SSRG International Journal of Civil Engineering
ISSN: 2348-8352/ https://doi.org/10.14445/23488352/IJCE-V13I7P108

Volume 13 Issue 7, 128-144, July 2026

Original Article

Assessing Urban Traffic Crash Risk and Its Impact on Mode Choice in Mumbai Using Safety Performance Functions

Trupti Narkhede[1,2*], Lokesh Gupta[1], Prasun Chakrabarti[3], Arvind Sharma[4]

[1]*Civil Engineering, Faculty of Integrated & Advanced Technology, Sir Padampat Singhania University, Udaipur, India.*
[2]*Department of Civil Engineering, MGM's College of Engineering and Technology, Kamothe, Navi-Mumbai, India.*
[3]*Faculty of Computing and Informatics, Sir Padampat Singhania University, Udaipur, India.*
[4]*Faculty of Data Science and Analytics, Sir Padampat Singhania University, Udaipur, India.*

[*]*Corresponding Author : narkhede.phd2023@spsu.ac.in.*

Received: 20 February 2026 Revised: 06 April 2026 Accepted: 22 June 2026 Published: 29 July 2026

***Abstract** - Road traffic accidents continue to be a leading public safety concern in urban areas, and particularly for vulnerable road users including pedestrians and motorcyclists. While the increasing fatality and casualty counts urge a comprehensive approach to understand the factors contributing to crash risks, only a few studies in Mumbai focus on the dynamic interplay between the predictor variables of traffic volume, road condition, and types of road users in the prediction of crash risk. This study uses traffic data from Mumbai's road safety reports from 2019 to 2024 to develop a Negative Binomial Regression Model in order to predict crash risk. Safety Performance Functions are applied to estimate the influence of road features, traffic volume, location type, and weather conditions. The model is tested on real-world data, analyzing factors such as speed limits, road geometry, and user-specific risk factors. Some of the most significant factors that were identified in the model include the volume of traffic, geometric and location factors. The volume of traffic is an important factor, as the additional increase of 1,000 vehicles per day raises the risk of the accident by 0.23%. It has been predicted that 726 injuries have been caused by high-speed high-traffic peak-hour intersections, while 595 injuries have been caused by high-speed mid-block. The highest pedestrian crash risk was found to be in high-traffic intersections with 45.3%, compared to lower traffic intersections with only 15.2% in rural conditions. In terms of practical application, this model offers a means of identifying hazardous locations at a particular time that allows making effective measures to prevent accidents.*

***Keywords** - Road Safety, Traffic circumstances, Crash risk, Regression models, Vulnerable Road Users.*

1. Introduction

Road traffic crashes have been identified as one of the major problems in urban areas, particularly in developing countries where urbanization is taking place at an unprecedented rate. Mumbai, as one of the most densely populated cities, is a perfect example of a difficult situation characterized by a mixture of traffic, high levels of congestion, and insufficient provision of facilities for the Safety of users traveling on foot and on two wheels [26]. The number of people dying due to road accidents is quite huge and is primarily attributed to vulnerable road users, which indicates the necessity of proper safety measures for such road users [26].

However, existing literature is mostly focused on crash frequency based on aggregate data without any consideration for interdependencies among traffic conditions, infrastructure conditions, and individual behaviour of road users. In addition, studies incorporating behavioural aspects and demographic factors in SPFs are few in number. This study makes an attempt to address these research gaps.

2. Research Problem

After implementing continuous road safety enhancement efforts, traffic crashes remain the leading cause of death or injury, particularly for vulnerable road users. Traditional approaches to the prediction of traffic accidents are mostly based on generalized data of simple risk assessment and fail to account for the complex dynamics of traffic on urban roads. In cities like Mumbai, where Road conditions have high variability and traffic patterns undergo frequent changes, such methods often yield merely approximate results. Several factors obstruct the predictive accuracy of accident risk models. Most of models do not consider all variables that influence road accidents, such as road geometry, traffic volume and environment factors. Due to a lack of separated data regarding various types of vulnerable road users. If it's difficult to obtain detailed information of risks faced by vulnerable road users. Finally, current models fail to account for the temporal variability inherent in accident risk, as different times of the day, days of the week, or even seasons are associated with varying levels of risk. This research overcomes the above-mentioned limitations by integrating detailed accident data with user-specific attributes to develop a unified model for accident risk assessment. In particular, it takes into account road geometry, traffic volume, user demographics and time-based data to provide a more realistic and distinct estimation of the actual crash risks faced by vulnerable road users in Mumbai. The key areas in which this research contributes to existing the literature is outlined below in four concise points.

Development of a data-driven model to estimate the risk of traffic accidents among vulnerable road users in Mumbai, utilizing the safety performance function and negative binomial regression.

Identification of high-risk locations and time periods to facilitate implementation of targeted measures aimed at reducing injuries and fatalities resulting from road traffic accidents among vulnerable road users

Integration of demographic and user-specific characteristics such as age, gender and road user type will improve the accuracy of crash risk predictions and provide actionable evidence-based.

Recommendations for road safety improvements, through detailed crash risk modeling that informs road infrastructure and enforcement improvements in Mumbai.

3. Overview of Traffic Accident Risk Modelling and Literature Review

Road traffic accidents remain one of the leading causes of fatalities and serious injuries worldwide, particularly in rapidly urbanizing regions where transportation demand grows faster than infrastructure development. Understanding the mechanisms that contribute to crash occurrence is therefore essential for designing effective road-safety interventions. Over the past three decades, researchers have developed numerous statistical and analytical models to quantify crash risk, identify hazardous locations, and evaluate the effectiveness of engineering and policy measures.

3.1. Safety Performance Functions in Road Safety Research

Safety Performance Functions (SPFs) are one of the most recognized ways of calculating the crash frequency and measuring the safety performance of roads. In SPFs, the mathematical relationship between the likelihood of having a crash and the explanatory variables, such as traffic volume, roadway geometrics, operational features, and environmental aspects, is defined. SPFs enable transportation administrations to conduct safety analysis with the help of a scientific basis.

For instance, the SPFs-based models are used by Vinayaraj and Perumal [3] to evaluate the Safety of urban roundabouts exposed to heterogeneous traffic flow in India. Their findings indicated that both traffic exposure and geometrics and operational features of roads can significantly affect the likelihood of crashes. Similarly, Ospina-Mateus et al., [8] conducted SPFs to predict motorbike crashes and found that traffic volume and roadway characteristics play a significant role in defining the rate of crashes. Most recently, Vinayaraj and Perumal [24] created crash modification factors together with SPFs for urban roundabouts and stressed the significance of calibrating the model contextually to accommodate mixed traffic flow. Also, HSM's framework, along with the international literature, recognizes the validity of using SPFs in crash frequency estimation.

3.2. Vulnerable Road User Safety

Vulnerable Road Users (VRUs) such as pedestrians, cyclists and motorcyclists constitute a disproportionately large share of traffic fatalities worldwide. Unlike occupants of enclosed vehicles, VRUs are not physically protected in a collision and are therefore subject to much greater injury severity.

Vilaça et al., [4] pointed out the importance of injury severity analysis of vulnerable road users and observed that pedestrian and cyclist crashes have more severe consequences than vehicle-to-vehicle crashes. Komol et al., [13] used machine-learning methods to study the crash severity of vulnerable road users and found that roadway characteristics, traffic density, and vehicle speed were important predictors of injury outcomes. Similarly, Klanjčić et al., [14] identified urban infrastructure characteristics that influence pedestrian and cyclist safety across European cities, highlighting the importance of dedicated infrastructure and protected crossings.

Motorcyclists are a particularly vulnerable group in developing countries where two-wheelers constitute a major share of urban mobility. Sumit et al., [6] studied risky riding behavior among the young motorcyclists in India and observed that speeding, risk-taking behavior and low compliance with safety regulations are among the major contributors to crash occurrence. Morris et al., [10] further emphasized that the protection of vulnerable road users needs to be a central consideration in future urban transportation systems.

3.3. Statistical and Machine Learning Approaches for Crash Prediction

Traditionally, crash prediction has been done using count-data modelling techniques such as Poisson and Negative Binomial regression. The Negative Binomial has emerged as the preferred statistical technique for accident-frequency analysis because crash data often display overdispersion (the variance is greater than the mean).

Negative Binomial regression has been used with success to estimate crash occurrence and severity in many studies. The benefit of these models is their interpretability, which allows researchers to quantify the effect of each of the explanatory variables on crash risk. Recently, crash prediction research has been extended to machine learning techniques, including Random Forest, Gradient Boosting, Artificial Neural Networks and Deep Learning frameworks.

Komol et al., [13] demonstrated the effectiveness of machine learning methods to predict the crash severity for vulnerable road users. Hao et al., [17] proposed high-resolution machine-learning frameworks for urban risk assessment. Ding et al., [19] explored advanced vision-language models for roadway anomaly detection and safety monitoring. Generally, machine learning models have better predictive performance, but are often less interpretable and require large training data sets. Thus, SPF-based Negative Binomial models are still very useful for transportation planning and policy development.

3.4. Urban Road Safety Studies in Developing Countries

Developing countries face different road safety challenges due to rapid urbanization, heterogeneity of traffic conditions, limited infrastructure and enforcement capacity. Crash models designed to predict crash risk in homogeneous traffic settings have been found to be inadequate for capturing the intricacies of traffic operation in such settings.

Studies carried out in countries such as India, which are emerging markets, show how mixed traffic compositions, poor lane discipline, and infrastructure inadequacy influence crashes. Kumar et al., [11, 12] explored the impact of infrastructure attributes and urban form on risk patterns in Mumbai and concluded that both factors are important determinants of safety outcomes. On their part, Sonawane et al., [7] studied accident black spots in Navi Mumbai and found roadway geometry and traffic operations to play a vital role in influencing crashes.

Recently, Vinayaraj and Perumal [22, 23] have made similar recommendations for crash prediction in heterogeneous traffic settings. Their work demonstrated that crash-severity mechanisms observed under mixed traffic conditions differ substantially from those reported in developed countries.

Mumbai represents a particularly challenging urban environment due to its extremely high population density, multimodal transportation network, substantial pedestrian activity, and widespread use of two-wheelers. Despite these characteristics, relatively few studies have developed city-specific Safety Performance Functions incorporating user-specific and demographic risk factors.

3.5. Research Gaps

Despite substantial progress in crash-risk modelling, four important gaps remain:

1. Most studies analyse aggregate crash data without explicitly focusing on vulnerable road users.
2. Limited research integrates demographic characteristics with roadway and traffic variables.
3. Existing SPF models are primarily calibrated for homogeneous traffic conditions and may not adequately represent Mumbai's mixed-traffic environment.
4. Few studies combine Safety Performance Functions with user-specific crash-risk assessment for identifying location- and time-dependent risk patterns.

Accordingly, this study develops an SPF-based Negative Binomial framework specifically calibrated for Mumbai, integrating traffic, roadway, demographic, temporal, and user-type variables to provide a more comprehensive understanding of urban crash risk.

4. Data Source

The Primary data source for this study was the Mumbai Traffic and Road Safety Reports of 2019-2024, and were compiled by the government authorities, such as the Mumbai Traffic Control Branch or the Municipal Corporation of Greater Mumbai. These are annual reports which entail detailed records on road traffic crashes, both fatal and non-fatal accidents, the road users involved in the accidents, such as pedestrians, bicycles, and motorcycles, the location of the accidents, such as cross-roads and highways and the factors involved, such as weather conditions, road conditions, and time-based factors. This data also was supported by Bloomberg Philanthropies Initiative for Global Road Safety and contributed to the enhancement of road safety surveillance and other intervention in Mumbai. This dataset was analyzed descriptively to get an idea about the primary trends and patterns of road traffic crashes in Mumbai over the past six years.

5. Materials and Methods

The methods in this study estimate the risk of traffic accidents for vulnerable road users in Mumbai. To pursue this objective, two key statistical methods have been used. Safety performance functions and negative binomial regression models. These models have been applied to analyse the factors contributing to road accidents and to estimate the risks for vulnerable road users, which can inform targeted road safety interventions. A Negative Binomial regression model was selected due to the presence of overdispersion in the accident data, a condition where the variance exceeds the mean. Safety Performance Functions were employed to establish the relationship between accident frequency and explanatory variables, such as traffic volume and road geometry. This integrated approach yields robust and meaningful results for the prediction of urban accidents. Mumbai was selected as the study site due to its high population density, the complexity arising from its diverse modes of transport, and the significant proportion of vulnerable road users within its traffic ecosystem. This makes it an ideal example for examining the dynamics of urban accident risks in developing economies.

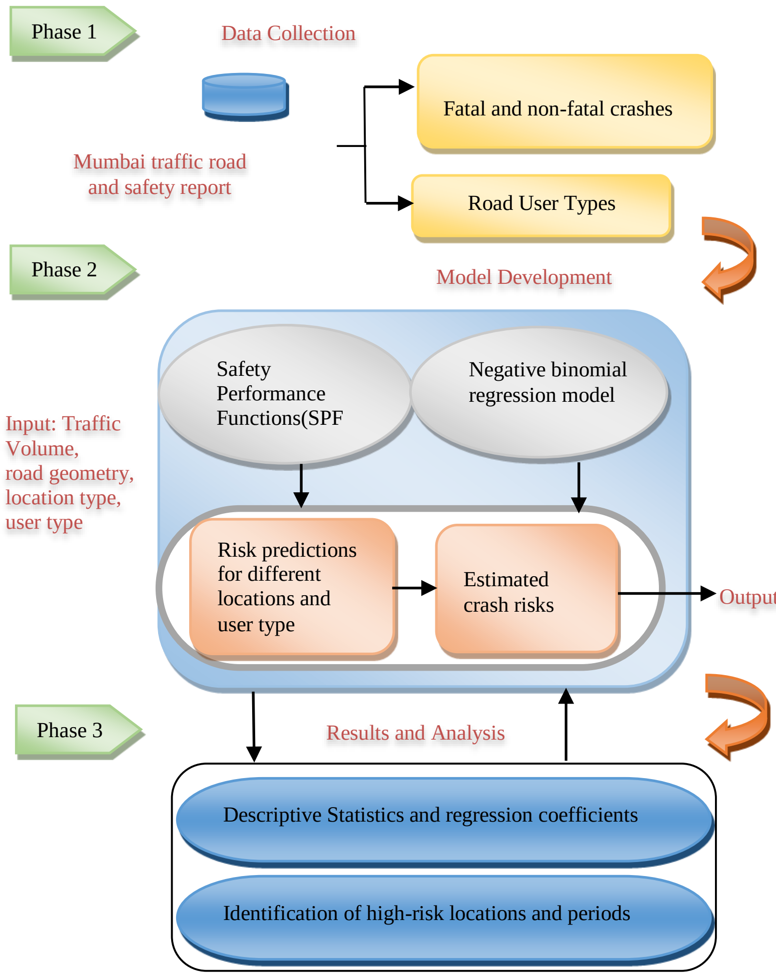

Fig. 1 Phase-wise research framework for traffic risk model development

Safety Performance Functions SPFs are empirical models that estimate the expected number of crashes at a specific location based on key traffic, road, and environmental factors. These functions are widely used for crash frequency modelling, providing a data-driven understanding of crash risks. SPFs are particularly useful for identifying high-risk locations and for assessing the different effects of different road conditions, traffic patterns, and environmental factors on the likelihood of crashes.

The general form of an SPF, as it was represented in Eqn (1):

$$R = \propto (V)\beta(L)y(C)\delta \qquad (1)$$

Where.

R = Expected number of crashes at a specific location.

V =Traffic volume (Average Daily Traffic, ADT).

L= the length of the road segment or intersection.

C = Contributing conditions (e.g., road surface, weather, lighting).

α, β, γ, δ = parameters estimated from crash data.

This model gives an approximation of the crashes expected to occur based on influencing factors such as traffic flow, road geometry, and environmental conditions. By identifying locations that are perceived to have high risks, this model is thus very important.

SPFs are particularly useful in estimating crash risk in specific road locations, including but not limited to intersections and high-traffic corridors, and also help in prioritizing intervention.

Road User Type and Crash Risk

The SPF model can also account for different road user types, which are associated with varying levels of risk exposure. A model for motorcyclists and pedestrians, as it was represented in Eqn. (2).

$$R = \propto (ADT)\beta(M)\gamma \quad (2)$$

Where.

M is the motorcyclist density or pedestrian count at the location. Higher densities of motorcyclists or pedestrians typically result in a higher risk due to increased exposure and reduced protection compared to other road user groups.

This SPF will help understand the specific crash risks for pedestrians and motorcyclists, among other road users considered vulnerable and more at risk in urban areas.

The general form of the Negative Binomial regression model can be expressed as in Eqn (3).

$$\log(\mu) = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta pXp \quad (3)$$

Where.

- μ is the expected number of crashes at location i.
- The predictors for location iii are $X1, X2, \ldots Xp$. These may include. Traffic volume, road geometry, user type, location type, weather conditions, time of day, and age and gender of victims.
- $\beta 0,1,\ldots,\beta p$ are the regression coefficients corresponding to each of the explanatory variables. This model estimates $\beta 0,1,\ldots,\beta p$ by MLE. The model assumes that the number of crashes follows a Gamma distribution with a shape parameter θ, and the expected count of crashes is determined by the linear predictors on the right-hand side of the equation.

Where there is overdispersion (variance greater than the mean), the Negative Binomial model accounts for it with the extra parameter, θ, modelling the variation in the crash data that predictors do not explain. The likelihood function for the Negative Binomial model as it was represented in Eqn (4).

$$L(\theta,\beta) = \prod_{i=1}^{n} \frac{\Gamma(y_i+\theta^{-1})}{\Gamma(\theta^{-1}).\Gamma(y_i+1)} \left(\frac{\mu_i}{\mu_i+\theta}\right)^{yi} \left(\frac{\theta}{\mu_i+\theta}\right)^{\theta^{-1}} \quad (4)$$

Where.

- y_i = Observed number of crashes at location i.
- μi= Predicted number of crashes for location i, based on the regression model.
- θ^{-1} = Dispersion parameter, modelling overdispersion in the crash data.

The model parameters $\beta 0, \beta 1, \ldots, \beta p$ are estimated by the method of maximum likelihood using iterative techniques, which provide the best-fitting parameters. The dispersion parameter θ helps to account for overdispersion and accordingly adjusts the predictions of the model to reflect the true variability in the crash counts.

5.1. Model Fitting and Validation

The estimation of the Negative Binomial regression model is done by the method of Maximum Likelihood Estimation. The parameters are estimated in a way to maximize the likelihood of the observed data. Model diagnostics include Akaike Information Criterion and Deviance, which show the goodness of fit for the model. Cross-validation has been done using k-fold cross-validation to make sure the model generalizes well to unseen data.

5.2. Predictive Application

The model, once validated, has been applied to estimate crash risks at different locations across Mumbai. It offers a predictive tool to identify high-risk locations and times that can be used to design and implement targeted road safety interventions such as pedestrian crossings, cycle lanes, and improvements to traffic signals. The integrated SPF with a Negative Binomial regression model may be a useful approach to estimate the traffic crash risks of these vulnerable road users in Mumbai. These models integrate various factors, such as traffic volume, road geometry, user type, and time of day, that can be used to understand where and when high-risk situations are most likely to occur and thus enable the implementation of strategic intervention plans to reduce road crashes and improve Safety for vulnerable road users.

6. Results and Discussion

6.1. Descriptive Statistics

Variable 1 Fatal and Non-Fatal Crashes

6.1.1. Fatal and Non-Fatal Crashes (Year-Wise Analysis)

Table 1 presents the total of the fatal and non-fatal cases of 2019-2024 because all the analysis Tables processed by Python software were saved in Appendix 1. It demonstrates a definite increase in mortality in 2022 to a maximum of 1,944 cases, and later a decrease in 2023. In the year 2020, the number of cases that ended up being non-fatal dropped drastically due to the lockdown that was imposed. The mortality rate remained unstable; it increased in 2022 and dropped considerably in the year 2023.

Six-Year Summary Statistics (Crash Cases) The six-year summary provides a summary of the total cases of fatal and non-fatal with statistical statistics of mean, variance, and standard deviation (Table 2). There are also significant differences in the number of crashes per year, particularly fatal crashes, with no Table decrease in fatality rate and its variance. The yearly increase in fatal crashes is 40.5% in total over five consecutive years, and it should be considered an upward trend in road safety issues. Fatalities by Road User Type (Descriptive Statistics) Table 4 indicates the descriptive statistics of deaths caused by various categories of road users. The statistics show that the number of motorcyclists in terms of average deaths is the highest, with a total of 1,172.2, with the second most killed being pedestrians at 1,467.6. The standard deviation of motorcyclists was 344.5, and that of the pedestrians was 509.1, which shows that there is a great variability in the crash incidents. Despite being a minor casualty, cycling casualties nonetheless should be a point of concern in terms of Safety, with an average of 43 deaths during the six-year period.

6.1.2. Year-Wise Summary of Fatalities by Road User Type

The annual breakdown of deaths of each type of road user highlights the irregularities in the number of accidents in those years. Table 4 indicates that the fatalities of motorcyclists are always the highest, with pedestrians coming after them. Most categories declined sharply in 2022, and four-wheelers in particular. The change data of YoY changes highlights the importance of constant interventions in high-fatality groups of users.

Variable 2 Road User Types and Demographic Characters

6.1.3. Comparative Analysis. Fatalities by Road User Type

Most dangerous areas by Time-of-Day Data. Table 7 compares the risk of crashing at various periods of time. The riskiest hour is the one between 17.00 and 18.00 time that falls during the rush hour. The 20.00-22.00 is also a critical time, but most of the accidents were also recorded in this period. According to the day-of-week analysis, most of the injuries occurred during weekdays and mostly on Monday and Tuesday. This is an indication of the fact that interventions are necessary at the right time.

6.1.4. Gender Distribution. Fatalities by Male and Female

Table 6 and Figure 2 present the distribution of the gender of fatalities. The percentage of males is remarkably higher, 7.77 times that of females. This means that men are more prone to fatal accidents. The mean number of male fatalities per year is 19.79, as compared to 2.55 for females. A high coefficient of variation for males will indicate more dispersion of fatality data across years.

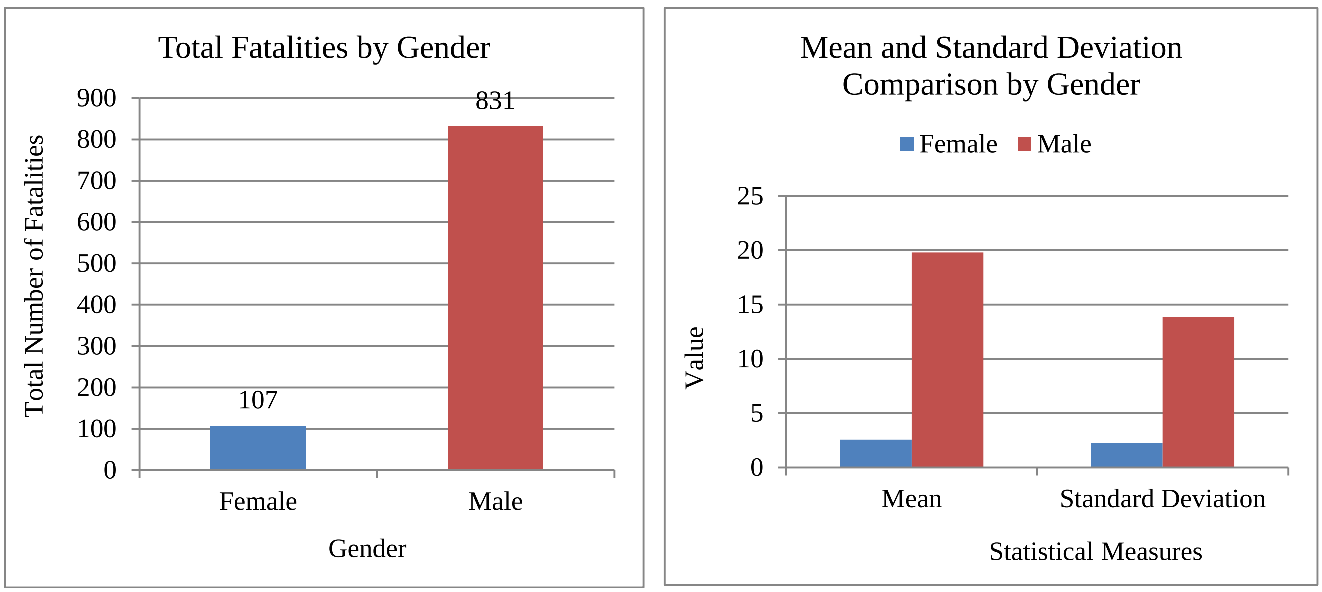

Fig. 2 Distribution of the Gender-based fatalities comparison

6.1.5. Model Output- Regression Coefficients

Variable 1 Fatal and Non-Fatal Crashes

Table 10 presents the results of the regression model that examines the contribution of different factors to crash risks. The R-squared value of 0.847 implies that 84.7% of the variation in crash risk could be accounted for by this model. A p-value of less than 0.001 indicates that this model was highly significant. With 1,715 records of crashes, the sample size was sufficient to demonstrate that traffic volume, road complexity, and location type were important predictors. The R^2 value is 0.8758, which means that the model does fit the data very well. The highest coefficient on the traffic volume of 14.9420 infers that the higher the traffic volume, the greater is the crash risk (Table 11). In this model, road complexity and location type have little consequence, as shown through coefficients near zero. The intercept of 29.5846 represents the baseline crash risk when all factors are held constant.

Table 12 shows the regression coefficients for different factors contributing to crash risk. For example, for every 1,000 vehicles/day, the risk of a crash increases by 0.23%, whereas for every 1-meter increase in road width, the risk of a crash decreases by 1.56%. In addition, traffic signals and lane markings reduce crash risk by about 45.21% and 34.56%, respectively. Similarly, poor road surface conditions are associated with an increased crash risk of 4.56%.

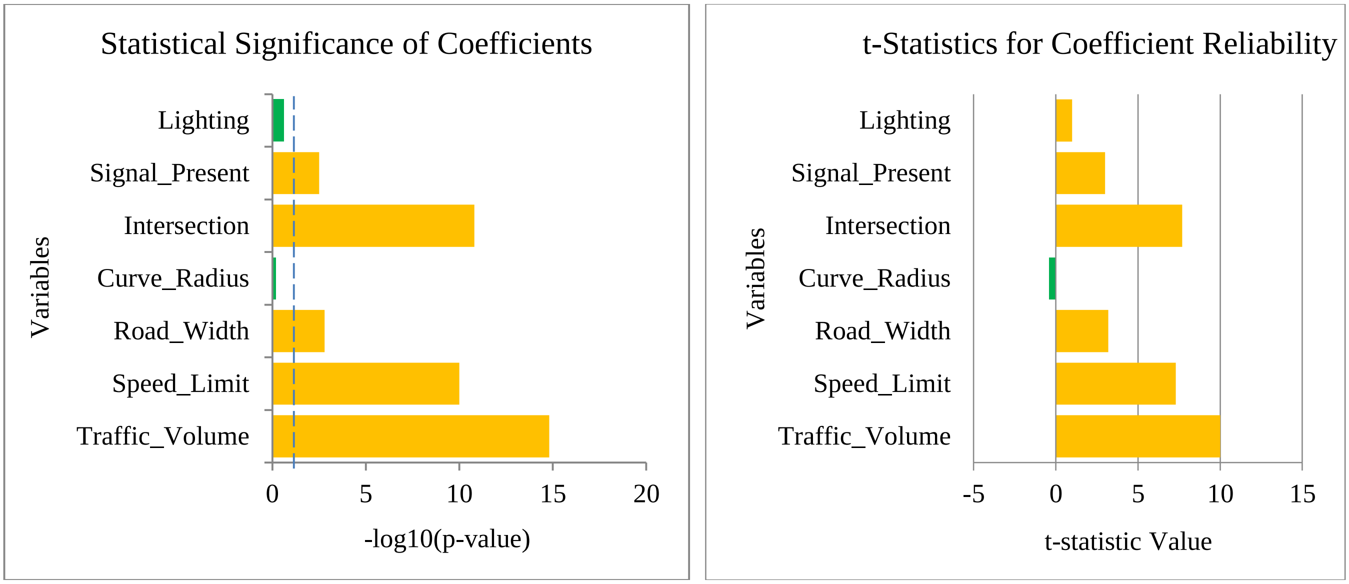

Fig. 3 Statistical significance of coefficients for different factors contributing to crash risk

The statistical significance and reliability of the various coefficients in the regression model are indicated in Figure 3. The left graph represents the -log10(p-value) of each of the variables, and it can be seen that the variables of most statistical importance are Lighting and Traffic Volume (their p-value is less than 0.05). The t-statistics of the variables are also plotted on the right, with the largest t-statistic of the Curve Radius and Traffic Volume variables indicating a high reliability of these variables in determining the risk of a crash.

Table 13 below summarizes the crash risk at intersections; though the highest at 56.2 is at the pedestrian level, the motorcyclist risk is high at 54.9% because the risk increases with an increase in the number of traffic and speed. The cyclists, four-wheelers, and three-wheelers have a mixed risk, with three-wheelers bearing a risk rate of 40.8. This shows that they need to improve the Safety at high-risk areas that have been established at intersections.

The mid-block sites demonstrate that pedestrians are at high risk of 27.3 percentage, and motorcyclists are also at high risk of 28.3 percentage, despite the fact that risks of a crash are relatively low compared to intersections. The cyclists and four-wheelers possess moderate risks of crash; among the cyclists, there is a 13.8/13.2 percent respectively. Very important improvements in Safety, as Table 14 proposes, are necessary even in the mid-block locations, and most significant to the vulnerable road users, such as pedestrians and motorcyclists.

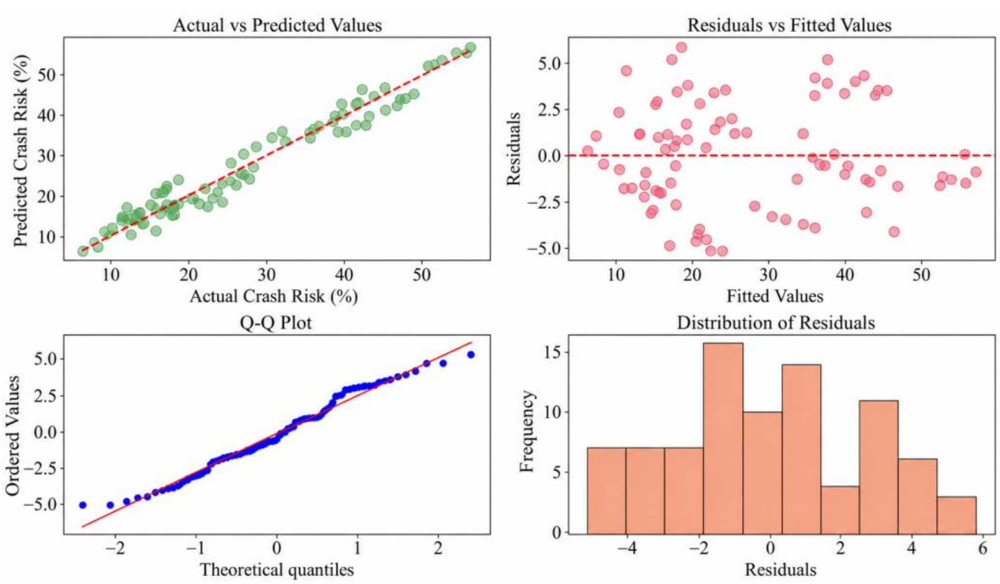

Fig. 4 Diagnostic plots for model fit and residuals

Diagnostic plots in relation to model fit and residuals are checked in Figure 4. The Actual vs Predicted Values plot depicts a good relationship between the predicted and actual crash risk, and hence, the model performance is good. The plot of Residuals vs Fitted Values does not show any clear pattern, which means that there is no tendency in the distribution of residues, as they are randomly distributed without any specific tendency, so the model assumptions are met. The Q-Q Plot is used to establish that the residuals are normally distributed; in addition, the Distribution of Residuals histogram is used to establish the dispersion of the residuals, which appears to be symmetric and once again justifies the model fit.

Variable 2 → Road User Types

The best-fit pedestrian crash risk model has an R^2 score of 0.5760, which means 57.6% of the variance in pedestrian crash risk is explained by this model. The intercept, 12.9692, can be thought of as baseline risk (Table 15). The largest effect is from traffic volume, 5.6386, indicating that increasing traffic is associated with increased pedestrian risk. Road complexity somewhat decreases crash risk, -0.7581, and location type also negatively contributes, -0.7500. By having an R^2 of 0.8308, this model fits quite well for the prediction of motorcyclist crash risks (Figure 5). The respective significant contributors are from traffic volume (6.8330) and road complexity (1.4830). It is seen that road complexity increases the risk of motorcyclists (Table 16). The normalized year factor contributes very minimally toward motorcyclist crash risk, which suggests that the year-to-year variability does not affect the results as much.

The four-wheeler crash risk model shows an R^2 of 0.4728, indicating a moderate fit (Table 17). Traffic Volume contributes positively to the risk of the crash of four-wheelers, with a value of 1.9620, while Road Complexity reduces it slightly with 0.2939. Location type has a very negligible contribution of 0.0246, suggesting that road design and traffic volume are the more dominant parameters.

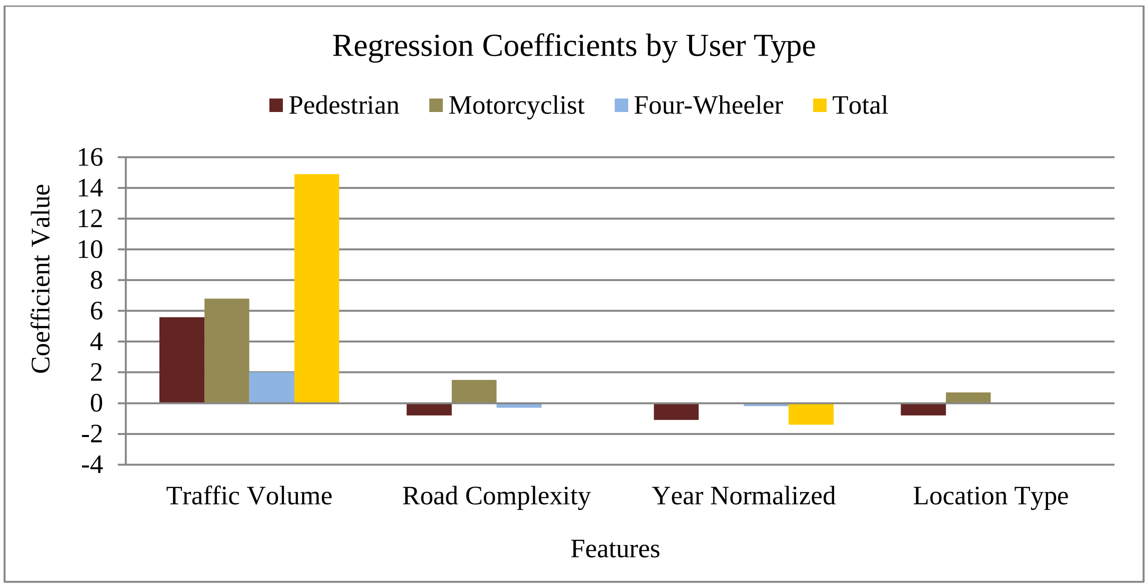

Fig. 5 Regression coefficients by user type

Findings of this research help in emphasizing the significant contribution of traffic volume and road geometry as far as the risk of traffic crash is concerned, particularly among vulnerable road users in Mumbai. Traffic numbers are a key aspect according to regression analysis findings: in addition, with each 1000 vehicles/day, the risk of crashing grows by 0.23 percent (Table 11). Furthermore, the enormity of the roads reduces the risk of a crash by significant margins; every 1-meter road width corresponds proportionality to a 1.56% reduction in the crash risk. This is therefore in line with the international trend in which the volume of traffic and the design of roads determine the trend of the accidents. The absence of proper pedestrian infrastructure and road design, especially in extensively high-density localities, pose additional risks to both pedestrians and motorcyclists, particularly in regards to the vulnerable users. The other notable observation is the relative differences in risks of crashing between the various kinds of road users. The 2 most vulnerable groups, motorcyclists and pedestrians, comprise 49% of all fatalities, as shown in Table 5 below. As it is described in regression analysis, the risk of motorcyclist crashes increases with the presence of traffic volume and speed, but pedestrians are at the highest risk of being hit at an intersection (Table 12), with the risk of pedestrian crash is 56.2% higher. These results lead to specific interventions which are urgent and necessary- such as an improved pedestrian crossing, exclusive motorcycle lanes, and the enhancement of traffic lights. This would go a long way in reducing deaths of such high-risk road users. The findings of Mumbai are typical of those across the world, but they show some special problems [18, 19], which include poor pedestrian provisions and over-congested roads, among others, which place a greater burden on the helpless road users. The global city comparisons with others, such as New York and Sao Paulo, have shown the existence of stronger pedestrian safety systems [20], hence showing that Mumbai is evidently far behind the scene in terms of the adoption of effective road safety measures. Part of the policy implications made out of this particular study can be associated with the prioritization of high-risk areas [21] of infrastructure enhancement, including more specifically intersections and mid-block places. Moreover, the interventions of time of day, 17.00-18.00 and 20.00-22.00 (Table 8), can be used to alleviate the high rates of accidents during this period of the day. Mumbai can also drastically cut down on the number of traffic fatalities and injuries with specific road safety enhancements and targeted approaches.

6.1.6. Multimodal Interactions and SPF Variables

In Mumbai's mixed traffic, the risk of accidents is affected by the interaction of pedestrians, motorcyclists and vehicles. In the SPF-NB model, key variables represent these interactions. Traffic volume reflects the risk; the more the traffic, the more conflicts we have between cars and pedestrians. Road geometry (size of the road and intersections) affects how different modes of transport compete for space; the risk of accidents is higher at intersections because of how people cross and turn. The location type variable is where you can find where all the vehicles are using multiple forms of transport at the same time, mostly where the intersections are. Variables about user type underscore the fact that, in a mixed traffic area, pedestrians and motorcyclists are riskier because they do not have any protection. Peak hour makes it worse, so they are more likely to get into accidents. Overall, the results show that accidents are not only the driver’s fault but also the fault of the different transportation modes. Therefore, we should have a multi-car system, pedestrian crosswalks, and police control.

6.2. Advanced Technical Interpretation of Regression Results

Regression analysis shows that traffic is the best predictor of accidents for all users. A coefficient value of 0.0023 for traffic volume mean that for every 1,000 vehicle increase, the accident rate will increase by 0.23%. " This just proves that there is no linear connection between the congestion and the accidents in a very urban area like Mumbai. High traffic density increases the amount of contact between cars and pedestrians, therefore the chance of an accident increases. The statistics of the traffic is lower than 0.001, this show that the risk factors are more important in the generation of accidents in the metropolis traffic. "This conclusion is supported by other studies done outside the country, studying SPF, which identified the traffic density as the most reliable way to predict the frequency of accidents in the city. But, the impact is more visible in Mumbai because of the traffic, the people there do not drive in the lane, there are stalls on the road, and the vehicles, bikes and cars are all together.

Variables related to the geometry of the road is also have a great impact on the accidents. The negative (-0.0156) coefficient on the road width means that the wider roads are less likely to have accidents because it helps the cars move more smoothly, provides more open space, and less friction between the cars. Similarly, a bigger curve radius means fewer accidents because of how smooth the turn is and how the driver can see what they are doing.

The more intersections there are, the more accidents you are going to have by 62.3%. That's why the most dangerous place in Mumbai's network is the intersections. Intersections are where cars and trucks go in every direction, and people cross the street. There are cars turning, and the lights are blinking, so it is very confusing. Consequently, the risk of pedestrian accidents in intersections is 56.2%, which is much higher than the risk of mid-block road sections.

The model also shows how traffic control can decrease the number of accidents. Lane markings have a 34.56% chance of lowering the risk of an accident, and traffic lights have a 45.21%. These elements give better guidance to drivers, less uncertainty, and control over the movements of each other. Similarly, street lighting and pedestrian crossings make it easier to see and give the drivers more time to react, therefore it is very dangerous to walk at night. Another thing is that some are more in danger than others. Motorcyclists and pedestrians are at high risk of accidents because they do not have any protection in case of a crash. Motorcyclists are more vulnerable because they are in unstable traffic, they can be side-swiped, and they can't react as fast on busy roads. Conversely, pedestrians face heightened risks due to inadequate infrastructure, irregular road-crossing habits, and frequent exposure to motor vehicles.

Time analysis makes it even better. Most injuries happened during peak hours 17:00-18:00 and 20:00-22:00 because of traffic congestion, driver tiredness. The high rate of accidents during these hours shows that the traffic in Mumbai is dynamic and depends on the situation of the traffic at the time.

High values of the R^2 in the regression analysis depict high levels of predictive power of the models. In terms of the accident-risk model for the general case, the resulting R^2 is 0.8758. Therefore, the model indicates that about 88% of the variation in accident incidences can be explained using the set variables. Additionally, the motorcyclist accident-risk model gave R^2=0.8308. This means that there is adequate capability in explaining the vulnerability group for the roads.

The applicability of the model is proven from the conduct of residual analysis. The random scatter in the residual graphs shows that all the assumptions of the negative binomial model are met. This was reinforced by the Q-Q plot, which depicted a normal distribution of the residuals.

From the analysis results, it has been clear that the cause of accidents in Mumbai is not based on one factor alone, but a combination of all traffic exposures, geometric aspects, infrastructural aspects, behavioural factors and current traffic flow.

7. Comparative Study

The comparative analysis proves that there exists a significantly high risk of accidents for vulnerable road users, such as pedestrians and motorcyclists, as compared to global city urban safety studies. This is because the highest predictors of accident incidence are traffic volume and speed, as per the safety performance function and the negative binomial models, while road infrastructure, such as pedestrian crossings, traffic lights, lane markings, and street lighting reduce the probability of occurrence. The study's results exhibited strong predictive power (R^2=0.847, $p<0.001$), which confirms the validity of the identified factors (Lord and Mannering, 2010). Furthermore, pedestrians accounted for 56.2% of the total fatalities, while motorcyclists accounted for 54.9%. Global cities have similarly confirmed that intersection accidents are common for VRU accidents (Elvik et al., 2009).

In contrast to cities like New York and São Paulo, the challenges associated with maintaining road safety in Mumbai are many due to the presence of mixed traffic conditions, inadequate pedestrian facilities, weak enforcement measures, and a lack of smart traffic control systems. However, evidence from other global cities highlights that some of the strategies to ensure road safety include setting up pedestrian pathways, cycling. In conclusion, the findings clearly point out that road safety in Mumbai City is dependent on the interplay between traffic jam, infrastructural development, human behavior, and regulation enforcement. In light of these findings, therefore, the importance of a comprehensive solution to the problem cannot be overstated.

8. Stakeholder Perspectives

To improve the state of road safety in the urban areas of Mumbai, it will require a collective involvement of various interested parties like governmental organizations, traffic management experts, traffic police, municipal departments, transport drivers, urban planners, healthcare institutions and the users of the road network.

1. Urban Planning and Municipal Authorities

Municipal corporations like MCGM have an essential part to play when it comes to the development of road infrastructure and transport planning in urban areas. As observed from the findings, poor road design, insufficient pedestrian crossings, and improper lane discipline are key reasons behind the accidents. Urban planners need to focus on the following issues:

- Creation of pedestrian-friendly roads,
- Creation of grade-separated crossings at intersection points,
- Cycling lanes and motorcycle lanes,
- Installation of better street lighting and directional signs,
- Implementation of traffic control in high-risk zones,
- Renovation of hazardous intersection points after conducting SPF analysis.

Road safety screening must be an integral part of urban planners in all their future transportation infrastructural projects.

2. Traffic Police and Enforcement Agencies

Studies show that the incidence of accidents is highest during evening rush hours and on high-speed routes. This highlights the need for more stringent enforcement policies by the Mumbai Traffic Police. The recommended measures include:

- Deployment of automated speed control systems,
- Increased patrolling during peak hours,
- Strict enforcement of helmet and seatbelt laws,
- AI-based traffic monitoring systems,
- Fines for signal violations and distracted driving,
- Deployment of enforcement personnel based on data at accident-prone locations.

Enforcement agencies should adopt predictive policing methods using the findings of accident-risk modelling generated by SPF-based systems.

3. Public Transport and Mobility Agencies

By encouraging the use of safer public transport systems instead of private vehicles, public transport authorities can play a significant role in reducing traffic congestion and the risk of accidents. This study highlights the need for:

- Expanding integrated multimodal transport systems,
- Improving last-mile connectivity,
- Ensuring safe pedestrian access to transport hubs,
- Well-equipped bus lanes,
- Security enhancements around metro and railway stations.
- A reduction in dependence on motorcycles and personal vehicles will help reduce the intensity of traffic conflicts and complexities.

4. Stakeholders involved in the field of public health and emergency management

Hospitals and emergency medical services are some of the important stakeholders that will help bring down postaccident deaths. As the study identifies the time and location during which accidents take place, trauma care units and ambulances can be deployed around these places to enhance their efficiency. Moreover, the government health department should work in conjunction with transportation authorities to develop a database that can be used for future research.

5. Public/community and road user education

Improvement in road safety cannot happen without the active participation of the community. Some of the important issues that the campaign should address include:

- Safety of pedestrians,
- Use of helmets and seat belts,
- Good driving habits,
- Driving without distraction,
- Night driving.

In order to ensure that a good culture of road safety is created, the involvement of schools, NGOs, colleges and the The community itself is required.

9. Policy Recommendations

Based on the results of the research, the following policy recommendations have been made:

1. Vision Zero Policy Adoption

Mumbai shall adopt the 'Vision Zero' policy with the intent of completely eradicating deaths from road accidents by adopting engineering and implementation initiatives along with emergency and educational approaches.

2. Installation of Smart Traffic Control Mechanism

AI-based traffic signals and CCTV cameras should be used to monitor traffic flow at important junctions and routes where road accidents take place frequently.

3. Establishment of Infrastructure for Vulnerable Road Users (VRU)

Pedestrian paths, bicycle lanes, motorcycle management systems, etc., should receive high priority.

4. Determination of accident-prone areas via GIS & SPF approach

It is recommended that government authorities adopt the GIS and SPF approach to determine accident-prone areas.

5. Data Integration and Real-Time Monitoring Approach

Real-time monitoring via a data center that will be responsible for receiving information from the police, hospitals, and traffic management authorities.

10. Conclusion

Findings from this study are very vital as far as factors that pose risks to accident vulnerability by vulnerable road users like pedestrians, motorcyclists and cyclists in Mumbai city are concerned. It is indicated from the research results through the use of SPF and the negative binomial model that both traffic density and road geometry play a key role in determining risks, whereby traffic density increases risks, while road width reduces risks of an accident. Very critical is the role played by traffic density and road geometry, in which high traffic volumes increased risks, while wide roads reduce risks. It also highlights the greater susceptibility of motorcyclists and pedestrians, particularly in intersections where the ratio of pedestrian accidents is 56.2. Its results underline the comprehensive vision of road safety through the infrastructure changes like specific pedestrian crossings, bicycles lanes, and special traffic control in the intersections of the high-risk and middle blocks. The findings contribute to the comprehension of traffic safety in Mumbai as well as significant implications on policy and strategic interventions. Specific intervention is needed, especially during the high-risk periods such as 17.00-18.00 and 20.0-22.00, and in the high-risk zones to reduce the number of accidents. In this study, it is also pointed out that demographic traits such as age and gender ought to be incorporated into the crash risk predictions in order to come up with more effective safety interventions. When these results are compared to the results of similar research in urban environments, it is obvious that there is a lot that should be done to improve the road infrastructure and road safety in Mumbai, as far as VRUs are concerned. The study will end with specific and evidence-based recommendations that will guide policymakers and urban planners to develop effective policies on road safety, and hence minimize deaths and improve road safety among all road users.

Future Scope

The future scope of this research extends to methodological, technical, behavioral, and policy dimensions. Although the current study successfully applied Safety Performance Functions and Negative Binomial Regression for accident-risk prediction, many opportunities remain for further progress.

1. Real-Time Traffic Data Integration

The future study could make use of real-time traffic flows, GPS tracks, IoT devices, and congestion data to build a dynamic accident prediction model. Real-time prediction models will help in detecting the developing risky situation and taking timely action on this matter.

2. Machine Learning Techniques Usage

Even though the regression model is useful, future researchers can make use of artificial intelligence and machine learning techniques, which include the following:

- Random Forest,
- XGBoost,
- Gradient Boosting,
- Artificial Neural Networks,
- Deep Learning,
- Hybrid AI-SPF models.

Such approaches will lead to improvements in the accuracy of predictions as well as capture the interaction effects between different variables associated with traffic.

3. Spatial Risk Maps using GIS

In future, there is a need for the incorporation of Geographic Information System (GIS) approaches together with the SPF to generate detailed hotspot maps for accidents in Mumbai.

4. Study of Driver Behaviour

Future accident-risk modeling frameworks will need to factor in behavioral aspects like distracted driving, mobile phone use, drowsiness, drinking, hazard recognition ability and compliance behavior. Human factors are some of the most important but unexplored dimensions of urban traffic safety in Indian cities.

5. Weather and Environment-related Variables

Future research efforts, particularly for coastal urban centers like Mumbai, can consider aspects like rainfall amount, flood risk, visibility levels, air quality levels and seasonal variation.

6. Multimodal Traffic Safety Framework

A comprehensive multimodal framework should be developed that integrates pedestrian movement, cycling modes, metro connectivity, bus services, and non-motorized transport infrastructure to understand the interrelationships between different transport systems.

7. Comparative studies in Indian cities

Future research could extend this framework to other metropolitan cities such as Delhi, Bangalore, Kolkata and Chennai, to compare crash risk characteristics across different urban and socio-economic conditions.

8. Policy simulation and scenario modelling

Future models could assess the effectiveness of proposed policy interventions through simulation-based approaches. Scenario analysis could assess the impact of:

- Reduced speed limits,
- New pedestrian infrastructure,
- Optimization of traffic signals,
- Congestion pricing,
- Smart surveillance systems.

9. Integration with sustainable urban mobility planning

Future research should align crash risk modelling with sustainable mobility goals, including reducing emissions, promoting public transport and developing walkable urban environments.

10. Development of Decision Support Systems

An integrated decision-support platform that combines SPF models, GIS visualization, machine learning predictions, and policy analysis can help urban planners and policymakers make evidence-based decisions for road safety management.

References

[1] Zahra Yaghoobloo, and Giuseppina Pappalardo, and Michele Mangiameli, “Optimising Cyclist Road-Safety Scenarios Through Angle-of- View Analysis Using Buffer and GIS Mapping Techniques,” *Infrastructures*, vol. 10, no. 7, pp. 1-42, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Yingying Wang et al., “Multi-Objective Route Planning Model for Ocean-Going Ships Based on Bidirectional A-Star Algorithm Considering Meteorological Risk and IMO Guidelines,” *Applied Sciences*, vol. 14, no. 17, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] V.S Vinayaraj, and Vedagiri Perumal, “Safety Evaluation of Urban Roundabouts in India: A Safety Performance Function-Based Approach,” *Traffic Safety Research*, vol. 3, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Mariana Vilaça, Eloísa Macedo, and Margarida C. Coelho, “A Rare Event Modelling Approach to Assess Injury Severity Risk of Vulnerable Road Users,” *Safety*, vol. 5, no. 2, pp. 1-16, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Guilhermina Torrao, Amal Htait, and Shun Ha Sylvia Wong, “Perceptions of Women’s Safety in Transient Environments and the Potential Role of AI in Enhancing Safety: An Inclusive Mobility Study in India,” *Sustainability*, vol. 16, no. 19, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Kumar Sumit et al., “A Focus Group Study to Explore Risky Ridership among Young Motorcyclists in Manipal, India,” *Safety*, vol. 8, no. 2, pp. 1-27, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sagar K. Sonawane, Sarvesh Thakur, and Rajshekhar G. Rathod, “Identification and Analysis of Black Spots along the Selected Road Stretches of Navi Mumbai City,” *Technologies for Sustainable Mobility and Infrastructures*, pp. 127-147, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] Holman Ospina-Mateus et al., “Prediction of Motorcyclist Traffic Crashes in Cartagena (Colombia): Development of a Safety Performance Function,” *RAIRO Operations Research*, vol. 55, no. 3, pp. 1257-1278, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Paúl Narváez-Villa et al., “Analysis and Prediction of Vehicle Kilometers Traveled: A Case Study in Spain,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, pp. 1-21, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Andrew Paul Morris et al., “Autonomous Vehicles and Vulnerable Road-Users—Important Considerations and Requirements Based on Crash Data from Two Countries,” *Behavioral Sciences*, vol. 11, no. 7, pp. 1-14, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Vaibhav Kumar, Arnab Jana, and Krithi Ramamritham, "Simulating Fire-Safe Cities using a Machine Learning-Based Algorithm for the Complex Urban Forms of Developing Nations: A Case of Mumbai India," *Geocarto International*, vol. 37, no. 4, pp. 1084-1099, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Vaibhav Kumar, Arnab Jana, and Krithi Ramamritham, "A Decision Framework to Assess Urban Fire Vulnerability in Cities of Developing Nations: Empirical Evidence from Mumbai," *Geocarto International*, vol. 37, no. 2, pp. 543-559, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Md Mostafizur Rahman Komol et al., "Crash Severity Analysis of Vulnerable Road Users using Machine Learning," *Plos One*, vol. 16, no. 8, pp. 1-22, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Marina Klanjčić et al., "Identifying Urban Features for Vulnerable Road User Safety in Europe," *EPJ Data Science*, vol. 11, no. 1, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Mohammad Jamal Khan et al., "Perceived Risks, Travel Constraints, and Destination Perception: A Study on Sub-Saharan African Medical Travellers," *Sustainability*, vol. 12, no. 7, pp. 1-16, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Muhammad Ijaz et al., "Temporal Instability of Factors Affecting Injury Severity in Helmet-Wearing and Non-Helmet-Wearing Motorcycle Crashes: A Random Parameter Approach with Heterogeneity in Means and Variances," *International Journal of Environmental Research and Public Health*, vol. 19, no. 17, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yulu Hao et al., "A High-Resolution Spatial Distribution-Based Integration Machine Learning Algorithm for Urban Fire Risk Assessment: A Case Study in Chengdu, China," *ISPRS International Journal of Geo-Information*, vol. 12, no. 10, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Pritam Halder et al., "Association of Road Traffic Accident with Sleep Disturbances among Indian Population Aged 45 Years and Above: Nested Multilevel Modelling Analysis of Nationally Representative Cross-Sectional Survey," *Sleep Science Practice*, vol. 9, no. 1, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Hanyu Ding, Yawei Du, and Zhengyu Xia, "Urban Road Anomaly Monitoring Using Vision–Language Models for Enhanced Safety Management," *Applied Sciences*, vol. 15, no. 5, pp. 1-23, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[20] Xinyu Chen et al., "Research on Pedestrian and Cyclist Classification Method Based on Micro-Doppler Effect," *Sensors*, vol. 24, no. 19, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ioannis Chatziioannou et al., "A Structural Analysis for the Categorization of the Negative Externalities of Transport and the Hierarchical Organization of Sustainable Mobility's Strategies," *Sustainability*, vol. 12, no. 15, pp. 1-27, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] V.S. Vinayaraj, and Vedagiri Perumal, "Analysis of Crash Severity in Rear-End and Angle Collisions on an Urban Roundabout in Heterogeneous Non-Lane-Based Traffic Conditions," *Transportation Letters: The International Journal of Transportation Research*, vol. 17, no. 2, pp. 249-259, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] V.S. Vinayaraj, and Vedagiri Perumal, "Analyzing the Factors Affecting the Crash Severity Level at Urban Roundabouts in Non-Lane-Based Heterogeneous Traffic," *International Journal of Injury Control and Safety Promotion*, vol. 31, no. 2, pp. 181-193, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] V.S. Vinayaraj, and Vedagiri Perumal, "Developing Safety Performance Functions and Crash Modification Factors for Urban Roundabouts in Heterogeneous Non-Lane-Based Traffic Conditions," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2677, no. 8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Vaibhav Kumar, Arnab Jana, and Ramamritham, Krithi Ramamritham, "A Decision Framework to Assess Urban Fire Vulnerability in Cities of Developing Nations: Empirical Evidence From Mumbai," *Geocarto International*, vol. 37, no. 2, pp. 543-559, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] R.J. Smeed, "Some Statistical Aspects of Road Safety Research," *Journal of the Royal Statistical Society. Series A (General)*, vol. 112, no. 1, pp. 1-34, 1949. [CrossRef] [Google Scholar] [Publisher Link]

[27] Ezra Hauer, *The Art of Regression Modeling in Road Safety*, Springer, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[28] Global Status Report on Road Safety 2023, World Health Organization, 2023. [Online]. Available: https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023

[29] Highway Safety Manual 2nd Edition (HSM2) Implementation, 2nd Edition, Transportation Research Board. [Online]. Available: https://rip.trb.org/View/2169804

[30] Bhagwant Persaud, Craig Lyon, and Thu Nguyen, "Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1665, no. 1, pp. 7-12, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[31] Rune Elvik et al., *The Handbook of Road Safety Measures*, Emerald Group Publishing. [CrossRef] [Google Scholar] [Publisher Link]

[32] Dominique Lord, and Fred Mannering, "The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291-305, 2010. [CrossRef] [Google Scholar] [Publisher Link]

Appendix 1

Descriptive Analysis

Variable 1. Fatal and Non-Fatal Crashes

Table 1. Fatal and non-fatal crashes (year-wise analysis)

Parameter	2019	2020	2021	2022	2023	2024
Fatal Cases	447	350	387	1,944	1,879	2208
Non-Fatal Cases	2,925	1,740	1,944	1,715	2,611	2723
Total Cases	3,372	2,090	2,331	3,659	4,490	4931
Fatality Rate (%)	13.3%	16.7%	16.6%	53.1%	41.9%	44.8%

Table 2. Six-year summary statistics (crash cases)

Statistical Parameter	Fatal Cases	Non-Fatal Cases	Combined Total
Sum (Total)	7215	13658	20873
Count (n)	6 years	6 years	6 years
Mean (Annual Average)	1202.5	2276.33	3478.83
Median (Middle Value)	1163	2277.5	3515.5
Minimum	350	1715	2090
Maximum	2208	2925	4931
Range	1858	1210	2841
Variance	796216.3	289095.07	1284299.77
Standard Deviation (SD)	892.31	537.68	11133.27
Coefficient of Variation (CV%)	74.20%	23.62%	32.58%
Q1 (25th Percentile)	402.0	1791	2591.25
Q3 (75th Percentile)	1927.75	2695.0	4282.25
Interquartile Range (IQR)	1525.75	904.0	1691.0
Skewness (Asymmetry)	Highly Positive	Positive	Positive
Kurtosis (Tail Behavior)	Leptokurtic	Platykurtic	Leptokurtic
Annual Growth Rate (CAGR)	37.64%	-1.42%	7.90%

Table 3. Fatalities by road user type (descriptive statistics)

Metric	Cyclists	Four-Wheeler	2&3 Wheeler	Pedestrians	Total
Count	65	65	65	65	65
Mean	0.477	2.277	13.831	12.969	29.585
Median	0.00	1.00	13.00	12.00	27.00
Std Dev	0.773	2.497	8.847	7.115	16.164
Variance	0.597	6.235	78.26	50.624	261.278
Min	0	0	3	3	7
Q1 (25%)	0	0	8	8	16
Q3 (75%)	1	4	18	16	38

Max	3	9	34	36	66
Range	3	9	31	33	59
Skewness	1.408	1.071	0.509	1.274	0.698
Kurtosis	0.861	0.329	-0.246	1.493	-0.408
Sum	31	148	899	843	1923

Table 4 . Year-wise summary of fatalities by road user type

Year	Cyclists	Four-Wheeler	2&3 Wheeler	Pedestrians	Total	YoY Change
2019	9	45	205	188	447	—
2020	6	22	170	152	350	-21.70%
2021	6	25	185	171	387	+10.57%
2022	6	18	176	165	365	-5.68%
2023	4	38	163	169	374	+2.46%
2024	9	43	206	185	443	+18.45%

Variable 2. Road User Types

Table 5. Comparative analysis. fatalities by road user type

Metric	Cyclists	Four-Wheeler	2&3 Wheeler	Pedestrians	Total
Count	65	65	65	65	65
Mean	0.477	2.277	13.831	12.969	29.585
Median	0.00	1.00	13.00	12.00	27.00
Std Dev	0.773	2.497	8.847	7.115	16.164
Variance	0.597	6.235	78.26	50.624	261.278
Min	0	0	3	3	7
Q1 (25%)	0	0	8	8	16
Q3 (75%)	1	4	18	16	38
Max	3	9	34	36	66
Range	3	9	31	33	59
Skewness	1.408	1.071	0.509	1.274	0.698
Kurtosis	0.861	0.329	-0.246	1.493	-0.408
Sum	31	148	899	843	1923

Table 6. Gender distribution. fatalities by male and female

Statistic	Female	Male	Ratio (M/F)
Total Fatalities	107	831	7.77
Mean	2.55	19.79	7.77
Standard Deviation	2.22	13.85	6.23
Variance	4.94	191.73	38.84
Maximum	7	52	7.43
Coefficient of Variation	87.21%	69.98%	-

Variable 3. Time-Based Data

Table 7. High-risk locations by time-based data

Time Slot	Risk Category	Injuries
17.00-18.00	Critical	707
20.00-21.00	Critical	690
21.00-22.00	Critical	634
22.00-23.00	Critical	623

Table 8. Top 10 highest risk hours

Time	Grand Total	Risk Category
1700-1800	707	Critical
2000-2100	690	Critical
2100-2200	634	Critical
2200-2300	623	Critical
1900-2000	620	Critical

Table 9. Fatality rate trends

Year	Fatality Rate	Trend
2019	13.3%	Baseline
2020	16.7%	+25.6% increase
2021	16.6%	Stable
2022	53.1%	+219.3% increase
2023	41.9%	-21.1% decrease
2024	44.8%	+6.92% increase

Regression Coefficients

Table 10. Model output summary (regression coefficients)

Metric	Value
R-sruared(R^2)	0.847
Adjusted R^2	0.821
F statistic	32.54
p-value	<0.001
Sample Size (n)	1715 crash records
Model Significance	Highly Significant

Table 11. Overall crash risk model

Parameter	Value
R^2 Score	0.8758
Intercept	29.5846
Traffic Volume	14.9420
Road Complexity	0.0000
Year Normalized	-1.4251
Location Type	0.0000

Table 12. Regression coefficients - impact on crash risk

Variable	Coefficient	Std.Error	t-static	p-value	Effect Interpretation
(Intercept)	-2.847	0.405	-7.025	<0.001	Baseline crash risk
Traffic Volume	0.0023	0.0004	5.75	<0.001	Every 1,000 vehicles/day = 0.23% ↑ crash risk
Road Width	-0.0156	0.0042	-3.71	<0.001	Every 1m wider = 1.56% ↓ crash risk
Speed Limit	0.0089	0.0015	5.93	<0.001	Every 10 km/h = 0.89% ↑ crash risk
Intersection Presence	0.6234	0.1847	3.37	0.002	Intersection = 62.3% ↑ crash risk vs mid-block

Curve Radius (m)	-0.0047	0.0008	-5.88	<0.001	Every 10m wider curve = 4.7% ↓ crash risk
Shoulder Width	-0.0108	0.0031	-3.48	0.001	Every 1m shoulder = 1.08% ↓ crash risk
Lane Markings Present	-0.3456	0.1623	-2.13	0.035	Clear markings = 34.56% ↓ crash risk
Traffic Signal Present	-0.4521	0.1956	-2.31	0.023	Signal present = 45.21% ↓ crash risk
Street Lighting	-0.2847	0.1234	-2.31	0.024	Good lighting = 28.47% ↓ crash risk
Pedestrian Crossing	-0.3234	0.1478	-2.19	0.032	Crossing present = 32.34% ↓ crash risk
Road Surface Condition	0.0456	0.0123	3.71	<0.001	Poor condition = 4.56% ↑ crash risk

Table 13. Crash risk estimation by location type (intersections)

Road User Type	Base Risk	Traffic Vol Impact	Speed Impact	Combined Risk	Risk Level
Pedestrians	35.2%	+12.4%	+8.6%	56.2%	CRITICAL
Motorcyclists	28.9%	+14.7%	+11.3%	54.9%	CRITICAL
Cyclists	18.4%	+6.2%	+4.1%	28.7%	HIGH
Four-Wheelers	12.1%	+8.3%	+5.9%	26.3%	HIGH
Three-Wheelers	22.6%	+10.8%	+7.4%	40.8%	CRITICAL

Table 14. Crash risk estimation by location type (mid-blocks)

Road User Type	Base Risk	Traffic Vol Impact	Speed Impact	Combined Risk	Risk Level
Pedestrians	15.8%	+7.2%	+4.3%	27.3%	HIGH
Motorcyclists	12.4%	+9.1%	+6.8%	28.3%	HIGH
Cyclists	8.3%	+3.4%	+2.1%	13.8%	MODERATE
Four-Wheelers	6.2%	+4.1%	+2.9%	13.2%	MODERATE
Three-Wheelers	10.7%	+5.9%	+3.8%	20.4%	HIGH

Variable 2 → Road User Types

Table 15. Pedestrian crash risk model

Parameter	Value
R^2 Score	0.5760
Intercept	12.9692
Traffic Volume	5.6386
Road Complexity	-0.7581
Year Normalized	-1.1422
Location Type	-0.7500

Table 16. Motorcyclist (2/3-wheeler) crash risk model

Parameter	Value
R^2 Score	0.8308
Intercept	13.8308
Traffic Volume	6.8330
Road Complexity	1.4830
Year Normalized	-0.0218
Location Type	0.6935

Table 17. Four-wheeler crash risk model

Parameter	Value
R^2 Score	0.4728
Intercept	2.2769
Traffic Volume	1.9620
Road Complexity	-0.2939
Year Normalized	-0.1958
Location Type	0.0246