Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–19

DOI: 10.1177/03611981251407917
journals.sagepub.com/home/trr
Sage

Segmentation of Risk Factors for Fatal Crashes at Urban Signalized Intersections: A Multi-Perspective Model Approach

Siddardha Koramati[1], Bandhan Bandhu Majumdar[2], Prasanta K. Sahu[2], Surojit Das[3], Aritro Ghosh[4], and Sabyasachi Biswas[4]

Abstract

Signalized intersections are frequently installed in developing countries to facilitate efficient traffic flow and seldom to increase traffic safety. As a result, fatal collisions still occur at intersections with signals. The purpose of this study is to gain a better understanding of signalized intersection safety by identifying and segmenting traffic and geometric risk factors associated with fatal crashes. For this purpose, a thorough road inventory survey—primary crash data—was used to analyze crashes at 67 signalized intersections in Hyderabad, an Indian metropolitan city. This paper proposes a multi-perspective model application and segmentation strategy that classifies a group of important crash factors determining crash fatality at urban signalized intersections by combining machine learning, data mining, and statistical modeling results. The proposed segmentation divided the crash parameters into three distinct categories: very high, high, and moderate risk factors. The key findings show that major road width, lack of right-turn protection, and absence of all-red time are the most influential factors contributing to fatal crashes at signalized intersections. Based on the findings, several policy recommendations were proposed. The segmentation of signalized intersection features would provide useful insights into the level of their influence and the impact of signalized intersection design on safety in developing countries. The study's findings and proposed policy insights may assist transportation officials in developing, prioritizing, and implementing specialized safety countermeasures for signalized intersections.

Keywords

signalized intersections, fatal crashes, multi-perspective model, statistical, machine learning, data mining

Introduction

Road crashes are a prominent cause of death among children, youth, and working-age groups, bringing significant health, social, and economic impacts to society around the world. The United Nations Decade of Action for Road Safety 2021–2030 aimed to reduce road traffic deaths globally by half; however, estimates show that fatalities have decreased by only 5% (*1*). Further, nine out of ten road crash deaths occur in low- and middle-income countries (LMICs) (*1*). India, an LMIC, has the world's highest reported annual road crash fatalities (*2*). The percentage of fatal crashes increased from 30.6 to 33.8 between 2018 and 2022 (*3*). The working age group of 18 to 60 accounted for 83.4% of total road crash deaths in 2022 (*3*). Further, in 2022, the proportion of fatalities that occurred in urban areas spiked from 31% to 32%, respectively. India is rapidly urbanizing, with over 600 million people predicted to reside in cities by 2036 (*4*). These data indicate that this will exacerbate cities' poor road safety. Multiple factors are to blame for this unsatisfactory situation, including a highly diverse

[1]Road Safety Program Manager, Sarvejana Foundation, Hyderabad, India
[2]Associate Professor, Department of Civil Engineering, BITS Pilani, Hyderabad Campus, India
[3]Transportation Research and Injury Prevention Centre, Indian Institute of Technology Delhi, Delhi, India
[4]Assistant Professor, Department of Civil Engineering, NIT Durgapur, West Bengal, India

Corresponding Author:
Bandhan Bandhu Majumdar, bandykolkata@gmail.com

traffic mix, unsafe operating conditions, inadequate safe access, and crossing facilities at intersections. Thus, there is a need to emphasize crash investigation at various levels (urban/rural) and road sections (intersections/midblock). Intersections are widely recognized as crash hotspots on road networks. Road intersections are regions where complicated conflicting traffic movements (merges, diverges) occur among distinct road users, making them prone to collisions. Although traffic light signals are thought to be the most effective approach to managing the flow of traffic, more than 2,238 people died in India at signalized intersections in 2022 (*3*). Finding the factors that influence fatal crashes is essential to propose effective countermeasures for lowering the fatality rate. Intersection-related crashes are influenced by various factors, including roadway, traffic, environmental, and driving attributes. Among these, roadway and traffic factors significantly affect injury severity. Road safety is an intricate problem that requires a multifaceted strategy to achieve the most effective results. Therefore, using a single modeling technique to predict crash outcomes may not adequately reveal the most critical risk elements. To address this challenge, a multi-perspective model approach may be employed to minimize shortcomings. Furthermore, a proper segmentation criterion could be utilized to gain insight into the features with the highest impact on traffic accidents. In this context, an investigation of the risk factors linked to fatal road crashes was done for Hyderabad City using statistical, machine learning, and data mining techniques. Further, a suitable segmentation criterion is proposed to classify the results of similar levels to identify the greatest risk factor. An appropriate segmentation criterion and segments of factors can be employed to identify the key risk factor segments to be considered and develop necessary strategies. The next section discusses key research that informs this study.

Literature Review

The association between crash fatalities in signalized intersections and variables that may influence the occurrence of crashes has been extensively explored and investigated in research based on aggregated data. Several factors, including total approach volumes, the number of phases in a cycle, the presence of a surveillance camera, free left-turn, permissive right-turn phases, the presence of a horizontal curve, the presence of a crosswalk, the number of lanes, cycle length, and so on, were found to have a significant impact on crashes (*5–9*). Several studies used statistical techniques, data mining, and machine learning models to predict and identify which factors would significantly influence severe crashes (*10–13*). Each of the methods mentioned above has its strengths and limitations. While statistical modeling techniques can determine independent variables having a greater impact on output, the application of artificial neural networks (ANNs) and association rules mining (ARM) can lead to a better understanding of the relationship between the causes and crash outcomes. Acknowledging that crash datasets are acutely imbalanced, no one method can be considered as an appropriate technique. Crash fatality is one of the road safety-related aspects that requires thorough investigation. However, to our knowledge, no study has used logistic regression, ANNs, or association rule mining and compared their results to establish a relationship between fatal crashes and intersection parameters in an urban setting in a developing country. Therefore, adopting different crash prediction models, including logistic regression, ANNs, and association rule mining, and comparing their outputs would better evaluate risk factors. Table 1 summarizes a few other significant past studies in this setting. The following sub-section presents the key research gaps and study approach to address them.

A brief review of existing research on the relevant area indicates several existing research and practice gaps to be addressed. First, the existing body of scholarly research on road safety, particularly signalized intersections, predominantly derives from high-income countries (HICs). This observation emphasizes a significant research void that needs to be addressed. The applicability of research findings from the HICs to the LMICs context may be limited attributable to distinct traffic conditions. Therefore, there is a need for conducting a detailed analysis of crash data along with key primary and secondary attributes to identify key risk factors leading to fatal crashes in a typical LMIC such as India. Understanding the factors leading to fatal crashes would help policymakers to plan proactive strategies to reduce traffic crashes. Second, the literature review clearly shows that several studies have explored the individual factors leading to road traffic crashes, but a lack of studies could be observed toward identifying risk factor segments particularly for signalized intersections. Thirdly, a combined application of robust statistical methods, machine learning, and data mining techniques for safety assessment is another unique contribution of the research. Therefore, development of a unique risk factor segmentation based on a combination of three different approaches related to fatal road traffic crashes at signalized intersections of a LMIC such as India remains the broad objective of the study. To demonstrate the proposed approach, Hyderabad, an Indian metropolitan city with significant crash history across various signalized intersections have been identified as the case study city. The next section presents the research methods used in this study.

Table 1. Summary of Existing Econometric, Artificial Neural Network (ANN) and Data Mining Studies in Road Safety

Researcher(s), year	Study area/database	Factors	Proposed technique	Key findings
Anjana & Anjaneyulu (*13*)	Traffic and geometric data from 32 intersections from three urban areas namely, Trivandrum, Ernakulam, and Kozhikode in the state of Kerala, India	Lane markings, median width, presence of countdown timer, presence of surveillance camera, exclusive left-turn lane, traffic volume, accused and victim's vehicle type	Logistic regression models and hierarchical Poisson regression	Exclusive left-turn lanes and countdown timer's presence of a surveillance camera, green time, median width, traffic volume, and proportion of two wheelers influence the safety
Xu et al. (*14*)	1889 pedestrian-related crashes at 318 signalized intersections between 2008 and 2012	Sex and age, time, crowded/obstructed footpath, heedless crossing, inattentive, road type, junction type, speed limit, and traffic congestion, weather, light, and road surface	Bayesian hierarchical logit model	Age>65, footpaths not obstructed/ overcrowded, heedless or inattentive crossing, crashes on the two-way carriageway, and those that occurred near tram or light-rail transit stops led to a significantly higher probability of pedestrians being killed
Xie et al. (*15*), Elvik (*16*)	898 pedestrian crashes at 262 signalized intersections during 2010–2012 in Hong Kong	Geometric design, traffic characteristics, signal control, built environment, along with the vehicle and pedestrian volumes	Poisson lognormal models	The number of crossing pedestrians, the number of passing vehicles, the presence of curb parking, and the presence of ground-floor shops were positively related with pedestrian crash
Mitra & Bhowmick (*17*)	52 signalized intersections of Kolkata, India. Crash records from 2011 - 2014	Road geometric features, infrastructural details, land use, and traffic control parameters	Binary logit model and negative binomial model	All-red time, protected right-turning phase, blocked carriageway, non- motorized traffic, and visibility of road markings affect the severity of crashes at signalized intersections
Haque & Ahmad Kidwai (*18*)	11 signalized intersections in New Delhi, India	Age, walking pace, mobile usage crossing path, stage crossing, carrying object waiting time crossing speed.	ANN and binary logit model	Pedestrian demographic, behavioral attributes, crossing state and mobile usage are notably influencing signal violations. The accuracy of ANN to predict signal violation behavior is found to be about 85% and is considerably higher than binary logistic regression model
Kuşkapan et al. (*19*)	Data from 197 crashes injury (2015–2019) at 57 intersections, Erzurum, Turkey	Age, gender, level of education, pedestrian density, traffic congestion vehicle type, presence of bus stops condition of road surface, width of the roads, number of lanes, time, day of week and weather	Artificial neural network (ANN)	The results of the sensitivity analysis revealed that the incidence of pedestrian-vehicle collisions is primarily influenced by traffic congestion, age, pedestrian density, and vehicle type.

(continued)

Table 1. (continued)

Researcher(s), year	Study area/database	Factors	Proposed technique	Key findings
Das et al. (*20*)	Five years (2014–2018) of fatal crash data, US.	Crash type, crash location, pedestrian position, motorist maneuver, intersection leg, mark crosswalk present	Associated rules mining—Apriori algorithm	key variable categories such as dark with lighting condition, vehicle going straight, vehicle turning, local municipality streets, pedestrian age range from 45 years and above are frequently presented in the developed rules
Samerei et al. (*21*)	Records of bus crashes in Victoria, Australia between 2006 and 2019	Traffic control, road geometry, road type, speed limit, light condition, surface type, weather, gender, age	Two-step clustering and association rules discovery	Darkness of roads in high-speed zones, pedestrian presence at highways, bus crashes with passenger car by a female bus driver, and the occurrence of multi-vehicle crashes in high-speed zones increase the risk of fatality

Methodology

The current study used a combination of statistical analysis, data mining, and ANNs to determine the elements that have the greatest influence on the outcome of road traffic fatalities. All these three techniques were independently applied to a common crash dataset derived from Hyderabad city for identifying top-most risk factors leading to road traffic fatality and were subsequently combined for road risk factor segmentation. The applied three methods are significantly different from each other relating to methodological basis and have their own respective strengths and weaknesses. First, the logistic regression technique can predict the probability of a binary outcome (e.g., fatal versus non-fatal). It has got a set of strengths such as easy interpretation, less modeling complexity and effectiveness with relatively smaller dataset. On the other hand, the logistic regression technique has several limitations. Firstly, it generally assumes a linear relationship between the independent variables and the log-odds of the outcome. Secondly, the logistic regression technique remains significantly sensitive to multicollinearity between independent variables as well. Third, the model remains restricted to use for binary type dependent variables only. Although this study has only two levels of crash severity, namely, fatal and non-fatal crashes, it is essential to use other methods which can handle other types of dependent variable as well as handle more complex patterns. Second, the ANN, which has also been employed in this study, provides substantially better predictive performance compared with traditional logistic regression techniques and can model complex non-linear relationships. ANN can identify hidden/complex patterns in road safety data as well, which makes it a more suitable choice for road safety modeling approach. However, both these techniques can identify an individual level attribute and its role in fatal crash severity only. For uncovering the interdependency between the individual attributes, and to check their combined impact on fatality that is overlooked by the regression model, Apriori algorithm—an ARM technique—can be used. Apriori algorithm can detect intriguing rules from a large dataset without preconceived notions. The primary benefit of the method is that it does not necessitate predetermined assumptions, unlike logistic regression models. Therefore, using Apriori algorithm, the ARM technique was employed to detect interactive associations among the specific factors. Therefore, the three methods with their respective strengths and weaknesses were applied to identify key risk variables resulting from each of the techniques and subsequently combined to demonstrate a multi-perspective model approach for risk factor prioritization and classification. The methodology's steps are outlined in Figure 1.

Study Area, Data Collection, and Database Development

Hyderabad is a city in Telangana, a state in central-south India. It serves as the primary urban hub for the whole south-central region of India and is Telangana's largest

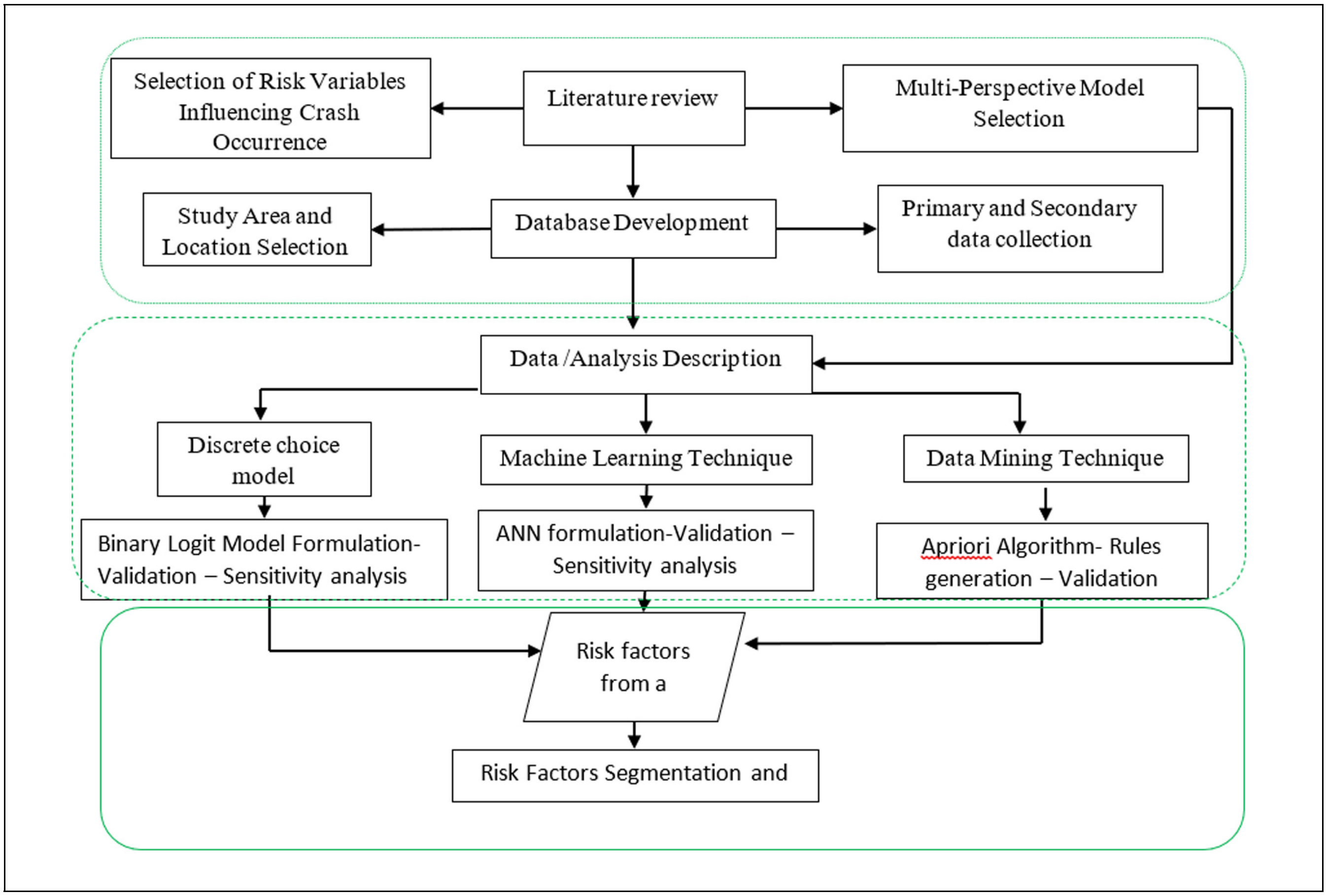

Figure 1. Methodological framework.
Note: ANN = artificial neural network.

and most populous metropolis (*22*). The number of crashes in Hyderabad increased 14.5% in 2023 compared with 2022 (*23*). The primary database was developed utilizing First Information Report (FIR) data collected from the Hyderabad Police Department between 2015 and 2019. Several critical pieces of information were missing from the data collected by the police agency. A road inventory survey (RSI) at 67 signalized intersections was conducted to collect the lack of road geometry and traffic information, such as the number of lanes, the presence of horizontal curves, cycle time, and so on. The following sections discuss the variables considered for the analysis. Table 2 shows the descriptive statistics for key variables. The dataset on fatal crashes exhibits a mean value of 0.49, ranging from a minimum of 0 to a maximum of 2, with a standard deviation of 0.74, indicating variability in the occurrence of fatal incidents. Average daily traffic (ADT) values ranged from 56,895 to 27,4281vehicles per day (vpd), with an average of 138,628. Table 3 also provides details of other factors considered in this study. Figure 2 provides a graphical representation of the fatal and non-fatal percentage share of each variable

Data Description

Number of lanes (NOL): According to the data, two-lane roadways appear to have the highest share of fatal crashes, accounting for 24%. Furthermore, the data show that four-lane roads were the second most common location for fatal crashes, accounting for 23% as shown in Figure 2.

Presence of horizontal curve (PHC): Descriptive investigation indicates that roads with the presence of horizontal curves are associated with a smaller share of fatal crashes (36%) than roads with no horizontal curve, followed by summer (64%).

ADT: The traffic volume is measured and represented as average daily traffic in this study. Descriptive investigation indicates that low ADT is associated with a greater share of fatal crashes (45%), followed by medium ADT (29%), and high ADT (26%).

Table 2. Descriptive Summary of Signalized Intersection Crash Database

Variable	Mean	Min.	Max.	SD
Fatal crashes	0.49	0	2	0.74
Non-fatal crashes	1.12	0	13	0.57
Number of lanes	3	2	8	1.23
Average daily traffic	138,628	56,895	274,281	47,113
Major road width (m)	11.91	5.5	48.4	3.69
Cycle time (s)	159	75	352	61.62
Presence of horizontal curve (1 = Yes,0 = No)	0.71	0	1	0.46
Right-turn protection (1 = Yes,0 = No)	0.21	0	1	0.41
Marking visibility (1 = Clear, 0 = Faded)	0.23	0	1	0.42
All-red time present (1 = Yes,0 = No)	0.11	0	1	0.31

Note: SD = standard deviation.

Table 3. Correlation Analysis Results

Variables	No of lanes	Log (ADT)	Major road width	Presence of right-turn protection	Visibility of road markings	Cycle time	Presence of horizontal curve
Fatal	0.162	0.132	0.52	0.131	0.242	0.244	0.101
No. of lanes	1	NA	NA	NA	NA	NA	NA
Log (ADT)	0.455	1	NA	NA	NA	NA	NA
Major road width	0.756	0.595	1	NA	NA	NA	NA
Presence of right-turn protection	0.039	0.042	0.099	1	NA	NA	NA
Visibility of road markings	0.022	0.048	0.066	0.047	1	NA	NA
Cycle time	0.133	0.370	0.115	0.025	0.010	1	NA
Horizontal curve	0.098	0.176	0.132	0.175	0.109	0.036	1

Note: ADT = average daily traffic; NA = not available.

Major road width (MRW): Data show that 38% of fatal crashes occur when MRW is medium (10.1 m - 13.4 m), followed by low MRW (37%) and high MRW (25%) as shown in Figure 2.

Presence of right-turn protection (RTP): One of the most crucial aspects of signalized intersections is the protected right turn, which enables vehicles to make a right turn only when permitted or completely prohibited to make a right turn at that intersection. According to the data, 12% of fatal crashes occurred when there was a protected right turn, whereas 88% of fatal crashes occurred when there was no right-turn protection.

Visibility of road markings *(RM):* Road markings are necessary to control, warn, guide, and inform road users. When the road markings are clear, 18% of fatal crashes occurred, and about 82% fatal crashes occurred near the intersections where the road markings are faded as shown in Figure 2.

Cycle time (CT): One of the key elements for proper traffic flow at the intersection is cycle time. Intersections with less cycle time (< 125 s) are found to be associated with more fatal crashes (40%), followed by moderate cycle time (35%) and high cycle time (25%).

Presence of all-red time present (ARTP): The data show that the highest number of intersections do not contain all-red time phases, which accounts for 94% of fatal crashes. The intersections with all-red time present resulted in only 6% of fatal crashes.

The next section provides a full overview of the data analysis for the multi-perspective model approach used to identify important risk variables leading to fatal crashes, which includes Binary logit, ANN, and Apriori algorithms.

Data Analysis

This section presents the comprehensive analysis employed in this study. Before beginning the analysis, a systematic cleaning of the data was carried out. The safer pre-processing method that prevents the creation of bias is the removal of missing values, which has been followed in this study. Firstly, a binary Logit model under the discrete choice modeling framework was used in statistical analysis to obtain variables that had the greatest influence on fatal crash outcomes. Secondly, ANNs were used to develop fatal crash prediction models. Thirdly,

Figure 2. Distribution of road traffic crashes (fatal and non-fatal).
Note: ADT – average daily traffic.

Table 4. Variance Inflation Factor Results

Independent variables	VIF values
No of lanes	2.51
Log (ADT)	1.02
Major road width	1.67
Presence of right-turn protection	2.75
Visibility of road markings	1.03
Cycle time	1.01

Note: ADT = average daily traffic; VIF = variance inflation factor.
The maximum VIF value is highlighted in **bold** to indicate acceptable multicollinearity among predictors.

Apriori algorithm was used to determine the optimal rules related to fatal crashes. Finally, based on the results of the three techniques, the risk variables were divided into several categories. The following sections briefly address each technique's data preprocessing, parameter selection, and model validation.

Binary Logistic Models

Data Preprocessing. In binary logistic regression, data preprocessing involves standardizing continuous variables, such as converting ADT to Log (ADT) and encoding categorical features into binary and ordinal form, before partitioning the dataset into training (80%) and testing (20%) sets. Another key component of the initial data investigation remains to check the correlation between dependent, independent and pair of independent variables.

Correlation Analysis. The correlation between the predictors is tested through spearman correlation analysis. A "Spearman correlation" test was conducted to establish the relationship between a) dependent and independent variables, and b) between a pair of independent variables. As the database contains both continuous and nominal variables, the non-parametric Spearman's rank order correlation has been found to be the most appropriate method to check the correlation between two variables (*17*). Table 3 presents the correlation analysis findings. Interestingly, among the variables, the road width was observed to be correlated with the number of lanes. Despite the present correlation, both were included in the model owing to the following factors. First, in typical Indian urban settings, there is a lack of uniformity among road types with respect to lane width and number of lanes. Therefore, both major road width and number of lanes are taken together to understand their respective roles on fatal crash. Several past studies also included both road width and number of lanes to understand their individual effect (*24*, *25*). Second, another statistical measure, variance inflation factor (VIF), inverse of the tolerance value can be used for measuring multicollinearity as well (*26*). VIF indicates the extent to which an indicator's variance is captured by the remaining indicators of a given construct. The instances of higher degrees of multicollinearity are reflected in lower tolerance values and higher VIF values (*26*). As reported by Kim (*27*) when the estimated VIF is higher than 5 to 10 or even within 10-30 range, multicollinearity is observed to be present with respect to that attribute. As the estimated values for both attributes are significantly lower than that as shown in Table 4, both attributes are included in the model. With respect to major road width and cycle time, the correlation between them can be termed as weakly correlated with each other (0.133 indicates a weak positive linear relationship) (*28*). Similarly, correlation among other pair of variables can be interpreted. Results clearly indicate all independent variables to be significant predictors of the dependent variable, and they were therefore included in the final model formulation.

Binary Logit Model Theory and Formulation. In this study, the probability of a crash being fatal (1) versus non-fatal (0) is investigated using a discrete choice model, more precisely a binary logit model. The logistic regression framework, which predicts the likelihood that a crash will be fatal based on several explanatory variables, is applicable because the dependent variable has a binary result. When the dependent variable is binary, and there are many observations, the binary logit model was the most suitable and preferred choice (9). Therefore, the binary logit (BLR) technique was utilized for modeling purposes as the response outcomes are two (fatal versus non-fatal). The dichotomous outcome variable yin is defined as below Equation 1:

$$y_{in} = \begin{cases} 1, \textit{if the severity outcome i sustained by observation n was fatal} \\ 0, \textit{if the severity outcome i sustained by observation n was non} - \textit{fatal} \end{cases} \tag{1}$$

The logistic function as presented by Manski and McFadden (*29*), can be defined as given in Equation 2 if the probability that an RTC casualty n will be fatal is $P(y_{in} = 1)$.

$$P(y_{in} = 1) = \frac{exp(\beta' X_{in})}{1 + exp(\beta' X_{in})} \tag{2}$$

where β is the vector of estimable coefficients, and X_{in} is the vector of independent variables. The model coefficients could be estimated by optimizing the log-likelihood function given in Equation 3 (*10*).

$$LL\left(\beta'\right) = \sum_{i=1}^{n} \{(y_{in} \ln(P(y_{in})) + (1 - y_{in})\ln(1 - P(y_{in})))\} \tag{3}$$

Table 5. Fatal Crash Binary Logit (BLR) Model-Signalized Intersections

Variable characteristics	Explanatory variables	Signalized intersection	
		Coefficient	Sig.
	Constant	5.61	0.000
Traffic-related factors	LOG (average daily traffic volume)	−2.053	0.000
	Cycle time (s)	−0.013	0.000
	Presence of right-turn protection (1/0)	−1.457	0.000
	Number of lanes (1,2,3)	−1.129	0.000
	Major road width (m)	0.488	0.000
	Presence of horizontal curve (1/0)	−0.975	0.000
	Visibility of road markings (1/0)	−2.053	0.000
Model summary	Log likelihood	796.932	
	Cox & Snell R square	0.251	
	Nagelkerke R square	0.389	
	Number of observations	1,068	

Table 6. Base Conditions of Variables

Base conditions (signalized intersections)	ADT	Cycle time (s)	Presence of right-turn protection (1/0)	Major road width (m)	Visibility of road markings (1/0)	Presence of horizontal curve (1/0)	No. of lanes
Base level	50,000	100	0	15	0	0	1
level 1	100,000	200	0.5	30	0.5	0.5	2
level 2	200,000	300	1	40	1	1	3

Note: ADT = average daily traffic.

The final model's summary is presented in Table 5. The binary logistic regression results for signalized intersections show that ADT ($\beta = -2.053$, $p = < 0.000$), cycle time ($\beta = -0.013$, $p = < 0.000$), presence of right-turn protection ($\beta = -1.457$, $p = < 0.000$), horizontal curve ($\beta = -0.975$, $p = < 0.000$), and the number of lanes ($\beta = -1.129$, $p = < 0.000$) has a positive effect toward reducing crashes. Several past studies found an inverse relation between traffic volume and fatal injury (*13*, *30*, *31*). The aforementioned findings are intuitive, as increased traffic volumes are likely to cause motorists to slow down, thereby reducing the risk of fatalities. Several studies have found that longer cycle times can enhance the safety of intersections (*32–35*). Generally, longer cycles result in fewer interruptions at traffic signals, and an increase in green times (mostly imposed by longer cycle lengths) provides more space for traffic and reduces vehicular conflicts. Several past studies have demonstrated that right-turn restrictions lower the number and severity of crashes (*17*, *32*, *36*, *37*). Turning movements that are restricted or prohibited reduce the number of turning conflict points at intersections, which is known to reduce crash risk (*38*). In India, motorists are required to drive on the left, so a right turn requires the greatest distance from opposing traffic to be completed safely (*17*). This unprotected right-turning causes vehicles to proceed with the turn despite a minor gap in the opposing traffic stream, resulting in conflicts as well as increasing a driver's workload for the need to decide, resulting in an increased likelihood of being associated with crashes.

Model Validation. The prediction performance of the logit model was validated using additional randomly selected 20% data, which was not included previously in the crash prediction models. The training and testing accuracy of predicting crash fatality risk was found to be 92.5% and 85.9%, respectively, which is reasonably acceptable as per past studies (*39*). Therefore, this model can analyze crash data and recommend safety measures for any other city with a similar dataset.

Sensitivity Analysis. To understand the significance of different factors influencing urban road crash fatality, a sensitivity analysis has been conducted. Three different levels of factors have been considered to calculate prediction probability. The three levels—base level, level 1, and level 2—are defined for each factor, as shown in Table 6. Considering all the data points, different levels are

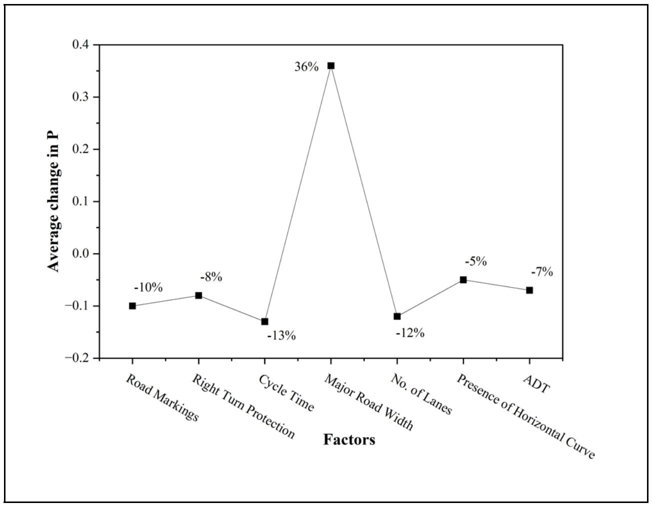

Figure 3. Influencing factors based on binary logistic models.
Note: ADT = average daily traffic.

considered based on the input data stream. For example, ADT has three levels, that is 50,000, 1lakh, and 2 lakhs, respectively, at base levels, level one and level two. Firstly, a base condition is designed where all attributes are at their certain base levels; for example, ADT is 50,000, major road width is 15 m, cycle time is 100 s, number of lanes is one, the presence of footpath, presence of horizontal curve and visibility of road markings is 0, and the probability is calculated. For the next scenario, the base level of ADT is changed from 50,000 to 1 lakh (level 1), but all other attributes are kept constant, and the probability is estimated. For the third case, all other attributes are kept at the base level, whereas the level for the ADT has changed to 2 lakhs (level 2), and the probability is estimated. Thus, the respective changes in probabilities are calculated as the levels of ADT are changed. Finally, the percentage of the average change in probability is derived. Similarly, the probability change estimation has been done for all factors in all cases, that is base level, level one and level two. According to sensitivity analysis as shown in Figure 3, the factors with the greatest impact on road traffic accidents (RTAs) are major road width (36%), cycle time (−13%), and no. of lanes (−12%).

Artificial Neural Network (ANN)

Data Preprocessing. In comparison to statistical models ANN models can be used to approximate the underlying nonlinear relationship between crash severity and safety predictors without any prior knowledge or assumptions on model structure [28]. Preprocessing steps such as feature scaling, which involves categorizing continuous variables such as average daily traffic, major road width, and cycle time into sub-variables low, medium, and high, as shown in Table 2, followed by encoding are performed. The data are also divided between training (80%) and validation (20%).

Model Formulation. ANN are frequently employed to forecast the results of traffic crashes. Simple computer function arrays known as neural networks are densely interconnected and substantially based on the organization of the human brain (*40*). In this investigation, a feed-forward neural network was used. Information only travels forward across this network's input, hidden, and output layers. The architecture of the ANN model in relation to the database is shown in Figure 4. Eight neurons make up the input layer (for eight independent variables), and each hidden layer neuron receives input from each of these nodes after being multiplied by a set of weights (WH_{11}, $WH_{12}...WH_{99}$). All weights and biases are assumed to have the same beginning values for the first iteration. Each hidden layer neuron is additionally given a bias value (bH_1, $bH_2...bH_m$), which serves as an input to the neuron. The outputs of each of these neurons are directed via an activation function. After that, the values are transferred to the output layer neuron by being multiplied by their corresponding weights. There is a connection between fixed bias and the output layer neuron (b_0). Various neural networks were trained with different numbers of hidden neurons to determine the respective models' accuracy and goodness of fit statistics. To enhance performance, the proposed ANN architecture uses various activation functions in the hidden and output layers. The hidden layer neurons use φ, which is the activation function investigated with ReLU, Sigmoid, and SELU during model tuning (Table 7). The final output layer uses the sigmoid activation function σ to restrict the output between 0 and 1. This combination of hidden-layer nonlinearity and output-layer sigmoid activation ensures that features are represented well while still being interpretable binary risk categories. Several combinations of optimizers (learning algorithms) and activation functions are used to develop ANN models. Adaptive moment estimation (Adam) is a learning rate method that determines different learning rates for various factors. Adam changes the learning rate for each neural network weight based on estimates of the gradient's first and second moments. Root-mean-square propagation (RMSProp) is a gradient-based learning method that normalizes the gradient after each iteration, as gradients of complex functions such as neural networks tend to balloon or collapse. Stochastic gradient descent (SGD) is a function that reduces the size of a variable (*41*). At each iteration, it determines the gradient of the cost function for a single case, and by selecting the most negligible loss at each iteration, it ultimately

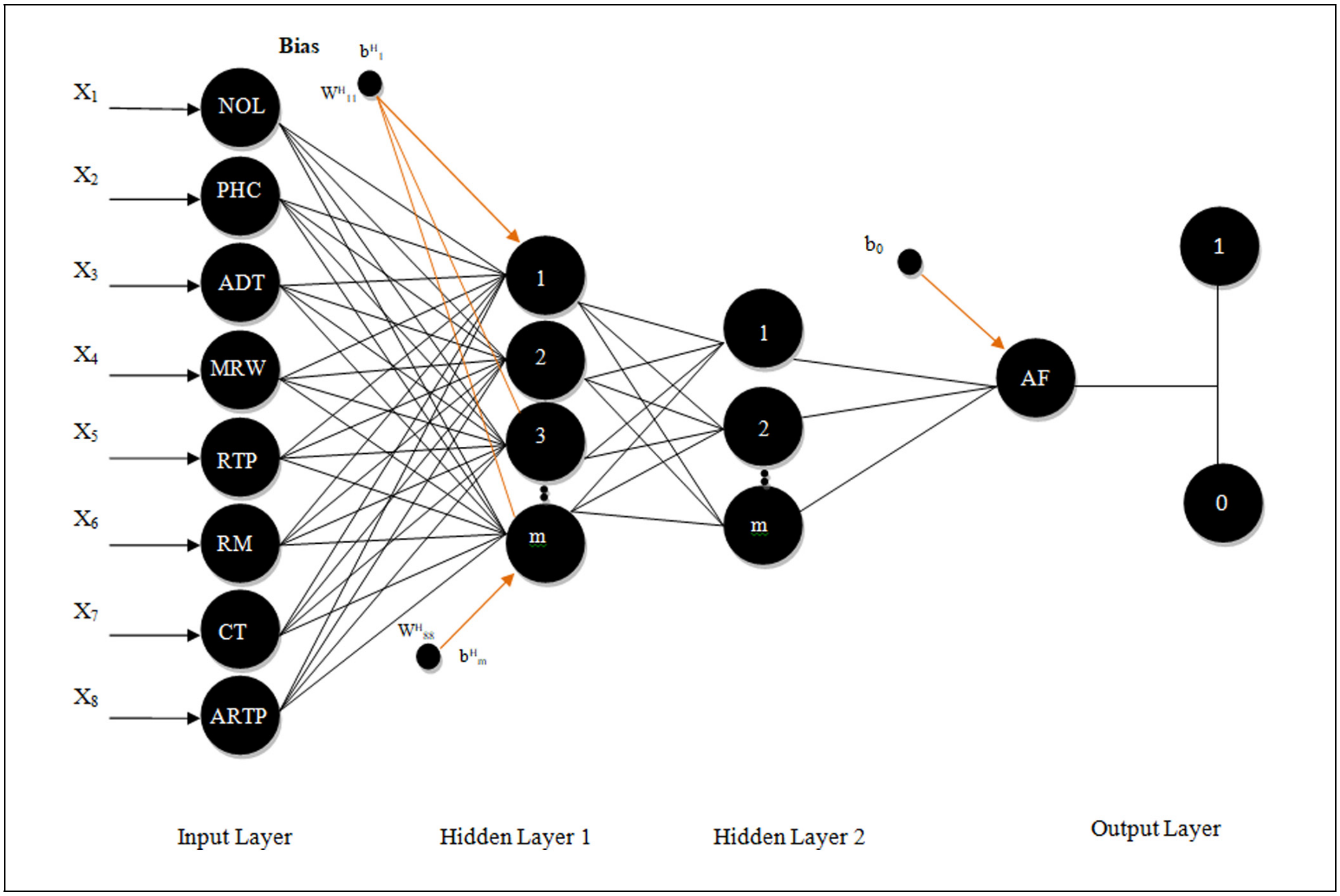

Figure 4. Model architecture for binary classification.
Note: ADT = average daily traffic; ARTP = all-red time present; CT = cycle time; MRW = major road width; NOL = number of lanes; PHC = presence of horizontal curve; RM = road markings; RTP = right-turn protection.

reaches the minimum. *Validation:* The models were evaluated using the following criteria: accuracy, loss, mean square error (MSE), root-mean-square error (RMSE), and mean square log error (MSLE). Table 7 shows the outcomes of a more comprehensive model exploration approach rather than a single optimizer comparison. To systematically analyze several ANN setups, we altered the optimizers (Adam, RMSProp, SGD), hidden-layer activation functions (ReLU, Sigmoid, SELU), number of neurons, and number of layers. The goal was to find the best-fit design for the dataset rather than to isolate the effect of a particular parameter. Across every instance, the output layer consistently used sigmoid activation to assure binary classification. The "best-fit" model was chosen largely for accuracy, with loss and error measures serving as supplementary indicators. The RMSProp optimizer and Sigmoid activation function model and with one hidden layer were found to be the best-fit model in this for the training with an accuracy of 90.27%. The model performance parameters for nine distinct ANN models for the training dataset are shown in Table 7. An ANN model's general mathematical equation is as follows (*41*):

$$A_N = \sigma\left[b_0 + \sum_{k=1}^{m}\left\{w_k\varphi\left(b_k + \sum_{i=1}^{n} w_{ik}x_i\right)\right\}\right] \tag{4}$$

where A_N denotes the model's normalized output, σ = sigmoid output activation function, φ = hidden-layer activation (ReLU, SELU, etc.), b_0 = output layer neuron's bias, w_k = weight between the output layer neuron and the hidden laye's kth neuron, b_k = bias associated with the hidden layer's kth neuron, w_{ik} = weight between ith input layer neuron and kth hidden layer neuron, x_i = input layer's ith variable (neuron) normalized, n = the number of input variables, m = number of neurons in the hidden layer

Validation of ANN Model. For a generalized application, the best-fit model, along with the other top four models identified from the training database, was used for crash

Table 7. Artificial Neural Network (ANN) Model Performance Parameters

Optimizer	Activation function	No. of neurons	Hidden layers	Accuracy	Loss	MSE	RMSE	MSLE
Adam	ReLu	20	2	0.8946	0.3267	0.0971	0.3116	0.0466
	Sigmoid	14	1	0.8964	0.4009	0.1043	0.3229	0.0501
	SELU	10	1	0.8176	2.7574	0.1690	0.4111	0.0812
RMSProp	ReLu	16	1	0.8090	0.8593	0.1870	0.4324	0.0898
	Sigmoid	**14**	**1**	**0.9027**	**0.3182**	**0.1294**	**0.3598**	**0.0622**
	SELU	10	1	0.7782	3.2437	0.1906	0.4366	0.0916
SGD	ReLu	12	1	0.3525	9.8606	0.6546	0.8091	0.3147
	Sigmoid	12	1	0.1014	13.635	0.8956	0.9464	0.4305
	SELU	11	1	0.8018	3.0226	0.1978	0.4447	0.0950

Note: The activation functions (ReLU, Sigmoid, and SELU) reported in Table 7 only apply to the hidden layers. The output layer always uses a sigmoid activation function (σ) for binary classification, regardless of the hidden-layer activation used; Bold values indicate the best-fit model.

Table 8. Testing Accuracies of Artificial Neural Network (ANN) Models

Optimizer	Activation function	No of neurons	Hidden layers	Accuracy	Loss	MSE	RMSE	MSLE
Adam	ReLu	20	2	0.8966	0.3269	0.0961	0.3100	0.0465
	sigmoid	14	1	0.8972	0.4012	0.1042	0.3228	0.0511
	SELU	10	1	0.8183	2.7582	0.1692	0.4114	0.0812
RMSProp	ReLu	16	1	0.8093	0.8599	0.1881	0.4337	0.0899
	sigmoid	**14**	**1**	**0.9098**	**0.3082**	**0.1184**	**0.3442**	**0.0591**

Note: MSE = mean square error; MSLE = mean square log error; RMSE = root mean square error. Bold values indicate the best-fit model.

Table 9. Sensitivity Analysis using the Garson Method

Factors	Number of lanes	Presence of horizontal curve	ADT	Major road width (m)	Visibility of road marking	Presence of right-turn protection	Cycle time (s)	Presence of all-red time
Relative influence	16.17	13.31	0.00008	5.94	15.82	25.13	0.12	23.51
Rank	3	5	8	6	4	1	7	2

Note: ADT = average daily traffic.

prediction based on a dataset containing crash information from 20% testing database. Table 8 summarizes the top five ANN models and their respective performance parameters for the new dataset. Results indicate that RMSProp optimizer and Sigmoid activation function model performed the best for testing data as well, with an accuracy of 90.98%.

Sensitivity Analysis Using the Garson Method

Sensitivity analysis was performed to determine the relative importance of the independent variables. The Garson (*42*) approach was utilized to determine the individual percentage of influence of each variable on crash outcome to streamline the weights of various factors employed in ANN analysis. The results of the sensitivity analysis are shown in Table 9. Each variable was ranked according to how much influence it had on the output variables. The top-ranked variable is right-turn protection, with a relative influence of 25.13%.

Apriori Algorithm

Data Preprocessing. Crash data were extracted and represented in a transactional categorical format, where each crash instance contained a list of attributes leading to a fatal/nonfatal crash, thus identifying frequent itemsets based on a minimum support threshold and generating association. Association rules analysis is sometimes called market basket analysis (*43*). Basket analysis is a strategy businesses use worldwide to detect which products are purchased together. Analysis entails searching

Table 10. Best Association Rules Based on Apriori Algorithm

Chain of factors			Confidence	Lift	Type of crash
Presence of horizontal curve [Yes]	Presence of right-turn protection [No = 78]	Presence of all-red time present [No = 76]	0.97	1.02	Fatal
Presence of right-turn protection [No]	Visibility of road markings [Faded = 102]	Presence of all-red time [No = 98]	0.96	1.01	Fatal
Presence of horizontal curve [Yes]	Visibility of road markings [Faded = 70]	Presence of all-red time [No = 67]	0.96	1.01	Fatal
ADT [low = 65]	Presence of all-red time [No = 64]	—	0.98	1.04	Fatal
Presence of right-turn protection [No = 126]	Presence of all-red time [No = 121]	—	0.96	1.01	Fatal
Presence of all-red time [No = 135]	Presence of right-turn protection [No = 121]	—	0.90	1.01	Fatal

Note: ADT = average daily traffic.
'—' indicates that no rule was generated for the corresponding entry.

Table 11. Factor Occurrence Count in Rules

Factors	Number of times appeared in rules
Presence of all-red time	6
Presence of right-turn protection	4
Visibility of road markings	2
Presence of horizontal curve	2
ADT	1
Cycle time (s)	0
Major road width (m)	0
Number of lanes	0

Note: ADT = average daily traffic.

supermarket statistics to identify a set of sold commodities. For example, if a consumer buys milk, how likely is it for them to buy bread or eggs during the same trip to the supermarket? This information can assist retailers undertake selective marketing and plan shelf space, resulting in greater sales (*44*). Apriori algorithm was introduced to mine common itemsets for Boolean association rules (*45*). Apriori algorithm utilizes a level-wise iterative search methodology that employs k-item sets to investigate (k + 1) item sets. Initially, the database is searched to collect the frequency of each item that meets the minimal support requirement. Afterward, the set L1 is obtained and utilized to determine L2, which consists of common 2-itemsets. Thus, the search is extended until no additional frequent *k*-item sets can be detected (*44*).

Validation. The extraction of best association rules and validation relies on indicators, specifically confidence and lift (*46*). A confidence value could measure the conditional probability that an item Y will occur if an item X also occurs. The indicator lift is used to quantify the association between item X and item Y. If the coefficient of lift is larger than 1, it indicates a positive correlation between the variables. The indicator lift is used to quantify the association between item X and item Y. If the coefficient of lift is larger than 1, it indicates a positive correlation between the variables. WEKA 3.8.6 software was used to implement Apriori algorithm to get the best association rules with fatal crashes. The current study considered the top six best rules (with three and two factors) based on a confidence value of at least 0.90 and a lift value of more than 1. Table 10 presents the best rules table, and Table 11 shows the number of times each factor appeared in significant rule.

The multi-perspective model segmentation strategy, which combines the findings from the three approaches, is covered in the following section.

Risk Factor Segmentation Procedure

The results of binary logistic models, ANN, and Apriori algorithms are combined in the current study, and a suitable segmentation criterion is adopted to prioritize the risk factor into three sections: very high risk, high risk, and moderate risk factors as presented in Table 13. For binary logit model, sensitivity weightage derived from the sensitivity analysis is used for ranking of attributes, for ANN model, sensitivity weightage derived from Garson's sensitivity analysis is utilized for ranking of the variables and for Apriori algorithm, the frequency of variables being present in respective rules are used as criterion for ranking as presented in Table 12. The following subsection presents the risk segmentation approach.

Risk Segmentation Approach: To integrate the results from the three models (ANN, Apriori, and binary logit), a simple rule-based technique was used for risk segmentation. The primary goal was to prioritize any factor recognized as most critical (rank = 1) by at least one

Table 12. Ranking of Risk Factors

Factors	Binary logit rank (sensitivity analysis)	ANN rank (sensitivity analysis)	Apriori algorithm rank (number of times appeared in rules)
Presence of all-red time (0/1)	—	2	1
Presence of right-turn protection (0/1)	5	1	2
Visibility of road markings (0/1)	4	4	3
Presence of horizontal curve (0/1)	6	5	3
ADT	5	8	4
Cycle time (s)	2	7	0
Major road width (m)	1	6	0
Number of lanes	3	3	0

Note: ADT = average daily traffic; ANN = artificial neural network.
'—' represents not applicable.

Table 13. Risk Factor Segmentation

Segmentation	Variables
Very high risk (rank 1)	Major road width, presence of right-turn protection, presence of all-red time
High risk (rank 2–3)	No. of lanes, cycle time, presence of horizontal curve, visibility of road markings
Moderate risk (> rank 4)	ADT

Note: ADT = average daily traffic.

approach, given each method employs a unique analytical mechanism.

- Very high risk: Assigned if any of the three models ranked the factor as 1.
- High risk: Assigned if a factor consistently ranked between 2 and 3 in at least two models but did not rank 1 in either.
- Medium risk: Assigned if a factor ranked more than 4 (e.g., 4–8) based on either or majority of the methods.

The above segmentation approach clusters the most critical attributes (top ranked attributes in either method), followed by the subsequent sets of critical attributes (within top three attributes in either method), followed by medium risk attributes incorporating attributes associated with relatively lower risk. For example, "Presence of right-turn protection" was assigned to very high risk despite having lower ranks in the other two models, namely rank 2 in Apriori and rank 5 in binary logit. This ensured that a factor identified as significant by any analytical framework was not neglected. This rule-based integration demonstrates the value of running three techniques concurrently: it takes advantage of the complementary perspectives of several modeling approaches and prioritizes critical safety elements even when only one model strongly flags them.

Very High Risk (Segment I)

This segment includes factors ranked 1 in all three techniques. The factors found in this category are major road width, presence of right-turn protection, and presence of all-red time. The major road width was found to have a significant positive association with fatal collisions. It has been accepted that wider roads can increase vehicular speed and driving maneuvers, increasing the probability of fatal crashes. Increased crossing distance and the time to cross may also result in increased pedestrian exposure to automobile traffic while crossing the road (*47*). Therefore, traffic authorities should review all listed speed limits on wide approach roads at intersections, reduce them, and use efficient methods to capture speeding vehicles and apply penalties. Protected right-turns increase safety and reduce collisions but at the expense of operational efficiency by reducing the green time available for through movements. Therefore, an alternative approach known as a part-time protected right-turn signal can also be adopted. A part-time protected right-turn signal is one in which permissive and fully protected right-turn phasing is used during peak and off-peak hours, respectively. Part-time protected right-turn signals provide a solid safety solution while avoiding the requirement for capacity enhancements to accommodate queued right-turners at signalized junctions (*48*). Red-light violations are fairly widespread in cities throughout India. The primary function of an all-red interval is to safeguard vehicles that may be in the midst of an intersection, including those who may have entered the intersection at the conclusion of the yellow change period. Adding an all-red interval after a yellow change interval would decrease right-angle crashes (*49*).

High Risk (Segment II)

This segment includes factors ranked between 2 and 3 in all three techniques. The factors that appear to be high risk are the number of lanes, visibility of road markings, presence of horizontal curve, and cycle time. As the number of lanes increased, fatal collisions decreased by 12%. This disparity in findings (relative to road width) could be attributed to irregularity in lane marking. During RSI, it was discovered that major roadways with sufficient width for three or four lanes were designed with only two and, in some cases, no lanes. The wider lane markings help narrow the roads for drivers and thus could lead drivers to slow down, as they perceive higher speeds while travelling within the lanes (*50*). As a result, it is possible to conclude that wide roadways without correct lane configuration are more dangerous, whereas wide roads with proper lane configuration are considerably safer. According to the binary logit model, the presence of clear *road markings* was found to reduce fatal crashes by 10%. This could be attributed to reduced speed behavior at intersections with clear markings compared with intersections with faded markings (*51*). To mitigate the speed of drivers approaching signalized intersections, some effective road marking treatments, such as dragon teeth markings, based on the principle of visual road narrowing, and raised median island, based on the principle of physical road narrowing, could be employed.

It appears that the existence of a horizontal curve has mixed effects on fatal crashes. Binary logit indicates that the presence of a horizontal curve can reduce fatal collisions by 5%, while association rule mining indicates that it can also result in fatal crashes when accompanied by faded road markings, no all-red time, and no protection for right turns. Therefore, irrespective of geometric design (straight or curved), it is important to take precautionary steps to reduce the speed of vehicles approaching the intersection. Advanced warning signs in conjunction with chevron sight boards and/or repeated arrows, as well as road marking treatments such as rumble strips and herringbone road marking, could offer a dependable reduction in speed and improved lane position ahead of horizontal curves. As cycle duration increased, the binary logit model predicted a 13% reduction in fatal crashes. Red-light running may decrease if drivers' expectations are satisfied by wide progression bands (through extended cycle lengths) and effective signal coordination (*34*). The increase in the yellow interval duration and provision of green extension is only possible with extended cycle lengths.

Moderate Risk (Segment III)

This segment includes factors ranked above 4 in all three techniques. *ADT* was identified as moderate risk factor. According to the binary logit model, increasing ADT rates reduces fatal crashes by 7%. Also, the data mining analysis showed that the possibility of fatal crashes increases when ADT is low, and all-red time is absent. This result could also be attributed to the red light running when the traffic volume is low. According to Schröter et al. (*52*), low red-light running is caused by the typical high motor vehicle traffic volumes and the lengthy crossing distances associated with intricate intersection designs. Since traffic volume is not a factor that can be controlled, countermeasures such as the presence of red-light cameras could greatly minimize crashes at intersections. Red-light cameras encourage more disciplined queuing of motorcyclists at the stop line, therefore reducing jump starts and red- light running on conflicting approaches (*53*).

Conclusions

This study aims to identify the key roadway and traffic variables that contribute to fatal crashes at signalized intersections in Hyderabad, an Indian metropolis. Subsequently, a hierarchy of criteria is developed to divide the risk factor into three distinct segments. This unique classification method would aid in prioritizing the risk factors and recommending necessary countermeasures. Inadequate information on the most/least essential components may impede improvement, increasing the chance of fatal crashes. In this regard, the study presents the following research contributions. Firstly, this study employed a multi-perspective model approach utilizing binary logit, ANN, and Apriori algorithms to identify significant risk factors leading to fatal crashes at signalized intersections. Such a multi-perspective model approach may contribute to comparing and understanding the different levels of influence each factor has on crash outcomes. The major contribution of this study does lies in using three different techniques to combine the variables and classify them into various groups with reference to their relative impact on crash severity at urban signalized intersections. All these three techniques were independently applied to a common crash dataset derived from Hyderabad city for identifying top-most risk factors leading to road traffic fatality and were subsequently combined for road risk factor segmentation. Three methods with their respective strengths and weaknesses were applied to identify key risk variables resulting from each of the techniques and subsequently combined to demonstrate a multi-model approach for risk factor prioritization and classification, which remains a novel addition to the existing road safety research literature. The models and results offer actionable recommendations for local policymakers and highlight the value of multi-model perspectives for advancing transportation safety research. Secondly, a unique segmentation approach was adopted to classify the risk factors into three groups based on their relative influence across the models. Each model was applied independently using a unique primary crash dataset from Hyderabad, a megacity in a developing country such as India, and the top-ranked variables influencing crash fatality from each analysis were identified and grouped post hoc to provide a transparent evaluation of risk factor importance across distinct methodological paradigms. Such interest groups would allow policymakers to make segment-specific and phased improvement recommendations while staying within budget. For example, the findings show that major road width, right-turn protection, and all-red time were classified as high-risk characteristics during the investigation. This conclusion proposes high-level interventions at intersections with wide major roads and modifying provisions such as right-turn protection and all-red time. Third, this study proposed specific actions focused on various risk factors that could lower the occurrence of fatal crashes in a LMIC such as India. The research on road safety in LMICs, which has been identified as an area of neglect (*54*), warrants increased attention and priority. According to a World Bank report, LMICs have witnessed a growing number of surgically curable chronic, congenital, and acquired disorders owing to crashes (*55*). However, since the majority of victims in LMICs come from low-income families, they lack proper access to medical care (*2*) and therefore cannot afford the additional costs (surgeries), which can lead to death. Therefore, the suggested methodological approach, identified risk factor segments and suggested policy-level mitigation actions can be implemented to establish short-term or long-term road safety initiatives to minimize fatal crashes and maximize the limited resources available. The study's results, conclusions, and recommendations are consistent with the United Nations' eleventh sustainable development goal (SDG-11).

Likewise other studies, there are certain limitations of the study, which need to be addressed as part of the future scope of work. First, the study examines the role of various factors leading to fatal/non-fatal crashes in Hyderabad. The data used in the study (Traffic Police data, 2015−2019) do not provide further crash classification such as fatal, grievous, simple, no-injury and property damage, others type of crashes in their crash database till 2018 data. Therefore, uses of all types of crash types and development of advanced prediction models remain one of the major scopes of further research for this study. Second, the primary source of information concerning road traffic collisions (RTCs) is derived from police records. However, the issue of under-reporting fatalities and injuries stemming from road traffic accidents is a significant global concern that affects several developing and developed countries across the globe. Therefore, to enhance the scope of this research, it is imperative to gather data from police, vehicle, driver registration, and hospital databases pertaining to the same population. Third, this study uses a post hoc grouping of the top-ranked variables that were independently determined by logistic regression, ANN, and Apriori algorithm. In future, a data-driven clustering for segmentation could be taken up as one of the key future scopes. Fourth, authors would like to state that the findings were not tested for transferability to similar cities. The developed models were data-dependent; therefore, the result may vary from city to city, which makes the results difficult to replicate. Therefore, it is essential to use similar datasets across cities and validate the results for generalized research results. Further, this study mostly covers the urban region; a separate study with a similar approach needs to be done for rural areas. However, the technique and selected factor categories are broad, making them suitable for emerging and replicable for cities in developing countries to improve road traffic safety.

Acknowledgments

Acknowledgment

We sincerely acknowledge the support from the Traffic Police Department, Hyderabad, for the detailed crash data.

Author Contributions

The authors confirm their contribution to the paper as follows: study conception and design: Majumdar and Sahu; data collection: Koramati; analysis and interpretation of results: Koramati, Majumdar, Das, and Ghosh; draft manuscript preparation: Koramati, Majumdar, Sahu, Das, Ghosh, and Biswas. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Prasanta Sahu is a member of Transportation Research Record's Editorial Board. All other authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was funded by the Department of Science and Technology, Government of India (DST SERB "SRG/ 2019/ 000725") and NFSG, BITS Pilani (N4/24/1022).

ORCID iDs

Siddardha Koramati https://orcid.org/0000-0002-0307-7906
Bandhan Bandhu Majumdar https://orcid.org/0000-0001-7721-8436
Prasanta K. Sahu https://orcid.org/0000-0002-4309-5631
Aritro Ghosh https://orcid.org/0009-0007-8944-7069
Sabyasachi Biswas https://orcid.org/0000-0003-1378-3561

Data Accessibility Statement

Data will be made available on request

References

1. WHO. *Global Status Report on Road Safety*. Geneva, Switzerland, 2023. Licence: CC BY-NC-SA 3.0 IGO.
2. World Bank. *Traffic Crash Injuries and Disabilities: The Burden on Indian Society*. (Vol. 1 of 7), World Bank Group, Washington, D.C., 2021.
3. MoRTH. *Road Accidents in India 2022*. Ministry of Road Transport and Highways, Government of India, New Delhi, 2023.
4. Athar, S., R. White, and H. Goyal. *Financing India's Urban Infrastructure Needs Constraints to Commercial Financing and Prospects for Policy Action*. World Bank, 2022. http://hdl.handle.net/10986/38306
5. Islam, M. R., S. Barua, S. Akter, M. Hadiuzzaman, and N. Haque. Impacts of Nongeometric Attributes on Crash Prediction at Urban Signalized Intersections of Developing Countries. *Journal of Transportation Safety and Security*, Vol. 12, No. 5, 2020, pp. 671–696. https://doi.org/10.1080/19439962.2018.1526840.
6. Khattak, M. W., A. Pirdavani, P. De Winne, T. Brijs, and H. De Backer. Estimation of Safety Performance Functions for Urban Intersections Using Various Functional Forms of the Negative Binomial Regression Model and a Generalized Poisson Regression Model. *Accident Analysis and Prevention*, Vol. 151, October 2020, 2021, p. 105964. https://doi.org/10.1016/j.aap.2020.105964.
7. Sobreira, L. T. P., and F. Cunto. Disaggregated Traffic Conditions and Road Crashes in Urban Signalized Intersections. *Journal of Safety Research*, Vol. 77, 2021, pp. 202–211. https://doi.org/10.1016/j.jsr.2021.03.003.
8. Ramezani-Khansari, E., A. A. Rassafi, and A.-S. Hashemiyan. Structural Modelling of Crashes in Signalized Intersections. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, Vol. 48, April 2024, pp. 1073–1082. https://doi.org/10.1007/s40996-023-01329-x.
9. Jashami, H., J. C. Anderson, H. A. Mohammed, D. P. Cobb, and D. S. Hurwitz. Contributing Factors to Right-Turn Crash Severity at Signalized Intersections: An Application of Econometric Modeling. *International Journal of Transportation Science and Technology*, Vol. 13, 2023, pp. 243–257. https://doi.org/10.1016/j.ijtst.2023.02.004.
10. Tamakloe, R., S. Das, E. Nimako Aidoo, and D. Park. Factors Affecting Motorcycle Crash Casualty Severity at Signalized and Non-Signalized Intersections in Ghana: Insights from a Data Mining and Binary Logit Regression Approach. *Accident Analysis and Prevention*, Vol. 165, 2022, p. 106517. https://doi.org/10.1016/j.aap.2021.106517.
11. Esenturk, E., D. Turley, A. Wallace, S. Khastgir, and P. Jennings. A Data Mining Approach for Traffic Accidents, Pattern Extraction and Test Scenario Generation for Autonomous Vehicles. *International Journal of Transportation Science and Technology*, Vol. 12, No. 4, 2023, pp. 955–972. https://doi.org/10.1016/j.ijtst.2022.10.002.
12. Kong, X., S. Das, and C. Yuan. In-Depth Understanding of Pedestrian–Vehicle Near-Crash Events at Signalized Intersections: An Interpretable Machine Learning Approach. *Transportation Research Record*, Vol. 2677, No. 5, 2022, pp. 747–759. https://doi.org/https://doi.org/10.1177/03611981221136138.
13. Anjana, S., and M. V. L. R. Anjaneyulu. Safety Analysis of Urban Signalized Intersections under Mixed Traffic. *Journal of Safety Research*, Vol. 52, 2015, pp. 9–14. https://doi.org/10.1016/j.jsr.2014.11.001.
14. Xu, X., S. Xie, S. C. Wong, P. Xu, H. Huang, and X. Pei. Severity of Pedestrian Injuries Due to Traffic Crashes at Signalized Intersections in Hong Kong: A Bayesian Spatial Logit Model. *Journal of Advanced Transportation*, Vol. 50, No. 8, 2016, pp. 2015–2028. https://doi.org/10.1002/atr.1442.
15. Xie, S. Q., N. Dong, S. C. Wong, H. Huang, and P. Xu. Bayesian Approach to Model Pedestrian Crashes at Signalized Intersections with Measurement Errors in Exposure. *Accident Analysis and Prevention*, Vol. 121, May 2018, pp. 285–294. https://doi.org/10.1016/j.aap.2018.09.030.
16. Elvik, R. Safety-in-Numbers: Estimates Based on a Sample of Pedestrian Crossings in Norway. *Accident Analysis and Prevention*, Vol. 91, 2016, pp. 175–182. https://doi.org/10.1016/j.aap.2016.03.005.
17. Mitra, S., and D. Bhowmick. Status of Signalized Intersection Safety-A Case Study of Kolkata. *Accident Analysis and Prevention*, Vol. 141, January 2020, p. 105525. https://doi.org/10.1016/j.aap.2020.105525.
18. Haque, F., and F. Ahmad Kidwai. Modeling Pedestrian Behavior at Urban Signalised Intersections Using Statistical-ANN Hybrid Approach – Case Study of New Delhi. *Case Studies on Transport Policy*, Vol. 13, 2023, p. 101038. https://doi.org/10.1016/j.cstp.2023.101038.
19. Kuşkapan, E., M. A. Sahraei, M. K. Çodur, and M. Y. Çodur. Pedestrian Safety at Signalized Intersections: Spatial and Machine Learning Approaches. *Journal of*

Transport and Health, Vol. 24, 2022, p. 101322. https://doi.org/10.1016/j.jth.2021.101322.
20. Das, S., R. Tamakloe, H. Zubaidi, I. Obaid, and A. Alnedawi. Fatal Pedestrian Crashes at Intersections: Trend Mining Using Association Rules. *Accident Analysis and Prevention*, Vol. 160, June 2021, p. 106306. https://doi.org/10.1016/j.aap.2021.106306.
21. Samerei, S. A., K. Aghabayk, A. Mohammadi, and N. Shiwakoti. Data Mining Approach to Model Bus Crash Severity in Australia. *Journal of Safety Research*, Vol. 76, 2021, pp. 73–82. https://doi.org/10.1016/j.jsr.2020.12.004.
22. *Britannica*. https://www.britannica.com/place/Hyderabad-India. Accessed June 29, 2024.
23. *Deccanchronicle*. https://www.deccanchronicle.com/nation/current-affairs/231223/hyderabad-witnessed-rise-in-road-accidents-42-of-road-victims-are-pe.html. Accessed June 29, 2024.
24. Noland, R. B., and L. Oh. The Effect of Infrastructure and Demographic Change on Traffic-Related Fatalities and Crashes: A Case Study of Illinois County-Level Data. *Accident Analysis and Prevention*, Vol. 36, No. 4, 2004, pp. 525–532. https://doi.org/10.1016/S0001-4575(03)00058-7.
25. Chimba, D., T. Sando, and V. Kwigizile. Effect of Bus Size and Operation to Crash Occurrences. *Accident Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 2063–2067. https://doi.org/10.1016/j.aap.2010.06.018.
26. Hair, J. F., C. M. Ringle, and M. Sarstedt. Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. *Long Range Planning*, Vol. 46, No. 1–2, 2013, pp. 1–12. https://doi.org/10.1016/j.lrp.2013.01.001.
27. Kim, J. H. Multicollinearity and Misleading Statistical Results. *The Korean Society of Anesthesiologists*, Vol. 72, No. 6, 2019, pp. 558–569. https://doi.org/https://doi.org/10.4097/kja.19087.
28. Ratner, B. The Correlation Coefficient: Its Values Range between 1/1, or Do They. *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 17, No. 2, 2009, pp. 139–142. https://doi.org/10.1057/jt.2009.5.
29. Manski, C. F., and D. McFadden. *A Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA; London, 1981.
30. Wong, S. C., N. N. Sze, and Y. C. Li. Contributory Factors to Traffic Crashes at Signalized Intersections in Hong Kong. *Accident Analysis and Prevention*, Vol. 39, No. 6, 2007, pp. 1107–1113. https://doi.org/10.1016/j.aap.2007.02.009.
31. Oh, J. T. Development of Severity Models for Vehicle Accident Injuries for Signalized Intersections in Rural Areas. *KSCE Journal of Civil Engineering*, Vol. 10, No. 3, 2006, pp. 219–225. https://doi.org/10.1007/bf02824064.
32. Turner, S. Safety Performance Functions for Traffic Signals: Phasing and Geometry. *Journal of the Australasian College of Road Safety*, Vol. 24, No. 3, 2013, pp. 30–40.
33. Stevanovic, A., J. Stevanovic, and C. Kergaye. Optimization of Traffic Signal Timings Based on Surrogate Measures of Safety. *Transportation Research Part C: Emerging Technologies*, Vol. 32, 2013, pp. 159–178. https://doi.org/10.1016/j.trc.2013.02.009.
34. James Bonneson, A., M. Brewer, and K. Zimmerman. *Review and Evaluation Off Actors That Affect the Frequency of Red-Light-Running*. 2000.
35. Quiroga, C., E. Kraus, I. Van Schalkwyk, and J. Bonneson. . *Red Light Running – A Policy Review. Report No. CTS-02/150206-1.Texas Transportation Institute, The Texas A&M University System, College Station, TX, 2003*.
36. Poch, M., and F. Mannering. Negative Binomial Analysis of Intersection-Accident Frequencies. *Journal of Transportation Engineering*, Vol. 122, No. 2, 1996, pp. 105–113. https://doi.org/10.1061/(ASCE)0733-947X(1996)122:2(105).
37. Wang, X., and M. Abdel-Aty. Analysis of Left-Turn Crash Injury Severity by Conflicting Pattern Using Partial Proportional Odds Models. *Accident Analysis and Prevention*, Vol. 40, No. 5, 2008, pp. 1674–1682. https://doi.org/10.1016/j.aap.2008.06.001.
38. Simodynes, T., T. Welch, and M. Kuntemeyer. Effects of Reducing Conflict Points on Reducing Accidents (Abstract Only). In *Proceedings of the Third National Access Management Conference*, Fort Lauderdale, FL, October 4–7, 1998, p. 141.
39. Ghasedi, M., M. Sarfjoo, and I. Bargegol. Prediction and Analysis of the Severity and Number of Suburban Accidents Using Logit Model, Factor Analysis and Machine Learning: A Case Study in a Developing Country. *SN Applied Sciences*, Vol. 3, No. 1, 2021, pp. 1–16. https://doi.org/10.1007/s42452-020-04081-3.
40. Shaik, M. E., M. M. Islam, and Q. S. Hossain. A Review on Neural Network Techniques for the Prediction of Road Traffic Accident Severity. *Asian Transport Studies*, Vol. 7, November 2021, p. 100040. https://doi.org/10.1016/j.eastsj.2021.100040.
41. Koramati, S., and A. Mukherjee. Development of Crash Prediction Model Using Artificial Neural Network (ANN): A Case Study of Hyderabad, India Follow the Regularized Leader. *Journal of The Institution of Engineers (India): Series A*, Vol. 104, 2022, pp. 63–80. https://doi.org/10.1007/s40030-022-00696-4.
42. Garson, G. D. Comparison of Neural Network Analysis of Social Science Data. *Social Science Computer Review*, Vol. 9, No. 3, 1991, pp. 399–434.
43. Bigham, B. S. Road Accident Data Analysis : A Data Mining Approach. *Indian Journal of Scientific Research*. Vol. 3, 2014, pp. 437–443.
44. Han, J., M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2011.
45. Agrawal, R., T. Imielinski, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, 1993.
46. Montella, A., R. de Oña, F. Mauriello, M. Rella Riccardi, and G. Silvestro. A Data Mining Approach to Investigate Patterns of Powered Two-Wheeler Crashes in Spain. *Accident Analysis and Prevention*, Vol. 134, June 2020, p. 105251. https://doi.org/10.1016/j.aap.2019.07.027.
47. Shehadeh, E. A., A. H. Al-Bayatti, and M. A. Bingöl. Effect of Roadway Environment Characteristics on Pedestrian Safety at Signalised Intersections in Amman. *Urban,*

Planning and Transport Research, Vol. 12, No. 1, 2024, pp. 1–21. https://doi.org/10.1080/21650020.2024.2317766.

48. Howlader, M. M., Y. Ali, A. Burbridge, and M. M. Haque. Before-after Safety Evaluation of Part-Time Protected Right-Turn Signals: An Extreme Value Theory Approach by Applying Artificial Intelligence-Based Video Analytics. *Accident Analysis and Prevention*, Vol. 194, 2024, p. 107341. https://doi.org/10.1016/j.aap.2023.107341.
49. Datta, T. K., K. Schattler, and S. Datta. Red Light Violations and Crashes at Urban Intersections. *Transportation Research Record*, Vol. 1734, 2000, pp. 52–58. https://doi.org/10.3141/1734-08.
50. Garach, L., F. Calvo, and J. De Oña. The Effect of Widening Longitudinal Road Markings on Driving Speed Perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 88, 2022, pp. 141–154. https://doi.org/10.1016/j.trf.2022.05.021.
51. Hussain, Q., W. K. M. Alhajyaseen, N. Reinolsmann, K. Brijs, A. Pirdavani, G. Wets, and T. Brijs. Optical Pavement Treatments and Their Impact on Speed and Lateral Position at Transition Zones: A Driving Simulator Study. *Accident Analysis and Prevention*, Vol. 150, 2021, p. 105916. https://doi.org/10.1016/j.aap.2020.105916.
52. Schröter, B., S. Hantschel, S. Huber, and R. Gerike. Determinants of Bicycle Crashes at Urban Signalized Intersections. *Journal of Safety Research*, Vol. 87, 2023, pp. 132–142. https://doi.org/10.1016/j.jsr.2023.09.011.
53. Haque, M. M., H. C. Chin, and H. Huang. Applying Bayesian Hierarchical Models to Examine Motorcycle Crashes at Signalized Intersections. *Accident Analysis and Prevention*, Vol. 42, No. 1, 2010, pp. 203–212. https://doi.org/10.1016/j.aap.2009.07.022.
54. Perel, P., K. Ker, R. Ivers, and K. Blackhall. Road Safety in Low- and Middle-Income Countries: A Neglected Research Area. *Injury Prevention*, Vol. 13, No. 4, 2007, p. 227. https://doi.org/10.1136/ip.2007.016527.
55. Debas, H. T., P. Donkor, A. Gawande, D. T. Jamison, M. E. Kruk, and C. N. Mock. *Essential Surgery*. *Disease Control Priorities*, 3rd ed., Volume 1. The International Bank for Reconstruction and Development/The World Bank, Washington, D.C., 2015.