Engineering and Applied Science Research 2024;51(4):452-461 Research Article

EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Prediction of fatal crashes based on various victim types on national highways passing through urban areas of developing countries

Srinivasa Rao Gandupalli*[1, 2], Purnanandam Kokkeragadda[1] and Mukund R. Dangeti[2]

[1]Department of Civil Engineering, University College of Engineering Kakinada (UCEK), Jawaharlal Nehru Technological University Kakinada (JNTUK), Kakinada, Andhra Pradesh, India
[2]Department of Civil Engineering, GITAM School of Technology, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India

Received 22 December 2023
Revised 23 April 2024
Accepted 2 May 2024

Abstract

The issue of safety on national roads that traverse urban areas in emerging countries has been a significant cause for concern. Statistical analysis and modelling techniques for road safety evaluation in metropolitan India are still evolving due to inadequate crash data, inventory, and traffic volume data records. This work aims to formulate safety performance functions (SPF) using negative binomial (NB) count data models to pinpoint the variables influencing total fatal crashes and other individual victim types, i.e., pedestrians, motorcyclists, and single vehicles on mid-block sections. Using four years of crash data (2014–2017) from Visakhapatnam City Police, India, the applicability of the current study framework has been established. Besides the geometric design elements, segment length, speed and average daily traffic, this study also focuses on information collected from road safety audits such as the provision of service roads, land use type, median opening, side access, pedestrian crossings, sight clearance to the driver, availability of earthen shoulders, proper signage, and good road markings. The study outcome reveals that road segment length, service road presence, and land use type are significant risk variables associated with fatal crashes in Visakhapatnam City, India. The length of the road segment positively correlated with frequencies of total fatal, pedestrian, motorcycle and single-vehicle fatal crashes. It increased the frequency of fatal crashes on each increment by 103%, 122%, 73% and 98%, respectively. Service roads increase crash frequencies, and road stretches with commercial/mixed land use types attract more crashes. This research emphasizes the essential safety precautions that transportation engineers and planners must implement to establish a more secure environment for all road users.

Keywords: Safety measures, Crash data, Mid-block sections, Negative binomial, Count data model

1. Introduction

Globally, around 1.35 million people yearly lose their lives, and 50 million people are injured in traffic accidents, the 8th leading cause of mortality across all age groups and the primary cause of death among people aged 5-29 [1]. Numerous factors contribute to this tendency, including increased urbanization, poor safety, a lack of enforcement, distracted or tired drivers, those under the influence of drugs or alcohol, speeding, and the failure to use seatbelts or helmets [1]. Most casualties in traffic accidents (11%) worldwide occur in India. About 450,000 accidents take place in India annually, killing almost 150,000 people. India is now seeing a rise in road accidents and fatalities due to the fast growth of the road network, the rising number of vehicles, and the higher average speed on roadways. It causes considerable human misery and diminishes the country's GDP by taking lakhs of economically productive lives [2]. 35 % of all fatalities and 30% of all traffic accidents occur on the national highways, which comprise about 2.1% of the entire road network and carry 40% of the traffic. In vehicle categories, motorcyclists and pedestrians accounted for the highest share of road accidents and deaths, and victims mainly constitute young people in the productive age groups of 15-49 [2].

Ensuring the safety of pedestrians and motorcyclists has been a significant issue in Indian urban areas. Despite the large proportion of motorcycles in Indian towns, there is a lack of proper infrastructure, including separated lanes and adequate pedestrian facilities. This deficiency contributes to frequent accidents and, tragically, fatalities. Previous studies reported fatality rates of megacities like Kolkata, 73% [3]; Delhi, the Capital of India, 76%; and Mumbai, 87%; some medium-sized cities comprise Agra 77%, Amritsar 67%, Bhopal 85%, Ludhiana 70%, Vadodara 73% and Visakhapatnam 78% fatal victims are pedestrian and motorcyclist [4]. Conversely, the fatalities from vulnerable road users (VRU) in high-income countries like Australia and the USA are significantly fewer [1]. The National Highways of India are facing issues with inadequate access control mechanisms. High-speed highways pass through urban and rural regions, where human populations are present on both sides of the National Highway (NH). VRUs in metropolitan areas often coexist with automobiles and heavy vehicles, resulting in increased contact and a higher likelihood of fatal collisions [5]. The Ministry of Road Transport and Highways (MoRTH) and the National Crime Records Bureau (NCRB) of India release annual reports based on police data. The study analyses police crash data to identify VRU groups and factors influencing fatal crashes on national highways passing through urban areas, aiming to develop countermeasures and enhance road safety. The findings will assist policymakers, urban

*Corresponding author.
Email address: sgandupa@gitam.edu
doi: 10.14456/easr.2024.42

planners, and traffic authorities implement targeted interventions and infrastructure improvements in Visakhapatnam and similar metropolitan areas.

There are various factors influencing fatalities of different victim types. Multiple studies have identified risk factors and developed remedies for fatality risk over the last few decades, mostly in urban areas of developed countries [6-9]. The findings may have limited relevance to emerging countries like India. Recently, researchers have been motivated to investigate the primary elements that affect the likelihood of fatal accidents in metropolitan areas of India [5, 10-12]. Visakhapatnam is categorized as a city of medium size, with a population of 1.73 million individuals. However, there are several obstacles to road safety. The city's swift transformations and population expansion have heightened traffic congestion, exposing vulnerable road user groups to an elevated likelihood of accidents. This study addresses the research gap by examining historical crash data (2014-2017) obtained from the Visakhapatnam police. It specifically identifies the main risk factors at the network level for dangerous routes that pose a significant risk of fatalities for various victim types, i.e., pedestrians, motorcyclists, and single-vehicle crashes. Single-vehicle crashes involve vehicle rollovers and collisions with road infrastructure, classifying them as a specific category of victims.

2. Literature survey

Safety Performance Functions (SPF) are crucial in road safety research, helping engineers and planners assess infrastructure safety. They establish a correlation between crash frequency and factors like traffic operating parameters, geographical features, and built environmental characteristics. Traditional traffic prediction models, like Poisson and Negative Binomial regression, are used to develop SPF. Poisson models are lacking for handling data with over-dispersion, indicating that the variance of the numbers exceeds the mean. This violates the underlying assumptions of Poisson regression, which assumes that the mean and variance are the same. Negative binomial models address this issue [6-9]. SPF helps identify critical factors contributing to collisions and evaluates their impact on road safety.

Significant problems in crash-frequency data were discovered by Lord and Mannering [6], Abdulhafedh [7], and Basu and Saha [8], as well as the benefits and drawbacks of the various methodological approaches used by researchers to address the issues with data characteristics. Singh et al. [9] evaluated numerous studies on crash prediction models used in India.

2.1 Review on the network level

Naqvi and Tiwari [10] developed SPFs using NB regression models to study fatal collisions in India's three NH sections. Roadside land use, Segment length, service roads, and topography type have significant statistical effects. Jain et al. [11] conducted road safety audits on a four-lane NH and offered steps to address problems. Unauthorized truck parking, median openings, missing road and median markings, speed, VRU facilities near habitats, access and service lanes need rapid renovation. Gandupalli et al. [5] used discrete modelling methods to predict VRU accident severity on 60 km of NH. Segment length, land use, driver visual clearance, crash season, crash time, accused vehicle type, curves, median openings, and pedestrian crossings affect safety. Rankavat and Tiwari [12] employed NB regression with or without flyover location categories to analyze fatal pedestrian collisions in Delhi, India. The results showed that arterial roads with more traffic, vehicle lanes, faster speed, and medians walled or elevated to prevent pedestrian crossings had more fatal accidents.

2.2 Review on various victim types

Obinguar and Iryo-Asano [13] used Metro Manila, Philippines pedestrian collision data to create pedestrian crash frequency models on national routes utilizing road characteristics and built environment factors. Significant pedestrian collision variables were principal roadways, footbridges, road portions with poor surface qualities, and land use. Kaplan and Prato [14] suggested a Copenhagen Region cyclist–motorist collision frequency and severity model. Results showed that biking exposure, bicycle traffic, and average daily automobile traffic contribute to this. Using NB and Bayes approaches, Schneider IV et al. [15] evaluated how horizontal curvature and other geometric factors affect single-vehicle motorcycle collisions on rural two-lane roads in Ohio. Xin et al. [16] examined how horizontal curve design and related factors affect the severity of single-motor cycle collision injuries in Florida. Rusli et al. [17] investigated rural mountainous highway single-vehicle collisions in Malaysia. Rainfall and speeding along downgrade slopes increase the likelihood, whereas shoulders and adequate delineation reduce crashes. Geedipally and Lord [18] evaluated single-vehicle, multi-vehicle, and combination approaches to modelling collisions on Texas multilane undivided roads. Martensen and Dupont [19] examined six European nations' single- and multi-vehicle fatal road collisions. Traffic flow, junctions, and medians between carriageways were the most critical factors in determining single- and multi-vehicle crashes. Mukherjee and Mitra [20] developed fixed-parameter NB models to investigate pedestrian fatality risk factors. Evidence suggests that vehicle speed, traffic volume, pedestrian interaction, overtaking tendency, and land-use type are essential. Difficulty using the crosswalk, lack of a pedestrian signal, footpath encroachment, wider roadway, and short sight distance are other fatal accident risk factors.

2.3 Review on traffic and exposure characteristics

Mitra et al. [21] evaluated five years of multilane NH collision data and used random parameter panel data models to uncover factors affecting head-on, rear-end, and total accidents. Although horizontal alignment, access control measures, and vehicle types have variable safety performances, segment length and ADT have consistent effects across crash types. Urban midblock crash analysis by Liu et al. [22] used zero-inflated NB regression. Lane count, yearly average daily traffic per lane, and segment length contribute. Krishnan et al. [23] created urban mid-block hierarchical SPFs. Road length, lane width, and roadside obstructions increase single-carriageway collision frequency. Traffic volume increases single- and dual-carriageway collision frequency. Chikkakrishna et al. [24] established SPFs for planning, designing, and operating factors on a four-lane NH-58 stretch in Uttarakhand, India. Crash rates increased with traffic volume, curvature change, and operating speeds. In Maharashtra, India, Mhetre and Thube [25] collected NH-48 highway crash data. The findings of NB regression modelling demonstrated that head-on collisions, crash timing (day/night), and weather variables impact road crash deaths.

2.4 Review on road infrastructure and land use characteristics

Bisht et al. [26] used random parameter NB modelling to assess road traffic collision probability on Indian intercity expressways. Hazards, vertical alignment length, access location and underpasses, horizontal curve length, horizontal alignment radius, and high AADT enhanced fatal collision probability. Villages, vertical alignment gradient, curve length, and speed decrease fatal crash risk. Dinu and Veeraragavan [27] predicted accidents on Indian two-lane undivided roads using a random parameter Poisson model. Important variables include traffic volume, vehicle composition, driveway density, and horizontal and vertical curvatures. Chikkakarishna et al. [28] found roadway geometry, traffic, temporal, environment, and land use factors linked with crash severity and frequency on nonurban four-lane highways in India. Using Poisson family regression, Chikkakrishna et al. [29] found traffic volume, access roads, median openings, and schools to be the most critical safety variables of an NH. Chatterjee and Mitra [30] examined how alignment indices affected 52 curve collision rates on an Indian multilane NH. Curve change rate, radius ratio, and degree of curvature are useful alignment metrics. Vayalamkuzhi and Amirthalingam [31] developed complete collision models for 4-lane divided roadways in heterogeneous traffic. According to crash data, operating speed, gradient, access point density, median opening, traffic volume, and curvature are relevant.

Sil et al. [32] examined vehicle speed dispersion and horizontal curve geometry. Statistics showed that passenger Cars were a critical vehicle category, and the best predictors were radius and degree of curvature. Hosseinpour et al. [33] examined head-on accident frequency and severity on 448 segments of five Malaysian federal roadways. Higher head-on accident rates were associated with horizontal curvature, heavy-vehicle traffic, terrain type, and access locations. Speed restriction and shoulder width also reduced collision frequency. Haghan et al. [34] used the Highway Safety Manual collision prediction model to study rural road crashes in Iran. They observed that segment length, AADT, shoulder width, roadside residential land use, and horizontal curve radius are essential. Vasalamkuzhi and Amirthalingam [35] examined how geometric design affects safety in heterogeneous traffic flow. Poisson regression and NB regression assessed crash safety. Minachi et al. [36] constructed NB regression and Empirical Bayes accident prediction models for urban intersection safety evaluation. Bhat et al. [37] modelled Bangalore urban road accident prediction. Key factors include carriageway width, shoulder type, frequency of minor crossings, land use, road condition, traffic stream average speed, and truck composition. The literature mentioned above clearly demonstrates that much study has been carried out on the safety of some victim types, such as pedestrians in India and abroad. According to the literature, the VRU is greatly limited in India, primarily focusing on large metropolitan cities like Delhi, Kolkata, Mumbai and Bengaluru. The present study addresses the identified gap in earlier studies by explicitly examining the crashes on NH passing through medium-sized cities like Visakhapatnam, India and determining the most significant risk variables linked to traffic exposure and operational infrastructure, roadway, and land use characteristics.

3. Research methodology

The study utilized a five-stage technique to achieve its research objective, ensuring a straightforward and adaptable approach to various locations. The research technique encompasses many vital components, namely: (a) collection of crash data, (b) descriptive analysis of crash data and identification of study segments followed by site visits, and field data collection (such as road safety audits, traffic volume surveys, speed studies, etc.), and (c) the development of an SPF using a reactive approach. (d) Utilizing statistical modelling tools to identify significant risk variables for different categories of victims, (e) implementing policy actions to enhance road safety [3, 20].

3.1 Crash data collection and identification study stretch

The police-recorded data is the exclusive source for the yearly crash reports published by the MoRTH and NCRB. For the current study, crash data was collected from Visakhapatnam traffic police, Andhra Pradesh, India, for 2014-2017. The police records contain information on the name of the police station, crime number, accident date, location, crash type, number of deaths, number of injuries, accident time, vehicles involved, and geographic location. However, they lack essential information such as the causes of the collision, the manner of the crash, the socio-demographic characteristics of the victim, the road infrastructure, and the geometric condition of the site. The present study examined fatal incidents only, as fatal crashes are more accurately reported nationwide for legal reasons, with a considerably lower chance of underreporting than non-fatal crashes [11]. The preliminary investigation revealed that most fatal crashes, approximately 80%, were concentrated along a 62-kilometre section of NH-16 between Tagarapuvalasa and Anakapalli that traversed the city of Visakhapatnam, selected as a study stretch and shown in Figure 1.

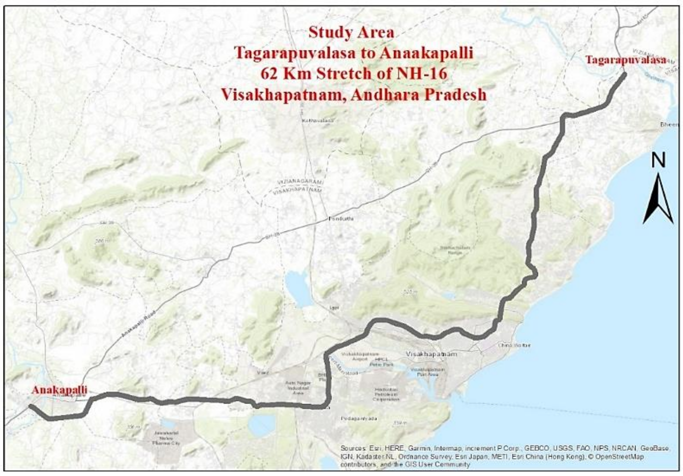

Figure 1 Study area of NH-16 passing through Visakhapatnam City

A "segment" refers to the distance between two consecutive intersections on a road. Further, the study stretch was divided into 92 distinct road segments, with lengths ranging from 0.1 to 2.9 kilometres. The study focused on non-intersection and mid-block crashes, excluding intersection crashes. Based on the literature review indicates that an intersection can significantly impact crashes up to 75m from its centre [21]. This study excluded crashes from an influence zone 100 meters from the intersection's centre. Finally, fatal crashes on mid-block segments were considered for further investigation. A total of 546 fatal crashes were recorded on a 62 km stretch from Tagarapuvalasa to Anakapalli from 2014 to 2017. Among these, 415 crashes have been classified as non-intersection mid-block fatal crashes, and all these 415 crashes have been mapped and analyzed.

3.2 Road safety audit data

The current study utilized data from road safety audits conducted in 2017 across the entire study stretch. Road safety audits provide information on road infrastructure and characteristics like segment length, curves, intersections, median openings, pedestrian crossings, service roads, side access roads, land use type, earthen shoulders, sight clearance, proper signage, and adequate road markings. The study involves a four-lane highway with a raised median, passing through urban and suburban areas, as depicted in Figure 2.

Cross section of sub-urban areas

Cross-section of an urban area

Figure 2 Typical Cross section of NH-16 Passing through Visakhapatnam City

3.3 Traffic volume and speed data

Using videography, traffic volume studies were conducted at four locations on the stretch during the weekday, under usual weather conditions. Traffic volume was extracted from the video, and the average daily traffic (ADT) and traffic composition were determined. Further, spot speed studies were conducted on different segments of the study stretch, and cumulative frequency distribution curves were plotted to determine the 85th percentile stream speed of the various segments.

The study aims to identify factors influencing fatal crashes among different victim types and propose effective measures to mitigate these risks. Table 1 displays the percentage of fatal crashes categorized by various victim types during the research period. This study is intended to identify the most vulnerable groups of road users towards fatal crashes. The study investigates fatal crashes involving pedestrians, motorcycles, and single vehicles as they are statistically more likely to occur. Table 2 provides a comprehensive overview of the identified site-specific response and predictor variables.

Table 1 The proportion of fatal crashes of various victim-type during the study period

Victim Type	Proportion of Fatal Crashes of various victim types
Pedestrian	53.0
Motorcycle	25.5
Car	2.2
Heavy vehicle	1.0
Three-wheeler	1.7
single vehicle involved	16.6

Table 2 Detailed summary of the study's response and predictor variables.

Variable	Variable Description	Mean	Std. Deviation	Minimum	Maximum
Response Variables					
Total Fatal Crashes	Number of reported fatal crashes at the site (2014-2017)	4.38	5.203	0	35
Pedestrian Fatal Crashes	Number of reported pedestrian fatal crashes at the site (2014-2017)	2.30	3.53	0	25
Motorcycle Fatal Crashes	Number of reported motorcycle fatal crashes at the site (2014-2017)	1.12	1.44	0	6
Single-vehicle Fatal Crashes	Number of reported single-vehicle fatal crashes at the site (2014-2017)	0.73	0.99	0	4
Predictor Variables					
Segment Length	Length of the road segment	0.669	0.627	0.11	2.92
Speed	Average 85th percentile speed	56.95	12.384	36	78
Service road	Service Road (0 if absent, 1 if present)	0.24	0.432	0	1
Land use	Type of Land use (1 if commercial/mixed, 0 open area/no land use)	0.58	0.498	0	1
Median openings	Number of Median openings over each segment	0.71	0.760	0	3
Curves	Number of Curves over each segment	1.21	1.893	0	8
Pedestrian Crossings	Number of Pedestrian crossings over each segment	0.77	1.390	0	8
Earthen shoulder	Presence of Earthen shoulder (0 if absent, 1 if present)	0.44	0.500	0	1
Sight clearance to the driver	Presence of Sight clearance to driver (0 if absent, 1 if present)	0.33	0.475	0	1
Proper signage	Presence of Proper signage (0 if absent, 1 if present)	0.70	0.463	0	1
Adequate road markings	Presence of adequate road markings (0 if absent, 1 if present)	0.82	0.389	0	1
Side access	Number of Side access roads over each segment	4.11	3.552	0	16
Log (ADT)	Log (ADT) on Segment	11.121	0.147	10.82	11.23

4. Data modelling

The occurrence of a crash is a discrete random event. Applying ordinary least square regression techniques for non-negative integer data like crash frequency is inappropriate. The Poisson regression model is a good choice for explaining the random occurrence of distinct events. The Poisson distribution may be stated as [6].

$$P[y_i] = \frac{e^{-m_i} m_i^{y_i}}{y_i!} \qquad y_i = 0, 1, 2, \ldots \tag{1}$$

Where $P[y_i]$ is the probability that y_i crashes will occur during the counting period on a road segment 'i' and m_i is the average number of incidents happening during a length of time t on a road segment 'i', which is the expected number of incidents throughout the counting period, $E[y_i]$.

Poisson regression models are estimated by describing the Poisson parameter m_i as a function of predictor variables.

$$m_i = e^{(\beta X_i)} \tag{2}$$

where X_i is a set of predictor variables, and 'β' is a set of estimable parameters.

It is important to remember that the Poisson distribution has an equal mean and variance. The Poisson distribution is inappropriate when the observed data show a variance/mean ratio that deviates significantly from 1.0. From Table 2, fatal crash frequencies of total, pedestrian, motorcycle, and single-vehicles were found to be over-dispersed, i.e., the ratio of variance/mean greater than 1.0. The NB model was selected to forecast crash frequency due to the over-dispersion in the data. The Poisson model is extended to include the idea that the Poisson parameter follows a gamma probability distribution. This model is known as the negative binomial (or Poisson-gamma) model. The distribution may be stated as

$$m_i = e^{(\beta X_i + \varepsilon_i)} \quad (3)$$

$$\log(m_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_n X_{in} + \varepsilon_i \quad (4)$$

The gamma-distributed error term ε_i has a mean of 1 and variance α. The accumulation of this term enables the variance to deviate from the mean.

$$\text{Variance} = \text{mean} + \alpha \times mean^2 \quad (5)$$

As α gets closer to zero, the NB model becomes the Poisson regression model. Hence, the choice among these two models depends on the value of α. The parameter α is frequently mentioned as the over-dispersion parameter. Obtaining a substantial value for the parameter α is crucial for a NB model. Finally, based on Akaike's Information Criterion (AIC) values, Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), and Mean Squared Predictive Error (MSPE), the proposed models were evaluated for goodness-of-fit. The AIC value is calculated using the formula [21]

$$AIC = -2LL(\beta) + 2k \quad (6)$$

k represents the model's estimated parameters, and the model with the lowest AIC is preferred.

$$MAD = \frac{1}{n}\sum_{i=1}^{n} |y_{predicted} - y_{observed}| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_{predicted} - y_{observed})^2} \quad (8)$$

$$MSPE = \frac{1}{n}\sum_{i=1}^{n}(y_{predicted} - y_{observed})^2 \quad (9)$$

Where $y_{predicted}$ and $y_{observed,}$ the predicted and observed number of crashes for each road segment, with n being the total number of road segments [38].

4.1 Correlation analysis

Correlation analysis was used as the first step in the statistical analysis of this study to determine the degree of correlation among predictor and response variables. The study found a moderate to strong relationship between significant predictive factors and fatal crashes, with correlation coefficients ranging from 0.3 to 0.7, with a 95% confidence interval. The critical findings of correlation analysis are presented in Table 3. The multicollinearity analysis consolidates predictor variables with a correlation coefficient below 0.3 with another predictor variable in the model [21]. Out of 12 model predictor variables, based on correlation coefficients and p-value. Eight variables were included in the analysis, viz. length of the road segment, the speed, the presence of service road, the type of land use, the number of median openings, the presence of proper signage, and the number of side access roads.

Table 3 Correlation analysis

Response Variables	Predictor Variables							
	Segment Length	**Speed**	**Service road**	**Land use**	**Median openings**	**Curves**	**Proper signage**	**Side access**
Total Fatal Crashes	.536***	-.291**	.431***	.271**	.330***	.207*	.241**	.365***
Pedestrian Fatal Crashes	.525***	-.217*	.466***	.226*	.210*	.259**	-.346***	.424***
Motorcycle Fatal Crashes	.429***	-.287**	.330***	.244**	.270**	.132	.199*	.263**
Single-vehicle Fatal Crashes	.317***	-.211**	0.09	.203*	.302**	.151	.061	0.146

*Significant at 90% CI
**Significant at 95% CI
***Significant at 99% CI

5. Model outcomes and discussion

The current investigation includes the development of four NB crash prediction models to examine the risk variables linked to fatal crashes encompassing total fatal, pedestrian fatal, motorcycle fatal, and single vehicle-involved fatal crashes. The 'backwards elimination' technique developed a comprehensive model with numerous predictor variables. The succeeding sections emphasize the importance of the model results from the standpoint of a developing nation.

5.1 Total fatal crash model

Multiple SPFs were developed to identify critical factors contributing to fatal collisions. Three of the 8 model predictor variables, namely segment length, presence of service road and land use type, were more critical to the total fatal crashes based on the backward elimination approach at a 90% confidence interval. Table 4 displays the analysis results for the total fatal crash model. The statistical model identified a substantial and favourable correlation between the length of the segment (β=0.71, $p<0.01$) and the frequency of fatal crashes. Each increase in segment length results in a 103% increase in the frequency of fatal collisions. This is because longer road segments are unaffected by intersections and serve as high-speed corridors. In addition, the model demonstrated a noteworthy association between fatal crashes and service roads (β=-0.53, $p<0.01$). As a result, there was a significant 41% decrease in fatal crashes on road segments without service roads compared to those with service roads. The reason for having a service road is to reduce the interaction between vehicles and side access roads, enable higher speeds, and, at the same time, widen pedestrian crossings. Mitra et al. [21] acknowledged that the length of a road segment and the presence of service lanes impacted accidents in urban mid-block sections. Liu et al. [22] showed a positive correlation between segment length and all types of collisions on mid-block portions in different states of the USA. Multiple research studies [23, 34] arrived at similar conclusions.

The presence of open areas/ no land use type is negatively correlated (β=-0.34, $p<0.1$) with fatal crash rates. This indicates that the likelihood of fatal crashes in segments without any specific land use type is 29% lower compared to parts with commercial or mixed land use. This is because vehicles travelling through the road segments with commercial or mixed land use have a higher level of interaction with other motorized traffic and pedestrians, resulting in more fatal crashes. Furthermore, the model summary indicated a significant reduction in the log-likelihood function of the total fatal accident model compared to the null model. Multiple studies have recognized speed as a primary factor in crashes [24, 31, 33], and in the current investigation, it is also found to be interconnected with land use. Haghani et al. [34] identified the potential hazards linked with residential land use along roadways in Iran. Bisht et al. [26] observed that the presence of a village or settlement and speed negatively correlated with fatal crashes. Chikkakrishna et al. [28] also studied the geometric, traffic, temporal, and environmental parameters and different land uses along a divided four-lane road facility in India.

Table 4 Results of total fatal crash model

Attributes	Model coefficient (β)	t-statistic	p-Value	Exp(β)
Intercept	1.328	6.067	0	
Segment length	0.71	5.284	0	2.03
Presence of Service Road -Not Available	-0.53	-2.481	0.01	0.59
Land use type- Open area/ no land use	-0.336	-1.645	0.1	0.71
Model Summary				
Dispersion parameter		3.41		
Log-likelihood function		-151.09		
Restricted Log-likelihood function (null model)		-220.377		
Akaike's Information Criterion (AIC.)		312.17		

5.2 Pedestrian fatal crash model

The study reveals that two of the 8 model predictor variables, namely the length of a road segment (β=0.798, $p<0.01$) and the presence of a service road (β =-0.868, $p<0.01$), significantly affect the frequency of fatal pedestrian crashes, based on the backward elimination approach at a 99% confidence interval presented in Table 5. The frequency of fatal pedestrian collisions increases 2.22 times for every unit increase in road segment length. Saheli and Effati [38] identified segment length as a contributing factor for pedestrian crashes on four-lane divided roads in Iran. In their study, Gandupalli et al. [5] discovered that increasing the length of a segment by one unit results in a 17.6% higher likelihood of fatal crashes involving a vulnerable road user (VRU). Portions without service roads have 58% fewer pedestrian fatal crash frequencies than segments with service roads. Several studies have recognized that service roads reduce the probability of crashes occurring at intersections [10, 21]. Nevertheless, service lanes along mid-block road segments encourage high speeds and increase the likelihood of fatal crashes. As a result, it is seen that service roads widen the area for pedestrians to cross, which in turn increases their susceptibility to accidents.

Table 5 Results of pedestrian fatal crash model

Attributes	Model coefficient (β)	t-statistic	p-Value	Exp(β)
Intercept	0.647	2.812	0.005	
Segment length	0.798	5.637	0	2.221
Presence of Service Road -Not Available	-0.868	-3.878	0	0.42
Model Summary				
Dispersion parameter		4.49		
Log-likelihood function		-114.07		
Restricted Log-likelihood function		-147.346		
Akaike's Information Criterion (AIC.)		236.15		

5.3 Motorcycle fatal crash model

The findings from the fatal model for motorcycle victims are shown in Table 6. Consistent with prior models, there is a positive correlation between the length of the road segment (β=0.547, $p<0.01$) and motorcycle fatal crashes. An increase of one unit in the length of the segment results in a 72.8% rise in fatal crash frequency for motorcycles. However, these values are lower when compared to the crash models that consider total crashes and crashes involving pedestrians. The study findings of the model align with prior research conducted by Mukherjee and Mitra [3] and Naqvi and Tiwari [10], indicating that as the length of the segment expands, the likelihood of crashes also increases.

In addition, the model also found a significant association between land use type and motorcycle fatal crashes (β=-0.473, $p<0.1$). The incidence of fatal crashes in open areas with no land use is 38% lower than in commercial/mixed land use areas, and this figure represents 29% in the total fatal crash model. Several researchers [5, 10, 20, 38] also reported residential, commercial/diverse land uses positively affect fatal crashes. In addition to the length of the segment and the kind of land use, fatal motorcycles collisions are also strongly linked to service roads. Consequently, the number of deadly accidents on road sections without service roads was 38% lower than those with service roads.

Table 6 Results of motorcycle fatal crash model

Attributes	Model coefficient (β)	t-statistic	p-Value	Exp(β)
Intercept	0.151	0.485	0.62	
Segment length	0.547	2.966	0.003	1.728
Presence of Service Road -Not Available	-0.485	-1.586	0.1	0.615
Land use type- Open area/ no land use	-0.473	-1.488	0.1	0.623
Model Summary				
Dispersion parameter		3.27		
Log-likelihood function		-88.975		
Restricted Log-likelihood function		-97.02		
Akaike's Information Criterion (AIC.)		187.95		

5.4 Single vehicle fatal crash model

Approximately 16.6% of fatalities are caused by single-vehicle rollovers or collisions with stationary objects. Hence, it is necessary to examine single-vehicle crashes, and predictive models were explicitly formulated for fatal crashes. The estimation results of the model are presented in Table 7. Similar to the previous models, the results validated a positive correlation between segment length (β=0.684, p<0.01) and fatal crashes. An increase of one unit in the segment length results in a 98% increase in fatal crash frequency. Extended road segments contribute to higher speeds, a higher tendency for overtaking, and a greater likelihood of fatal crashes involving a single vehicle. Xin et al. [16] analyzed single motorcycle accidents in Florida, while Schneider et al. [15] examined similar incidents in Ohio. According to their analysis, speed and segments with curves are significant contributors to the rise in fatalities. Furthermore, the study results are consistent with the single-vehicle crash models formulated by other previous researchers [17, 18].

Table 7 Results of single vehicle fatal crash model

Attributes	Model coefficient (β)	t-statistic	p-Value	Exp(β)
Intercept	-0.894	-3.885	0	
Segment length	0.684	3.996	0	1.979
Model Summary				
Dispersion parameter		22.16		
Log-likelihood function		-71.429		
Restricted Log-likelihood function		-78.822		
Akaike's Information Criterion (AIC.)		148.84		

The study focused on Visakhapatnam city due to available data from authorities. It was unable to compare models with other towns of similar size. Nevertheless, 25% of the data was omitted from the modelling process and utilized as controlled data to assess model validity.

6. Model validation

Validation was conducted on 26 road segments, accounting for 25% of study road segments in Visakhapatnam city. The models' prediction accuracy was confirmed in over 25% of these segments, which were not previously accounted for in accident prediction models. The descriptive statistics of the 26 road segments are given in Table 8.

Table 8 Descriptive Statistics (Model Validation)

Variable	Mean	Std. Deviation	Minimum	Maximum
Response Variables				
Total Fatal Crashes	4.85	3.79	0	16
Pedestrian Fatal Crashes	2.62	2.5	0	11
Motorcycle Fatal Crashes	1.23	1.51	0	7
Single-vehicle Fatal Crashes	0.81	1.06	0	4
Predictor Variables				
Segment Length	0.7	0.609	0.08	1.97
Speed	55.62	10.97	36	78
Service road	0.23	0.43	0	1
Land use	0.69	0.471	0	1
Median openings	0.62	0.571	0	2
Curves	1.23	1.557	0	5
Pedestrian Crossings	1.23	1.751	0	8
Earthen shoulder	0.38	0.496	0	1
Sight clearance to the driver	0.12	0.326	0	1
Proper signage	0.69	0.471	0	1
Adequate road markings	0.88	0.326	0	1
Side access	4.38	4.46	0	17
Log (ADT)	11.07	0.155	10.82	11.23

The effectiveness of the model's predictive capability has been confirmed on these 26 mid-block road segments. The model was evaluated using crash data acquired from police reports. The AIC, MAD, RMSE, and MSPE values were calculated to verify the accuracy of the developed models using the validation dataset. A clear difference in the validation parameters, namely AIC, MAD, RMSE, and MSPE, between the total fatal crash model and other models for specific victim types, such as pedestrian, motorcycle, and single-vehicle fatal accident models presented in Table 9.

Table 9 Validation results for the models

Model	AIC	MAD	RMSE	MSPE
Total fatal crash model	312.17	2.62	3.51	12.33
Pedestrian Fatal crash model	236.15	1.44	2.09	4.36
Motorcycle fatal crash model	187.95	1.02	1.5	2.25
Single vehicle fatal crash model	148.84	0.79	1.05	1.1

The study shows that individual victim-type models, specifically pedestrian, motorcycle, and single-vehicle models, have lower AIC, MAD, RMSE, and MSPE values than the total fatal accident model. The recommendation is to develop individual victim-type models to identify factors influencing fatal crashes and propose countermeasures to enhance the safety of each victim type.

7. Conclusions

The issue of road safety is becoming highly significant in medium-sized cities in developing countries like Visakhapatnam. Nevertheless, the lack of systematic gathering of accident data and the limited access to maintenance records challenge researchers. This study examines the factors that influence the frequency of crashes involving different categories of victims on a four-lane divided national highway traversing through an urban region. The analysis is conducted using NB count data models. Models were created to analyze crash data over four years on a 62-kilometre stretch of a four-lane divided highway that runs through Visakhapatnam city (NH-16). The findings of this study can assist policymakers, decision-makers, and stakeholders in road safety in efficiently allocating funds and implementing beneficial safety measures for all road users.

The present work utilizes the R Programming software package to develop prediction models for total fatal crashes, pedestrian accidents, motorcycle accidents, and single-vehicle accidents. These models are used to analyze fatal crashes in the study region. Following the correlation analysis, the study included eight predictor variables: segment length, speed, presence of service road, land use, median openings, proper signage, and side access roads. The backward elimination approach created a comprehensive model incorporating several predictor variables. The prior literature has identified segment length, presence of a service road, and land use type as risk variables. The outcomes of this study confirm these conclusions.

The length of the segment positively influences the fatal crash frequency of total and other individual victim-type crashes, such as pedestrian, motorcycle, and single-vehicle crashes. Each unit increase in segment length results in a varied accident frequency increase ranging from 73% to 122%. Road segments with greater distances between intersections promote higher speeds, escalating fatal incidents. The presence of service roads positively correlated with total, pedestrian and motorcycle fatal crash frequencies. The frequency of fatal crashes on road segments without service roads was consistently 38% to 58% lower compared to those with service roads. Service roads often reduce the probability of crashes at intersections, but in mid-block portions, they encourage high speeds and raise the risk of fatal accidents.

Additionally, they increase the width of the road pedestrians need to traverse. Land use type is an essential indicator of the frequency of fatal crashes, both for total crashes and specifically for pedestrian crashes. The crash rates in open areas with no land use are 29% and 39% lower for total and pedestrian fatal crashes, respectively, compared to commercial/mixed land use categories.

This study's findings suggest that to enhance safety, it is necessary to implement specific actions based on the results obtained from SPFs, particularly in regulation and design.

1. Implementing dedicated lanes to separate vulnerable road users from other vehicles.
2. Implementing appropriate speed restrictions, traffic calming measures, and utilizing speed cameras to monitor vehicle speed on high-speed roads within metropolitan areas.
3. Provision for dedicated pedestrian crossings with refugee islands and a separate signal phase for pedestrians near bus stops and school zones. Also, erecting barriers in the median will restrict the undesignated pedestrian entry.
4. Establishing enough infrastructural facilities and control mechanisms for commercial and mixed land-use areas.

8. Acknowledgement

The authors would like to express their gratitude to the Andhra Pradesh State Transport Department and the Visakhapatnam City Traffic Police Authorities for their cooperation with the data component of the project.

9. References

[1] World Health Organization. Global status report on road safety 2018. Geneva: WHO; 2018.
[2] Ministry of Road Transport & Highways. Road accidents in India 2017. New Delhi: Transport Research Wing; 2019.
[3] Mukherjee D, Mitra S. Development of a systematic methodology to enhance the safety of vulnerable road users in developing countries. Transp Dev Econ. 2022;8:1-17.
[4] Mohan D, Tiwari G, Mukherjee S. Urban traffic safety assessment: a case study of six Indian cities. IATSS Res. 2016;39(2):95-101.
[5] Gandupalli SR, Kokkeragadda P, Dangeti MR. Analysis and modelling of crash severity of vulnerable road users through discrete methods: a case study approach. Innov Infrastruct Solut. 2023;8:298.
[6] Lord D, Mannering F. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp Res A Policy Pract. 2010;44(5):291-305.
[7] Abdulhafedh A. Road crash prediction models: different statistical modeling approaches. J Transp Technol. 2017;7(2):190-205.

[8] Basu S, Saha P. Regression models of highway traffic crashes: a review of recent research and future research needs. Procedia Eng. 2017;187:59-66.

[9] Singh G, Sachdeva SN, Pal M. Predictive modelling of road accidents in India: a review. Indian Highways. 2017;45(6):29-38.

[10] Naqvi HM, Tiwari G. Safety performance functions for fatal crashes on national highways in India. Transp Res Procedia. 2020;48:1185-99.

[11] Jain SS, Singh PK, Parida M. Road safety audit for four lane national highways. The 3rd International Conference on Road Safety and Simulation; 2011 Sep 14-16; Washington, United State. Washington: TRB; 2011. p. 1-22.

[12] Rankavat S, Tiwari G. Association between built environment and pedestrian fatal crash risk in Delhi, India. Transp Res Rec. 2015;2519(1):61-6.

[13] Obinguar DD, Iryo-Asano M. Macroscopic analysis on the frequency and severity of pedestrian crashes on National Roads in Metro Manila, Philippines. IATSS Res. 2021;45(4):521-9.

[14] Kaplan S, Prato CG. A spatial analysis of land use and network effects on frequency and severity of cyclist–motorist crashes in the Copenhagen region. Traffic Inj Prev. 2015;16(7):724-31.

[15] Schneider WH, Savolainen PT, Moore DN. Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways. Transp Res Rec. 2010;2194(1):91-8.

[16] Xin C, Wang Z, Lee C, Lin PS. Modeling safety effects of horizontal curve design on injury severity of single-motorcycle crashes with mixed-effects logistic model. Transp Res Rec. 2017;2637(1):38-46.

[17] Rusli R, Haque MM, King M,Voon WS. Single-vehicle crashes along rural mountainous highways in Malaysia: an application of random parameters negative binomial model. Accid Anal Prev. 2017;102:153-64.

[18] Geedipally SR, Lord D. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson–gamma model. Accid Anal Prev. 2010;42(4):1273-82.

[19] Martensen H, Dupont E. Comparing single vehicle and multi-vehicle fatal road crashes: a jointanalysis of road conditions, time variables and driver characteristics. Accid Anal Prev. 2013;60:466-71.

[20] Mukherjee D, Mitra S. Identification of pedestrian risk factors using negative Binomial model. Transp Dev Econ. 2020;6:1-14.

[21] Mitra S, Haque M, King MJ. Effects of access, geometric design, and heterogeneous traffic on safety performance of divided multilane highways in India. J Transp Saf Secur. 2017;9(1):216-35.

[22] Liu C, Zhao M, Li W, Sharma A. Multivariate random parameters zero-inflated negative binomial regression for analyzing urban midblock crashes. Accid Anal Prev. 2018;17:32-46.

[23] Krishnan MJ, Anjana S, Anjaneyulu MVLR. Development of hierarchical safety performance functions for urban mid-blocks. Procedia Soc Behav Sci. 2013;104:1078-87.

[24] ChikkaKrishna NK, Parida M, Jain SS. Calibration of safety performance function for crashes on inter-city four lane highways in India. Cogent Eng. 2015;2(1):1-18.

[25] Mhetre KV, Thube AD. Count data modeling for predicting crash severity on Indian highways. Eng Technol Appl Sci Res. 2023;13(5):11816-20.

[26] Bisht LS, Tiwari G, Rao KR, Jha KN. Exploratory assessment of road traffic crashes on the intercity expressway in India. Adv Transp Stud Int J. 2022;3(SI):45-58.

[27] Dinu RR, Veeraragavan A. Random parameter models for accident prediction on two-lane undivided highways in India. J Saf Res. 2011;42(1):39-42.

[28] ChikkaKrishna NK, Parida M, Jain SS. Identifying safety factors associated with crash frequency and severity on nonurban four-lane highway stretch in India. J Transp Saf Secur. 2017;9(1):6-32.

[29] ChikkaKrishna NK, Parida M, Jain SS. Poisson family regression techniques for prediction of crash counts using bayesian inference. Procedia Soc Behav Sci. 2013;104:982-91.

[30] Chatterjee S, Mitra M. Alignment indices as a tool to evaluate safety on horizontal curves of multilane highways in India. Proc Eastern Asia Soc Transp Stud. 2021;13:1-12.

[31] Vayalamkuzhi P, Amirthalingam V. Development of comprehensive crash models for four-lane divided highways in heterogeneous traffic condition. Transp Res Procedia. 2016;17:626-35.

[32] Sil G, Nama S, Maji A, Maurya AK. Effect of horizontal curve geometry on vehicle speed distribution: a four-lane divided highway study. Transp Lett. 2020;12(10):713-22.

[33] Hosseinpour M, Yahaya AS, Sadullah AF. 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;62:209-22.

[34] Haghani M, Jalalkamali R, Berangi M. Assigning crashes to road segments in developing countries. Proc Inst Civ Eng – Transp. 2019;172(5):299-307.

[35] Vayalamkuzhi P, Amirthalingam V. Influence of geometric design characteristics on safety under heterogeneous traffic flow. J Traffic Transp Eng (Engl Ed). 2016;3(6):559-70.

[36] Minachi C, Gladsen J, Kalaanidhi S, Gunasekaran K. Development of accident prediction models for safety evaluation of urban intersections. Indian Highways, Indian Roads Congress. 2015;43(5):9-13.

[37] Bhat P, Hebbani L , Ramav A, Kolhar P. Accident prediction modelling for an urban road of Bangalore. Int J Res Eng Technol. 2013;2:137-41.

[38] Saheli MV, Effati M. Segment-Based count regression geospatial modeling of the efect of roadside land uses on pedestrian crash frequency in rural roads. Int J Intell Transp Syst Res. 2021;19:347-65.