Comparative Study on Spatial Clustering Methods for Identifying Traffic Accident Hotspots

Mohd. Minhajuddin Aquil[1*], Mir Iqbal Faheem[2]

[1] Research Scholar, School of Engineering, Career Point University, Rajasthan, India
[2] Research Supervisor, School of Engineering, Career Point University, Rajasthan, India

*Corresponding author: aquilminhaj@gmail.com

doi: https://doi.org/10.21467/proceedings.112.64

ABSTRACT

Traffic accidents in an urban road network are inevitable as a result claims and disputes arise among different road users. It is imperative to estimate the likelihood of traffic accidents resulting from different factors that contribute to loss of life, property and health of road users. There is a pressing need to reduce traffic accidents by identifying the location of accident hotspots using suitable analysis methods and examining them which is essential for the safety of road users. In this research traffic accident hotspots are identified using two spatial clustering analysis methods namely Getis-Ord Gi* and Nearest Neighborhood Hierarchy (NNH). These methods are compared and evaluated using the Prediction Accuracy Index (PAI) for their degree of accuracy. In this study, a cumulative traffic accident data of Hyderabad city of Telangana state over four years is researched upon and considered. Getis-Ord Gi* analysis measures the concentration ratio based on Z score identified as high (positive Z-values) and low values (negative Z-values). NNH analysis is another spatial clustering method which displays hotspot regions in the form of Convex hulls and Ellipses. The choice of the above two clustering methods represents the significance of the precision required. The findings of the study reveal that NNH method performed better compared to Getis-Ord Gi* method in its ability to detect hotspots. The above research methodology can be performed to any size of road network area globally having relevant accident data for the identification of hotspots for reducing the traffic accidents.

Keywords: Urban Road Network, Accident Hotspots, Nearest Neighborhood Hierarchy, Getis-Ord Gi*, Prediction Accuracy Index.

1 INTRODUCTION

Transportation system is imperative for economic activities and facilitates the movement of both passenger's and freights. Traffic accidents are not a modern phenomenon and transport hazards have been occurring from the early stages of human movement. Traffic safety management is a significant topic around the world, both at the national and global level. It is a matter of serious concern that more than 2,31,000 people are killed in road traffic crashes in India every year. Approximately half of all deaths on the country's roads are among vulnerable road users such as motorcyclists, pedestrians and cyclists (WHO 2013). The road transport system is the most important infrastructural system which has a vital role in the performance of other activities. One of the aspects of a road transport system that must be considered is road safety. (Mohammad Ali Aghajani, 2017). Therefore, it is urgently necessary to understand the factors that lead to pedestrian traffic accidents and to identify their traffic accident patterns to provide a better and safer pedestrian travel environment. (Lin Hu, 2020). There is a pressing need to reduce traffic accidents by identifying the location of accident hotspots using suitable analysis methods and examining them which is essential for the safety of road users.

Proceedings DOI: 10.21467/proceedings.112; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-947843-3-3

A geographic information system (GIS) creates useful information to help decision making. It is a computer-based tool which captures, stores, analyze, and displays geospatial data. Considering the evolution of geomatics in recent years in the field of road infrastructure, many researchers have carried out studies focusing on the development and implementation of geographic information-based tools for effective planning (Niloofar Haji Mirza Aghasi, 2017). Additionally, GIS has enabled researchers to develop diverse traffic accident analysis and provide more functions for analyzing locations with high accident rates (Jau-Ming Su, 2018). Since traffic accidents have geographic and spatial properties, GIS tool can be used to identify dangerous locations and thus formulate road safety policies by understanding the spatial positions of traffic accidents and related factors. Generally, traffic safety practitioners would focus on traffic accidents at some particular "hotspot" first, since it may not be realistic for each city to tackle all roadway accidents problems simultaneously due to limited resources (Yang, 2013). Recently the numbers of studies for analyzing accident pattern and improving road design have increased considerably. Among these tools, Geographical Information Systems (GIS) stand out for their ability to perform and display complex spatial analysis for a given road network. (Romi Satria, 2016).

The purpose of this research is to identify the high-risk potential locations using two spatial clustering analysis methods namely Getis-Ord Gi* and Nearest Neighborhood Hierarchy (NNH). Subsequently, these methods are compared and evaluated using the Prediction Accuracy Index (PAI) for their degree of accuracy and ability to detect hotspots. A cumulative traffic accident data of Hyderabad city of Telangana state over four years is researched upon and considered.

1.1 EARLIER STUDIES

In this section, the previous work carried out on profiling hotspots using different methods and models with the overview of the relevant researchers works is discussed. (Deepthi J.K, 2010) identified the accident-prone zones in Kannur district, Kerala. the road accident data for three years and parameters like date, location, type of vehicle involved, number of persons injured or died were included in the study. GIS tool was used, simple and Kernel densities were applied in identifying the accident patterns. The road geometry was measured in the frequently occurring accident locations to find out the causes for the accident. Based on the result, suggestions were provided to decrease the accidents in the future. (Kuo P.F, 2011) compared the police dispatch time with two conditions such as Police patrol routes with organized hotspots and Police patrol route patterns without focusing on hotspots. Crime and road accident data were collected within the city limits serviced by the College Station Police Department from January 2005 and September 2010. The results indicated that using DDACTS principles police dispatch time can potentially be reduce by 13% and 17%, using the top five, and top 10 hot-spot routes, respectively.

(Blazquez et al., 2013) has researched and found seven critical areas in Santiago with high child pedestrian crashes using GIS employing kernel density estimation and Moran's I index test. Also, it was identified a positive spatial autocorrelation on crash contributing factors while a random spatial pattern was found for crashes related to the age attribute. With respect to gender, weekday, and month of the year no statistical significance in the spatial relationship was obtained in child pedestrian crashes.

(Hao Ye, 2014) presented GIS based approach for the production of digital hotspot maps, based on a historical accident dataset and geospatial methods in GIS. In this approach, two methods were used such as Kernel Density Estimation (KDE) method and Percent Volume Contour (PVC) method to identify hotspot distribution, finally the map layers of hotspot patterns were integrated with classical navigation maps. The derivation of hotspot property data is also discussed following a description for geospatial hotspot production.

(Vivek & Saini, 2015) discussed the accident pattern in India. WSI method was employed for profiling hotspot locations on National Highway - 3 in the District Una, State of Himachal Pradesh, India. (Choudhary J, 2016) carried out spatial clustering of accidents and spatial densities of hot spots of Varanasi city, India using Moran's I method. The statistical techniques such as Getis-Ord Gi* statistics and Kernel Density Estimation were employed and compared using five years of accident data. Results showed that the identification of hot spots by K and Gi* using three of spatial relationships like fixed distance band, inverse distance and inverse square distance were widely similar.

(Shafabakhsh G.A, 2017) studied four clustering analyses techniques using GIS to have a better understanding of collision prone locations in road network. KDE method was used in ARCMAP and two other analyses tools were used in SANET software. Finally, both results of network analysis can be compared with traditional KDE method. This study observed accidents were more clustered than expected by random chance in the selected region. This study also highlighted the importance of usage of GIS tool as a management system for accident analysis by combination of spatial-statistical methods. (Bunnarong S & Upala P, 2018) identified pedestrian crash zones of primary schools and secondary schools in Bangkok, Thailand using GIS. Moran's I statistic and the Kernel Density Estimation (KDE) were applied and results shown that accident locations within school zone were a clustered pattern considering Moran's Index and Pedestrian crash zones results were different with KDE. The researcher found that, the most substantial areas were urban crowded networks, intersections of arterials and local roads about 508 meters from the school center. (Omari A, 2019) predicted accident hot spots in different locations using Weighted Overlay Method (WOM) and Fuzzy Overlay Method (FOM), which are widely used in decision making and alternatives analysis on the basis of results found from Analytic Hierarchy Process (AHP). Additionally, accident hot spots were predicted performing GIS and Fuzzy logic. Point Density (PD) method was also used to verify hot spots in urban region that resulted from the mentioned two methods. Accident data of Irbid City in Jordan was researched between the year 2013 and 2015. Both WOM and FOM showed to be successful in finding hot spots in parts of study area when verified to PD surface.

(Hu L, 2020) evaluated pedestrian crashes that occurred frequently in the main urban area of Changsha city, China, from the year 2014 to 2016. The results demonstrated that there were several clusters of pedestrian crashes in urban areas, which are related to the population, road network, regional functional zoning and social and economic features. Specially, the severity of pedestrian casualties has strong relationships with parameters like darkness, lighting conditions, road isolation facilities and pedestrian age and behavior. It has been observed that casualties are more severe at night than during the day, and school-age children and elderly pedestrians tend to suffer more.

Road traffic accident analysis studies aim at the identification of frequently occurring accident locations and safety deficient areas. This literature observed that road traffic accidents are major social problem across the globe. Hence detailed investigations are required to provide solutions and to propose effective strategies for avoiding future accidents.

1.2 RESEARCH GAP IDENTIFIED

As many authors developed different models using various methods such as KDE, Kriging and NNH and contributed their work during their research. Literature survey showed that most of the work is concentrated on KDE. It is also observed that least work has been carried out for identification of accident hot spots comparing NNH and Getis-Ord Gi* methods on the selected stretch in the urban areas representing a further gap in research.

2 RESEARCH METHODOLOGY

In this research traffic accident hotspots are identified using two spatial clustering analysis methods namely Getis-Ord Gi* and Nearest Neighborhood Hierarchy (NNH). No previous research has analyzed road traffic accident hotspots using these methods in the selected study area. Outline of the research methodology for the intended study is presented in Figure 1.

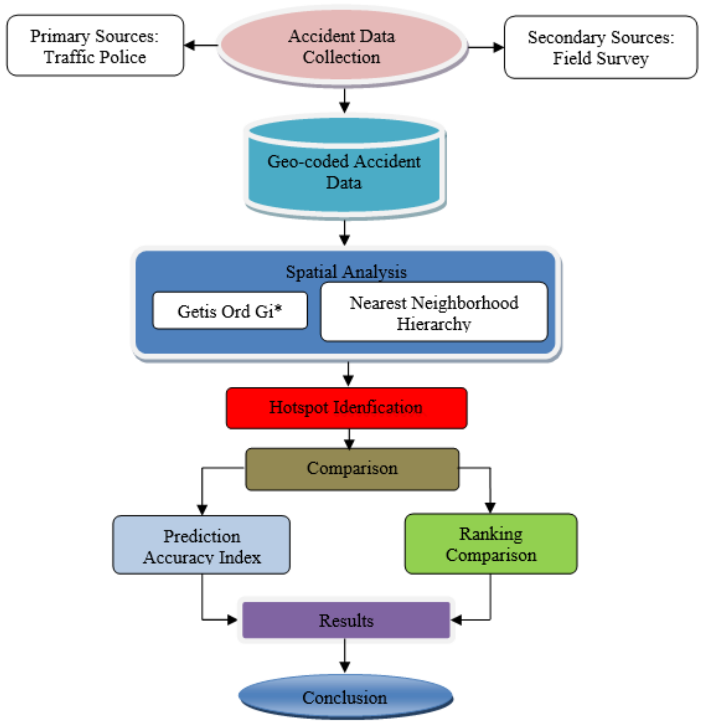

Figure 1. Research process

2.1 Getis-Ord Gi* Method:

Getis-Ord GI* function analyze mapping of clusters for selected traffic accident data. Gi analysis measures the concentration ratio based on Z score identified as high values (positive Z-values) and low values (negative Z-values). It is used to study the indication of identifiable spatial patterns. Equation 1 shows the calculation of G(d) values.

$$G(d) = \frac{\sum\sum wij(d)xixj}{\sum\sum xixj} \quad i \neq j \tag{1}$$

Where x_i is the value at location i, x_j is the value at location j, and if j is within d of I, and w_{ij} (d) is the spatial weight. The weights are frequently maintained by some weighted distance. This process works by observing at each feature within the context of neighbouring features. A feature with a high value is exciting but may not be a significant hot spot scores and values as shown in Figure 2. The resultant z-scores and p-values tell you where spatial features lie with respect to threshold associated based on statistical significance whether it is high value or low value clusters. This method helps in identifying strategically significant spatial clusters as hot spots and cold spots.

Proceedings DOI: 10.21467/proceedings.112
ISBN: 978-81-947843-3-3
Series: AIJR Proceedings
ISSN: 2582-3922

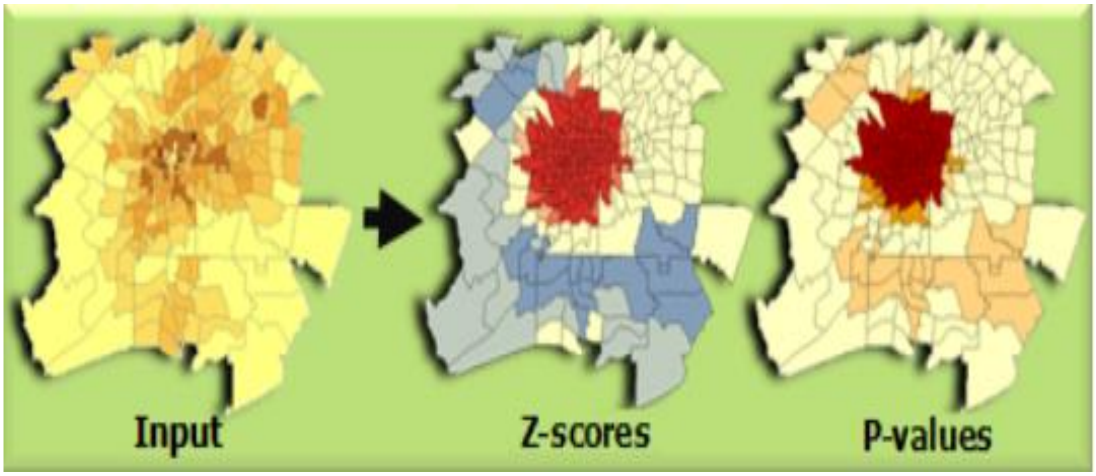

Figure 2. Hotspot scores and values

A high z-score and low p-value specify a spatial clustering of low values. The upper the z-score, the stronger is that the clustering. The z-score is predicted on the randomization null hypothesis calculation. Getis-Ord Gi* is employed to feature statistical significance of hotspot analysis and for predicting where accidents occurs.

2.2 Nearest Neighborhood Hierarchical Method:

In this study Nearest Neighborhood Hierarchical (NNH) Clustering is obtained using CrimeStat software that outputs clusters in two formats i.e., convex hulls and ellipses. It is another method of spatial clustering which is used to detect hotspots in a road network. The user can visualize the clusters in these two diverse formats. Convex hulls are polygons that covers exactly all the clustered points. For a detailed analysis convex hulls are ideal when compared to ellipses as convex hull is more precise and can define the actual area where the hot spot occurs. Meanwhile, Ellipse is more like a symbolic representation of the cluster and looks better on a map eventually easily understood by users. As can be seen in the Figure 3, the ellipses can extend beyond the actual sites but also can remove part of the hot spot area, they are a mathematical abstraction (Levine, 2010).

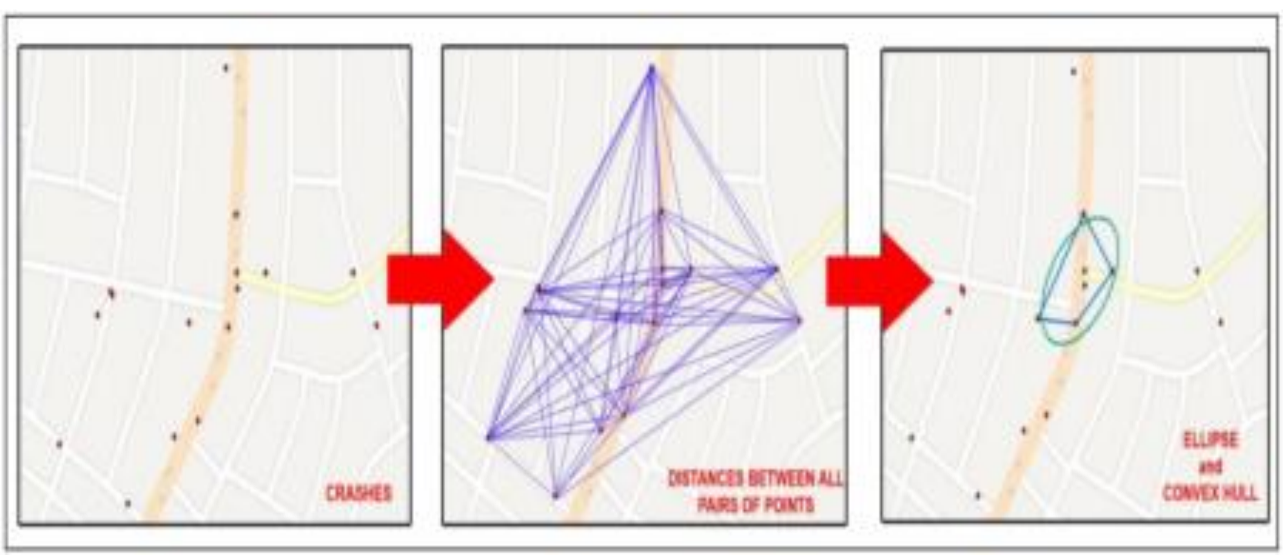

Figure 3. NNH Clustering

The hotspots are identified by selecting a specific threshold distance (d), the point pairs with smaller distances are clustered together. If defined numbers of observations are closer than the threshold value, a new cluster is generated. Calculation of first order clusters, the second and high order clusters are carried out that can form with the same manner until only one cluster is left or the threshold conditions fails.

2.3 CASE STUDY

More recently road traffic accidents became a major problem in India, predominantly in Hyderabad. The city records around 2,500 accidents annually, though the number of deaths due to accidents has reduced but there has been no reduction in the accident numbers yet (Faheem, 2019). With the growing number of top education institutes, Giant companies like IKEA, AMAZON and US-based IT firms like Google, Microsoft, Samsung, Motorola opting for Hyderabad city as their trade Centre. As many people are migrating here ending up making the city commercial which is ironic. With the boost of the IT industry, a lot of job opportunities have opened up for both technical and non-technical positions. Due to urbanization, citizens are facing problems in day-to-day life like traffic congestion, air and noise pollution, and could even worsen if the infrastructure is not upgraded to meet population demands. Hyderabad has a dense road network with a greater number of road traffic accident cases, making it an ideal location for the intended study as shown in Figure 4.

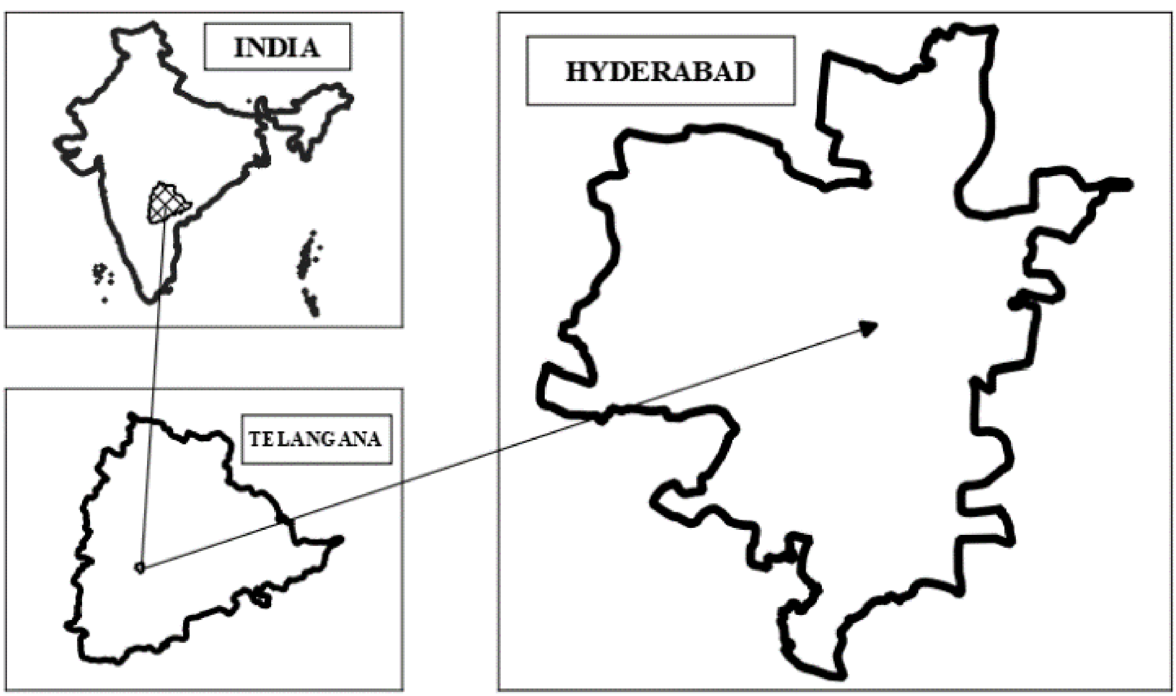

Figure 4. Study Area Location

2.4 ANALYSIS

The research process presented in Figure 1 is demonstrated through a case study shown in Figure 4 of a road network of city of Hyderabad in the State of Telangana adopting the two distinctive Spatial clustering methods Getis-Ord Gi* and NNH for identification of traffic accident hotspots. QGIS and CrimeStat softwares are used to evaluate critical accident locations. Four-year accident data was collected from Hyderabad Traffic Police Department for analysis propose.

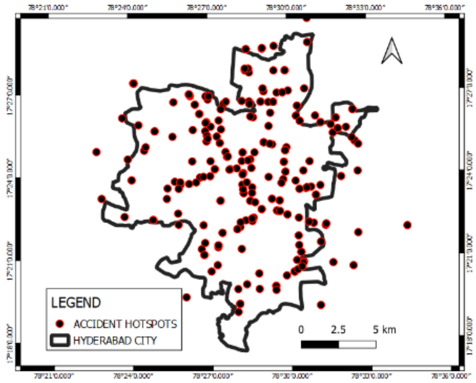

Figure 5. Accident Location in Area under Study

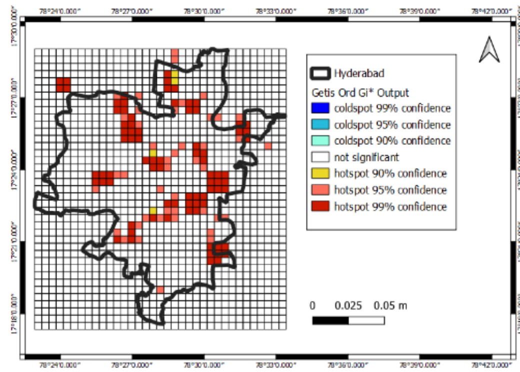

Figure 6. Output of Getis-Ord Gi* using QGIS

Tabulation of accident data is carried out followed by Geocoding and files are saved in .csv format. It is then added in vector layer and attribute table is listed. After accident data is uploaded into the software, immediately spots or points are displayed as shown in Figure 5. These points are accident locations that have occurred in the past four years. The accident data is represented graphically through image in QGIS. Since no country or region has a specific regular shape, hence grids are formed to have a uniform pattern for the analysis of the study area. Once the grids are formed, the analysis is run using local Getis-Ord Gi* tool. Output of Getis-Ord Gi* is shown in the Figure 6. It represents strategically significant hotspots. The degree of redness indicates high risk accident locations.

Another Spatial clustering method is NNH clustering method which is obtained from CrimeStat Software. In this study the hotspots are identified by selecting 1km threshold distance (d), the point pairs with smaller distances are clustered together. Firstly, accident data shape file is uploaded as primary file and measurement parameters such as area of the study and length of the stretch of each road are specified. Hotspot analysis is then run using NNH tool. Number of standard deviations for the ellipse is assumed to be 1x. Calculation of first order clusters, the second and high order clusters are carried out that can form with the same manner until only one cluster is left or the threshold conditions fails.

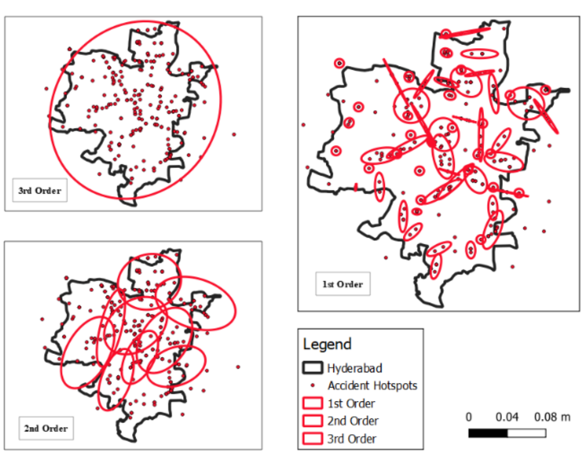

Figure 7. Accident Hotspot Locations using NNH (Ellipses)

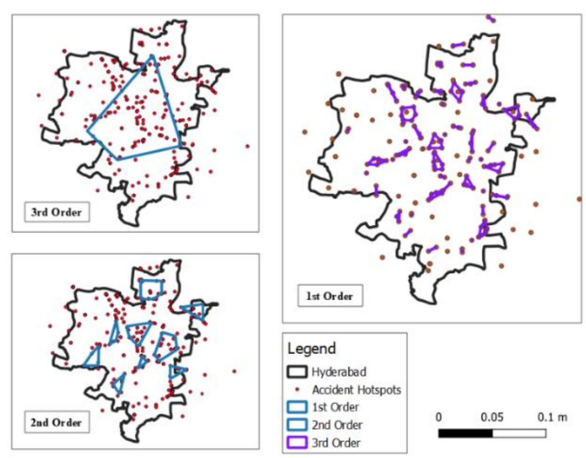

Figure 8. Accident Hotspot Locations using NNH (Convex Hulls)

It can be clearly observed that the regions having a greater number of accident hotspot locations are clustered as shown in Figure 7 and 8. As stated earlier the user can visualize NNH output in two formats, Convex Hulls are polygons that covers exactly all the clustered points and Ellipses are more like a symbolic representation of the cluster and looks better on a map eventually easily understood by users. NNH routine have six outputs such as Order of cluster pattern, Number of points, Number of clusters, Standard deviation of ellipse, area of ellipse and density of cluster as shown in Figure 7 and 8.

2.5 COMPARISON OF SPATIAL CLUSTERING METHODS

Prediction accuracy index (PAI) is used to compare and evaluate the performance of the two proposed methods for their degree of accuracy. The equation for calculating PAI is given in equation (2).

$$\mathrm{PAI} = \frac{n/N\ X100}{m/MX100} \tag{2}$$

where n is the number of accidents in hotspots, N is the total number of accidents, m is the area covered by hotspot, and M is the total area covered. The higher the PAI value indicates the better performance of the proposed method. The findings of the study reveal that NNH method performed better compared to Getis-Ord Gi* method in its ability to detect hotspots.

3 RESULTS AND DISCUSSIONS

The NNH and Getis-Ord Gi* spatial clustering models were developed and employed for analyzing the traffic accident hotspot considering the road network of city of Hyderabad. The results obtained by Getis-Ord Gi* model resulted in high and low clustering values. This classification provides colour coded map that gives a clear visualization of road traffic accidents hot spots. Increase in the degree of redness indicates the higher risk of accidents prone regions as shown in Figure 6. NNH shows clusters of first, second and third order as presented in Figure 7 and 8. Further, boundary of convex hulls is specified from the points identified in hotspot as shown in Figure 8. On the other hand, ellipses not only cover the hotspots but also beyond it as shown in Figure 7. Finally, both methods depict similar clustering pattern and according to the study the areas under high risk of road traffic accident hotspots were identified as follows: 99% confidence of road traffic accident are occurring frequently are Amberpet, Golnaka X roads, Fever hospital X road, Chaderghat Signal, Chaderghat Chaman, Afzalgunj Bus stop, old P.S Santoshnagar, DRDO road, Attapur Pillar No. 32 to 40 & 47 to 50, Panjagutta junction, Near ESI hospital Erragadda and Bowenpally as shown in Table 1. Research questions raise which method of spatial clustering need to be adopted for more accuracy which resulted in high necessity to assert using PAI Index. Prediction Accuracy Index (PAI) indicated NNH method performed better compared to Getis-Ord Gi* method in its ability to detect hotspots.

Table 1. Traffic Accident Hotspot Locations

Parameter	Spatial Clustering Analysis	
	Getis-Ord Gi*	NNH
High and Low Value Clusters	Identified as referred in Figure 6	-
Convex Hulls and Polygons	-	Identified as referred in Figure 7 and 8

ISBN: 978-81-947843-3-3
ISSN: 2582-3922

4 CONCLUSION

This research explored the qualitative and quantitative traffic parameters influencing the accident hotspots on the hazardous locations. Spatial Clustering methods are employed to evaluate the vehicular road accident in complex networks in area under study. This research concluded as follows: 1) Traffic congestions is a core problem in urban transport management due to increase number of trips and travel time. 2) Potholes is another cause of accidents because these are not visible clearly during night time for two-wheelers. It is recommended to consider injuries and fatalities for reducing urban road traffic accident hotspots. Other spatial analysis methods can be employed for further research like IDW, Kriging, etc.

5 ACKNOWLEDGEMENT

The authors wish to acknowledge the Hyderabad traffic police department for providing data which is used in this research. Also, anonymous review of this paper is gratefully thanked.

How to Cite this Article:

Aquil, M. M., & Faheem, M. I. (2021). Comparative Study on Spatial Clustering Methods for Identifying Traffic Accident Hotspots. *AIJR Proceedings*, 535-543.

REFERENCES

Anitha S. (2016). "*Identification of Hotspots of Traffic Accidents Using GIS*", International Journal of Advances in Engineering and Technology, 7(3),13-16.

Bhagyaiah M. and B. Shrinagesh. "*Traffic Analysis and Road Accidents: A Case Study of Hyderabad using GIS*", In: 7th IGRSM International Remote Sensing & GIS Conference and Exhibition, pp. 1-9, IOP Publishing, (2014).

Carola A. Blazquez and Marcela S. (2013) "*A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile*", Journal of Accident Analysis and Prevention, 304 – 311.

Faheem M.I and Taheer., (2019) "*Identification of Blackspots and Development of Prediction Model using Multiple Linear Regression Model*", International Journal of Research and Analytical Reviews, 6(2), 680-687.

Gebretensay B. F. (2018) "*Road Traffic Accident Analysis and Prediction Model: A Case Study of Vadodara City*", International Research Journal of Engineering and Technology, 5(1), 191-196.

Jau-Ming Su, Yu-Ming Wang, Chih-hung Chang and Pei-Ju Wu (2018) "*Application of a Geographic Information System to Analyze Traffic Accidents Using Nantou County, Taiwan, as an Example*", Journal of the Indian Society of Remote Sensing, Springer.

Lin Hua, Xianhui Wu, Jing Huang, Yong Peng and Weiguo Liu (2020) "*Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China*" Safety Science 127, 104710.

Mohammad Ali Aghajani, Reza Shahni Dezfoulian, Abdolreza Rezaee Arjroody and Mohammadreza Rezaei "*Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics (case study: Ilam province, Iran)*" In: World Conference on Transport Research – WCTR, Shanghai, Transportation Research Procedia 25C, 2131–2143, (2017).

Niloofar Haji Mirza Aghasi (2017), "*Introducing GIS as legitimate instrument to deal with road accident data: a case study of Iran, Tehran*", Spatial Information Research, Springer.

Pandey. (2017) "*Identification and Improvement of Accident Black Spots on N.H.86 District Sagar, Madhya Pradesh*", International Research Journal of Engineering and Technology, 4(9), 740-742.

Romi Satria and María Castro (2016) "*GIS tool for Analyzing Accidents and Road Design: a Review*", In: XII Conference on Transport Engineering, Valencia, Spain, Transportation Research Procedia 18, 242 – 247 (2016).

Shu Yang, Sijia Lu and Yao-Jan Wu. "*GIS-based Economic Cost Estimation of Traffic Accidents in St. Louis, Missouri*" In: 13th COTA International Conference of Transportation Professionals, Procedia - Social and Behavioural Sciences 96, 2907 – 2915, (2013).

S. Sarifah. (2018) "*Determining Hotspots of Road Accidents Using Spatial Analysis*", Indonesian Journal of Electrical Engineering and Computer Science, 9(1), 146-151.

Thakali L, Kwon T.J, Liping Fu, (2015) "*Identification of crash hotspots using kernel density estimation and kriging methods: a comparison*", J. Mod. Transport.