advances in Technology

*Adv. Technol.* **2021**, *1(1), 191-206*

## Research Paper

# Accident Hotspots in Southern Expressway of Sri Lanka: Interpolation Evaluation using GIS

Nalaka P. Kodippili,[a] M. S. R. Akther,[b] G. Naveendrakumar[c,*]

[a] Faculty of Graduate Studies, University of Sri Jayewardenepura, Nugegoda 10250, Sri Lanka

[b] Postgraduate Institute of Agriculture, University of Peradeniya, Peradeniya 20400, Sri Lanka

[c] Faculty of Applied Science, Vavuniya Campus of the University of Jaffna, Vavniya 43000, Sri Lanka

Email correspondence : gnaveendra@vau.jfn.ac.lk (G. Naveendrakumar)

Received: 7 March 2021; Revised: 9 April 2021; Accepted: 14 May 2021; Published: 31 May 2021

**Abstract**

In this study, the southern expressway, which is the first and lengthiest E class highway (126 km) in Sri Lanka, was analysed for roadside accident incidences. The primary objective of this paper is to identify the best-fit interpolation techniques for the hotspots' most distinctive causes of vehicular crashes. The accident details were collected from the Police Headquarters consisting of 966 accidents that took place during the period from 2015 to 2017. To identify accident hotspots, GIS-based interpolation techniques such as Ordinary Kriging, Kernel Density Estimation (KDE), Inverse Distance Weighting (IDW), and Nearest Neighbour Interpolation methods were used. The spatial interpolation outcome of the four methods was compared based on the standard Prediction Accuracy Index (PAI). The analysis was executed using QGIS and GeoDa. Results of PAI revealed that an IDW and KDE outperformed the other two interpolation methods. The left and right lanes of the expressway, spotted with 11 and 20 hotspots, respectively, indicate the right lane was 50% more prone to accidents than the left lane. Notably, nearly 5% of the entire road stretch is estimated as accident-prone spots in both lanes. Peak accidents were recorded during afternoon and evening hours, and buses were the most active vehicle type. Uncontrolled speeding was the primary reason for more than 50% of the accidents, while unsuccessful overtake accounted for more than 20% of the accidents on the highway. The road design modifications and warning sign placements at appropriate places may be recommended as countermeasures.

**Keywords:** Highway, hotspot analysis, kernel density estimation, prediction accuracy index, vehicle collision

---

https://doi.org/10.31357/ait.v1i1.4835

## Introduction

Road accidents are in the increasing trend worldwide, majorly due to the development of transportation infrastructure fails to keep pace with other sectors. Therefore, road traffic accidents are the leading factor in human deaths or illnesses worldwide. The importance of addressing issues related to road traffic injuries is being neglected irrespective of their severity globally. It accounted for over 1 million deaths and 20 to 50 million injuries each year [1]. Considering trends and factors responsible for the roaring incidences, low and middle-income countries experience more effects having a solid link to their level of economic and social developments [1], [2]. Being a lower-middle-income country in South Asia, Sri Lanka has a significant burden of road traffic incidents and their consequences. In Sri Lanka, a traffic-related accident showed an ever-increasing trend and resulted in many fatalities due to an alarming number of vehicles and the inferior quality of road infrastructure [3]. Human error, mainly the carelessness of the drivers or pedestrians, is the primary reason for this kind of accident. Hence, the probability of accident occurrence and its severity can be overcome by systematically analysing the incident scenario and implementing relevant solutions like applying proper traffic control devices, proper roadway design practices, and effective traffic rules enforcement [4]. According to Somasundaraswaran [3], the total number of accidents is on the increasing verge in Sri Lanka due to the growing pattern of vehicle ownership at an alarming rate. A study on roadside traffic incidence stated that the pattern of road traffic crashes suffered a substantial increase from the overall perspective from 1938 to 2013 [5].

Vehicle accidents that happen due to common causes should be prevented, considering the risk of fatalities and property damages [6]. However, devising effective solutions requires analysis of spatial and temporal patterns in the zonation of traffic accidents, which can be effectively achieved by applying geospatial technology [7]–[9]. The hotspot identification, time, reason, and vehicle type involved in those accidents are ultimate factors to be considered to develop an appropriate strategy in this regard. A geographic Information System (GIS) is a computer system used for storing, querying, analyzing, and displaying geographic data. Spatial thinking helps identify the patterns and explore the reasons for the identified pattern characteristics, unlike conventional methods. GIS technology has been a popular tool for the visualization of accident data and analysis of hotspots [10]–[12]. The QGIS (Quantum GIS) is an open-source geographical information system software that can be acquired free of charge from many reliable sources. GIS platform is having numerous methods of analyzing spatial patterns of point events like accident data. Hotspot analysis based on Kernel Density Estimation (KDE) is one of the most common methods utilized by researchers [13]–[15]. It calculates the density of crashes in a given search bandwidth, also known as its neighbourhood. KDE develops a continuous surface of density estimates of separate crashes within a search bandwidth, thus providing aesthetically and statistically significant results [16]. The spots with a significantly higher number of vehicle crashes are known as accident hotspots [17] were geo-statistically analysed.

In Sri Lanka, lack of road accident research and the limited availability of relevant data, the policymakers were deprived of formulating appropriate policy interventions to prevent road accidents. Few attempts being made to determine the hotspots for road accidents and trends were analysed in previous studies in the region using statistical methods are focused mainly on a given interpolation technique. In this study, the Southern Expressway, which is the first and lengthiest E class highway (126 km) in Sri Lanka, has been analysed for vehicular accidents in terms of counts. This expressway serves to link the capital Colombo with many cities in the down south of Sri Lanka [18]. The main objective of this study is to find out the best fit interpolation technique to demarcate the most accident-prone locations as it is an integral part of any road safety program. The sub-objectives of the paper are to identify distinct locations of the vehicle crash hotspots and the leading causes of accidents in the Southern Expressway of Sri Lanka using the best-fitted interpolation technique.

This study necessitates the strict enforcement of road safety regulations to reduce vehicle accidents, especially on highways, in the future. Identification of accident hotspots is beneficial to take countermeasures regarding design modification, placement of warning signs, and to strict enforcement of expressway road rules and regulations.

## Methods

This section presents the study area, data, methods of interpolation, hotspot selection basis, and comparison of interpolation accuracy.

### *Study Area*

The southern expressway, the first and lengthiest E-class highway (126 km) in Sri Lanka, was analysed for roadside accidents. It links the capital with many cities in the down south of Sri Lanka [18]. This expressway starts at latitudes and longitudes 6.839974°, 79.981489° and ends from 5.976286°, 80.518329°. It is a bidirectional expressway where the left lane carries the traffic towards the capital, and the right lane carries towards the down south of Sri Lanka. There are ten interchanges available along the expressway from start to end. Nowadays, many people use this expressway daily for office works and commute between Colombo and southern coastal cities relatively within a short duration. Figure 1 illustrates the study area.

### *Data*

The expressway accident details collected from the Police Headquarters consist of 966 accidents that took place during the period from 2015 to 2017 years. The geo-coded accident records have been cleaned for errors, missings, incomplete, and inconsistencies. Such data include the details of the location, date and time, vehicle type, reason, and damage level (Table 1). The road trace, interchanges, and distance information were collected from the Road Development Authority of Sri Lanka.

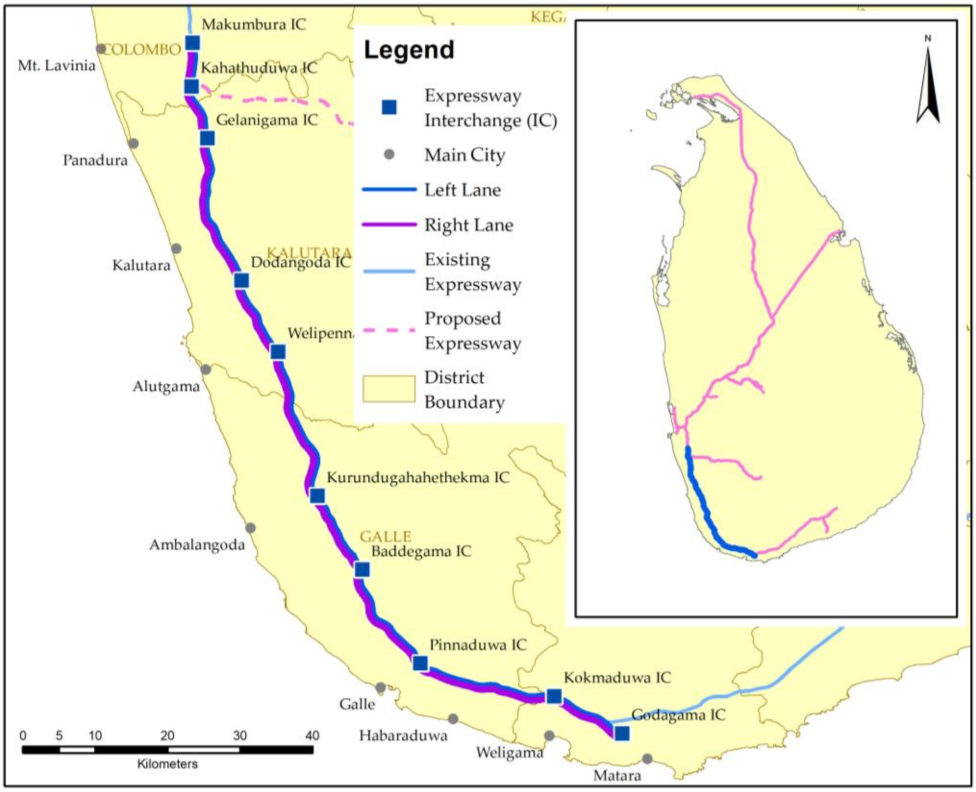


**Figure 1.** Study area map shows the location of the expressway stretch in Sri Lanka

**Table 1.** Description of accident data

| Variable | Attributes | Implication |
|---|---|---|
| Location | Latitude and longitude | Maps the accident location |
| Date and time | Date and time | Find the peak accident vulnerable date-time of the day |
| Vehicle | Light, heavy, and long vehicle | Identifies the most accident-prone vehicle category |
| Reason | Out of control, speeding, skidding, and reckless driving | Necessary to enforce stricter regulations |
| Damage | Slightly wounded, critically wounded, mortal, and property damage | Damage level identification |

***Spatial Autocorrelation: Moran's I Method***

Moran's I used to evaluates whether the pattern of given data with its associated attributes distributed as clustered, dispersed, or random. Its one of the oldest techniques and widely used to determine the spatial correlation [19]. Spatial Autocorrelation analysis was performed using GeoDa opensource software. It can be computed using Equation 1.

$$I = \frac{N \sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{W \sum_I (x_i - \bar{x})^2} \qquad \text{Equation 1}$$

where $N$ is the number of cases, $x_i$ is the variable value at a particular location, $x_j$ is the variable value at another location, $x$ is the mean of the variable, and $W$ is a weight applied to the comparison between location $i$ and location $j$. The $w_{ij}$ is a distance-based weight matrix, which is the inverse distance between locations $i$ and $j$ $(1/dij)$.

***Interpolation Techniques***

To identify accident hotspots, a GIS (Geographic Information System) based analysis technique was used. Initially generated a point layer using the MS Excel data. Four different spatial interpolation techniques, the Ordinary Kriging (OK), Kernel Density Estimation (KDE), Inverse Distance Weighting (IDW), and Nearest Neighbour Interpolation (NNI), were used to delineate the bidirectional lanes to identify the hotspots which have the higher probability of accidents. These interpolation techniques were selected based on their popularity in the relevant application context. Since there is no universal threshold value to demarcate the hotspot, a cut-off value that filters a relatively higher risk area is arbitrarily selected [20]. The hotspot and related spatial analysis were carried out using the free and open-source software QGIS.

A prediction accuracy index (PAI) developed by [21] was used to compare the four interpolation techniques mentioned above. The PAI is estimated as the ratio between the percentage of accident rate and the percentage of hotspot area (see Equation 2). In this paper, all the PAI values were estimated concerning the area, although it could be estimated on a length basis. The number of accidents and hotspot areas were normalized prior to the application to increase the ability to identifying risk zones; the higher the PAI, the better the method's performance.

$$PAI = \frac{\frac{n}{N}\times 100}{\frac{m}{M}\times 100} \qquad \text{Equation 2}$$

Where n is the number of accidents hotspots, N is the total number of accidents, m is the area of road involved in accident hotspots, and M is the total area of the road.

Bi-directional lanes of the expressway were interpolated separately since incoming and outgoing traffic behaved differently. Usually, the high-risk zones in which a significantly higher number of accidents in a small place are identified as the 'hotspot' [17]. Spatial interpolation methodologies are used to predict the distribution of variables of interest in different disciplines [22]. Ordinary Kriging, Kernel Density Estimation, Inverse Distance Weighting, and Nearest Neighbour Interpolation have been used widely in hotspot analysis.

***Ordinary Kriging (OK)***

The OK is one of the commonly used kriging techniques. It is used to predict values from known observations, where the mean is assumed unknown, constant, and is estimated from the local neighbourhood [23]. The value of the unmeasured position can be estimated [24]–[26] using Equation 3.

$$Z^* (x_0 ) = \sum_{i=1}^{n} \lambda_i \, Z\,(x_i) \qquad \text{Equation 3}$$

$Z^* (x_0 )$ - predicted value at the unmeasured position $x_0$

$Z(x_i)$ - measured value at a position $x_i$

$\lambda_i$ - weighting coefficient from the measured position to $x_0$

$n$ - number of positions within the neighbourhood searching

***Kernel Density Estimation (KDE)***

The KDE is one of the most popularly used methods in order to analyse the properties of a point event distribution [27], [28]. It has been used widely in the analysis of traffic accident 'hotspots' and detection. The density at a particular location can be computed [29] using Equation 4.

$$\lambda\,(s) = \sum_{i=1}^{n} \frac{1}{\pi r^2} k\,(\frac{d_{is}}{r}) \qquad \text{Equation 4}$$

$\lambda(s)$ - density at location $s$

$r$ - search radius (bandwidth) of the KDE (only points within $r$ are used to estimate $\lambda(s)$)

$k$ - weight of a point $i$ at distance $d_{is}$ to location $s$. The $k$ is usually modelled as a function (called kernel function) of the ratio between $d_{is}$ and $r$.

***Inverse Distance Weighting (IDW)***

The IDW is a commonly used interpolation method that predicts unknown values based on their distance from known values [30]. It enforces that the estimated value of a point is influenced more by nearby known points than those farther. The estimated value of the point at a particular location can be computed [31] using Equation 5.

$$z_0 = \frac{\sum_{i=1}^{s} z_i \frac{1}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}} \qquad \text{Equation 5}$$

$z_0$ - estimated value at point 0

$z_i$ - measured value at point $i$

$s$ - number of points used to estimate the unknown value

$d_i$ - distance between points $i$ and 0

$k$ - power identifying the influence of distance

***Nearest Neighbour Interpolation (NNI)***

The NNI is a common and straightforward interpolation technique in spatial analysis. The approach finds the closest subset of input samples to a query point and applies weights to them based on proportionate areas [32]. The nearest neighbour algorithm selects the value of the nearest point/observation to a specific grid cell and does not consider the values of neighbouring points at all. The algorithm is straightforward to implement.

### *Heat Map using KDE*

Heat map plugin in QGIS was exploited to delineate the hotspots using the KDE algorithm. This interpolation technique was identified as a best-fit model from the PAI analysis. Heatmap plugin was used mainly because of its sophisticated and well-organized algorithmic nature.

## Results and Discussion

### *Spatial Autocorrelation*

The results of the point distribution of accident incidence indicate the ideally perfect clustering pattern in both directional lanes (Figure 2). Results of the autocorrelations of the right and left lanes were 0.98 and 0.97, respectively. It indicates that most of the accident incidences could be grouped in one location.

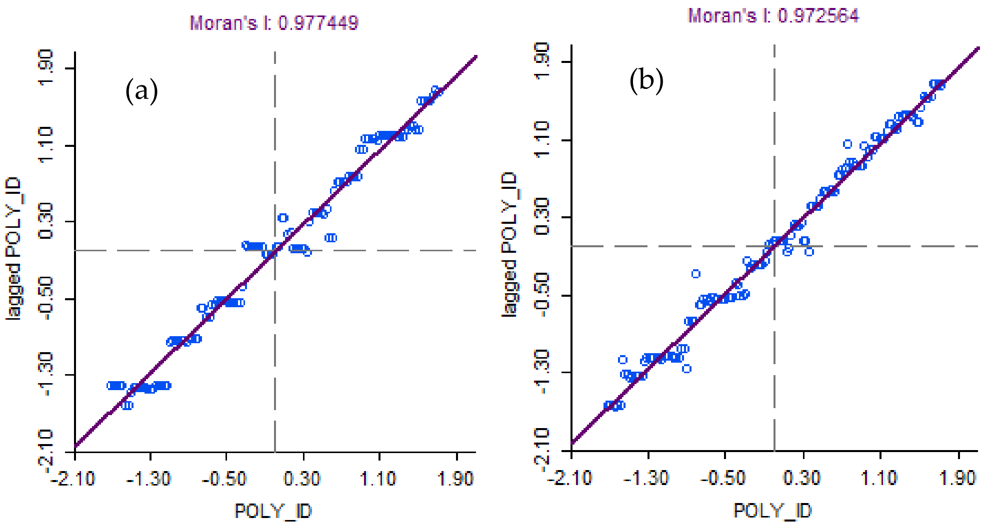


**Figure 2.** Moran's I autocorrelation scatterplot showing the relationship between a variable and the average value at neighbouring locations in the (a) left and (b) right lanes of the Southern Expressway of Sri Lanka (period of 2015-2017)

### *Prediction Accuracy Index (PAI)*

The results of the hotspot analysis using five different spatial interpolation techniques, namely Ordinary Kriging (OK), Kernel Density Estimation (KDE), Inverse Distance Weighting (IDW), and Nearest Neighbour Interpolation (NNI), were compared. Table 2 compares the outcomes of the five interpolation techniques in terms of area-based PAI.

A more substantial variation was observed in the left lane than the right lane in terms of PAI. The PAI values for the left and right lanes varied between 1.48-2.29 and 1.42-1.70, respectively. Though several hotspot stretches detected in the right lane were comparatively high in all four methods, the PAI values observed in the left were slightly higher than in the right lane. In the left lane, the IDW demonstrated a higher PAI (2.29), while for the right lane, KDE showed the highest PAI value (1.70). However, the highest PAI value corresponded with the highest number of hotspots for the left lane, the right lane corresponded with the lowest number of hotspots. Relatively higher PAI values better explain the ability of a method to locate most accident-prone locations in a small road stretch, which helps road authorities to manage the incidents efficiently. In terms of PAI, IDW and KDE outperformed the other interpolation techniques, which support the finding in a similar context [20].

### *Accident Hotspots*

The left and right strips of the expressway were spotted with 11 and 20 hotspots, respectively. Figure 3 shows the hotspots identified in the southern expressway of Sri Lanka. The results further indicate that the right lane was 50% more prone to accidents than the left lane. Notably, nearly 5% of the entire road stretch is estimated as accident-prone spots of the whole expressway stretch. This is important to narrow down the focus area in terms of formulating counter-strategies to reduce accidents in the southern expressway of Sri Lanka.

**Table 2**. Outcome comparison of the four interpolation techniques

| **Method** | **Number of accidents happened in hotspots ($n$)** | **Total accidents ($N$)** | **Area covering hotspot in km² ($m$)** | **Area of highway in km² ($M$)** | **PAI** |
|---|---|---|---|---|---|
| Left lane | | | | | |
| OK | 112 | 525 | 0.120263 | 1.009342 | 1.79 |
| KDE | 111 | 525 | 0.110165 | 1.009342 | 1.94 |
| IDW | 248 | 525 | 0.208362 | 1.009342 | 2.29 |
| NNI | 241 | 525 | 0.314065 | 1.009342 | 1.48 |
| Right lane | | | | | |
| OK | 192 | 444 | 0.279445 | 1.009446 | 1.56 |
| KDE | 183 | 444 | 0.245149 | 1.009446 | 1.70 |
| IDW | 211 | 444 | 0.299843 | 1.009446 | 1.60 |
| NNI | 204 | 444 | 0.326005 | 1.009446 | 1.42 |

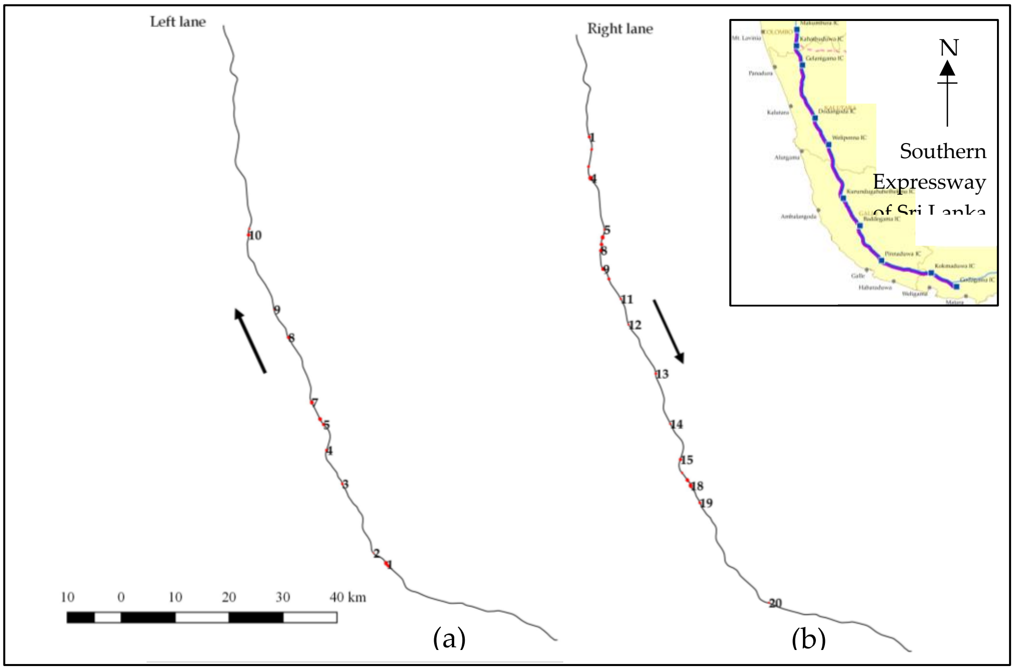


**Figure 3.** Accidents hotspots in the (a) left and (b) right lanes of the Southern Expressway of Sri Lanka (period of 2015-2017)

The time of accidents revealed that afternoon and evening hours were the peak accident recorded time of a day (Figure 4). Buses were the most active heavy vehicle type causing the majority of the collisions in the hotspots. Uncontrolled speeding was the primary reason for more than 50% of the accidents, while unsuccessful overtake accounted for more than 20% of the accidents in the identified hotspots (Figure 5). Though the right lane was more prone to accidents, comparatively higher critical injuries in the left lane, necessitate more emergency medical aid.

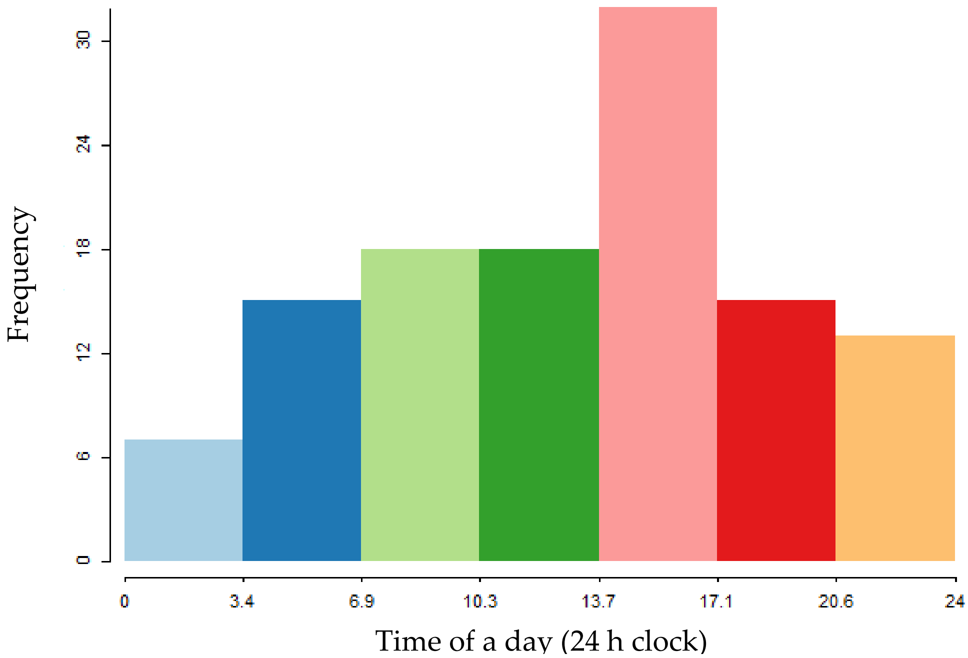


**Figure 4.** Time of accidents in the left and right lanes of the Southern Expressway of Sri Lanka (period of 2015-2017)

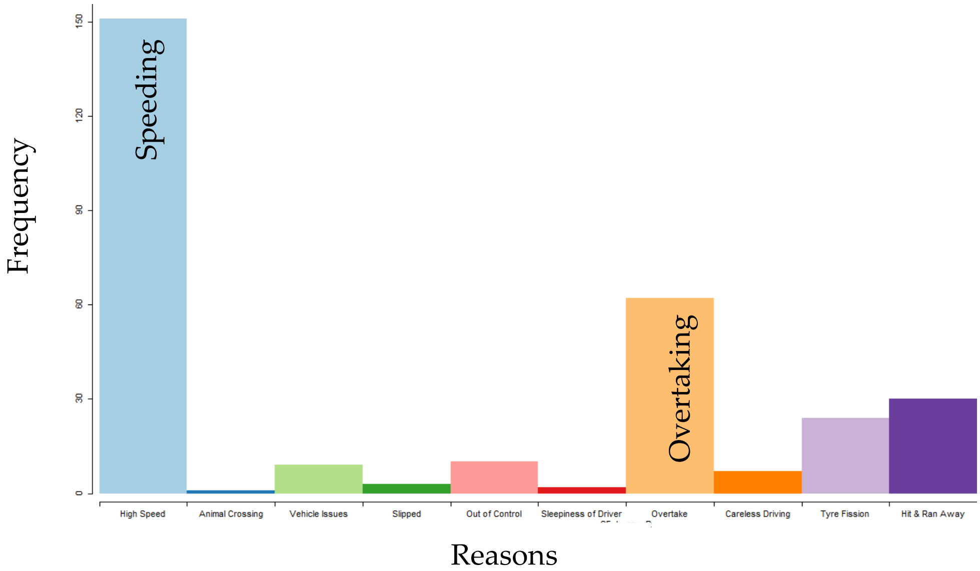


**Figure 5.** Reasons for the accidents in the left and right lanes of the Southern Expressway of Sri Lanka (period of 2015-2017)

## Conclusion

The roadside accidents in Southern Expressway showed a clustering pattern. Based on the results obtained, IDW and KDE algorithms are the best fit model to demarcate the hotspot locations. To ensure road safety, the accident hotspots identified in this investigation are recommended for the countermeasures regarding the design modification, placement of warning signs, and strict enforcement of expressway road rules and regulations. This study may be limited due to the inadequacy of temporal accident data and the non-accessibility fused with cost factor impede extended, in-depth analysis. Supplementary research experiment by segmenting the expressway into distinct categories such as straight, curve, and interchanges required to discern the operational causes for accidents. A further detailed study with temporal traffic data and incorporating topographical and climatic/weather conditions to analyse the factors affecting the distribution would significantly impact research results of relevant context. Moreover, it is also suggested to provide attentiveness by taking easy actions such as placing signboards in appropriate places to reduce accidents in the identified hotspots.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

[1] WHO. *World report on road traffic injury prevention.* Report of the World Health Organization (WHO). **2004**.

[2] WHO. *Global Status Report on Road Safety: Time for Action.* Report of the World Health Organization (WHO). **2009**.

[3] A. K. Somasundaraswaran. Accident Statistics in Sri Lanka. *IATSS Res.*, **2006**, 30, *115–117*.

[4] V. Prasannakumar, H. Vijith, R. Charutha, N. Geetha. Spatio-temporal clustering of road accidents: GIS based analysis and assessment. In *Procedia - Social and Behavioral Sciences.* **2011**, 21, *317–325*.

[5] S. D. Dharmaratne, A. U. Jayatilleke, A. C. Jayatilleke. Road traffic crashes, injury and fatality trends in Sri Lanka: 1938–2013. *Bull. World Health Organ*. **2015**, 93, *640–647*.

[6] I. A. G. Edirimanne. *Strategies to minimize deficiencies on transport related insurance claims*, University of Moratuwa, **2010**.

[7] D. L. Harkey. Evaluation of truck crashes using a GIS-based crash referencing and analysis system. *Transp. Res. Rec*. **1999**, 1686, *13–21*.

[8] N. Levine, K. E. Kim, L. H. Nitz. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accid. Anal. Prev*. **1995**, 27, *675–685*.

[9] L. Li, L. Zhu, D. Z. Sui. A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. *J. Transp. Geogr*. **2007**, 15, *274–285*.

[10] H. J. Miller. Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T). *Geogr. Anal*. **2010**, 31, *373–399*.

[11] W. D. Cook, A. Kazakov, B. N. Persaud. Prioritising highway accident sites: A data envelopment analysis model. *J. Oper. Res. Soc*. **2001**, 52, *303–309*.

[12] S. K. Ghosh, M. Parida, J.K. Uraon. Traffic Accident Analysis for Dehradun City Using GIS. *ITPI J*. **2004**, 1, *40–54*.

[13] S. Erdogan, I. Yilmaz, T. Baybura, M. Gullu. Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Accid. Anal. Prev*. **2008**, 40, *174–181*.

[14] T. K. Anderson, Kernel density estimation and K-means clustering to profile road accident hotspots. *Accid. Anal. Prev*. **2009**, 41, *359–364*.

[15] C. A. Blazquez, M. S. Celis. A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. *Accid. Anal. Prev*. **2013**, 50, *304–311*.

[16] J. Choudhary, A. Ohri, B. Kumar Identification of Road Accidents Hot Spots in Varanasi using QGIS Identification of Road Accidents Hot Spots in Varanasi using. In *National Conference on Open Source GIS: Opportunities and Challenges*. **2015**, 7–13.

[17] M. A. Aghajani, R. S. Dezfoulian, A. R. Arjroody, M. Rezaei. Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics (case study: Ilam Province, Iran). *Transp. Res. Procedia*. **2017**, 25, *2126–2138*.

[18] RDA. E01 - Southern Expressway (SE). Official Website of Road Development Authority (RDA) of Sri Lanka. **2014**, Available: http://www.exway.rda.gov.lk/index.php?page=expressway_network/e01. [Online].

[19] R. Haining. *Spatial Data Analysis*. Cambridge University Press, **2003**.

[20] L. Thakali, T. J. Kwon, L. Fu. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. *J. Mod. Transp*. **2015**, 23, *93–106*.

[21] S. Chainey, L. Tompson, S. Uhlig. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Secur. J*. **2008**, 21, *4–28*.

[22] J. Li, A. D Heap. A Review of Spatial Interpolation Methods for Environmental Scientists. *Aust. Geol. Surv. Organ*. **2008**. 154.

[23] M. A. Oliver, R. Webster. *Basic Steps in Geostatistics: The Variogram and Kriging*. Springer, **2015**.

[24] C. M. Rutter, E. H. Isaaks, R. M. Srivastava. An Introduction to Applied Geostatistics. *J. Am. Stat. Assoc*. **1991**, 86, *548*.

[25] N. Cressie. Statiscs for Spatial Data. *Terra Nov*. **1992**, 4, *613–617*.

[26] T. Hengl. *A Practical Guide to Geostatistical Mapping*. University of Amsterdam, **2009**, *291*.

[27] B. W. Silverman. *Density estimation: For statistics and data analysis*. London: Chapman Hall, **1986**.

[28] T. C. Bailey ,A. C. Gatrell. Interactive spatial data analysis. *Interact. Spat. data Anal*. **1995**, 77, *1642*.

[29] Z. Xie, J. Yan. Kernel Density Estimation of traffic accidents in a network space. *Comput. Environ. Urban Syst*. **2008**, 32, *396–406*.

[30] M. R. Mehdi, M. Kim, J. C. Seong, M. H. Arsalan. Spatio-temporal patterns of road traffic noise pollution in Karachi, Pakistan. *Environ. Int*., **2011**, 37, *97–104*.

[31] S. M. Ansari, K. V. Kale. Methods for Crime Analysis Using GIS. *Int. J. Sci. Eng.*

*Res*. **2014**, 5, *1330–1336*.

[32] R. A. Sibson. Brief Description of Natural Neighbour Interpolation. in *Interpreting multivariate data*, **1981**.