Comprehensive Analysis of Accident Data and Development of Road Crash Model in the Road Networks of Kerala

B.S. Jisha[a,*], M. Satyakumar[b]

[a] *Research Scholar, Department of Civil Engineering, Mar Baselios College of Engineering and Technology, Thiruvananthapuram 695015, Kerala, India,*

Assistant Professor, TKM Institute of Technology

[b]*Professor, Department of Civil Engineering, Mar Baselios College of Engineering and Tech-nology, Thiruvananthapuram 695015, Kerala, India.*

Abstract

Worldwide, more than 1.2 million persons killed and more than 50 million injured each year in road crashes. In India an increase of about 47% in road fatalities will be expected in the next 20 years. Kerala is the second most accident prone state in India. Kerala is a densely populated state, with ribbon development and encroachment. Road safety becomes more and more important every year as the annual growth rate of traffic is 10% in Kerala. The challenging factors in the traffic condition existing in Indian road networks are the mixed traffic condition and vulnerable road users. A comprehensive road safety plan is required to reduce the accidents and the severity of accidents. In this study the accident data were collected from the State Crime Records Bureau, Kerala and the crash model was developed based on the recorded accident data. The study is entirely based on the data recorded by police department. A comprehensive analysis of accident data was done and the Accident Severity Score based on various roadway characteristics, geometric characteristics, environmental characteristics, and vehicle and driver characteristics was found. Negative binomial regression was used to develop crash model and influence of different factors on Accident Severity was found.

Keywords: Road Crash Model, Accident data, Accident Severity Score

1. Introduction

Road accidents create the grave problem of public safety, causing fatal injury, mortality, and monetary loss all over the world. Determination of accurate cause to the accident is a crucial factor in the development of effective prevention. Predictive modelling provides a scientifically sound means for identifying patterns and correlations between accidents and causative variables such as roadway conditions, vehicle characteristics, road user behaviors, and climatic conditions.

The current study provides an accident prediction model derived from a range of factors impacting accident risk and severity. The model is expected to calculate the impact of various factors utilizing statistical techniques. Policymakers, traffic engineers, and city planners can leverage the data generated from this prediction model to deploy evidence-based safety measures, optimize roadway design, and improve transportation systems. The findings provide valuable information that can be used to reduce the frequency of accidents and lower their severity, resulting in improved road safety for all road users

1.1 Objectives

The main objectives of the study are to find the Accident Severity Score of accidents based on different factors and to develop a road crash model based on the recorded accident data.

2. Literature Review

Accident prediction models are important tools in traffic safety analysis since they provide a forecast of the probability and severity of road accidents using a combination of causative factors. The models assist in the detection of high-risk locations, quantification of the influence of roadway and environmental conditions, and the creation of data-driven safety countermeasures. Researchers and traffic engineers have employed various statistical and machine learning methods in modeling accident occurrence and severity since the beginning. The different accident forecasting models, divided into traditional statistical and more recent machine learning models are discussed below.

2.1 Traditional Statistical Models

Traditional statistical models have been widely used in accident prediction due to their strong theoretical foundations, interpretability, and ability to quantify relationships between accident occurrences and influencing factors. The most commonly used traditional models include Poisson Regression, Negative Binomial Regression, Logistic Regression, and Generalized Linear Models (GLMs). Studies have shown that negative binomial regression outperforms Poisson regression in handling accident frequency data (Lord, Washington, & Ivan, 2005). It is widely used for crash modeling, especially when data exhibit high variance (Park & Lord, 2007). Logistic regression has been extensively used in accident severity studies due to its ability to model binary (fatal vs. non-fatal) or multinomial (minor, major, fatal) outcomes (Milton, Shankar, & Mannering, 2008). Poisson regression has a major limitation of over dispersion, where the variance of accident counts is higher than the mean. This often occurs in real-world accident data, making Poisson regression less effective in some cases (Lord & Mannering, 2010)

Traditional models, such as regression-based models, have remained in favour due to interpretability, satisfactory theoretical rationale and the ease of application. However, they assume linear relationships and struggle with complex, non-linear interactions between variables. This limitation has led to the emergence of machine learning approaches, which can capture more intricate patterns in accident data (Xie, Zhang, & Liang, 2009).

2.2 Machine Learning Approaches for Accident Prediction

Machine learning (ML) approaches have emerged as powerful alternatives to traditional statistical models for accident prediction. Unlike conventional models, ML techniques can handle nonlinear relationships, high-dimensional datasets, and complex interactions between variables, leading to improved prediction accuracy. ML methods are widely used for accident severity classification, accident frequency modelling, and risk assessment. Machine learning models can be broadly categorized into supervised learning, unsupervised learning, and ensemble learning methods.

Machine learning techniques have been applied in various road safety studies:

1. Crash Frequency Prediction – Using Random Forests and Gradient Boosting to model accident frequency on highways (Abdel-Aty & Pande, 2007).
2. Accident Severity Classification – Using SVM and Neural Networks to classify accident severity (Chang & Chen, 2005).
3. High-Risk Zone Identification – Using K-Means clustering to detect accident-prone areas (Wen et al., 2021).

Machine learning has revolutionized accident prediction by enabling more accurate, flexible, and scalable models than traditional statistical approaches. However, challenges such as interpretability, computational complexity, and data requirements must be addressed to maximize their effectiveness.

3. Materials and Methods

3.1 Mixed Modelling Approach

Accident severity prediction plays a crucial role in traffic safety research, enabling policymakers and transportation engineers to implement data-driven interventions. In this study a traditional statistical model, Negative Binomial Regression (NBR), was used for accident severity modelling due to their interpretability and ability to provide insights into key contributing factors. Traditional models often assume linearity and independence of variables**,** which may not always hold in real-world accident data. Similarly machine learning models often lack interpretability**,** making them less suitable for drawing policy-related conclusions. To address these limitations, this study proposes a mixed modelling approach, integrating traditional statistical models with machine learning techniques for accident severity prediction. The core element of this approach is the Accident Severity Score (ASS), which converts categorical severity outcomes into a continuous dependent variable, allowing for improved modelling flexibility. This paper explores the framework for developing a hybrid model that **balances** interpretability and predictive performance, enabling more effective accident risk assessment.

3.2 Accident Severity Score (ASS)

Traditional accident severity classification relies on discrete categories such as Death, Grievous Injury, and Minor Injury, which can introduce limitations when applying regression-based methods. To overcome this, the Accident Severity Score (ASS) is formulated as a continuous measure by assigning weights to different severity levels:

$$ASS= (W_d \times Death)+(W_g \times Grievous\ Injuries)+(W_m \times Minor\ Injuries)$$

where: W_d, W_g, W_m represent the assigned weights for death, grievous injury, and minor injury, respectively (10 for Death, 5 for Grievous Injury, and 1 for Minor Injury). The weights are calibrated based on historical accident data and expert judgment. Since accident severity often follows a count-based distribution with over dispersion, the Negative Binomial

Regression (NBR) model is used in the study. The NBR model is expressed as:

$$ASS=e^{\beta 0+\sum \beta i Xi}+\epsilon$$

$$\mathbf{ASS_{final}=\alpha_1(Linear\ Regression)+\alpha_2(Random\ Forest)+\alpha_3(XGBoost)}$$

where αi are optimized weights based on cross-validation performance.

4. Data collection and selection of variables

The dataset used in this study is based on road accident records from the different roads in Kollam district for the consecutive 3 years 2017, 2018 & 2019. The data were collected from State Crime Records Bureau Trivandrum. In order to facilitate modelling, an Accident Severity Score (ASS) has been prepared based on casualty data. A systematic selection process was employed to identify key factors influencing accident severity. The dataset consists of independent variables (predictors) categorized into accident characteristics, roadway conditions, environmental factors, vehicle-related attributes, and driver-related information. These variables are explained below:

i. Accident Characteristics consist of locational identifiers based on accident registration details, temporal attributes, accident type, number of fatalities, type of injuries and the persons involved. These characteristics are further are illustrated below:

- Location identifiers contain details of district, zone, range, subdivision, circle, police station (PS)
- Official accident registration details consist of FIR No, date of report and time of report
- Temporal attributes affecting accident frequency include date of accident, time of accident
- Accident Type consists of collision, hit-and-run, rollover, etc.
- Casualty Details include death and the number of fatalities
- Nature of in Injuries consist of severe injuries requiring hospitalization and minor Injuries not requiring extensive medical care
- Involved persons cover driver, passenger, pedestrian, cyclist and other persons.

ii. Roadway Characteristics

- Type of Area – Classification of location as urban, semi-urban, or rural.
- Ongoing road works – Indicator of construction zones.
- City/Town/Village – Classification based on population density and infrastructure.
- Number of Lanes on Road – Indicator of road capacity.
- Presence of Road Divider – Whether the road has a median divider.
- Accident Spot Details – Identification of accident-prone zones.
- Speed Limit – Legal speed restrictions at the accident location.

- Road Surface Condition – Pavement type and maintenance status.
- Junction Type (T-junction, Crossroad, etc.) – Impact of intersections.
- Road Chainage (Kilometer Markers) – Location tracking based on road length.
- Geospatial Data (Latitude & Longitude) – Exact accident coordinates.

iii. Environmental Factors

- Weather Conditions – Clear, rainy, foggy, or extreme weather events.
- Visibility – Impact of lighting and atmospheric conditions.

iv. Vehicle-Related Factors

- Vehicle Type – Two-wheeler, four-wheeler, heavy vehicle, etc.
- Accused/Victim Vehicle Details – Identification of responsible and affected vehicles.
- Vehicle Registration Number & Year of Manufacture – Vehicle age and condition.
- Load Condition & Load Category – Overloading as a contributing factor.
- Mechanical Failure – Brake failure, tire burst, or other mechanical defects.
- Insurance & Fitness Certificate – Compliance with regulatory standards.
- Disposition After Accident – Whether the vehicle was towed, repaired, or scrapped.

v. Driver-Related Factors

- Age, Gender, Occupation, Qualification
- License.
- Traffic Violation History – Past offenses and reckless driving records.
- Safety Device Usage – Helmet or seatbelt compliance.
- Alcohol/Drug Influence – Whether intoxication contributed to the accident.
- Head Injury Indicator – Presence of head trauma in accident victims.

vi. Dependent Variable: Accident Severity Score (ASS)

To facilitate a regression-based approach, a continuous severity score was developed using casualty data. The dataset compiled for Kollam district (2017, 2018, and 2019) provides a comprehensive representation of accident occurrences by incorporating key accident risk factors. The inclusion of Accident Severity Score (ASS) as a dependent variable enables the integration of both traditional statistical models and machine learning approaches to predict accident severity. The variables selected ensure a general approach to accident modelling, covering roadway conditions, vehicle attributes, environmental influences, and driver characteristics

5. Model Building Process

Here an outline is given to the steps involved in constructing accident prediction models using both traditional statistical approaches and machine learning techniques. The model aims to predict accident severity using the Accident Severity Score (ASS) as a dependent variable, derived from the number of fatalities, grievous injuries, and minor injuries. The independent variables include roadway characteristics, vehicle attributes, environmental conditions, and driver-related factors.

5.1 Comparison of Average Accident Severity Score (ASS) Across Independent Variables

A comparative analysis of average ASS across different categories of independent variables to identify significant patterns and high-risk factors was done. This comparison helps to pinpoint the factors contributing to the most severe accidents. It helps in feature selection by identifying which variables have the strongest impact on ASS, thereby improving the reliability of statistical and machine learning models. Here the ASS based on the different independent variables described above were determined and one way ANOVA test or an independent sample t test was conducted to examine to check whether it is statistically significant.

Table 5.1 presents the descriptive statistics and ANOVA results comparing the Accident Severity Score (ASS) among different categories of safety device usage. The mean ASS varies considerably depending on whether individuals were using protective safety equipment at the time of the accident. The results indicate that individuals not wearing a seatbelt (M = 8.05, SD = 9.29) and those not wearing helmet (M = 7.98, SD = 4.58) had the highest accident severity scores. Conversely, individuals wearing a seatbelt (M = 5.76, SD = 3.41) or helmet (M = 5.72, SD = 2.60) had significantly lower accident severity scores. The category labelled "NA" (M = 7.27, SD = 7.12), which likely includes pedestrians, cyclists, or other road users who are not required to use seatbelts or helmets, also exhibited relatively high accident severity. A one-way analysis of variance (ANOVA) was conducted to determine whether the differences in accident severity scores among the safety device usage groups were statistically significant. The ANOVA results reveal a highly significant effect of safety device usage on accident severity, $F(4, 3998) = 35.372$, $p < 0.001$. Given that the p-value is below 0.05, the null hypothesis—stating that safety device usage does not affect accident severity—is rejected. These results confirm that non-usage of safety devices significantly increases accident severity.

Table 5.1 Comparison of Accident Severity Score (ASS) by Use of Safety Devices

Use of Safety Devices	Number of Accidents	Mean ASS	Std. Deviation	Std. Error Mean
NA	822	7.2652	7.11898	.24830
Seat Belt	622	5.7605	3.41006	.13673
Wearing Helmet	1491	5.7163	2.60137	.06737
Without wearing Helmet	541	7.9778	4.57807	.19683
Without wearing seatbelt	517	8.0503	9.28747	.40846
Total	3993	6.6506	5.45545	.08633
ANOVA		F(4,3988)=35.372, p=0.000		

Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017,2018, 2019

Table 5.2 presents the descriptive statistics and independent sample t-test results comparing the Accident Severity Score (ASS) between male and female accident victims. The mean ASS for female victims ($M = 7.75$, $SD = 7.52$) is higher than that for male victims ($M = 6.50$, $SD = 5.09$). This suggests that, on average, accidents involving female victims tend to result in more severe outcomes compared to those involving male victims. The independent sample t-test was conducted to examine whether this difference in mean ASS is statistically significant. The results indicate a statistically significant difference in accident severity between male and female victims ($t(548.197) = 3.549$, $p < .001$). Since the p-value is less than .05, the null hypothesis (which assumes no difference in ASS between genders) is rejected, confirming that gender has a significant impact on accident severity. It is important to note that the equal variance assumption was not met, as indicated by the Levene's test for equality of variances. Therefore, the Welch t-test, which does not assume equal variances, was used for statistical comparison. The findings imply that females may be at a higher risk of severe injuries or fatalities in road accidents compared to males.

Table 5.2 Comparison of Accident Severity Score (ASS) by Gender of Driver

Gender	Number of Accidents	Mean ASS	Std. Deviation	Std. Error Mean
Female	486	7.7469	7.51793	.34102
Male	3507	6.4987	5.08717	.08590
Total	3993	6.6506	5.45545	.08633
Independent Sample t-test*		t (548.197)=3.549, p=0.000		

** Equal Variance not assumed; Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017, 2018, 2019*

Table 5.3 provides an overview of accident outcomes categorized by severity, including fatalities, grievous injuries, minor injuries, and accident types. A vast majority of accidents (86.9%) did not result in any fatalities, while 11.5% led to a single death, and only a small fraction (1.7%) recorded multiple fatalities. Grievous injuries were more common, with 70.3% of accidents involving at least one grievous injury, while 18.2% had none. A smaller percentage (11.7%) of accidents resulted in multiple grievous injuries, with some cases reporting up to six. Similarly, minor injuries followed a similar trend, with 81.6% of accidents not causing any minor injuries, while 14.7% led to one minor injury. A small proportion (3.7%) involved multiple minor injuries. The classification of accident types reveals that 13.1% of accidents were fatal, 78.2% resulted in grievous injuries, 6.6% involved minor injuries, and only 2.1% were non-injury incidents. This distribution highlights the significant impact of road accidents, with a considerable proportion leading to severe or grievous injuries, reinforcing the necessity for stronger road safety measures and enforcement. Similarly the analysis based on all the independent variables was done.

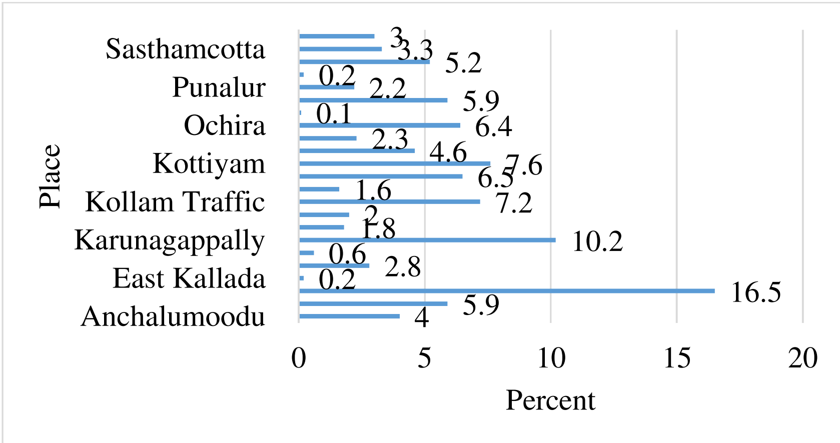
Table 5.3 Distribution of Accident Outcomes by Severity

Severity	Number of Occurrence in an Accident	Frequency	Percent
Death	0	3471	86.9
	1	458	11.5
	2	30	0.8
	3	34	0.9
Grievous	0	725	18.2
	1	2808	70.3
	2	397	9.9
	3	13	0.3
	4	26	0.7
	5	6	0.2
	6	18	0.5
Minor	0	3259	81.6
	1	587	14.7
	2	104	2.6
	3	26	0.7
	4	17	0.4
Accident Type	Fatal	522	13.1
	Grievous Injury	3122	78.2
	Minor Injury	264	6.6
	Non-Injury	85	2.1
Total		3993	100.0

Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2018

Figure 5.1 presents the distribution of road accidents across various locations. Among the listed areas, Chavara recorded the highest number of accidents, accounting for 16.5% of the total. The data indicates that accidents are more concentrated in urban and high-traffic areas, likely due to higher vehicular movement, congestion, and road conditions.

Figure 5.1 Distribution of Accident by Place



Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017, 2018, 2019

5.2 Negative Binomial (NB) Regression

In this study, the dependent variable, the number of deaths per accident, represents a count variable with non-negative integer values (0, 1, 2, 3). To determine the most appropriate regression model, it is crucial to assess the distribution of this variable. The descriptive statistics reveal that the mean number of deaths per accident is 0.155, while the variance is 0.197 (Table 5.4). Since the variance is slightly greater than the mean, this suggests the presence of over dispersion, where the variability in the data exceeds what is expected under a standard Poisson distribution. Given that the variance exceeds the mean, a Negative Binomial Regression (NBR) model is a more suitable alternative. The Negative Binomial model accounts for over dispersion by introducing an additional dispersion parameter, thereby providing more accurate estimates and improving model fit. Based on the observed over dispersion in the data, the Negative Binomial Regression model is the preferred choice for modelling the number of deaths per accident.

Table 5.4 Descriptive Status of Number of Deaths

Descriptive Statistics						
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Death	3993	.0	3.0	.155	.4442	.197

Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017, 2018, 2019

Table 5.5(a) presents the categorical independent variables used in the model. These include gender, safety device usage, area type, presence of roadwork, road divider, road surface type, junction type, road features, visibility, traffic control measures, traffic violations, driver's license status, and weather conditions. The distribution of cases within each category is also reported. Table 5.5(b) displays descriptive statistics for the continuous independent variables: age of the driver, age of the vehicle, number of lanes on the road, speed limit, and load condition of the vehicle. The mean age of drivers involved in accidents is 37.34 years (SD = 14.80), while the average age of the vehicles is 6.23 years (SD = 4.87).

Table 5.5 (a) Negative Binomial Regression Categorical Variable Information

Variables (code)		N	Percent
Gender	Female (1)	188	10.5%
	Male (2)	1602	89.5%
Safety Device	Seat Belt (1)	357	19.9%
	Wearing Helmet (2)	832	46.5%
	Without wearing Helmet (3)	304	17.0%
	Without wearing seatbelt (4)	297	16.6%
Area	City (1)	404	22.6%
	Town (2)	392	21.9%
	Village (3)	994	55.5%
Ongoing Roadwork	No (1)	1725	96.4%
	Yes (2)	65	3.6%
Divider	No (1)	1578	88.2%
	Yes (2)	212	11.8%
Road Surface	Metalled (1)	105	5.9%
	Surfaced (2)	1685	94.1%
Junction	Four arm junction (1)	216	12.1%
	Round about junction (2)	40	2.2%
	Staggered junction (3)	340	19.0%
	T- Junction (4)	931	52.0%
	Y- junction (5)	263	14.7%
Road Features	Bridge (1)	27	1.5%
	Culvert (2)	6	0.3%

	Curved Road (3)	276	15.4%
	Ongoing Road Works/Under Construction (4)	2	0.1%
	Others (5)	29	1.6%
	Straight Road (6)	1450	81.0%
Visibility	Poor (1)	167	9.3%
	Good (2)	1623	90.7%
Traffic Control	Flashing signal/blinker (1)	124	6.9%
	Police controlled (2)	339	18.9%
	Stop sign (3)	191	10.7%
	Traffic light signal (4)	111	6.2%
	Uncontrolled (5)	1025	57.3%
Traffic Violation	Driving on Wrong Side (1)	39	2.2%
	Drunken driving (2)	5	0.3%
	No Violation (3)	870	48.6%
	Over Speeding (4)	871	48.7%
	Use of mobile phone (5)	5	0.3%
Licence	Learner License (1)	5	0.3%
	Valid Permanent License (2)	1754	98.0%
	Without License (3)	31	1.7%
Weather	Cloudy (1)	234	13.1%
	Dust Storm (2)	1	0.1%
	Heavy rain (3)	2	0.1%
	Light rain (4)	92	5.1%
	Mist/Fog (5)	53	3.0%
	Other (6)	211	11.8%
	Sunny/Clear (7)	1131	63.2%
	Very Cold (8)	1	0.1%
	Very Hot (9)	65	3.6%
	Total	1790	100.0%

Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017, 2018, 2019

Table 5.5 (b) Negative Binomial Regression: Continuous Variable Information

		N	Minimum	Maximum	Mean	SD
Dependent Variable	Death	1790	.0	3.0	.140	.3950
Covariate	Age	1790	.0	76.0	37.342	14.7980
	Age of vehicle	1790	.0	33.0	6.227	4.8734
	Lanes Road	1790	1	3	1.71	.507
	Speed Limit	1790	1	4	2.96	.397
	Load Condition	1790	1	6	3.86	.704

Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017, 2018, 2019

The model's goodness-of-fit statistics (Table 5.5(c)) suggest an adequate fit. The deviance (644.943, df = 1,747, Value/df = 0.369) and Pearson Chi-Square (1,906.981, df = 1,747, Value/df = 1.092) indicate that the model appropriately accounts for the variability in the data. Additionally, the likelihood ratio chi-square test ($\chi^2 = 225.281$, df = 42, $p < .001$) confirms that the full model significantly improves over an intercept-only model. The model selection criteria, including the Akaike Information Criterion (AIC = 1,378.370) and Bayesian Information Criterion (BIC = 1,614.438), provide further evidence of model adequacy.

Table 5.5(c) Negative Binomial Regression: Goodness of Fit

	Value	df	Value/df
Deviance	644.943	1747	.369
Scaled Deviance	644.943	1747	
Pearson Chi-Square	1906.981	1747	1.092
Scaled Pearson Chi-Square	1906.981	1747	
Log Likelihood	-646.185		
Akaike's Information Criterion (AIC)	1378.370		
Finite Sample Corrected AIC (AICC)	1380.537		

Bayesian Information Criterion (BIC)	1614.438		
Consistent AIC (CAIC)	1657.438		

Omnibus Test[a]		
Likelihood Ratio Chi-Square	df	Sig.
225.281	42	.000
a. Compares the fitted model against the intercept-only model.		

Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017, 2018, 2019

Table 5.5(d) presents the Wald Chi-Square test results, assessing the significance of each independent variable. The findings indicate that:

Significant Predictors ($p < .05$):

- Safety device use ($\chi^2 = 39.501$, $p < .001$): The type of safety device worn (seatbelt, helmet, or none) significantly impacts the number of deaths.
- Presence of a road divider ($\chi^2 = 5.334$, $p = .021$): Roads with dividers are associated with different fatality outcomes.
- Road surface type ($\chi^2 = 6.594$, $p = .010$): Surfaced roads differ in accident severity compared to metaled roads.
- Junction type ($\chi^2 = 26.071$, $p < .001$): The type of intersection significantly affects accident severity.
- Traffic control measures ($\chi^2 = 29.460$, $p < .001$): The presence of different traffic control mechanisms (e.g., traffic signals, stop signs) influences fatalities.
- Driver's license status ($\chi^2 = 6.199$, $p = .013$): Those driving without a valid license experience different fatality rates.
- Weather conditions ($\chi^2 = 48.959$, $p < .001$): Certain weather conditions (e.g., rain, fog) are associated with increased fatalities.

Non-Significant Predictors ($p > .05$):

- Gender ($\chi^2 = 3.136$, $p = .077$): There is no statistically significant difference in accident fatality rates between males and females.
- Ongoing roadwork ($\chi^2 = 2.585$, $p = .108$): The presence of road construction does not significantly influence fatality rates.

- Visibility ($\chi^2 = 0.750$, p = .386): The impact of poor visibility on accident severity is not statistically significant.
- Traffic violations ($\chi^2 = 1.311$, p = .519): Specific traffic violations such as driving on the wrong side or over-speeding do not significantly predict fatalities in this model.
- Age of driver ($\chi^2 = 1.378$, p = .240) and age of vehicle ($\chi^2 = 0.354$, p = .552) are not significant predictors of accident deaths.

Table 5.5 (d) Negative Binomial Regression: Tests of Model Effects

Source	**Type III**		
	Wald Chi-Square	df	Sig.
(Intercept)	13.762	1	.000
Gender	3.136	1	.077
Safety Device	39.501	3	.000
Area	5.296	2	.071
Road work ongoing	2.585	1	.108
Divider	5.334	1	.021
Road Surface	6.594	1	.010
Junction	26.071	4	.000
Road Features	1.252	3	.740
Visibility	.750	1	.386
Traffic control	29.460	4	.000
Traffic Violation	1.311	2	.519
Licence	6.199	1	.013
Weather	48.959	7	.000
Age	1.378	1	.240
Age of vehicle	.354	1	.552
Lanes Road	.000	1	.993
Speed Limit	3.254	1	.071
Load Condition	.117	1	.732
Dependent Variable: Death			

Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017, 2018, 2019

Parameter Estimates and Interpretation: The parameter estimates (B-values in Table 5.5e) provide how each independent variable affects the dependent variable (i.e., the number of deaths per accident). Key interpretations include:

- Safety Device Use: Compared to individuals without seat belts, those who wore seat belts (B = -0.783, p = .001) and helmets (B = -1.034, p < .001) had significantly lower death rates. Those without helmets did not show a significant difference in fatalities compared to the reference category.
- Junction Type: Staggered junctions (B = -1.240, p < .001) significantly reduce accident fatalities compared to Y-junctions.
- Traffic Control: Accidents occurring at intersections with flashing signals (B = -0.974, p = .006), police-controlled signals (B = -0.587, p = .009), and stop signs (B = -1.429, p < .001) had significantly lower death rates than those at uncontrolled intersections.
- Weather Conditions: Accidents in dust storms (B = 6.783, p = .003) and light rain (B = 3.067, p = .004) were associated with higher fatalities. Other extreme weather conditions (e.g., very cold or heavy rain) were found to have non-significant effects.

Table 5.5 (e) Negative Binomial Regression Parameter Estimates

Parameter	B	Std. Error	Hypothesis Test		
			Wald Chi-Square	df	Sig.
(Intercept)	-27.732	1.5820	307.269	1	.000
[Gender=1]	-.447	.2522	3.136	1	.077
[Gender=2]	0[a]	.	.	.	.
[Safety Device=1]	-.783	.2341	11.192	1	.001
[Safety Device=2]	-1.034	.2011	26.432	1	.000
[Safety Device=3]	-.005	.2082	.001	1	.979
[Safety Device=4]	0[a]	.	.	.	.
[Area=1]	-.039	.2173	.031	1	.859
[Area=2]	.393	.1900	4.274	1	.039
[Area=3]	0[a]	.	.	.	.
[Road work On going=1]	1.227	.7633	2.585	1	.108

[Road work ongoing=2]	0[a]	.	.	.	.
[Divider=1]	.790	.3419	5.334	1	.021
[Divider=2]	0[a]	.	.	.	.
[Road Surface=1]	.821	.3196	6.594	1	.010
[Road Surface=3]	0[a]	.	.	.	.
[Junction=1]	.041	.2951	.019	1	.891
[Junction=2]	.311	.4875	.406	1	.524
[Junction=3]	-1.240	.3436	13.016	1	.000
[Junction=4]	.255	.2188	1.354	1	.245
[Junction=5]	0[a]	.	.	.	.
[Road Features=1]	-.999	1.0643	.880	1	.348
[Road Features=2]	.080	1.3961	.003	1	.954
[Road Features=3]	-.159	.2402	.439	1	.508
[Road Features=4]	-18.121	231874.7271	.000	1	1.000
[Road Features=5]	-22.372	45871.9268	.000	1	1.000
[Road Features=6]	0[a]	.	.	.	.
[Visibility=1]	-.305	.3520	.750	1	.386
[Visibility=2]	0[a]	.	.	.	.
[Traffic control=1]	-.974	.3535	7.597	1	.006
[Traffic control=2]	-.587	.2262	6.735	1	.009
[Traffic control=3]	-1.429	.3424	17.406	1	.000
[Traffic control=4]	-.850	.3514	5.856	1	.016
[Traffic control=5]	0[a]	.	.	.	.
[Traffic Violation=1]	21.997	.5713	1482.751	1	.000
[Traffic Violation=2]	.342	97121.6696	.000	1	1.000
[Traffic Violation=3]	22.175	.1507	21664.792	1	.000
[Traffic Violation=4]	22.333[b]	.	.	.	.
[Traffic Violation=5]	0[a]	.	.	.	.
[Licence=1]	-24.223	130151.9428	.000	1	1.000
[Licence=2]	-1.158	.4651	6.199	1	.013
[Licence=3]	0[a]	.	.	.	.

[Weather=1]	.615	1.0856	.321	1	.571
[Weather=2]	6.783	2.3150	8.584	1	.003
[Weather=3]	-19.882	205827.2519	.000	1	1.000
[Weather=4]	3.067	1.0694	8.225	1	.004
[Weather=5]	3.386	1.0882	9.684	1	.002
[Weather=6]	2.262	1.0470	4.666	1	.031
[Weather=7]	2.091	1.0352	4.081	1	.043
[Weather=8]	4.443	1.7940	6.134	1	.013
[Weather=9]	0[a]	.	.	.	.
Age	.006	.0050	1.378	1	.240
Age of vehicle	-.011	.0179	.354	1	.552
Lanes Road	-.002	.1658	.000	1	.993
Speed Limit	.363	.2014	3.254	1	.071
Load Condition	.043	.1256	.117	1	.732
(Scale)	1[c]				
(Negative binomial)	1[c]				

Dependent Variable: Death

Model: (Intercept), Gender, Safety Device, Area, Road work ongoing, Divider, Road Surface, Junction, Road Features, Visibility, Traffic control, Traffic Violation, Licence, Weather, Age, Age of vehicle, Lanes Road, Speed Limit, Load Condition

a. Set to zero because this parameter is redundant.

b. Hessian matrix singularity is caused by this parameter. The parameter estimate at the last iteration is displayed.

c. Fixed at the displayed value.

Source: Estimated from Kerala Police Department – First Information Reports (FIRs) and accident reports, 2017, 2018, 2019

6. Conclusion

The negative binomial regression analysis confirms that several factors significantly influence the number of deaths per accident. The use of safety devices, traffic control measures, road surface conditions, and weather conditions are among the strongest predictors of fatality rates. However, variables such as gender, visibility, and certain traffic violations do not show significant relationships with accident deaths. These findings emphasize the importance of enforcing seatbelt and helmet laws, improving road infrastructure, and strengthening traffic control measures to reduce fatalities.

This regression model can be used to predict the number of deaths in an accident based on multiple independent variables. The estimated equation indicated that safety device use, area type, road surface, traffic control, speed limit, and weather conditions significantly contributed to variations in fatality risks.

References

1. Chang, L. Y., & Chen, W. C. (2005). Data mining of tree-based models to analyse freeway accident frequency. *Journal of safety research*, *36*(4), 365-375.
2. Abdel-Aty, M., & Pande, A. (2007). Crash data analysis: Collective vs. individual crash level approach. *Journal of safety research*, *38*(5), 581-587.
3. Xie, Y., Zhang, Y., & Liang, F. (2009). Crash injury severity analysis using Bayesian ordered probit models. *Journal of Transportation Engineering*, *135*(1), 18-25.
4. Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation research part A: policy and practice*, *44*(5), 291-305.
5. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
6. Kumar, M., Singhal, S., Shekhar, S., Sharma, B., & Srivastava, G. (2022). Optimized stacking ensemble learning model for breast cancer detection and classification using machine learning. *Sustainability*, *14*(21), 13998.