Analysis of Vehicle Accidents using Spatio-Temporal Tools in ArcGIS; A Case Study of Hayatabad, Peshawar

Muhammad Babar Ali Rabbani[1], Prof. Dr. Sher Afzal Khan[2], Dr. Qaiser Iqbal[3], Engineer Qamar Zaman[4]

[1,2,3,4] Sarhad University of Science & Information Technology, Peshawar

babaralirabbani@yahoo.com[1], hod.civil@suit.edu.pk[2], qi.civil@suit.edu.pk[3], qamar.civil@suit.edu.pk[4]

Abstract— Identification of traffic accident spots play a pivotal role in planning of roads and application of effective strategies in order to minimize the traffic accidents. This study puts into use the spatial distribution of the traffic accidents scattered throughout the area using spatial analysis and statistical approaches. The purpose of this research study is to analyze the traffic accidents occurring in the Hayatabad area of Peshawar. The fundamental objective of this study is to detect accidents hotspot in an observed area by a complex statistical algorithm. A methodology was developed in ArcGIS 10.2 to analyze the spatial patterns of traffic accidents and to identify hotspots. This study has conducted NNHA spatial clustering method in CrimeSTAT for the identification of hotspot clusters for accidents points in ArcGIS. Moreover, based on the detected hotspots, spatio-temporal tool like Kernel Density Estimation (KDE) analysis was performed in Crime STAT to create a temporal map of RTAs hotspots in ArcGIS. A geostatistical method known as Kriging Interpolation method (KI) was also used to assess the results computed by KDE. The results indicated that the roundabouts located in this area are the major hotspot of accidents, which includes Bagh-e-Naran roundabout, Phase-6 roundabout, Tatara Park roundabout and Jamrud road. Comparison of KDE and KI was performed and it was found that KI outperforms KDE in identifying hotspots. It has been concluded that these hotspots lacked the basic traffic controlling devices, which are necessary for controlling the speed and converging or merging of vehicles at these locations.

Keywords— Accident hotspots, spatial analysis, clustering, kernel density estimation, kriging interpolation.

I. INTRODUCTION

Nowadays, accident is one of the essential leading causes of deaths worldwide. It is estimated that the Road Traffic Accidents (RTAs) is the 8th major cause of deaths worldwide in 2016 and is even surging to the top despite of modern use of preventive measures. Most accidents conclude in fatalities and severe injuries, which results in financial burden like medical costs, car damage and the injured are often left with permanent defects. Above 90% accidents take place in middle income or low-income countries [11]

According to a report published by [12] about 27,081 deaths occurred in Pakistan in 2017 due to RTAs, hence pushing Pakistan up to the rank of 104th country in the list of highest number of deaths due to the RTAs happen in any country. This constitutes of the total 15 deaths daily in Pakistan. It was also reported that the lives lost in RTAs exceeds than the lives lost due to terrorism in the same country. This abhorrent situation leads to the urgent need for awareness in public about the compliance of traffic rules. One of those safety provisions to be provided is to pinpoint the hotspots of RTAs. This can only be achieved by congregating the RTAs data of a certain area. About, 7000 RTAs takes place yearly in Pakistan reported to the Police stations in 2006, which has become an epidemic in the country [1]. Pakistan has seen a significant rise in accidents for the past 15 years, therefore, its evaluation has become an integral issue that is vital to be addressed [13]. According to the Global Status report on Health Safety by [12], Pakistan is regarded as a middle-income country with the majority of people are unable to afford the advanced pre-crash vehicle accident system, which includes anti-lock braking system and electronic stability control. Even Pakistan has no strict law for vehicles seat belt and there is an absence of statistics about the person's seat belt wearing ratio. Meanwhile, only 10% of the bike riders throughout the country uses helmet as a safety precaution. As an example, Karachi RTAs crash damage was assessed in terms of money damage. In collecting the data, most of the companies rejected to share their details of accidents, which puts a negative impact on the conclusions [7]. Crime STAT was put into application in order to pinpoint the highest crime happening spots. Similarly, Crime STAT can be efficiently used to detect other hotspot provided that the data is spatially located [8]. A study was conducted to assess the conventional analysis and statistical analysis called as Kernel Density Estimation, in which it was proven that results were identical [9]. A plethora of programmable applications are available for spatial analysis, which can be used for statistical analysis because they utilize specific algorithms irrespective of coding language hence visualizing akin results when plotted in ArcGIS [10]. It was found that ordinary KDE performs its analysis on counting the naturally occurring clusters, which is more efficient than the conventional tally count method. The majority of the clusters were found near a residential area, entertainment area, hotels and hospital area. It was also stressed that the implementation of spatial analysis for improving traffic safety is in its development

stages. This industry needs new implementing methods, precise data collection using the modern intelligent transportation system to prevent the loss of valuable data, which can ensure precise results [18]. A research study was conducted in Manhattan Area in New York State using shapefile provided by New York GIS office, which successfully pointed out the point of clustered accidents using spatial and temporal tools available in 3rd party GIS-linked software. They proposed that spatio-temporal nature of accidents can efficiently be analyzed by the method of KDE [19]. KDE helps in estimating the traffic accident hotspots due to its ability to assess the land usage of an area as well as its ability to evaluate the current road system in real time [2]. In Western Australia, a spatial analysis was performed to detect heavy vehicle crashes. The analysis was performed using KDE method. It was expected that the clusters would be near to the Perth metropolitan area and the results also generated the same model, which proved the consistency in the results [4]. Kriging method is seldomly used as an evaluating technique but the Kriging method was found more superior and promising as compared to Kernel Estimation because the nature of autocorrelation in its algorithms. Another positive outcome that can be obtained using Kriging method is that instead of using the Personality Assessment Inventory (PAI) as a severity measure of the accident, Kriging method can be incorporated into the analysis and it can also produce akin results as of conventional method approach [16].

According to Pakistan Bureau of Statistics in 2018, accidents became a major nuisance in Pakistan. In 2018, the number of accidents taking place was the highest ever recorded, escalating to 11121 accidents. The trends of accidents became more dominant in the year 2014-2015 and has ever been on rise shown in Fig 1.
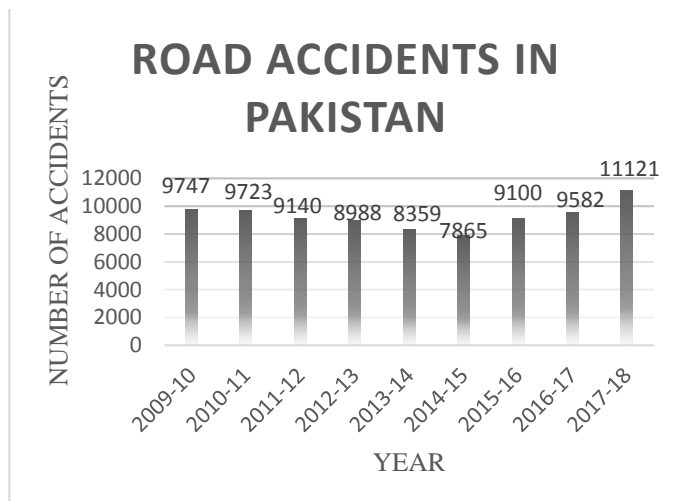

Fig 1: Roads Accidents counts in Pakistan by Pakistan Bureau of Statistics

The objectives of this study are:

1) To convert the raw textual data into numerical values in order to identify the highest frequency accident spots in Hayatabad area.

2) To analyze the accidents points and produce the most clustered accident areas.

3) To compare Kernel Density and Kriging results regarding accident hotspots.

II. MEHODOLOGY

ArcGIS operates using maps, which can easily be downloaded via its native option called Open Street Maps (OSM). Afterwards, hotspot analysis is performed to locate the different spots by various methods based on its diverse algorithms. Spatio-temporal tools like KDE method of analysis and KI algorithm is performed to identify accident hotspots.

A. Accident data collection:

Hayatabad area is situated in the North West of Peshawar, which is the capital of the Khyber Pakhtunkhwa province of Pakistan. Traffic accident analysis and its staggering results depend on the correct input of accidents location and the accident coordinates. In order to achieve the reliable results, it is necessary to collect the data from authentic accident data collecting agency, which is maintained by the area's traffic police department and the local police stations.

Fig 2 data indicates the number of accidents with respect to the accident type. These graphs signify the numbers of accidents leading to death, accidents resulted in injuries and non-fatal accidents. The data collected is of the past 10 years, that is, from 2009 to 2018.
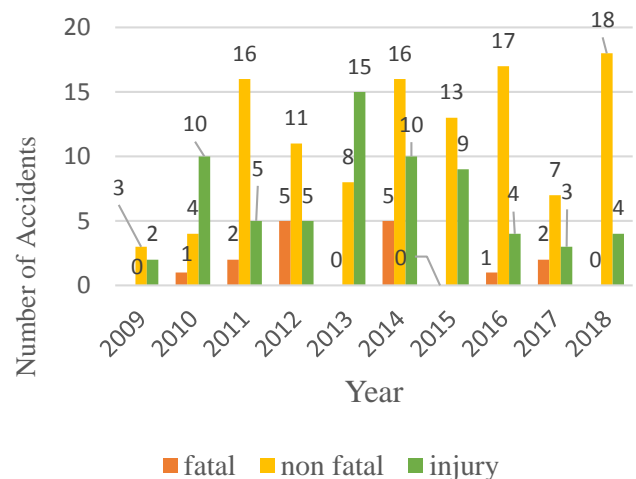

Fig 2: Overall accidents taking place in past 10 years

B. Accident data analysis approach:

Hotspot analysis:

Nearest Neighbor Hierarchical Analysis (NNHA) for spatial clustering estimation is a method used conveniently for analyzing the spatially distributed data. It determines the cluster of accidents on the bases of any two nearest distance accidents taking place in the studied phenomenon. The user must identify a value for a fixed distance, which can marginally affect the outcomes of the result because it becomes more of a subjective rather than objective oriented. The unique advantage of NNHA is that if the weight density and the intensity of the traffic

accident is known, the risk analysis of nearby traffic accident spots can also be assessed and it can detect risky areas, which is called as risk adjusted NNHA. This method can also be performed using SANET and other spatial analyst tools [15].

Kernel Density Estimation:

This method is used to detect accidents points and then it calculates the density using numerical statistical algorithms, hence the riskiest areas are detected in an area under consideration. In other words, this method underlying principle is that it defines the density of the nearest neighboring accident points by creating density and then calculating the distance of another accident from the initial reference accident point. Finally, combining the values from highest frequency to lowest of each point to that reference point. The KDE algorithm is shown below using (1), which was formulated by [14]:

$$f(x,y) = \frac{1}{nh^2} \sum_{i=1}^{n} K \begin{pmatrix} di \\ h \end{pmatrix} \dots (1)$$

Where K is the kernel constant and di is distance among the accident points from the ith point of any under the observation point of accident, n is a number of points and h is kernel width. There are two methods to determine the density, one is conventional and the other is a kernel. The former creates a large number of arbitrary mesh cells and perform its analysis while the later one creates a bandwidth of a user defined value then a circular area of known magnitude from highest to lowest frequency is encircled and finally a numerical analysis is performed [20].

Kriging Interpolation Method:

Kriging is also used for spatial interpolation. Kriging method is sometimes called as a smoothening technique. The underlying concept behind the kriging method is that the outputs are the weighted mean of the data instead of density in KDE. The weighted pattern is predicted in such a way that the weighted average is different to every point and a lag distance is predicted between the known point and predict point. It can account for the missing points that KDE might have left due to the overlapping of some points. Kriging is useful because it discourages the repetition of the points if they have similar coordinates. The mathematical form, given by [5] of kriging model is stated below in (2).

$$Z(x) = m(x) + \sum_{i=1}^{n} \lambda i \ [Z(x_i) - m(x_i)] \dots (2)$$

Where,

λi is a Kriging weight allocated to $Z(x_i)$ for prediction of accidents frequency at a location where $m(x)$ and $m(x_i)$ are the predicted values of the variables $Z(x)$ and $Z(x_i)$, at any location x.

III. RESULTS AND DISCUSSION

A. Hotspot Analysis:

Nearest Neighbor Hierarchical Analysis (NNHA) performance is based on the number of simulations and minimum clusters to be needed in the output. NNHA determined

the risky areas for the accidents using the police provided data record of accidents.

NNHA is executed using the radius as Nearest Neighbor distance, keeping the search radius in the middle. Minimum cluster points were selected as 10 and the units of those clusters were selected in Miles as shown in Table 1. Simulations runs were selected as 30. Higher the simulation runs, higher will be the accuracy of the results. Cluster identification output file are saved in convex hull shapes and ellipses.

Table 1: Risk adjusted Nearest Neighbor Hierarchical

Risk adjusted NNHA		
Name	**Risk adjusted NNHA**	**Fixed distance NNHA**
Search Radius	Medium	10
Minimum points per cluster	10	10
Units	Miles	Miles
Simulation runs	30	40
Number of Accident points	196	196

NNHA is vital for further analysis due to the fact that this step of analysis indicates the riskiest roads in an area, thereby, enabling the researchers to know about the hotspots in further steps of this analysis. In other words, this step is also a check for statistical analysis if the hotspot were not clustered at these riskiest roads.

The riskiest areas in Hayatabad are identified at different locations by blue ellipses as shown in Fig 3. The riskiest area is Bagh-e-Naran round about, second risky area is Jamrud Road and Hayatabad entrance road came third riskiest area in this method for accidents to happen, Fig 3
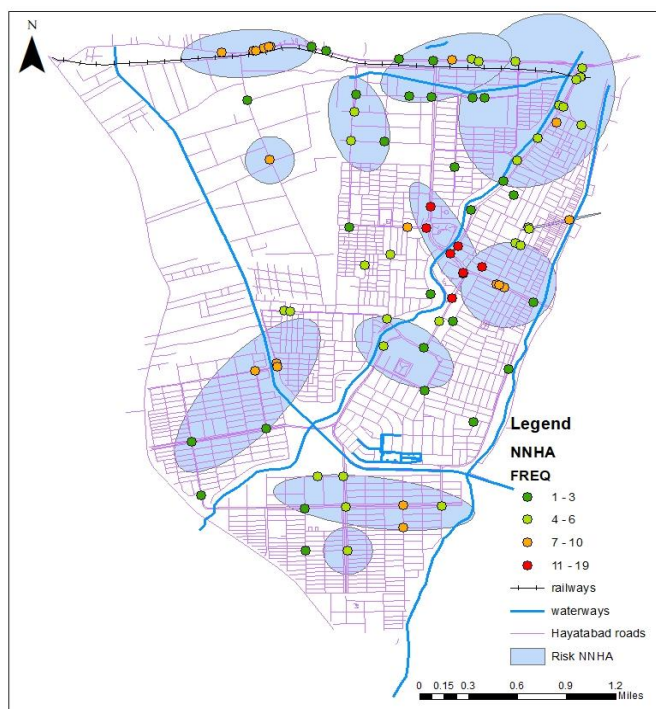
Fig 3: Hotspot identification by NNHA

B. Kernel Density Estimation:

The KDE spatial analyst tool is a powerful tool when density estimation is required. KDE has shown Bagh-e-Naran roundabout as a potential risk for accidents. Bagh-e-Naran roundabout and Tatara Park roundabout are near, it they both were regarded as a major hotspot. Here Phase-6 round about is regarded as a cold spot with 95% confidence results and standard deviation range of 9.5-12.7 shown in Fig 4. Jamrud road linking to the Industrial Estate road is also a major hotspot as shown by spatial analyst tool of KDE. Table 2 shows the parameters selected for analysis in CrimeSTAT and its visualization in ArcGIS.

Table 2: KDE parameters selected for analysis

Kernel Density Estimation in Crime STAT & ArcGIS	
Name	*Kernel Parameters selected*
Kernel Estimate method	single
Interpolation method	normal
Bandwidth type	adaptive
Minimum sample size	100
Area (units)	square miles
Output Density	absolute densities
Output cell size	15.386
Search Radius	50

Figure 4 indicates the hotspot with red color. The road accidents are more clustered towards the roundabout roads of the area.
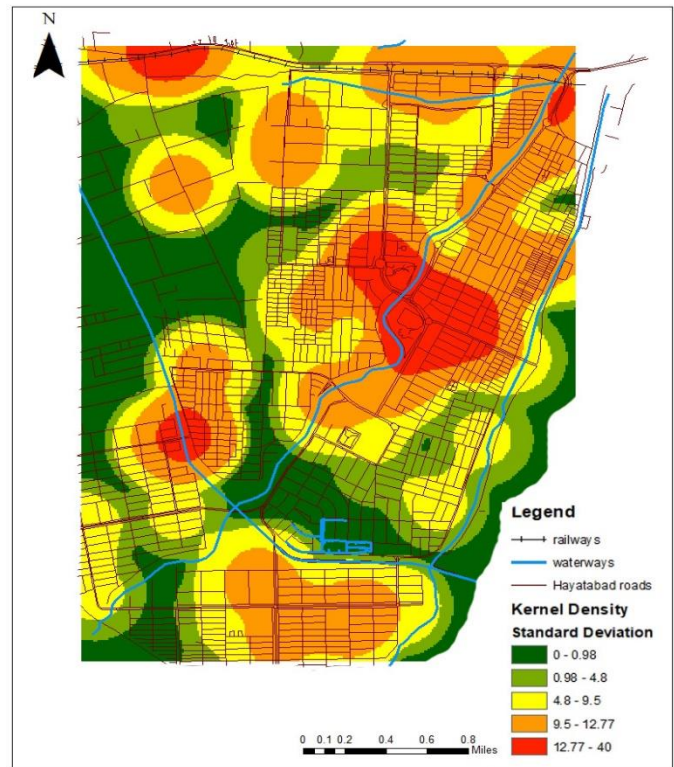

Fig 4: Kernel Density Estimation of accident hotspots

C. Kriging Interpolation Method:

The estimation of output of Kriging is more visually user friendly and easy to interpret. Besides Bagh-e-Naran round about an essential hotspot, Kriging has combined Tatara Park round about as one body and identified these locations as one body of the hotspot. The results predicted by the KDE and KI methods are marginally different from each other, therefore, indicating the need for a more accurate method to be implemented for traffic accidents hotspot identification. This result difference does not indicate the discrepancy in any other method but this difference has its own place in the analysis due to the reason that each method is correct on its statistical algorithmic computation. Table 3 shows the parameters for Kriging Interpolation

Table 3: Kriging parameters for analysis

Kriging Interpolation in CrimeSTAT & ArcGIS	
Name	*Kriging Parameters selected*
Number of points in radius	12
Kriging model	Universal
Search radius	Variable

Jamrud Road was detected as second hotspot but its standard deviation was marginally less than the Bagh-e-Naran round about hotspot and Tatara Park round about. Phase-6 round about was regarded as a cold spot by Kernel density because of the repetition of same coordinates, which Kriging treats as one point as seen in below Fig 5
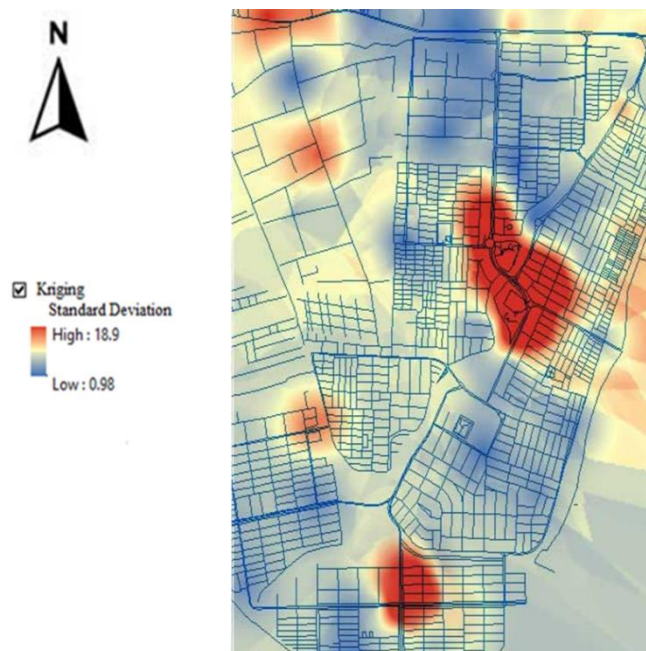
Fig 5: Kriging Interpolation output of accident hotspots

D. Comparison of Kernel Density and Kriging Interpolation:

These two methods are different in analysis techniques, difference in algorithm and produced marginally different results in hotspot detection. Therefore, the comparison of these two methods is essential. The mathematical comparison is performed using an equation known as Prediction Accuracy Index (PAI) used initially by [3]. This equation is the ratio of percentage of crashes occurring inside any hotspot to the percentage of length covered. This (3) was previously used in crime investigation mapping to identify the hotspots by [17], [16] and [6].

$$PAI = \frac{\frac{n}{N}*100}{\frac{m}{M}*100} \ \dots \ (3)$$

Where,

n is the number of accidents in a hotspot, N is the total number of accidents, m is defined as the length of road inside a hotspot and M is the overall length of roads in area consideration. The numerator is regarded as the Hit rate and denominator as the Length covered.

The higher value of PAI shows the power of a certain method to identify riskiest accident's location in a proportionally smaller area, which can help the traffic agencies to introduce economical accident prevention resources. Evident from the Table 4, KI method outperforms the KDE for hotspot detection. Kriging method is often neglected for analysis purposes; however, it can be used as a promising alternative to the Kernel method.

Table 4: PAI comparison of KDE and Kriging Method

Kernel Density versus Kriging					
Kernel Density Estimation					
Accident inside hotspot	Total Accidents	Length of Road (km)	Length of hotspot (km)	Area (km²)	PAI
15	196	312	20	24	1.19
Kriging Interpolation					
43	196	312	50	24	1.36

CONFLICT OF INTEREST

The contents of this study are free from plagiarism and therefore the study is original and is not copied from anywhere. Previous work of original authors has also been referenced.

CONCLUSION

This study has concluded that:

- Hotspots are mostly clustered at roundabouts in Hayatabad area.

- Accidents at these hotspots happen due to illegal parking at roundabouts, failure to observe traffic laws, illegal immediate turn instead of going around the roundabout in order to change the direction of the vehicle, point of converging and diverging at roundabout.

- Tatara park roundabout accidents happen due to the traffic generated by a park nearby, whose condition worsens in public holidays and two lanes of road is reduced to one lane.

- Bagh-e-Naran roundabout is a major hotspot, which is due to the merging and diverging of vehicles at roundabout and absence of traffic controlling devices.

- Jamrud road is also identified as a major hotspot due to the absence of observing traffic rules, improper road width decreased due to encroachment, poor road maintenance and unchecked operations of cargo and freight vehicle movement at this spot.

- The comparison of Kernel Density Estimation method and Kriging interpolation shows that the latter outweighs the former in the hotspot detection. Kriging interpolation, which is often ignored in the analysis, is capable of producing promising results.

Future work:

The future work should be focused on KDE of network street distance of the Hayatabad City instead of Euclidean distance and its results must be compared with Moran's I statistics. In order to assess the degree of correctness of both the implemented methods, regression models and Empirical Bayes model followed by Monte Carlo Simulation should be implemented.

RECOMMENDATIONS

It is recommended through this study that the Jamrud road carries freight transportation, therefore the its route should be changed to avoid it mixing with residential traffic. Tatara roundabout creates a lot of traffic congestion due to the presence

of park near the roundabout, therefore, a dedicated road should be constructed that enters inside the park, which will avoid the vehicles to stop near the roundabout and therefore, the traffic lane would not be reduced. The presence of service road in the Phase-6 roundabout creates havoc during rush hour. As the service road opens in roundabout, this opening should be closed and the opening must be shifted at a considerable distance from the roundabout to deter road users to take opposite turns in the roundabout. Bagh-e-Naran roundabout needs to be installed tire busters' spikes to stop users to park their vehicles in the roundabout. Most accidents occur at this spot is due to the merging and diverging of vehicles, therefore, keeping in view the traffic volume at this roundabout, traffic controlling devices like traffic signals should be installed.

REFERENCES

[1] Ahmed, Aizaza. "Road safety in Pakistan." National Road Safety Secretariat, Islamabad, 142, 2007.

[2] Chen, Yen-Chi. "A tutorial on kernel density estimation and recent advances." Biostatistics & Epidemiology 1.1, 161-187, 2017.

[3] Chainey, Spencer, and Jerry Ratcliffe. GIS and crime mapping. John Wiley & Sons, 2013

[4] Gudes, Ori, and Richard Varhol. "Using a spatial analysis approach to investigate Articulated Heavy Vehicle." Journal of Transport Geography 31, 64-71. 2015

[5] Goovaerts, Pierre. Geostatistics for natural resources evaluation. Oxford University Press on Demand, 1997.

[6] Hart, Timothy C., and Paul A. Zandbergen. "Effects of data quality on predictive hotspot mapping." Final report submitted to the National Institute of Justice 1, 2012.

[7] Kazmi, Jamil H., and Salman Zubair. "Estimation of vehicle damage cost involved in road traffic accidents in Karachi, Pakistan: a geospatial perspective." Procedia engineering 77, 70-78, 2014

[8] McCullagh, Michael J. "Detecting hotspots in time and space." ISG06 349 : 1-17, 2006

[9] Nakaya, Tomoki, and Keiji Yano. "Visualising crime clusters in a space‐time cube: An exploratory data‐analysis approach using space‐time kernel density estimation and scan statistics." Transactions in GIS 14.3: 223-239, 2010

[10] Prasannakumar, V., et al. "Spatio-temporal clustering of road accidents: GIS based analysis and assessment." Procedia-Social and Behavioral Sciences 21, 317-325, 2011

[11] World Health Organization. Global status report on road safety 2015. World Health Organization, 2015.

[12] World Health Organization. Global status report on road safety 2018. World Health Organization, 2018.

[13] Razzak, Junaid Abdul, et al. "A successful model of road traffic injury surveillance in a developing country: process and lessons learnt." BMC public health 12.1, 357.2012

[14] Rosenblatt, Murray. "Remarks on some nonparametric estimates of a density function." The Annals of Mathematical Statistics 832-837, 1956

[15] Shafabakhsh, Gholam Ali, Afshin Famili, and Mohammad Sadegh Bahadori. "GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran." Journal of traffic and transportation engineering (English edition) 4.3, 290-299, 2017

[16] Thakali, Lalita, Tae J. Kwon, and Liping Fu. "Identification of crash hotspots using kernel density estimation and kriging methods: a comparison." Journal of Modern Transportation23.2, 93-106, 2015

[17] Van Patten, Isaac T., Jennifer McKeldin-Coner, and Deana Cox. "A microspatial analysis of robbery: Prospective hot spotting in a small city." Crime Mapping: A journal of research and practice 1.1, 7-32., 2009

[18] Jia, Ruo, Anish Khadka, and Inhi Kim. "Traffic crash analysis with point-of-interest spatial clustering." Accident Analysis & Prevention 121, 223-230, 2018

[19] Romano, Benjamin, and Zhe Jiang. "Visualizing traffic accident hotspots based on spatial-temporal network kernel density estimation." Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2017.

[20] Rushton, G., and C. Tiwari. "Spatial filtering/kernel density estimation.", 359-364. 2009

Muhammad Babar Ali Rabbani, the author, is resident of Peshawar. He completed in bachelor degree of Civil Engineering from University of Engineering & Technology, Peshawar in 2013. The author is pursuing his master's degree in Transportation Engineering from Sarhad University of Science & Information Technology, Peshawar. This paper is his research study for the master's degree.