Crash risk factor identification using association rules in Nagpur city, Maharashtra, India

Bahuguna Dalai* **and Vishrut S. Landge**

Civil Engineering Department, Visvesvaraya National Institute of Technology, South Ambazari Road, Nagpur 440 010, India

The increase in traffic volume in urban road networks poses a significant challenge to transportation safety. It is evident that different traffic zones experience unique crash patterns and severities. The different factors that affect crash rates are caused by the various characteristics of the drivers, weather conditions, design of roadside infrastructure and driving behaviour. Although studies have shown that various factors can affect crash rates, there are insufficient studies on the exact categorization of these factors. Accordingly, the present study focuses on traffic crashes on streets where the risks of an accident occurrence are higher, using Nagpur city, Maharashtra, India as a case study. Three levels of risk zones were selected, i.e. zone-I (low risk), zone-II (medium risk) and zone-III (high risk). The risk zones are created in ArcGIS software using the kernel density estimator function. The association rule was then used to find out the various crash risk factors within the zone. The results of the study reveal that the risk of pedestrian fatalities is higher in areas where the speed limit is more than 40 km/h and day-to-day pedestrian activity is present. Based on the results, we propose a lower speed limit in zone-I, in addition to providing pedestrian-crossing facilities such as zebra crossings or refuge islands for crosswalks. Moreover, we propose implementing an awareness campaign for road traffic safety aimed at educating road users on how to follow road discipline, especially with regard to utilizing pedestrian facilities, aggressive young motorcyclists, lane changing and overtaking manoeuvres.

Keywords: Association rules, driver characteristics, risk factors, traffic crash, urban roads.

THE rapid growth of India's population and the increasing number of motor vehicles are expected to increase the travel demand. According to the 2011 census, 31.2% of Indians live in urban areas, while 68.8% live in rural areas. The rapid urbanization and increasing number of motor vehicles have made people more vulnerable to accidents. This has resulted in an alarming number of fatalities, injuries and disabilities. The population growth and consequent increase in traffic volume have the potential to lead to further consequential impacts on road safety and the operational characteristics of road networks. According to the Ministry of Road Transport and Highways (MoRTH), Government of India (GoI), in 2018, 151,417 people lost their lives and 467,044 were injured due to road traffic accidents in the country. Of these, 40.9% occurred in urban areas and 59.1% in rural areas[1].

The densely populated urban areas and the corresponding increase in traffic volume have significantly influenced the incidence of road-traffic accidents. Urban areas have more conflict points than rural areas. Drivers require additional capabilities and better performance under different circumstances. Figure 1 illustrates aspects of a driver's capacity and performance. In this model, if the driver's capabilities surpass the task demand, he/she can move forward safely. If the driver's task demands exceed his/her capabilities, a collision or loss of control can happen. However, exceptions can be made if the unseen road user makes a compensatory action, such as jumping out of the way of the oncoming vehicle. Such movement by road users effectively changes the task demand at a critical moment[2].

Over the years, various statistical methods have been used to analyse the crash-frequency data. For example, Lord and Mannering[3] discussed the methodological issues and characteristics of crash-frequency data. Several other studies also address methodological issues[4]. Specific issues and deficiencies in safety research have been suggested by various researchers. Regarding data collection, one of the difficulties faced is the existence of unobserved heterogeneity, i.e. where factors affecting the crash frequency are either not observed or are not possible to record[5]. In addition, there are potential dangers with regard to bias and erroneous statistical inferences when these observed factors are correlated with unobserved factors[6]. However, in recent years, enormous progress has been made in developing traffic safety models incorporating unobserved heterogeneity[5]. In some studies, two or more methods were used to find the best-performing model. For instance, Miranda-Moreno *et al.*[7] analysed three models, including a heterogeneous negative binomial model, a negative binomial model and the Poisson lognormal model, to evaluate the relative performance and ranking of location for safety improvements. Wu *et al.*[8] used an association rule (AR) and fault tree analysis to study the crash severities at various junctions.

*For correspondence. (e-mail: dbahuguna@students.vnit.ac.in)

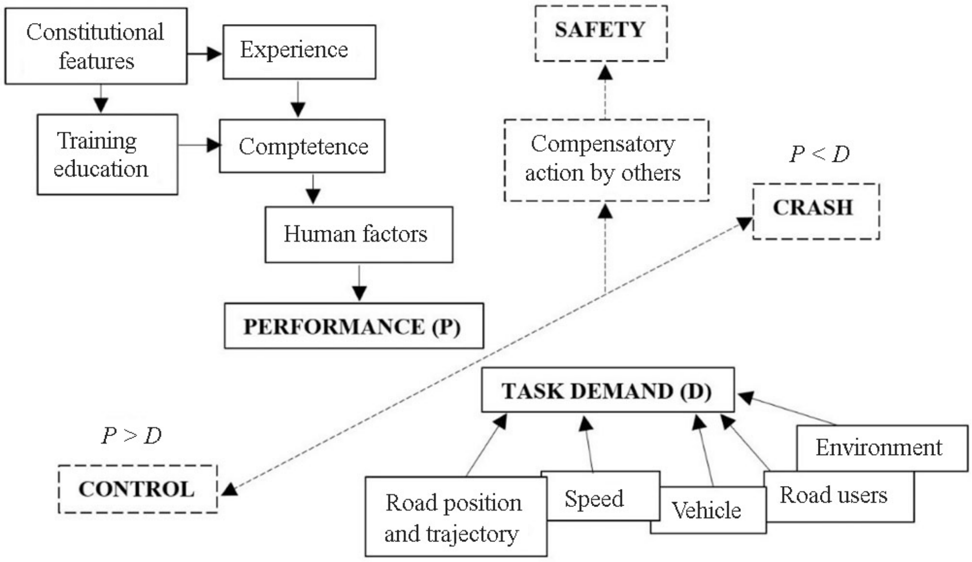

Figure 1. Task capability model (adapted from Fuller[2]).

Other studies have examined the impacts of various factors associated with crashes in urban areas. These studies have identified various crash risk factors, which can be grouped into four main categories: human factors, vehicle factors, environmental factors and road conditions.

Young drivers are more prone to severe injuries and are twice as likely to suffer from severe crashes[9,10]. Older drivers (>65 years) suffer from greater injury severity[11]. The likelihood of male drivers being involved in a severe accident is higher than their female counterparts, and they are also more careless while driving[12,13]. In the vehicle category, trucks are more dangerous than cars to cyclists at intersections[14]. In terms of severity, motorcycle crashes are associated with higher injury severity[15]. When comparing different road geometries, curved roads have a higher probability of fatal injuries[16,17]. The design and layout of local streets and arterial roads play a significant role in the severity of crashes[18,19]. Facilities such as divisions between carriageways and footpaths greatly influence crash severity. Poor pavement conditions and side friction are associated with more severe crashes[20,21]. In addition, street lights play a crucial role, as previous studies have shown that crash severity is likely to increase (at least slightly) at crossroads with no streetlights[22]. Some drivers lose control of their vehicle in the dark and the probability of severe injuries increases under darkness[23,24]. Different periods of the day are also associated with crash risk factors[25]. The number of single-vehicle crashes during weekends is higher than during weekdays[10]. Rainy conditions and land-use patterns also influence the severity of injuries[26].

In recent decades, several analysis methods have been proposed for identifying traffic accident 'hotspots' zones[27,28]. Geographic information system (GIS) is a tool used for traffic crash analysis; it considers several parameters, namely the number of fatalities and property damage in road traffic crashes[29,30]. GIS helps in the spatial analysis and identification of 'black spots' zones[31,32]. The most common method used for spatial analysis in a GIS environment is kernel density estimation (KDE)[33,34].

The black zones identified using methods include more than one road segment and should be contiguous[35,36]. A segment of a road or junction with high traffic crashes is considered a black spot or hotspot[37]. In other words, a black zone is a set of contiguous road segments experiencing a higher proportion of traffic crashes[38]. A density map (raster format) is generated using KDE. This method can be used to estimate the concentration of traffic accidents in a road network or any region[39]. Once the hotspot locations are identified, the corresponding crash risk factors can be removed or ameliorated using countermeasures[40].

ARs, which were introduced in 1993, are a popular data-mining method[41]. It allows researchers to explore the relationship between various risk factors and crash severity. In recent years, ARs have been used in data mining to find rules of patterns in various fields, such as market basket analysis[42], medical science[43], urban landscapes[8] and traffic safety[44]. Pande and Abdel-Aty[45] used AR mining techniques to analyse crash data and determine the correlations between the various sets of attributes in the dataset. Marukatat[46] applied the AR method for analysing the real traffic crash data collected from the local police station. The outcome of the study provided insights into the improvement of traffic safety.

Earlier studies on traffic crash evaluation have provided a comprehensive introduction to hotspot analysis and the traffic crash analysis method. However, most of these studies on traffic crashes focused on analysing the spatial and temporal changes in traffic accidents without taking into account the various factors that affect the characteristics of individual zones and the crash risk factors related to these zones. To fill this gap, in the present study we have developed an accident hotspot analysis technique that combines the temporal and spatial patterns of traffic crashes with the AR method to find the crash risk factors associated with a zone.

The main objective of this study is to identify the most significant crash risk factors responsible for the severity of crashes in high-risk urban zones. The high-risk crash severity pattern for these zones was obtained using the AR algorithm. With knowledge of the pattern of crash severity in different zones, we can develop effective safety countermeasures and driving strategies. The following method was used in the present study: first, in ESRI ArcGIS software, applying kernel density function, a black zone map was generated. Second, AR was used to identify the high-risk factors associated with each zone.

The contribution of the present study is as follows: first, the crash risk factors of different zones in urban areas were analysed, and the relationship between risk factors and urban zones was established. Second, a comparison of various zones was obtained from AR, which will help in traffic risk management. Last, this study explored the various advantages of using the AR method. The results show more understanding of the pattern of crash risk in an urban area of the high-risk zone. This can help traffic engineers improve road safety more effectively within the limited budget.

Methodology

A new method of traffic accident reporting system was introduced in India in 2017, allowing for the classification of the severity of traffic accidents (such as 'fatal' or 'grievously injured') in the official statistics. Each Police Department records accident information in a standard reporting format. Subsequently, all the States and Union Territories report their data to MoRTH, GoI, which annually updates the accident database[1]. The recorded information consists of several types of data, such as crash information (including accidents, accident severities and vehicle types), field information (such as urban and rural roads, lane markings, traffic signal control types, and road geometries and locations) and miscellaneous information (such as weather and time of the accident).

In the present study, two years of accident data, from 2018 to 2019, were acquired from the Nagpur Police Department. In these two years, 3268 crashes occurred in Nagpur city. Only 2243 of these had complete information, resulting in data for 1343 high-risk and 675 low-risk zone crashes. Despite having a large sample size, the total number of accidents included in this study was small due to insufficient information. Nonetheless, the study will give some valuable insights into the factors contributing to crashes in various zones of urban areas. The information available in the dataset was scrutinized and 11 factors were selected for analysis. After the initial analysis, three factors were excluded from the final analysis since they were not statistically significant. Table 1 shows the descriptive statistics of the eight factors used in the final model.

Data inclusion criteria

For data analysis, the criteria for utilizing traffic crashes from the database are as follows: (a) For fatal crashes, the criteria include the death of the driver, occupant or non-occupant within 30 days of an accident. (b) For injury crashes, the criteria include the victims who need to be hospitalized or at least first aid is required after the incident.

Spatial analysis

Using ArcGIS, a spatial distribution map was developed. The objective was to establish a traffic accident map showing the spatial distribution of accidents. The method described here aims to identify crashes with different concentrations for various zones. As mentioned above, the road crash data of Nagpur city from 2018 to 2019 were used for analysis. The geographical coordinates of the crash sites were recorded using a handheld GPS recorder (ETrex 20x). The black zones were then categorized into three zones: zone-I (high concentration levels), zone-II (medium concentration levels) and zone-III (low concentration levels).

Association rules

ARs find interesting relations or associations among variables based on different rules. ARs comprise a data-mining technique and were initially developed for market basket analyses. Data mining is a method that determines the logical pattern in crash data. It utilizes various techniques such as statistical analysis, machine learning, modelling method and data management to extract valuable information.

To illustrate the use of ARs for traffic safety, an example of a crash database consisting of different categories of accidents was considered. The crash categories were denoted as items. A set of items was referred to as an item set. $I = i_1, i_2, ..., i_m$ denotes a set of items and $C = c_1, c_2, ..., c_n$ denotes the crash information set in a database, where a record of crash c_i contains item sets that are a subset of the item set I. The item set containing k items was considered a 'k-item set'. The sample crash information shown in Table 2 contains an eight-item set, as there are eight variables for describing the crash characteristics. The AR analysis effectively identified the item sets occurring simultaneously in an event (crash). The number of times an item occurs alone or in combination with another set of

Table 1. Risk factors with crash frequency

Location	Factors	Frequency	Percentage
Junction type	T-junction	229	20.4
	Y-junction	161	14.3
	Four-arm junction	163	14.5
	Staggered junction	149	13.2
	Roundabout	142	12.6
	More than four-arms	156	13.8
	Others	126	11.2
Type of traffic control	Traffic light signal	183	16.2
	Police controlled signal	176	15.6
	Stop sign	166	14.7
	Flashing signal/blinker	170	14.9
	Uncontrolled	435	38.6
Collision type	Head-on collision	234	20.8
	Rear-end collision	103	9.1
	Hit and run	130	11.5
	Parked on street	35	3.1
	Vehicle to vehicle	311	27.6
	Vehicle to pedestrian	151	13.4
	Vehicle to non-motorized vehicle	105	9.3
	Vehicle to others	59	5.2
Age (yrs)	Less than 18	72	6.4
	18–25	251	22.3
	26–35	302	26.8
	36–45	256	21.8
	46–60	168	14.9
	Above 60	64	5.7
	Not known	24	2.1
Gender	Male	980	87
	Female	147	13
Time interval (h)	6:00–8:59	126	11.2
	9:00–11:59	185	16.4
	12:00–14:59	178	15.8
	15:00–17:59	197	17.4
	18:00–20:59	195	17.3
	21:00–23:59	115	10.2
	24:00–2:59	61	5.4
	3:00–5:59	46	4.1
	Not known	25	2.2

Source: Traffic Department, Nagpur, Maharashtra, India.

Table 2. Sample crash risk factors

Location	Junction control type	Collision type	Time (hours)	Day of week	Driver age (yrs)	Gender	Severity
Zone-I	Uncontrolled	Side collision	18–21	Weekday	25–35	Male	Injury
Zone-I	Uncontrolled	Side collision	15–18	Weekend	25–35	Male	Injury
Zone-I	Uncontrolled	Head-on collision	18–21	Weekday	18–25	Male	Fatal
Zone-I	Uncontrolled	Head-on collision	12–15	Weekday	35–45	Female	Injury
Zone-II	Traffic light	Vehicle to pedestrian	18–21	Weekday	35–45	Male	Injury
Zone-II	Traffic light	Vehicle to object	21–12	Weekday	18–25	Male	Injury
Zone-II	Traffic light	Head-on collision	21–24	Weekend	25–35	Male	Injury
Zone-III	Uncontrolled	Side collision	21–24	Weekend	25–35	Male	Injury
Zone-III	Uncontrolled	Head-on collision	18–21	Weekday	25–35	Female	Injury
Zone-III	Stop sign	Head-on collision	12–03	Weekend	35–40	Male	Fatal

items was used to determine its significance in the database.

The ARs were represented by $A \rightarrow B$, where A denotes antecedent and B denotes consequent. The rules were then filtered using the concepts of lift, confidence and support. The support refers to the percentage of specified cases in the data containing both A and B. The confidence is the percentage of cases that contain A, which also contain

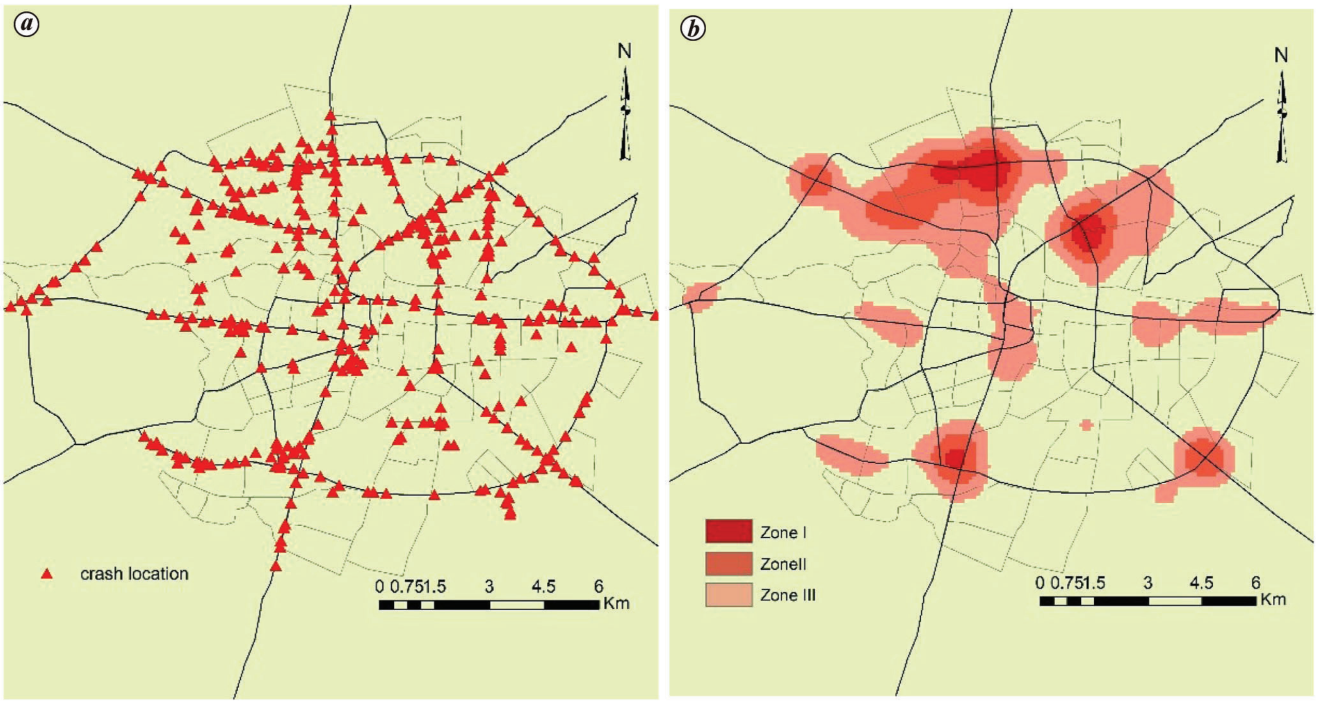

Figure 2. ***a***, Spatial distribution of crash frequency. ***b***, Black zone map.

B. The lift is the ratio between the confidence and the percentage of cases that contain *B*. The method of calculation is given below

$$\text{Support } (A \rightarrow B) = P(A \cup B).$$

$$\text{Confidence} = (A \rightarrow B) = A|B.$$

$$\text{Lift } (A \rightarrow B) = \frac{\text{Confidence } (A \rightarrow B)}{P(B)}.$$

If the lift value is greater than 1, it indicates a positive interdependency. If the lift value is less than 1, it indicates a negative interdependency, whereas a value equal to 1 indicates independence. The higher lift value shows the greater strength of ARs. For strong associations between variables, the limiting values assigned for lift (L), support (S), and confidence (C) were as follows: $L > 1$, $S > 40\%$ and $C > 60\%$. The analyses were conducted using Weka 3.8.5.

AR uses many algorithms to compute different sets of rules to extract useful information patterns from a dataset. One of the most widely used techniques for finding these patterns is the apriori algorithm[47]. Through an iteration process, the apriori algorithm identifies frequent item sets from a database. It generates many possible item sets, computes the rule's support and prunes the item sets to the most frequent one in each iteration. It also generates several item sets automatically and continues till the minimum support value, i.e. the threshold fixed for the study, is reached[48].

Results and discussion

In order to analyse the data related to fatal and injury crashes, a spatial analysis was conducted using ArcGIS. Figure 2 *a* and *b* represents the accident spatial distribution map and black zone map of Nagpur city respectively. The patterns of the crashes show that the majority of the accidents occur in the northern part of the city. A unique numerical value is assigned to the network to identify the zone. The area is divided into three categories based on inclusion criteria.

The plot shows that these clusters change within the city, which affects spatial variation. For instance, the western and northern portions of Nagpur city are more prone to experiencing traffic accidents. The clusters are formed most likely by the traffic characteristics such as the presence of more heavy vehicles and high speed limits, and also other environmental factors. The red colour in Figure 2 *b* shows the high-intensity zones and the orange colour shows the low-intensity zones. The northern part of Nagpur is observed to have a greater number of fatalities. This may be due to the high-speed vehicles entering the city from the highways as a result of the 'speed adaptation effect', which is seen in driver who continuously drives at higher speeds on the highways. When a driver continuously drives at high speed for a long time, a slow speed seems to appear much slower than the actual speed[49]. Thus, if a vehicle is travelling at high speed while exiting a highway, this will most likely lead to going through a curve or street at a higher speed.

Table 3. Association rules for two-cause risk factors

Antecedent	Consequent	Support	Confidence	Lift	Count
Zone type – Zone-I	Collision type – Head-on collision	0.031	0.873	1.58	36
Zone type – Zone-I	Collision type – Rear end	0.196	0.866	1.57	309
Junction type – Uncontrolled	Vehicle type – Car–motorcycle	0.011	0.853	1.55	13
Collision type – Sideswipe	Vehicle type – Truck–car	0.162	0.837	1.54	186
Collision type – Rear end	Driver gender – Male	0.16	0.842	1.48	184
Junction type – Four arms	Severity – Injury	0.021	0.834	1.46	24
Collision type – Fixed object	Time – 18–21 hours	0.021	0.833	1.45	12
Collision type – Sideswipe	Severity – Injury	0.025	0.745	1.43	29
Zone type – Zone-III	Severity – Injury	0.025	0.786	1.42	24
Zone type – Zone-II	Severity – Injury	0.026	0.788	1.41	30
Zone type – Zone-II	Collision type – Sideswipe	0.068	0.783	1.4	78
Zone type – Zone-I	Severity – Fatal	0.017	0.716	1.38	20
Driver age – 25–35 yrs	Severity – Injury	0.029	0.756	1.36	33
Junction type – Y-junction	Zone type – Zone-III	0.027	0.815	1.35	31
Collision type – Sideswipe	Zone type – Zone-III	0.074	0.858	1.35	85
Collision type – Rear end	Zone type – Zone-II	0.165	0.857	1.35	19
Junction type – Blinker	Zone type – Zone-I	0.011	0.856	1.35	13
Collision type – Sideswipe	Zone type – Zone-II	0.019	0.853	1.34	22
Collision type – Rear end	Junction type – Four arms	0.028	0.847	1.33	32
Time – 00–3 hours	Collision type – Vehicle–vehicle	0.012	0.625	1.27	14
Driver gender – Male	Zone type – Zone-I	0.02	0.788	1.25	23
Driver gender – Male	Day of week – Weekend	0.023	0.618	1.24	26
Driver gender – Male	Time – 3–6 hours	0.025	0.742	1.24	29
Day of week – Weekday	Driver gender – Female	0.122	0.627	1.24	140
Day of week – Weekend	Driver gender – Male	0.064	0.626	1.23	74
Collision type – Vehicle–pedestrian	Zone type – Zone-I	0.055	0.668	1.22	63
Collision type – Vehicle-vehicle	Zone type – Zone-I	0.019	0.76	1.22	22
Time – 21–24 hours	Zone type – Zone-I	0.013	0.753	1.21	15
Collision type – Rear end	Driver gender – Male	0.029	0.757	1.2	33
Collision type – Rear end	Junction type – Four arms	0.355	0.653	1.18	409

Crash risk factors identification

As mentioned above, the apriori algorithm was used in this study to frame ARs for crash severity and influencing risk factors for the three study zones. Tables 3–5 show the frequencies of the item sets generated by ARs. These rules have lift > 1, confidence > 60% and support greater than 30%. Among the rules generated, only the top 30 are shown for the two-item rules, and only the top 20 rules are shown for the three- and four-item rules. Maximum rules are generated for the two-factor sets. In many instances, a higher number of item sets can be difficult to interpret. Therefore, analysis of the results is limited to the four-factor rules for easy interpretation.

Table 3 shows the first 30 rules for the two-factor sets, together with their degrees of lift, confidence and support. As can be seen from the results, the crash risk pattern comprises different combinations of two-item rules. The combinations of crash proportions differ compared to all the crashes at each crash severity level in a specific zone. The top six frequent items in the dataset are zone type = zone-l, driver gender = male, time = 18–21 h, severity = injury, junction type = uncontrolled and collision type = vehicle–pedestrian. This finding indicates that the association of these risk factors is strongly connected to the crashes occurring in urban zone-I. More male drivers are involved in crashes compared to female drivers. This may be because a greater number of drivers are male, particularly in commercial vehicles. Previous research has also found a similar trend[50]. Male drivers drive dangerously, such as exceeding the speed limit, drunken driving and overtaking. In addition, if a junction comprises an uncontrolled signal or stop sign, additional conflict points are generated. This may result in a chaotic situation, which can, in turn, be responsible for the additional crashes at the junction.

At the junctions in zone-I, the risk is more with high support value while driving at night, especially for young drivers (aged between 18 and 25 years). This can be attributed to their lack of driving experience, and aggressive and risky driving behaviours[51]. These rules show that the zone-1 crashes involving ‘vehicle to vehicle’ collisions and ‘head-on’ collisions are frequent. The variable {severity = fatal} is associated with zone-1, whereas the variable {severity = injury} is associated with zone-II and zone-III among the top ten rules. This confirms that the roadway segment in zone-1 is closely associated with vehicle–vehicle crashes at night. Nearly 58% of the night-time crashes occur on zone-1 streets. In zone-1, more rules were associated with different risk categories. The AR {collision type = rear end} $\Rightarrow$ {zone type = zone-I} shows a high support value (0.196), having a frequency value of 309. Zone-I is more crash-prone during nighttime than

Table 4. Association rules for three-cause factors

Antecedent	Consequent	Support	Confidence	Lift	Count
Zone type = Zone-I Junction type = Stop sign	Collision type – Vehicle–vehicle	0.023	0.982	1.662	15
Zone type = Zone-I Severity = Fatal	Collision type – Car–motorcycle	0.063	0.981	1.659	84
Zone type = Zone-I Junction type = Uncontrolled	Day of week – Weekday	0.047	0.973	1.649	18
Day of week = Weekend Junction type = Uncontrolled	Collision type – Sideswipe	0.026	0.891	1.543	13
Zone type = Zone-II Driver age = 18–24 yrs	Collision type – Vehicle–pedestrian	0.022	0.885	1.538	55
Zone type = Zone-II Collision type = Vehicle–pedestrian	Day of week – Weekend	0.037	0.875	1.516	26
Zone type = Zone-II Driver gender = Male	Junction type – Uncontrolled	0.048	0.856	1.499	67
Zone type = Zone-III Collision type = Sideswipe	Day of week – Weekday	0.022	0.852	1.489	77
Zone type = Zone-III Severity = Injury	Collision type – Rear end	0.059	0.842	1.485	24
Zone type = Zone-III Junction type = Stop sign	Collision type – Vehicle–vehicle	0.037	0.838	1.484	97
Collision type = Sideswipe Driver gender = Male	Collision type – Vehicle–vehicle	0.067	0.832	1.471	106
Day of week = Weekend Junction type = Uncontrolled	Collision type – Vehicle–vehicle	0.021	0.829	1.467	211
Collision type = Right angle Driver gender = Female	Collision type – Vehicle–vehicle	0.085	0.824	1.456	43
Collision type = Head on Driver gender = Male	Severity – Fatal	0.092	0.821	1.451	12
Collision type = Rear end Driver age = Below 18 yrs	Zone type – Zone-II	0.184	0.818	1.448	32
Collision type = Rear end Driver age = 25–34 yrs	Severity – Injury	0.037	0.812	1.446	44
Collision type = Sideswipe Driver gender = Male	Severity – Injury	0.011	0.812	1.441	124
Collision type = Rear end Severity = Injury	Time period – 18–21 hours	0.028	0.809	1.436	34
Collision type = Rear end Driver gender = Male	Zone type – Zone-II	0.038	0.804	1.332	20
Collision type = Rear end Driver gender = Female	Collision type = Vehicle–vehicle	0.108	0.801	1.327	68

daytime. Female driver crashes are rare in these zones, as many drivers are males.

Table 4 lists the first 20 rules for the three-cause factors Rule 1, i.e. zone type = zone-I, junction type = stop sign ⇒ collision type = vehicle–vehicle, has the highest lift value. Gender = male, vehicle type = truck car, and junction type = uncontrolled are three risky factors associated with severe crashes in zone-II. The younger drivers are correlated with crashes involving motorcycles and other vehicles. The ARs indicate that male drivers have a high support value and low lift value, whereas female drivers have a high lift value and low support value. This indicates that male drivers are more frequently involved with particular risk factors, and female drivers are rarely involved in crashes; nevertheless, they strongly associate with the rule.

Table 5 presents the four-factor rules. The rule zone type = zone-l, junction type = uncontrolled, driver age = 25–35 yrs ⇒ collision type = vehicle–vehicle indicates the high lift value. The succeeding rule is connected with severity = injury and collision type = pedestrians–vehicle. Another interesting rule is zone type = zone-III, severity = injury ⇒ gender = female. The same kind of rule is correlated with male drivers having high support and low lift values. Such types of rules indicate that different zones are associated with different combinations of risk factors. Collision type = vehicle–vehicle, exist in 60% of the first 20 four-factor rules. The variable zone type = zone-I occurs more regularly, comprising nearly 20% of the four-factor rules. Overall, the ARs mentioned above suggest potential associations for crashes in zones. After studying the ARs based on two, three and four factors, some variable categories are perceived as more influential for zonal crashes. In addition, the risk factor combination in all the zone types is complex and dissimilar. This suggests that the different

Table 5. Association rules for four-cause risk factors

Antecedent	Consequent	Support	Confidence	Lift	Count
Zone type = Zone-I Junction type = Uncontrolled Driver age = 25–34 yrs	Collision type = Vehicle–vehicle	0.015	0.939	1.692	15
Zone type = Zone-I Junction type = Uncontrolled Severity = Injury	Collision type = Vehicle–pedestrian	0.027	0.932	1.675	32
Zone type = Zone-I Collision type = Vehicle–pedestrian Severity = Fatal	Junction type = Stop sign	0.038	0.928	1.673	42
Day of week = Weekend Time = 21–24 Driver gender = Male	Collision type = Vehicle–vehicle	0.022	0.926	1.668	24
Collision type = Vehicle–vehicle Driver age = 18–24 yrs Severity = Injury	Day of week = Weekend	0.052	0.962	1.508	60
Zone type = Zone-III Day of week = Weekday Severity = Injury	Collision type = Vehicle–vehicle	0.025	0.951	1.507	29
Zone type = Zone-III Day of week = Weekday Severity = Fatal	Collision type = Vehicle–vehicle	0.018	0.943	1.507	20
Zone type = Zone-III Collision type = Vehicle–vehicle Severity = Injury	Junction type = Uncontrolled	0.013	0.851	1.506	34
Zone type = Zone-III Severity = Injury Driver gender = Female	Collision type = Vehicle–vehicle	0.016	0.848	1.504	19
Zone type = Zone-III Driver age = 18–24 yrs Severity = Injury	Collision type = Vehicle–vehicle	0.014	0.842	1.51	46
Zone type = Zone-II Collision type = Vehicle–pedestrian Severity = Injury	Day of week = Weekday	0.011	0.836	1.499	12
Zone type = Zone-II Day of Week = Weekday Driver age = 25–34 yrs	Collision type = Vehicle–vehicle	0.082	0.832	1.499	94
Day of week = Weekend Zone type = Zone-III Driver age = 18–24 yrs	Collision type = Vehicle–vehicle	0.023	0.832	1.494	35
Zone type = Zone-I Severity = Injury Driver gender = Male	Collision type = Vehicle–vehicle	0.024	0.831	1.494	27
Zone type = Zone-I Driver gender = Male Driver age = 18–24 yrs	Collision type = Vehicle–vehicle	0.051	0.829	1.492	59
Collision type = Vehicle–Vehicle Severity = Injury Driver gender = Male	Junction type = Uncontrolled	0.108	0.825	1.49	125
Zone type = Zone-II Driver gender = Male Driver age = 35–44 yrs	Collision type = Vehicle–vehicle	0.024	0.822	1.488	28
Zone type = Zone-II Time = 15–16 Severity = Injury	Collision type = Vehicle–pedestrian	0.078	0.817	1.487	90
Zone type = Zone-II Severity = Injury Driver gender = Male	Collision type = Vehicle–vehicle	0.022	0.814	1.485	26
Collision type = Vehicle–pedestrian Time = 9–12 hours Severity = Injury	Day of week = Weekday	0.021	0.811	1.476	24

types of urban zones should be treated differently while implementing safety improvement measures. Through the use of crash severity analysis and identifying risk patterns of zones, traffic management authorities can identify potential crash risks in urban areas where they can implement effective safety measures (e.g. introducing overtaking prohibited zones, traffic signal systems at uncontrolled junctions, and developing efficient pedestrian facilities) in risky zones.

Conclusion

In this study, AR analysis was used to examine the crash risk levels of the road traffic network of three urban zones. The analysis evaluated the significant contributory factors and patterns of crash severity in the form of rules, such as those based on two-cause, three-cause and four-cause factors. These findings can help in establishing priorities for formulating and enforcing corrective countermeasures. By comparing the AR analysis results of high-risk factors, we can suggest remedies that require immediate attention. The pattern of crash severity obtained from the AR analysis provides an understanding of the interdependencies of multiple factors with regard to each other. The findings can be summarized as follows:

- The fatal crash probability at night is greater compared to the daytime.
- The overall injury crashes probability is greater than that of fatal crashes.
- The average probability of night-time crashes is slightly higher than daytime crashes.
- The relative crash risk levels of uncontrolled and stop sign junctions are higher.
- The risk levels of various factors differ according to the crash severity patterns and location, i.e. the type of zone. Certain risk factors, such as young male drivers increase the fatality rate in high-risk zones.
- Other factors, such as pedestrian facilities and uncontrolled intersections, should be given attention based on priority.

The following suggestions are relevant with regard to the present study.

- Drivers need to improve their performance on urban streets, especially in regions where heavy vehicles enter the city from highways.
- Automatic messaging feedback should be provided to drivers, e.g. with reference to their driving performance.
- Based on the effects of speed adaptation, the drivers travelling on the highway may approach junctions and curve at a higher speed.
- Measures should be taken to avoid conflicts and attention when presenting critical information (e.g. other sources of light close to traffic signals, and advertisement hoarding near hazard warning and directional signboards).

In conclusion, the high-risk crash zones of different areas and the road network in Nagpur city revealed unique characteristic patterns of crash severity. Therefore, the AR algorithm is suitable for identifying crash risk associated with each zone. In the present study, it has been shown that the crash severity analysis by AR generates more objective results, and thus has a great potential of being successfully used in large datasets. Future research with regard to the application of AR in safety-related topics is encouraging and parameters estimated by ARs could be further explored. Since the present dataset lacks some of the valuable information related to crashes, further studies are required with greater emphasis on micro-level data to obtain specific potential crash risk factors in a zone, such as a driver's behaviour, vehicle speed and traffic volume. The road safety studies of urban areas in developing countries like India need to prioritize building a national crash database, similar to those in developed countries. Since most of the minor crashes are not reported in developing countries, the crash risk analysis of such data will significantly affect the final result. Also, compared to developed countries, the traffic in India is heterogeneous. So, traffic safety problems in urban zones will remain a field of future research in the country.

Conflict of interest: The authors declare that they have no conflict of interest.

1. Transport Research Wing, *Road Accident in India*, Transport Research Wing, Ministry of Road Transport and Highways, Government of India, New Delhi, 2018.
2. Fuller, R., The task–capability interface model of the driving process. *Recherche-Transports-Sécurité*, 2000, **66**, 47–57.
3. Lord, D. and Mannering, F., The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. Part A*, 2010, **44**, 291–305.
4. Savolainen, P. T., Mannering, F. L., Lord, D. and Quddus, M. A., The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accid. Anal. Prev.*, 2011, **43**, 1666–1676.
5. Mannering, F. L. and Bhat, C. R., Analytic methods in accident research: Methodological frontier and future directions. *Anal. Methods Accid. Res.*, 2014, **1**, 1–22.
6. Christoforou, Z., Cohen, S. and Karlaftis, M. G., Identifying crash type propensity using real-time traffic data on freeways. *J. Safety Res.*, 2011, **42**, 43–50.
7. Miranda-Moreno, L. F., Fu, L., Saccomanno, F. F. and Labbe, A., Alternative risk models for ranking locations for safety improvement. *Transp. Res. Rec.*, 2005, **1908**, 1–8.
8. Wu, P., Meng, X., Song, L. and Zuo, W., Crash risk evaluation and crash severity pattern analysis for different types of urban junctions: fault tree analysis and association rules approaches. *Transp. Res. Rec.*, 2019, **2673**, 403–416.
9. de Melo, W. A., Alarcão, A. C. J., de Oliveira, A. P. R., Pelloso, S. M. and Carvalho, M. D. D. B., Age-related risk factors with nonfatal traffic accidents in urban areas in Maringá, Paraná, Brazil. *Traffic Inj. Prev.*, 2017, **18**, 157–163.
10. Martensen, H. and Dupont, E., Comparing single vehicle and multivehicle fatal road crashes: a joint analysis of road conditions, time variables and driver characteristics. *Accid. Anal. Prev.*, 2013, **60**, 466–471.
11. Yasmin, S., Eluru, N., Bhat, C. R. and Tay, R., A latent segmentation based generalized ordered logit model to examine factors

influencing driver injury severity. *Anal. Methods Accid. Res.*, 2014, **1**, 23–38.
12. Zhang, G., Yau, K. K. W. and Gong, X., Traffic violations in Guangdong Province of China: speeding and drunk driving. *Accid. Anal. Prev.*, 2014, **64**, 30–40.
13. Yan, X., Radwan, E. and Abdel-Aty, M., Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accid. Anal. Prev.*, 2005, **37**, 983–995.
14. Bíl, M., Bílová, M., Dobiáš, M. and Andrášik, R., Circumstances and causes of fatal cycling crashes in the Czech Republic. *Traffic Inj. Prev.*, 2016, **17**, 394–399.
15. Rezapour, M., Moomen, M. and Ksaibati, K., Ordered logistic models of influencing factors on crash injury severity of single and multiple-vehicle downgrade crashes: a case study in Wyoming. *J. Safety Res.*, 2019, **68**, 107–118.
16. Kim, J.-K., Kim, S., Ulfarsson, G. F. and Porrello, L. A., Bicyclist injury severities in bicycle–motor vehicle accidents. *Accid. Anal. Prev.*, 2007, **39**, 238–251.
17. Khasawneh, M. A., Al-Omari, A. A. and Oditallah, M., Assessing speed of passenger cars at urban channelized right-turn roadways of signalized intersections. *Arab. J. Sci. Eng.*, 2019, **44**, 5057–5073.
18. Eluru, N., Bhat, C. R. and Hensher, D. A., A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accid. Anal. Prev.*, 2008, **40**, 1033–1054.
19. Rifaat, S. M., Tay, R. and De Barros, A., Effect of street pattern on the severity of crashes involving vulnerable road users. *Accid. Anal. Prev.*, 2011, **43**, 276–283.
20. Li, Y., Liu, C. and Ding, L., Impact of pavement conditions on crash severity. *Accid. Anal. Prev.*, 2013, **59**, 399–406.
21. Hosseinpour, M., Yahaya, A. S. and Sadullah, A. F., Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: case studies from Malaysian Federal Roads. *Accid. Anal. Prev.*, 2014, **62**, 209–222.
22. Depaire, B., Wets, G. and Vanhoof, K., Traffic accident segmentation by means of latent class clustering. *Accid. Anal. Prev.*, 2008, **40**, 1257–1266.
23. Clarke, D. D., Ward, P., Bartle, C. and Truman, W., Young driver accidents in the UK: The influence of age, experience, and time of day. *Accid. Anal. Prev.*, 2006, **38**, 871–878.
24. Yau, K. K. W., Lo, H.-P. and Fung, S. H. H., Multiple-vehicle traffic accidents in Hong Kong. *Accid. Anal. Prev.*, 2006, **38**, 1157–1161.
25. Khorashadi, A., Niemeier, D., Shankar, V. and Mannering, F., Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. *Accid. Anal. Prev.*, 2005, **37**, 910–921.
26. Keay, K. and Simmonds, I., Road accidents and rainfall in a large Australian city. *Accid. Anal. Prev.*, 2006, **38**, 445–454.
27. Loo, B. P. Y. and Yao, S., The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment. *Comput. Environ. Urban Syst.*, 2013, **41**, 249–261.
28. Eckley, D. C. and Curtin, K. M., Evaluating the spatiotemporal clustering of traffic incidents. *Comput. Environ. Urban Syst.*, 2013, **37**, 70–81.
29. Gundogdu, I. B., Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: case study of Konya. *Saf. Sci.*, 2010, **48**, 763–769.
30. Truong, L. T. and Somenahalli, S. V. C., Using GIS to identify pedestrian–vehicle crash hot spots and unsafe bus stops. *J. Public Transp.*, 2011, **14**, 6.
31. Erdogan, S., A comparison of interpolation methods for producing digital elevation models at the field scale. *Earth Surf. Process. Landf.*, 2009, **34**, 366–376.
32. Vandenbulcke, G., Thomas, I. and Panis, L. I., Predicting cycling accident risk in Brussels: a spatial case-control approach. *Accid. Anal. Prev.*, 2014, **62**, 341–357.
33. Thakali, L., Kwon, T. J. and Fu, L., Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. *J. Mod. Transp.*, 2015, **23**, 93–106.
34. Xie, Z. and Yan, J., Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urban Syst.*, 2008, **32**, 396–406.
35. Moons, E., Brijs, T. and Wets, G., Improving Moran's index to identify hot spots in traffic safety. In *Geocomputation and Urban Planning*, Springer, 2009, pp. 117–132.
36. Cheng, W. and Washington, S. P., Experimental evaluation of hotspot identification methods. *Accid. Anal. Prev.*, 2005, **37**, 870–881.
37. Okabe, A., Satoh, T. and Sugihara, K., A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int. J. Geogr. Inf. Sci.*, 2009, **23**, 7–32.
38. Flahaut, B., Mouchart, M., San Martin, E. and Thomas, I., The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. *Accid. Anal. Prev.*, 2003, **35**, 991–1004.
39. Anderson, T. K., Kernel density estimation and *K*-means clustering to profile road accident hotspots. *Accid. Anal. Prev.*, 2009, **41**, 359–364.
40. Steenberghen, T., Aerts, K. and Thomas, I., Spatial clustering of events on a network. *J. Transp. Geogr.*, 2010, **18**, 411–418.
41. Agrawal, R., Imielinski, T. and Swami, A., Database mining: a performance perspective. *IEEE Trans. Knowl. Data Eng.*, 1993, **5**, 914–925.
42. Kaur, M. and Kang, S., Market basket analysis: identify the changing trends of market data using association rule mining. *Procedia Comput. Sci.*, 2016, **85**, 78–85.
43. Kim, J., Bang, C., Hwang, H., Kim, D., Park, C. and Park, S., IMA: Identifying disease-related genes using MeSH terms and association rules. *J. Biomed. Inform.*, 2017, **76**, 110–123.
44. Montella, A., Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. *Accid. Anal. Prev.*, 2011, **43**, 1451–1463.
45. Pande, A. and Abdel-Aty, M., Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Saf. Sci.*, 2009, **47**, 145–154.
46. Marukatat, R., Structure-based rule selection framework for association rule mining of traffic accident data. In *International Conference on Computational and Information Science*, Springer, 2006, pp. 231–239.
47. Agrawal, R., Imielinski, T., and Swami, A., Mining association in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 1993, pp. 207–216.
48. Yabing, J., Research of an improved a priori algorithm in data mining association rules. *Int. J. Comput. Commun. Eng.*, 2013, **2**, 25–27.
49. Schmidt, F. and Tiffin, J., Distortion of drivers' estimates of automobile speed as a function of speed adaptation. *J. Appl. Psychol.*, 1969, **53**, 536.
50. Fountas, G., Anastasopoulos, P. C. and Abdel-Aty, M., Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. *Anal. Methods Accid. Res.*, 2018, **18**, 57–68.
51. Paleti, R., Eluru, N. and Bhat, C. R., Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accid. Anal. Prev.*, 2010, **42**, 1839–1854.

ACKNOWLEDGEMENT. We thank the Nagpur Police Department for providing the accident data of Nagpur city.

Received 15 March 2021; revised accepted 14 July 2022

doi: 10.18520/cs/v123/i6/781-790