per unit area by measuring the distance from each crash location to a reference location using a mathematical function, producing a continuous surface that highlights areas of high and low accident density [4].

The technique was used to estimate the density of crashes within a specified bandwidth (search radius) around each point across the study area. It calculates the magnitude per unit area for each hotspot feature, effectively capturing the spread of risk around crash clusters [5][6].

The mathematical form of each location,

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

Here, f(x,y) represents the density measure at a specific location (x, y), h is the bandwidth, k is the kernel function that depends on the bandwidth and distance, and the distance between points (x, y) is denoted as d_i.

2.4 Incremental Spatial Autocorrelation

The Incremental Spatial Autocorrelation (ISA) tool in ArcGIS runs the Spatial Autocorrelation (Moran's I) tool across a series of increasing distances to measure the intensity of spatial clustering at each distance. Moran's I is a correlation coefficient that quantifies spatial autocorrelation. A Moran's I value closer to +1 indicates clustering, while a value closer to -1 signifies dispersion. The statistical significance of the I index is evaluated using the associated p-values and z-scores [7]. The statistical equation can be written as,

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sigma^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

Here, the variable's value at accident location i and its neighboring location j is represented, while n denotes the total number of accidents. The row-standardized weights quantify the proximity relationship between location i and its neighboring location j, ensuring the spatial influence is appropriately scaled.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = \text{variance}$$

2.5 Getis-Ord Gi*

The G statistics family is particularly useful for measuring the spatial dependence of distributed variables and is often used in conjunction with Moran's I.

Introduced by Getis and Ord, these statistics helps to identify statistically significant spatial clustering. The output of the Gi* statistic includes the Gi z-score and Gi p-value. A high Gi z-score indicates clusters of high values (hotspots), while a low Gi z-score indicates clusters of low values (cold spots) and a Gi z-score close to zero signifies a random distribution of values across the study area [8], the mathematical representation if Gi* is,

$$Gi*(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j}$$

The Gi* statistic describes the spatial dependency of incident i across all n accident locations. Here, d represents the threshold distance, and x_j is the number of accidents at location j is considered. The spatial weight quantifies the relationship between two accident locations, i and j, based on a fixed distance band [1]. The Getis-Ord Gi* statistic is used to identify hotspots of traffic accidents. A high Gi* value indicates clusters of high index values (hotspots), while a low Gi* value signifies clusters of low index values (cold spots) [9]. This tool analyzes each feature in relation to its neighboring features. If a feature has a high value and its neighboring features also have high values, it is considered part of a hotspot [7]. The Getis-Ord Gi* statistic enhances the understanding of spatial relationships and helps identify localized clusters of dependence [10].

III. RESULTS AND DISCUSSIONS

Three years accidents records, 2021 to 2023 were collected from the Department of Traffic Police Station, in the Hyderabad. The total number of accidents are 7785. As noted, 11.97% of Fatal, 76.89% of Non Fatal and 11.13% of Property damage only accidents were occurred. Accidents rate were increased by 31.7% from the year 2021 to 2023.

Majority of the accidents occurred on Monday. Two wheeler and Pedestrian are found to be most Victims whereas Two wheeler and Four wheeler are found to be most Accused. Fig.4 illustrate that Males are most involved than Females in the Accidents. Fig. 7 illustrate that the accidents rates were more from the month of October to December in the seasonal analysis of accidents.

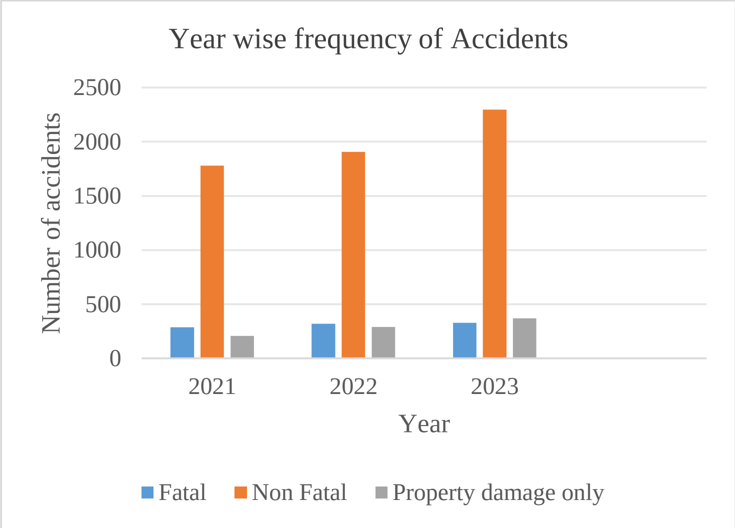


Fig.2 Total number of accidents from the year 2021 to 2023.

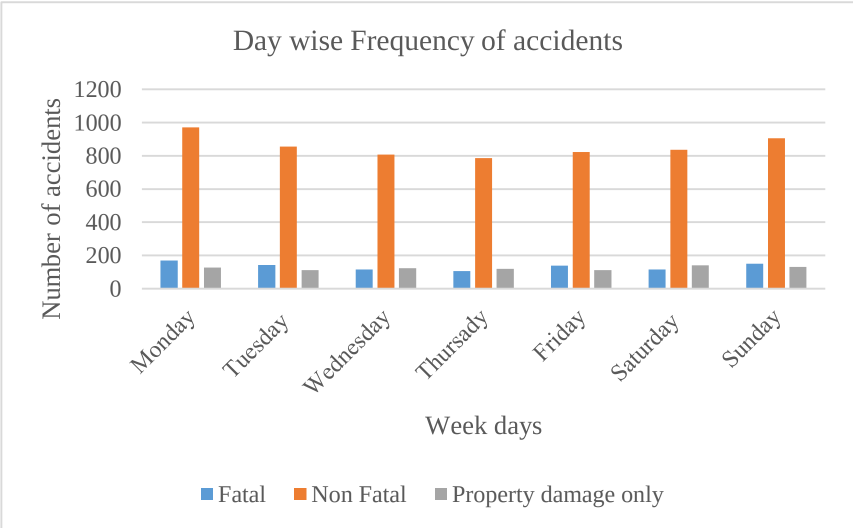


Fig.3 Day wise average frequency of accidents from the year 2021 to 2023.

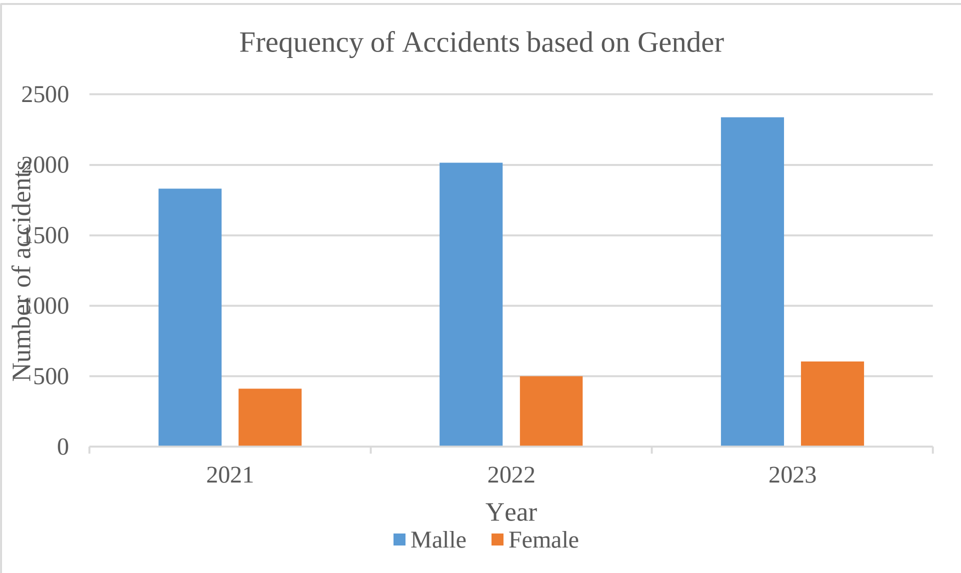


Fig. 4 Average frequency of accidents from 2021 to 2023 based on the Gender.

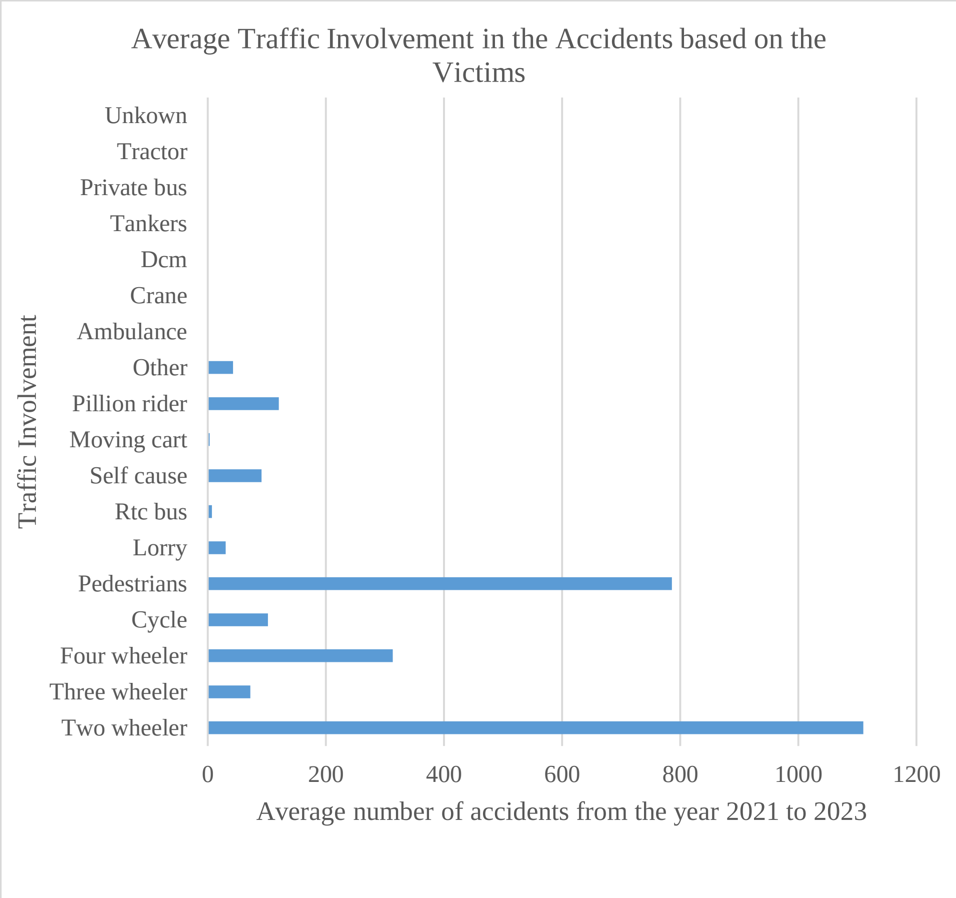


Fig. 5 Average Traffic involvement in the accidents from 2021 to 2023 based on the Victims.

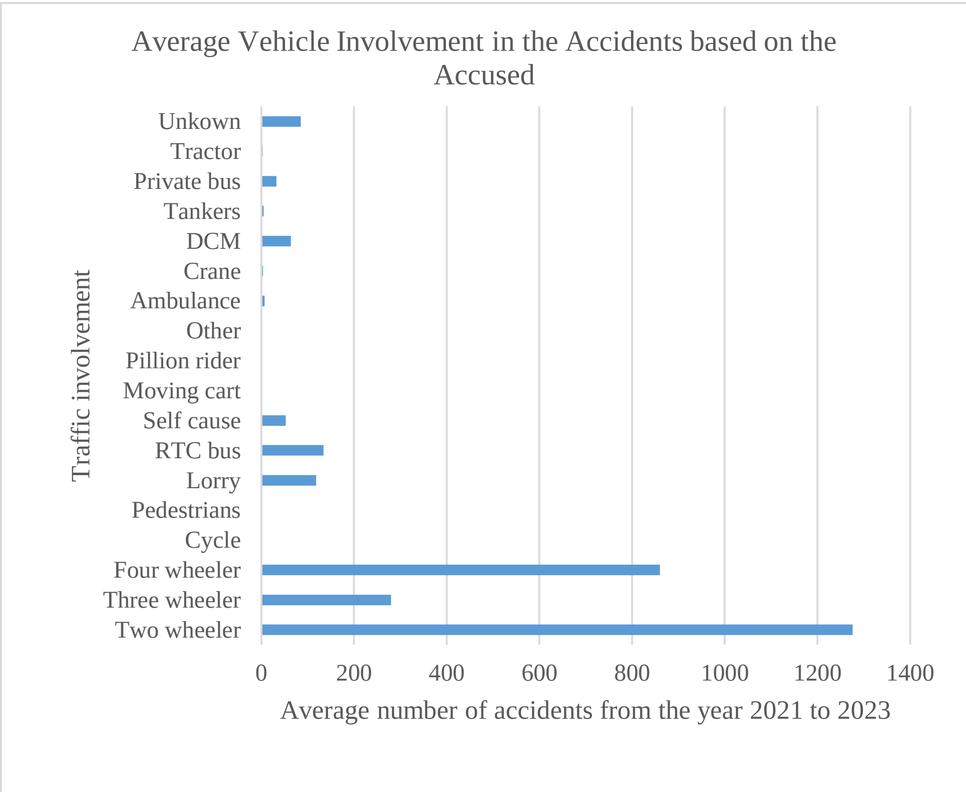


Fig. 6 Average Traffic involvement in the accidents from 2021 to 2023 based on the Accused.

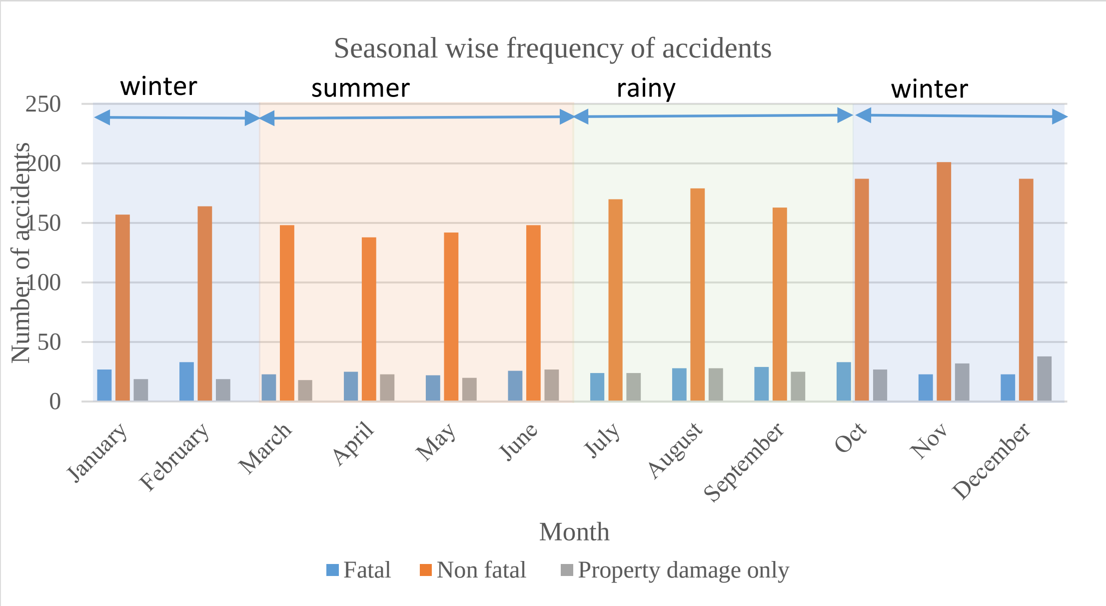


Fig. 7 Seasonal wise average frequency of accidents from the year 2021 to 2023.

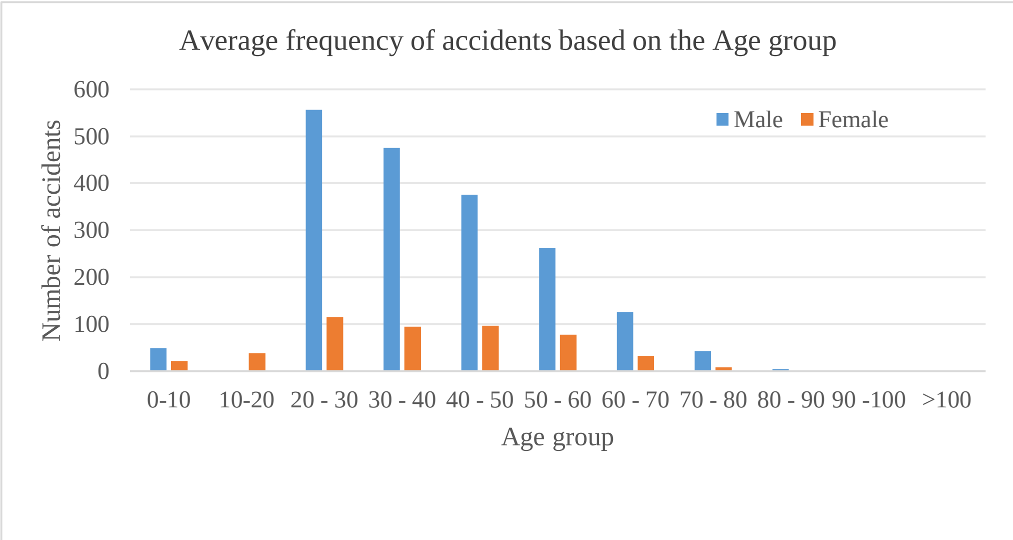


Fig. 8 Average frequency of accidents from the year 2021 to 2023 based on the Age group.

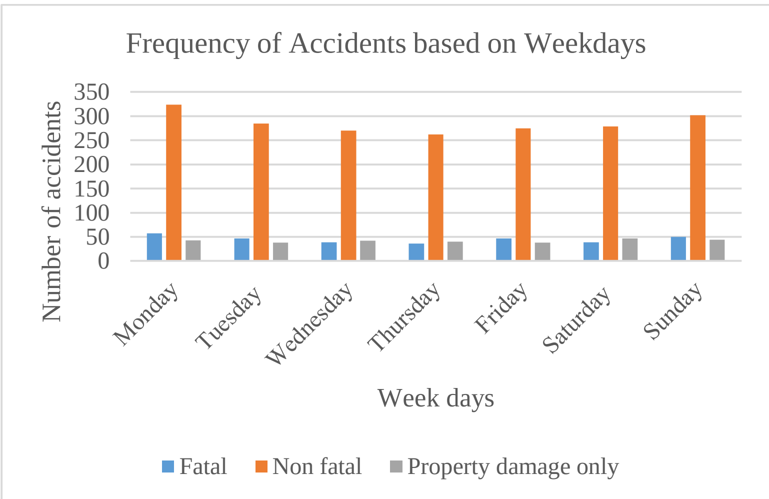
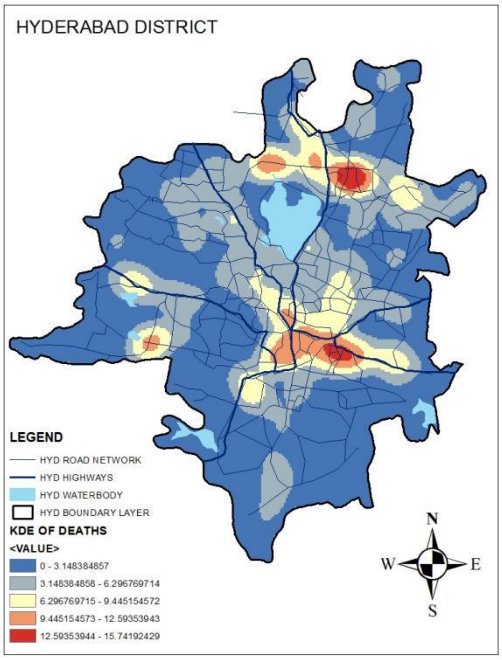


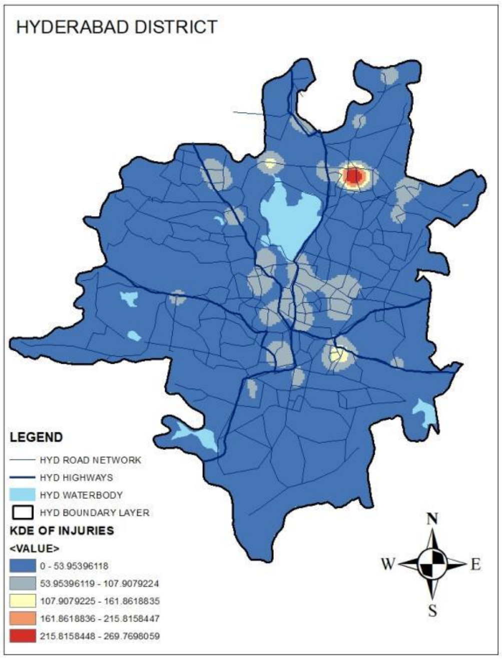
Fig. 9 Average number of accidents based on week days from the year 2021 to 2023.

3.1 Kernel Density Estimation

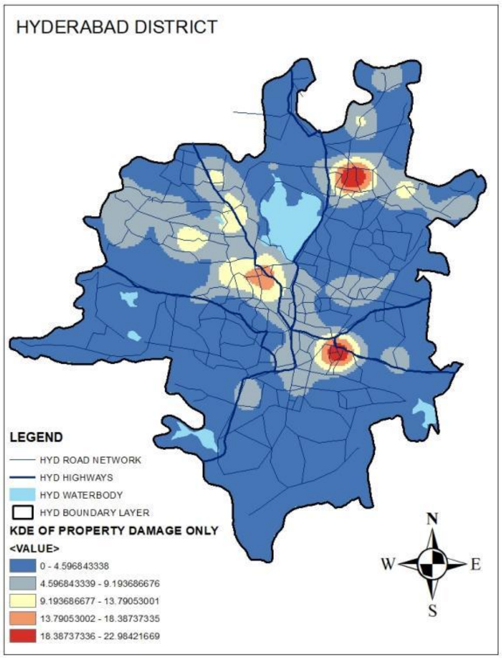
The result of Kernel Density Estimation (KDE) in ArcGIS is a continuous surface (a raster) that represents the concentration or density of point across a geographic area. KDE calculates the density of features per unit area by spreading the influence of each point using a kernel function, creating a smooth gradient of values. High-density areas are shown with warmer colours (e.g., red or yellow), indicating "hotspots" where features, such as Deaths, Injuries, Property damage only and Overall accidents concentrated. Low-density areas are represented with cooler colours (e.g., blue), indicating fewer occurrences. This output helps identify patterns, clusters, or trends in spatial data, enabling better decision-making for resource allocation, planning, or mitigation strategies. Fig. 10 illustrates the KDE for different category of accidents,



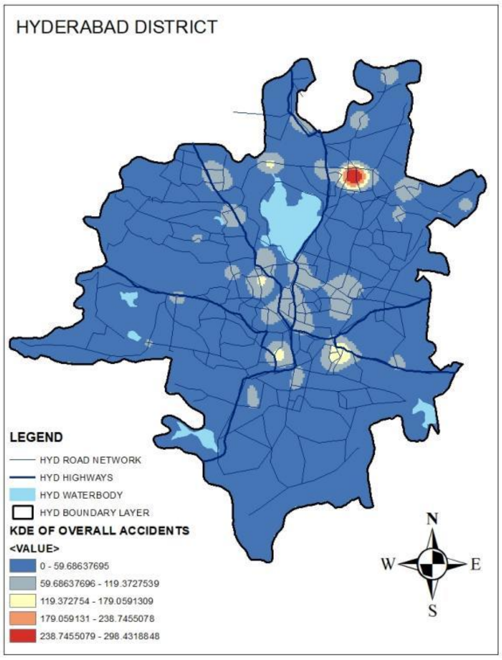
(a)



(b)



(c)



(d)

Fig. 10 KDE of different category of accidents, Deaths(a), Injuries(b), Property damages only(c) and Overall accidents(d).

3.2 Incremental Spatial Autocorrelation(ISA)

The result of Incremental Spatial Autocorrelation in ArcGIS is a graph and associated values that show the strength of spatial clustering for a given dataset at multiple distances. The tool calculates Moran's I index at incremental distances and identifies where spatial clustering (hotspots or patterns) is most statistically significant. The output graph displays peaks, indicating the distances at which clustering is strongest. This analysis helps determine the appropriate scale or distance for further spatial analysis, such as hotspot analysis, ensuring that patterns are analyzed at the most relevant spatial resolution.

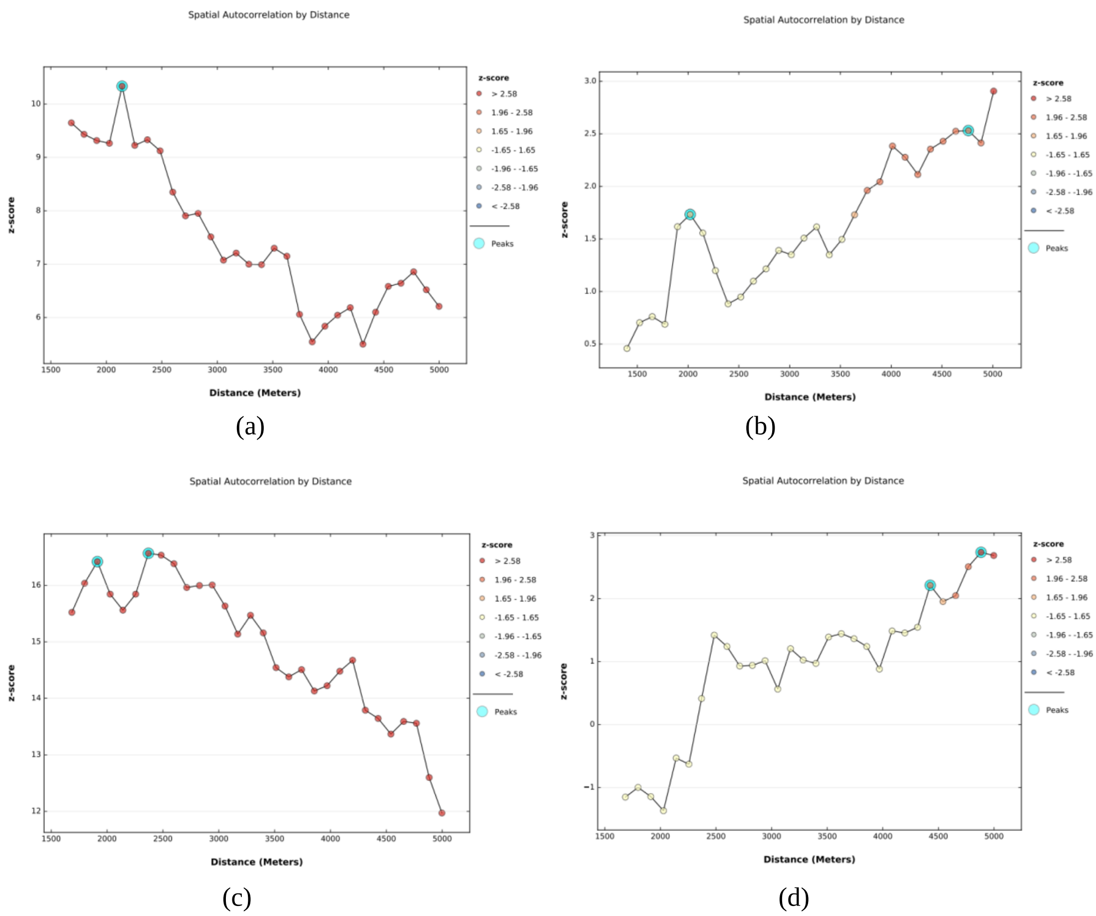


Fig. 11 Incremental Spatial Autocorrelation for different category of accidents, Deaths(a), Injuries(b), Property damage only(c) and Overall accidents(d).

The Incremental Spatial Autocorrelation tool can help identify the optimal threshold distance by analyzing clustering strength at multiple distances. Selecting the appropriate threshold distance ensures accurate identification of statistically significant hotspots and cold spots in the dataset. The threshold distance in Hotspot Analysis (Getis-Ord Gi* statistic) in ArcGIS determines the spatial extent within which features influence each other. It defines the distance at which neighboring features are included in the analysis to calculate the degree of clustering of high or low values.

Fig. 11 shows the Z-Score values for different category of accidents for series of increasing in distance. In total 30 increments were chosen, the peak value shows the highest Z-Score value which is used as a threshold distance for Hotspot analysis. Table. 1 shows the threshold distance for different accident category,

Table. 1 Threshold distance for Different category of accidents.

Accident category	Threshold distance(m)
Deaths	2141
Injuries	4759
Property damage only	2370
Overall accidents	4883

3.3 Hotspot analysis(Getis-Ord Gi*)

The results of the Getis-Ord Gi* statistic in ArcGIS produce a hotspot map that identifies statistically significant clusters of high and low values within the study area. The output is a raster or feature layer where each feature is assigned a z-score and p-value, indicating the intensity and significance of clustering. Features with high positive z-scores represent hotspots (clusters of high values), while features with low negative z-scores indicate cold spots (clusters of low values). Statistically insignificant areas are shown as neutral. This analysis helps reveal spatial patterns, such as accident-prone zones supporting targeted decision-making and resource allocation. Fig. 12 illustrates the hotspots of different category of accidents,

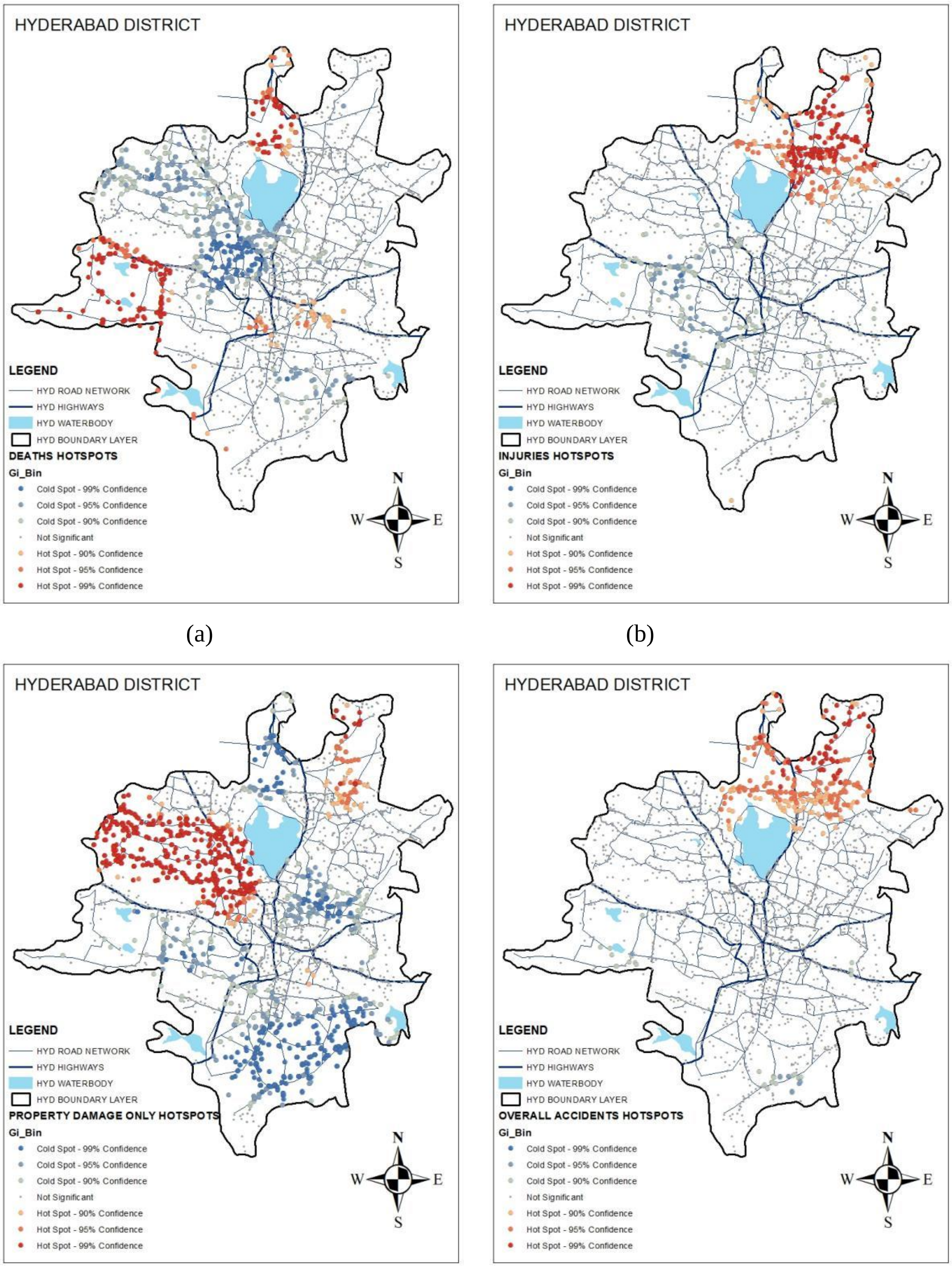


Fig. 12 Hotspot analysis of different category of accidents, Deaths(a), Injuries(b), Property damage only(c) and Overall accidents(d).

When comparing density values from KDE (Kernel Density Estimation) and z-scores from the Getis-Ord Gi* statistic for identifying statistically significant hotspots, the two approaches provide complementary insights into spatial patterns. Density values from KDE represent the concentration of events or features in a continuous spatial surface. These values indicate areas with high or low densities but do not inherently account for statistical significance. KDE simply visualizes where events are concentrated based on a smoothing parameter, offering a descriptive understanding of the data.

In contrast, z-scores from the Getis-Ord Gi* statistic quantify the statistical significance of spatial clustering. Positive z-scores indicate potential hotspots, with higher scores representing more intense clustering, while negative z-scores point to cold spots. Importantly, z-scores are accompanied by p-values, which determine whether the clustering is statistically significant or could occur by random chance. The ranking of the hotspots is shown the following tables,

Table. 2 Ranking of Hotspots of Deaths

Rank	Location	KDE	Z-Score
1	Chaderghat	13.88	2.53
2	Afzalgunj	12.15	2.35
3	Langerhouse	11.43	3.389
4	Begumpet	10.64	4.816
5	Tolichowki	8.84	4.664

Table. 3 Ranking of Hotspots of Injuries

Rank	Location	KDE	Z-Score
1	Secunderabad	267.45	3.14
2	Begumpet	114.52	2.16
3	Mettuguda	100.86	2.66
4	Bowenpally	82.07	1.88
5	Thirumalgherry	71.53	2.85

Table. 4 Ranking of Hotspots of Property damage only

Rank	Location	KDE	Z-Score
1	Secunderabad	22.98	2.53
2	Malakpet	21.63	1.94
3	Lakdikapool	16.31	3.95
4	Panjagutta	12.53	5.57
5	Banjarahills	11.54	4.95

Table. 5 Ranking of Hotspots of Overall accidents

Rank	Location	KDE	Z-Score
1	Secunderabad	298.431	2.04
2	Begumpet	125.52	2.56
3	Bowenpally	90.50	2.43
4	Karkhana	87.45	3.01
5	Thirumalgherry	81.77	3.40

IV. CONCLUSIONS

The accident rates were continuously increased from the year 2021 to 2023. The higher number of traffic accidents on Mondays can be attributed to the stress and fatigue of transitioning back to work routines, increased traffic congestion during morning commutes, and reduced focus from weekend activities. These factors make Monday a particularly risk-prone day for road users.

This paper used combination of spatial statistical methods for the purpose of different category of accidents(Deaths, Injuries, Property damage only and Overall accidents) spatial clustering. KDE were used to get the density values of accident occurrence and Getis-Ord Gi* is adopted to get the statistical significant hotspots from the results of KDE.

KDE effectively identifies accident-prone areas by highlighting regions with high accident concentrations on a continuous density map. This visualization helps prioritize safety measures and allocate resources to critical hotspots. Getis-Ord Gi* statistic identifies hotspots by analyzing spatial clustering and determining areas with statistically significant concentrations of accidents. It highlights regions with unusually high or low occurrences, providing a robust basis for targeted safety interventions.

These ranked locations act as a strategic tool for transportation authorities, guiding their investigative efforts to identify risk factors at specific sites. This approach not only highlights high-accident areas but also ensures the statistical reliability of these patterns, enabling the efficient use of limited financial resources to reduce both accident frequency and severity. The findings offer valuable evidence for various agencies, supporting the prioritization of critical locations, improved planning, and the development of targeted management strategies to mitigate traffic accidents.

REFERENCES

[1] ICMR, "Development of a Feasibility module for road traffic injuries surveillance.," *Ind Council Med Res Bul*, pp. 42-50, 2009.

[2] Yiwei Feng, "Formulating an Innovation Spatial-autocorrelation-based method for identifying road accidents Hot zones," in *IOP Conference Series: Earth and Environment Science*, pp. 1-7, 2020.

[3] Lakshmi Srikanth, "A case study on Kerenl Density Estimation and Hotspot Analysis methods in traffic safety management," in *12th International Conference on Communication Systems and Networks*, pp. 99-104, 2020.

[4] Silverman B, "Density Estimation for Statistic and Data Analysis," 1986.

[5] Anuradha K.C, "Spatial Analysis of Road Traffic Crashes Hotspots in Kathmandu Valley, Nepal," in *Proceedings of 15th IOE Graduate Conference*, pp. 326-333, 2024.

[6] A Stewart Fotheringham, "Quantitative Geography: Perspectives on Spatial Data Analysis," 2020.

[7] PrasannaKumar, "Spatiotemporal Clustering of road accidents: GIS based analysis and assessment," *Procedia Social Behaviour Sciences* 2011, pp. 317-325.

[8] Songchitruska, "Getis_Ord Gi* Statistics for Identifying Hotspots using Incidents management data," in *Proceedings of Transportation Research Board 89th Anuual Meeting*, 2010.

[9] Romi Satria, " GIS tools for Analyzing accidents and Road Design," *Transportation Research Procedia 18*, pp. 242-247, 2016.

[10] Elizabeth Hovenden, "Use of Spatial Analysis Techniques to Identify Statistical Significance Crash Hotspots in Metropolitan Melbourne," *Journal of Road Safety* , pp. 36-58, 2020.