Exploring Pedestrian Safety on Interstates with Bayesian and Machine Learning Models

Sheikh M. Usman
Graduate Research Assistant
Department of Civil & Environmental Engineering,
The University of Tennessee, TN, 37996, USA
susman1@vols.utk.edu
ORCID: https://orcid.org/0000-0003-1299-5159

Asad J. Khattak
Beaman Professor & Transportation Program Coordinator
Department of Civil & Environmental Engineering,
The University of Tennessee, TN, 37996, USA
akhattak@utk.edu
ORCID: https://orcid.org/0000-0002-0790-7794

Word Count: 6485 words text + 4 tables x 250 words (each) = 7485 words

ABSTRACT

Despite being prohibited from walking on freeways per federal laws, 14 to 17% of all pedestrian crashes in the United States happen on the interstates. Examining these crashes within the context of the safe systems approach is essential with an emphasis on the mitigation of safety risks for all road users. This study investigates the correlates of pedestrian crash injury severity on interstates in North Carolina, focusing on pedestrian actions, roadway conditions, and the type of vehicles involved in the crashes. The study utilizes police-reported pedestrian crash data from 2007 to 2022, coded by the Pedestrian and Bicycle Crash Analysis Tool (PBCAT) which provides unique and comprehensive crash descriptors. The analysis considers 882 pedestrian crash observations on freeways. The dependent variable, pedestrian injury severity, is categorized into distinct binary outcomes, fatal and severe injuries versus minor injuries. The study applies frequentist and Bayesian binary logit models with various prior specifications, along with a robust machine learning algorithm Random Forest for their ability to provide reliable estimates even with a limited sample size. The results show that pedestrian crashes are more predominant in rural (47%) than urban freeways (40%) in North Carolina. Key findings indicate that pedestrians crossing the roadway, crashes on dark unlit roads, and crashes involving pedestrian alcohol impairment are associated with a higher risk of severe injuries to pedestrians. The study findings highlight the need for faster roadside assistance to stranded pedestrians, stringent penalties for alcohol impairment, and the implementation of physical barriers preventing pedestrian access on freeways.

Keywords: Freeway, Access-Controlled Roads, Crossing Freeway, Pedestrian Safety, Machine Learning.

1. INTRODUCTION

Pedestrian safety on interstate highways remains a critical concern in the United States, despite regulations prohibiting pedestrian access to these controlled-access routes (*1*). According to the National Highway Traffic Safety Administration (NHTSA), pedestrian fatalities surged by 53% from 2009 to 2018, constituting 17% of roadway crash fatalities in the United States by 2018 (*2*). This alarming trend is mirrored on interstates and other freeways, which saw a 60% increase in pedestrian fatalities over the same period. From 2015 to 2018, an average of over 800 pedestrians annually lost their lives on these high-speed roadways. The Fatality Analysis and Reporting System (FARS) database indicates that in 2021 alone, 941 pedestrians died on interstates, representing 12.7% of all pedestrian fatalities that year (*3*). These alarming pedestrian fatality crash rates on freeways demand immediate attention and a comprehensive investigation to develop effective pedestrian safety solutions.

Despite the severity of the issue, it has mainly gone underexplored as very few studies have explored pedestrian crashes on high-speed controlled-access roadways. The literature indicates that "unintended pedestrians"—those who exit their vehicles due to incidents or malfunctions—are particularly vulnerable on freeways (*4*). Such individuals face extreme risks due to high vehicle speeds, with speed limits ranging from 55 mph to 85 mph, where the risk of pedestrian fatality increases significantly (*5*). Environmental factors such as darkness and poor lighting further exacerbate these dangers (*6*). It has been found that fatally injured pedestrians on interstates are commonly male, aged 20–39, and often have high blood alcohol concentrations (*7*). The main contributing factors identified in recent analyses include pedestrian and driver alcohol use and dark conditions (*8*). Given the changing characteristics of pedestrian fatalities, such as an increase in urban over rural road fatalities and more incidents occurring in the dark, it is crucial to understand how these factors contribute to pedestrian crash severity.

This study aims to identify critical factors related to pedestrian actions, roadway conditions, and vehicle characteristics that influence pedestrian crash injury severity on freeway pedestrian crashes. The study uses police-reported pedestrian crash data from North Carolina from 2007 to 2022. This research seeks to enhance the understanding of pedestrian safety on freeways by employing a frequentist binary logit model alongside Bayesian binary logit models with varying priors and a Random Forest machine learning algorithm. The study relates to the Safe Systems approach by addressing its safe road users and safe infrastructure components, ultimately aiming to reduce pedestrian fatalities and injuries on freeways. The study's contribution lies in its ability to provide valuable insights into the dynamics of pedestrian crashes on high-speed access-controlled roadways, which remains an underexplored avenue of research. The study findings emphasize the need for comprehensive safety measures and policy implications to protect pedestrians from freeway crashes.

2. LITERATURE REVIEW

Pedestrian-vehicle collisions are complex events influenced by numerous factors, including the age and gender of the pedestrian (*9–12*), lighting conditions (*13–15*), posted speed limits (*16*), built environment characteristics (*17*), spatiotemporal variables (*18*), and driver attributes (*19*). Several studies have focused on pedestrian injury severity concerning the functional classification of roadways, such as local streets, collectors, arterials, and interstates, to understand the impact of different roadway types on pedestrian safety (*20, 21*). A previous study compared pedestrian safety in urban and rural areas in Alabama and found that pedestrians were more likely to sustain severe injuries in crashes on rural county roads (*20*). Arterial roads, with their high traffic volumes and speeds, were found associated with an increased likelihood of pedestrian crashes, particularly at uncontrolled crossings in urban areas (*21*).

Despite the importance of the issue, limited research has specifically examined pedestrian crashes on interstates. An early study analyzed 394 fatal pedestrian crash reports from Texas, Missouri, and North Carolina between 1991 and 1993, revealing that most crashes occurred in dark conditions and involved pedestrians entering or crossing the highway. Additionally, 20% of these incidents were hit-and-run crashes, and 32% involved unintended pedestrians (*4*). Interstate highways were significantly associated with an increased risk of fatal pedestrian injuries in a study that utilized FARS data to examine pedestrian crash injury severity on interstates (*22*). A study using FARS data from 2015 to 2017 to analyze factors contributing to fatal pedestrian crashes on interstates found that pedestrian crashes that occurred in dark, unlit conditions and involved male pedestrians aged 20–44 were more frequent. Additionally, pedestrian alcohol impairment and incidents involving pedestrians crossing the freeway and disabled vehicles were more prevalent (*23*). Dark conditions and hit-and-run incidents emerged as the key correlates of pedestrian crashes on interstates in another study that analyzed pedestrian crashes on interstates in Louisiana from 2010 to 2019 (*24*).

The USDOT has set an ambitious target to eliminate all traffic fatalities and severe injuries under its Vision Zero goal. This target is to be achieved by implementing the Safe Systems Approach, which emphasizes designing a transportation system that accommodates human error and minimizes crash severity (*25*). The Safe Systems Approach focuses on creating forgiving road environments, improving vehicle safety, and promoting safer road user behaviors. It recognizes that while human errors are inevitable, the transportation system should be designed to ensure that these errors do not result in serious harm. Although these studies provide valuable insights, there is a need to thoroughly examine pedestrian safety on interstates focusing on the underlying human, vehicle, and roadway characteristics that influence pedestrian crash injury severity on freeways. This study aims to address this gap by seeking to capture the impact of pedestrian actions, roadway conditions, and vehicle characteristics on pedestrian crashes on freeways and the resulting injury severity. The study findings can be instrumental in informing policy and devising intervention strategies to mitigate pedestrian fatalities and severe injuries in crashes on freeways.

3. CONCEPTUAL FRAMEWORK

The primary objective of this study is to identify and analyze the key human, vehicle, and roadway characteristics that contribute to the severity of pedestrian injuries in crashes occurring on freeways. Utilizing pedestrian crash data from 2007 to 2022 coded by the Pedestrian and Bicycle Crash Analysis Tool (PBCAT), the study seeks to enhance our understanding of pedestrian safety on high-speed access-controlled roadways. Specifically, the study seeks to find answers to the following research questions:

1. How do pedestrian actions and behaviors, such as crossing the freeway and standing or walking along the road, influence the severity of pedestrian injuries in crashes on interstates?

2. What roles do roadway and vehicle characteristics, such as lighting conditions, road surface conditions, and vehicle types, play in determining the severity of pedestrian injuries in interstate crashes?

These research questions focus on identifying which pedestrian actions, roadway, and vehicle-related factors are most strongly associated with fatal and severe pedestrian injuries in interstate crashes. Figure 1 provides a visual illustration of the conceptual framework for the study.

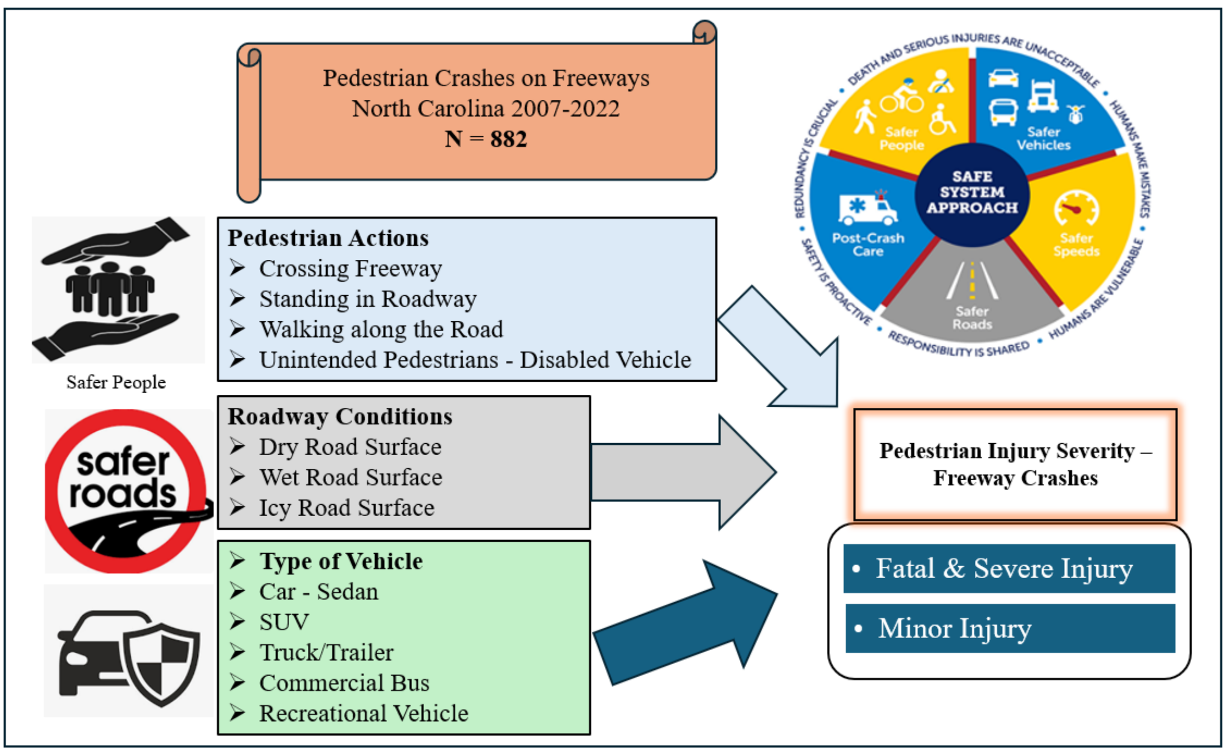


Figure 1. Study Conceptual Framework

4. DATA DESCRIPTION

Pedestrian crash information for this study was obtained from police-reported pedestrian crashes on interstates in North Carolina, meticulously coded through the Pedestrian and Bicycle Crash Analysis Tool (PBCAT) (*26*). While police-reported data can sometimes suffer from issues such as inconsistency or lack of detail, the PBCAT crash typing tool provides detailed crash type descriptors which enhances the completeness of these records, resulting in a unique and comprehensive high-quality pedestrian crash database. This database contains valuable information on pre-crash pedestrian and driver behaviors leading to vehicle-pedestrian crashes. The dataset used in this study spans from 2007 to 2022 and consists of 882 records of pedestrian crashes on interstates. The key variables present in the PBCAT database include pedestrian and driver demographic details (such as age, gender, and race), risky behaviors of drivers and pedestrians involved in pedestrian-vehicle crashes, driver and pedestrian impairment (due to alcohol or drugs), roadway factors (including roadway types/classes, roadway surface condition), lighting conditions, types of surrounding developments, weather conditions, and geocoded locations of crashes. This wealth of information provides critical insights into the contributing factors of pedestrian crashes on interstates. Pedestrian crash injury severity was selected as the dependent variable for the analysis. This variable is reported on a five-level KABCO scale of injury classification in the database, with the distribution as follows: 35.94% fatal injury, 19.16% severe injury, 25.28% minor injury, 16.21% possible injury, and 3.40% no injury crashes. Given the higher importance of fatal and severe injury pedestrian crashes on freeways, the dependent variable was consolidated into two categories: "Fatal and Severe Injury" (combining the fatal and severe injury classes) and "Minor Injury" (combining the minor, possible, and no injury classes).

Figure 2 presents the spatial distribution of the crashes. Figure 2 indicates that most pedestrian crashes occurred on interstates connecting Charlotte, Raleigh, Fayetteville, Greensboro, and Durham. The distribution of the dependent variable and the key explanatory variables is presented in Table 1.

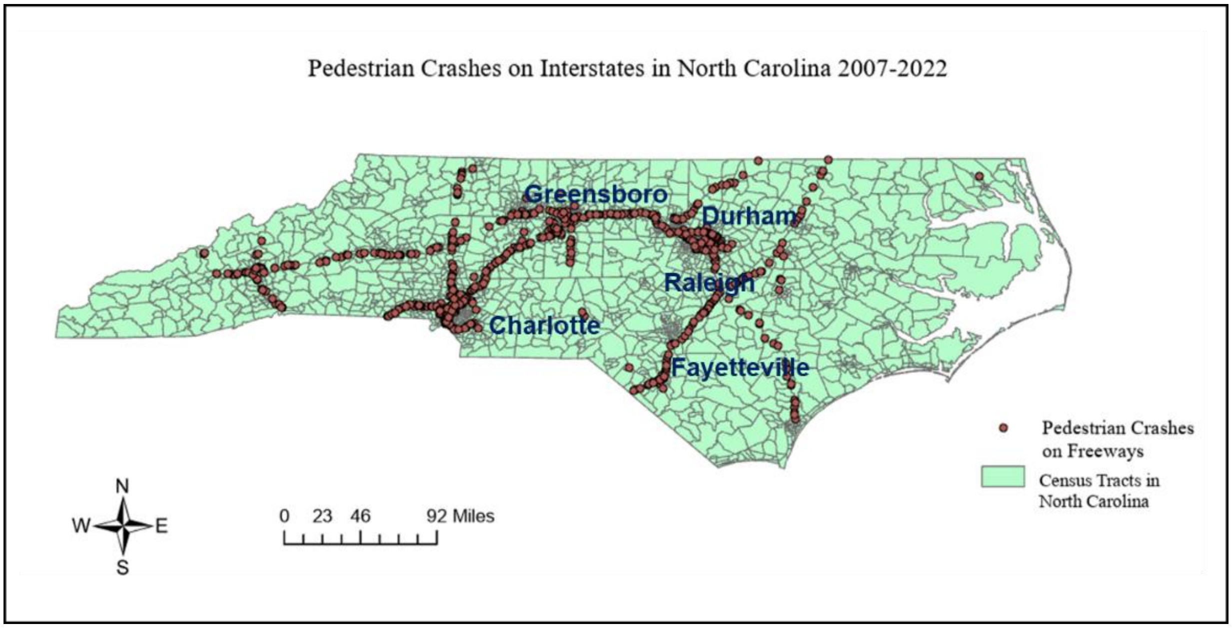


Figure 2. Spatial Distribution of Pedestrian Crashes on Interstates in North Carolina

TABLE 1. Descriptive Statistics of Key Variables (N = 882)

Variables	Categories	Frequency	Percentage
Pedestrian Injury Severity	Fatal & Severe Injury	486	55.10
(Dependent Variable)	Minor Injury	396	44.89
Pedestrian Gender	Male Pedestrian	691	78.34
	Female Pedestrian	191	21.66
Pedestrian Age	Pedestrian Age below 20	47	5.33
	Pedestrian Age between 20-60	750	85.03
	Pedestrian Aged above 60	85	9.64
Pedestrian Alcohol Impairment	Yes	209	23.70
	No	673	76.30
Driver Gender	Male Driver	670	75.96
	Female Driver	212	24.04
Driver Alcohol Impairment	Yes	49	5.56
	No	833	94.44
Pedestrian Actions/Crash Types	Crossing Freeway	248	28.12
	Lying in Roadway	20	2.27
	Disabled Vehicle Related	306	34.69
	Vehicle Loss of Control	56	6.35
	Ped Dash/ Dart	19	2.15
	Pedestrian Loss of Control	29	3.29
	Standing in Road	44	4.99
	Walking along the road in the direction of traffic	96	10.88
	Walking along the road against traffic	31	3.51
Light Conditions	Daylight	307	34.81
	Dark with Road Lights	141	15.99
	Dark without Road Lights	434	49.21
Vehicle Type	Commercial Bus	16	1.81
	Recreational Vehicle	53	6.01
	Passenger Car	414	46.94
	SUV	139	15.76
	Truck/ Trailer	260	29.48
Roadway Surface Condition	Dry Road Surface	697	79.02
	Wet Road Surface	146	16.55
	Icy Road Surface	39	4.42
Type of Freeway	Rural	412	46.71
	Urban	357	40.47
	Mixed	113	12.81
Weather Conditions	Clear Weather	636	72.11
	Cloudy Weather	128	14.51
	Foggy Weather	8	0.91
	Rain	90	10.20
	Snow	20	2.27
Posted Speed Limit	50-55 mph	107	12.13

	60-75 mph	775	87.87

All the crash-related variables used for analysis in this study were categorical. Hence, the reported descriptive statistics in Table 1 include the crash frequencies and percentages associated with the categories of the categorical variables. Dummy variables (indicator variables with binary values 0 and 1) were created for all categorical variables in the dataset. Of the 882 crashes, 486 (55.10%) resulted in fatal or severe injuries, while 396 (44.89%) resulted in minor injuries. Most pedestrians involved in crashes were male (78.34%), with female pedestrians accounting for 21.66%. Most pedestrians were aged between 20 and 60 years (85.03%). Pedestrians below 20 years of age accounted for 5.33%, while 9.64% of pedestrians were above 60. 23.70% of pedestrians involved in crashes were alcohol-impaired whereas 5.56% of drivers were found to be alcohol-impaired. Male drivers were involved in 75.96% of the crashes, while female drivers were involved in 24.04%. Regarding pedestrian actions or crash patterns on freeways, crossing the freeway was common, occurring in 28.12% of crashes. Pedestrians walking along the road in the direction of traffic (10.88%), standing in the roadway (4.99%), and lying in the roadway (2.27%) were other notable pre-crash actions. 34.69% of crashes involved disabled vehicles with unintended pedestrians on the freeways. Nearly half of the crashes occurred in dark conditions without road lights (49.21%), followed by daylight (34.81%) and dark conditions with road lights (15.99%). Passenger cars were involved in 46.94% of crashes, followed by trucks/trailers (29.48%) and SUVs (15.76%). Recreational vehicles (6.01%) and commercial buses (1.81%) were less commonly involved. Most crashes occurred on dry road surfaces (79.02%) followed by wet (16.55%) and icy road surfaces (4.42%). Most crashes occurred in clear weather (72.11%) followed by cloudy weather (14.51%), and rainy weather (10.20%). Crashes in snowy conditions (2.27%) and foggy weather (0.91%) were less frequent.

5. METHODOLOGY

Given the binary nature of the response variable, pedestrian injury severity, a binary logit model was employed for the analysis. Additionally, due to the limited sample size of pedestrian crash data (N = 882), Bayesian inference-based models with various prior specifications were estimated to enhance the robustness of the results. The Bayesian approach combines prior knowledge with observed data (likelihood function) to create posterior distributions of estimable parameters, which are then used to draw inferences. This approach offers several advantages over the conventional frequentist methods, including the ability to estimate discrete outcome models with smaller sample sizes, the incorporation of hierarchical structures in the modeling framework, and the provision of deeper insights through parameter distributions and credible intervals (*27*). In addition to the statistical parametric models, the study applied a supervised machine learning algorithm, Random Forest, to predict pedestrian injury severity outcomes given the crashes. A description of the applied methods and their performance evaluation criteria is given in the subsections below.

5.1 Binary Logit Regression

The binary logit model is suitable for modeling the outcome of a binary dependent variable based on one or more predictor variables (*28*). The model can be expressed as follows:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \quad (1)$$

where:

P(Y=1) is the probability of fatal and severe pedestrian injury given a crash on the freeway.

β_o is the intercept term.

$\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients for the predictor variables $X_1, X_2, \ldots, X_k$.

The probability of fatal and severe pedestrian injury in a freeway crash can be derived from the logistic function as follows:

$$P(Y=1) = \frac{1}{1+e^{-(\beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}} \quad (2)$$

The predictor variables $X_1, X_2, \ldots, X_k$ represent pedestrian actions, roadway, and vehicle characteristics.

5.2 Bayesian Binary Logit Models

Bayesian models incorporate prior knowledge or beliefs into the analysis through prior distributions, combined with the likelihood of the observed data to produce posterior distributions for the model parameters. The Bayesian posterior distributions are derived from both the likelihood and the prior distributions. It is crucial to specify prior distributions for the key crash-specific regression parameters. Three types of prior distributions can be used (*29*):

Strong Informative Priors: These are based on expert knowledge or previous literature and provide strong guidance to the posterior distribution.

Weak Informative Priors: These do not significantly influence the posterior distribution but help prevent inappropriate inferences.

Non-informative (Vague) Priors: These allow the data (likelihood) to primarily influence the results, providing a more flexible probabilistic interpretation.

The Bayesian approach in this study employs different prior specifications to estimate the models. These specifications include a) non-informative or flat prior that provides minimal prior information, allowing the data to predominantly influence the posterior distribution b) weak normal prior with a mean of 0 and a large variance of 1000 c) strong normal prior with a mean of

0 and a small variance of 10. These different prior specifications allow for a comprehensive assessment of the robustness and sensitivity of the model results to prior assumptions.

4.3 Random Forest Classifier

A random forest consists of multiple independent basic classifiers known as decision trees. Each classifier operates on its own when presented with a test sample, and the category label of the sample is determined by aggregating the voting results from each classification. The following steps outline the process of building a random forest classifier in the context of this study:

1. Variable Selection: An appropriate value for the variable "M," representing the number of features in each feature subset is selected. In this study, the variable "M" represents the number of features selected from the dataset, such as pedestrian actions, roadway conditions, and vehicle characteristics.

2. Random Feature Subsets: The algorithm randomly selects a new feature subset "θ_k" from the entire set of features, according to the chosen "M" value. Each subset "θ_k" is independent of the other subsets in the sequence θ_1, θ_2, ..., θ_k.

3. Decision Tree Construction: The training dataset is created using the selected feature subset and a decision tree for each group in the training set is estimated. Each classifier can be represented as $f(X, \theta_k)$, where X represents the explanatory variables i.e., pedestrian, roadway, and vehicle attributes.

4. Iteration: The iterative procedure is repeated by selecting a new feature subset "θ_k" and training the data with the feature subset until all the feature subsets have been traversed. This completes the construction of a random forest classifier.

5. Classification: Now using the test dataset, each observation is classified based on the aggregated voting results from each classification.

Incorporating random operations in the random forest greatly improves its classification performance (*30–34*). Its ability to handle complex, non-linear relationships and interactions among predictor variables makes it an effective complement to traditional statistical models.

5.4 Performance Measures

Four performance measures are used to evaluate the predictive performance of the statistical and ML models using the test dataset as a holdout sample in this study. These measures include accuracy, sensitivity, specificity, and balanced accuracy.

5.4.1 Accuracy

Classification accuracy is the ratio of correctly classified observations by the model to the total number of observations in the dataset. It can be calculated using **Equation 3**.

$$Accuracy = \frac{Number\ of\ correctly\ classified\ observations}{Total\ number\ of\ observations} \tag{3}$$

5.4.2 Sensitivity

Sensitivity for a specific class of the response variable is the ratio of the correctly classified observations in a class to the sum of the correctly and incorrectly classified observations related to the target class (True positives + False negatives). Sensitivity is obtained using **Equation 4**.

$$Sensitivity\ = \frac{Number\ of\ correctly\ classified\ observations\ in\ a\ class\ (True\ positives)}{True\ positives + False\ negatives} \tag{4}$$

5.4.3 Specificity

Specificity for a specific class of the response variable is the ratio of the correctly classified observations in all other classes (True negatives) to the sum of the True negatives and incorrectly classified observations in the target class (False positives). Specificity is obtained from **Equation 5**.

$$Specificity\ = \frac{Number\ of\ correctly\ classified\ observations\ in\ all\ other\ classes\ (True\ negatives)}{True\ negatives + False\ positives} \tag{5}$$

5.4.4 Balanced Accuracy

Balanced Accuracy considers both sensitivity and specificity in the model's performance evaluation and provides a balanced overall model performance measure. Balanced Accuracy is calculated using **Equation 6**.

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \tag{6}$$

6. RESULTS

6.1 Inference-based Results

6.1.1 Frequentist Binary Logit Model

The results of the frequentist binary logit model, estimated to examine the factors influencing pedestrian injury severity on interstates, are summarized in Table 2. The model was developed using the training dataset, which consisted of 80% of the total observations (N_{train} = 715), while the remaining 20% was reserved for testing (N_{test} = 167) to determine the predictive performance of the models. The explanatory variables included in the final model were found statistically significant at a 95% and higher confidence level. The model results indicate that pedestrians crossing the freeway were found more likely to sustain fatal and severe injuries, as indicated by the positive coefficient of 0.498. The marginal effect for this variable indicates that the probability of fatal and severe injury to pedestrians increases by 0.113 in crashes involving pedestrians crossing the freeway, keeping all other factors constant. Pedestrians walking along the road in the direction of traffic were found to have a higher probability of fatal and severe injuries, with a positive coefficient of 0.575. The marginal effect for this variable indicates that the probability of fatal and severe injury to pedestrians increases by 0.119 in crashes involving pedestrians walking along the road in the direction of traffic, keeping all other factors constant. Pedestrians standing in the roadway were found more likely to undergo fatal and severe injuries, with a coefficient of 0.875. The marginal effect for this variable indicates that the probability of fatal and severe injury to pedestrians increases by 0.182 in crashes involving pedestrians standing on the roadway, keeping all other factors constant. Alcohol impairment in pedestrians was a significant predictor of fatal and severe injury outcomes, with a coefficient of 0.919. Pedestrian crashes on interstates involving disabled vehicles on the roadway were found to be associated with a higher risk of fatal and severe injuries, with a coefficient of 0.567. Crashes involving male drivers were found more likely to result in fatal and severe pedestrian injuries, with a coefficient of 0.557. Referring to the roadway conditions, poor lighting conditions on roads were significantly associated with an increased likelihood of fatal and severe pedestrian injuries given a crash, with a coefficient of 0.694. The presence of icy road surfaces was found to significantly increase the likelihood of fatal and severe pedestrian injuries, with a coefficient of 1.04. Regarding vehicle types, crashes involving SUVs were found more likely to result in fatal and severe pedestrian injuries, with a coefficient of 0.467.

The model goodness of fit statistics indicate a reasonable model fit, with a log-likelihood at convergence value of -386.78 and a Pseudo R-squared value of 0.21, as suggested by the literature (*35*–*38*). The model is overall statistically significant (Prob > $\chi^2(10)$ = 0.0000), and the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values of the model are 793.56 and 839.28, respectively.

TABLE 2. Results of the Frequentist Binary Logit Model for Train Data

Variables	Coefficient	t-stat	p-value	Marginal Effect
Constant	-0.15	-2.620	0.009	------
Pedestrian Actions/Crash Types				
Crossing Freeway indicator	0.498	2.94	0.003	0.103
Walking along the Road in the Traffic Direction indicator	0.575	3.09	0.002	0.119
Standing in the Roadway indicator	0.875	2.12	0.034	0.182
Disabled Vehicle-Related Crash indicator	0.567	3.26	0.001	0.118
Pedestrian Alcohol Impairment Indicator	0.919	4.38	0.000	0.191
Roadway Conditions				
Dark Road Not Lighted	0.694	4.12	0.000	0.144
Icy Road	1.04	2.41	0.016	0.217
Vehicle Type				
SUV	0.467	2.03	0.042	0.097
Driver Sociodemographic				
Male Driver	0.557	2.85	0.004	0.116
Summary Statistics				
Number of Observations			715	
Log Likelihood at Convergence			-386.783	
Log Likelihood at zero			-491.227	
Pseudo R square			0.21	
LR $\chi^2(9)$			208.88	
Prob (> χ^2)			0.000	
AIC			793.56	
BIC			839.28	

Note: LR = Likelihood Ratio, AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria

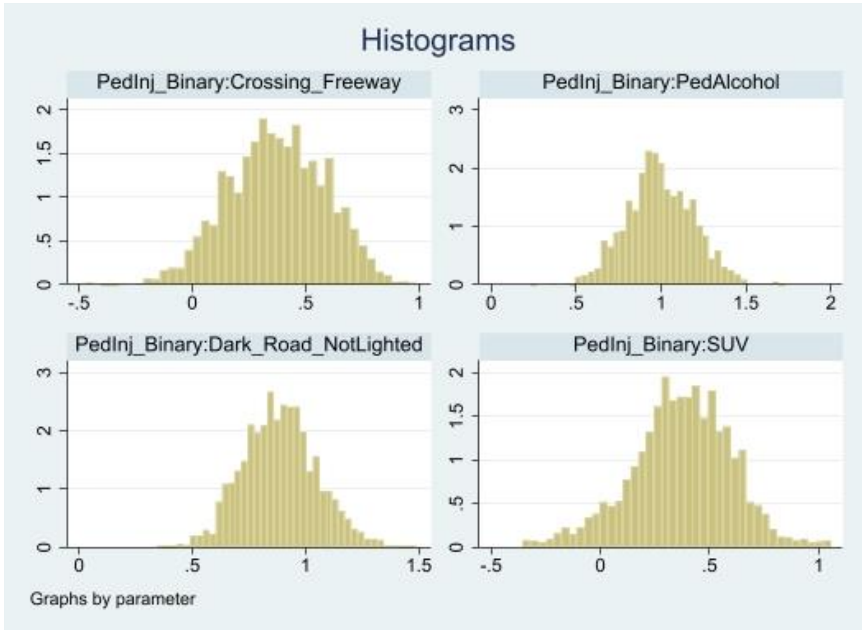
6.1.2 Results of Bayesian Models with Different Prior Specifications

Collectively the Bayesian models can be designated as binary logit models with flat or uniform prior (Model 1), with weak normal prior ~N(0, 1000) (Model 2), and with relatively strong normal prior ~N(0, 10) (Model 3). Table 3 summarizes the results of these models. The posterior means for pedestrians crossing the freeway are 0.37 (Model 1), 0.36 (Model 2), and 0.32 (Model 3). The risk associated with crossing freeways is consistently significant across all models, with a slight increase in the estimate when using flat priors. Model 1 indicates that the odds of fatal and severe injury to pedestrians are 1.45 (exp 0.37) times higher with pedestrians crossing the freeway. The posterior means for walking along the road in the direction of traffic are 0.53 (Model 1), 0.53 (Model 2), and 0.56 (Model 3). The estimates are consistent, with Model 3 showing a slightly higher risk of fatal and severe injury to pedestrians (odds ratio = 1.75). The posterior means for standing in the roadway are 0.89 (Model 1), 0.90 (Model 2), and 0.83 (Model 3). This variable consistently shows a higher risk for pedestrians. Model 2 provides the highest posterior mean estimate for the indicator and implies that the odds of fatal and severe injury to pedestrians are 2.46 (exp 0.90) times higher with pedestrians standing in the roadway.

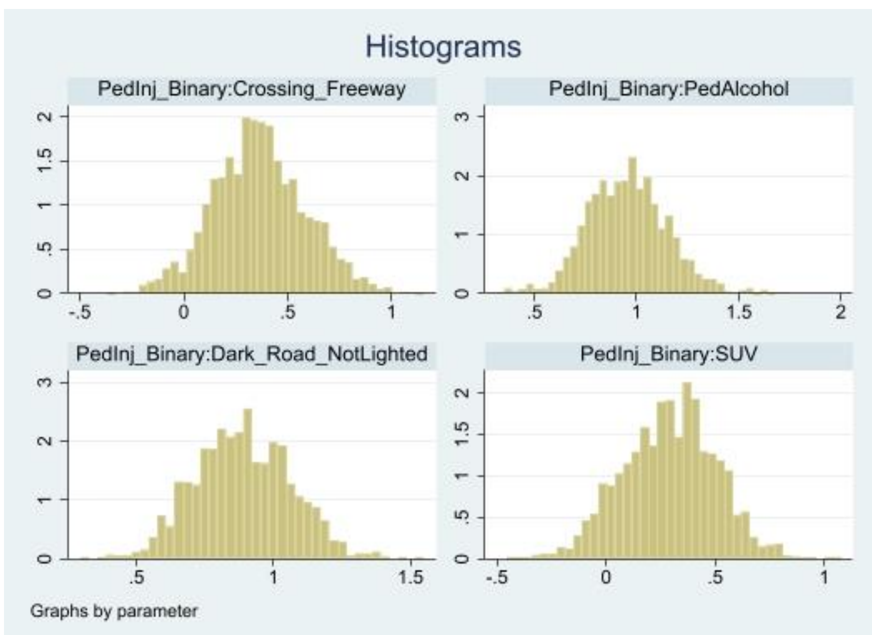
Referring to the roadway characteristics, the posterior means for dark roads without lights indicator are positive in all models, indicative of an increased likelihood of fatal and severe injuries to pedestrians given a crash. Model 1 shows the highest posterior mean (0.89) which relates to an odds ratio of 2.44 (exp 0.89). The posterior means for icy road conditions are 1.15 (Model 1), 1.06 (Model 2), and 1.10 (Model 3). All models consistently show a significant increase in the likelihood of fatal and severe injuries under icy conditions, with Model 1 showing the highest posterior mean estimate (1.15) which implies an odds ratio of 3.16.

Similarly, for the vehicle characteristics, the posterior means of the SUV indicator are positive in all models, showing a higher likelihood of resulting in fatal and severe pedestrian injuries given a freeway crash. The posterior means for the SUV involvement indicator are 0.37 (Model 1), 0.29 (Model 2), and 0.36 (Model 3). Model 1 gives the highest estimate of the posterior mean (0.37) indicating that the odds of fatal and severe pedestrian injury increase 1.45 (exp 0.37) times in pedestrian crashes on interstates involving SUVs.

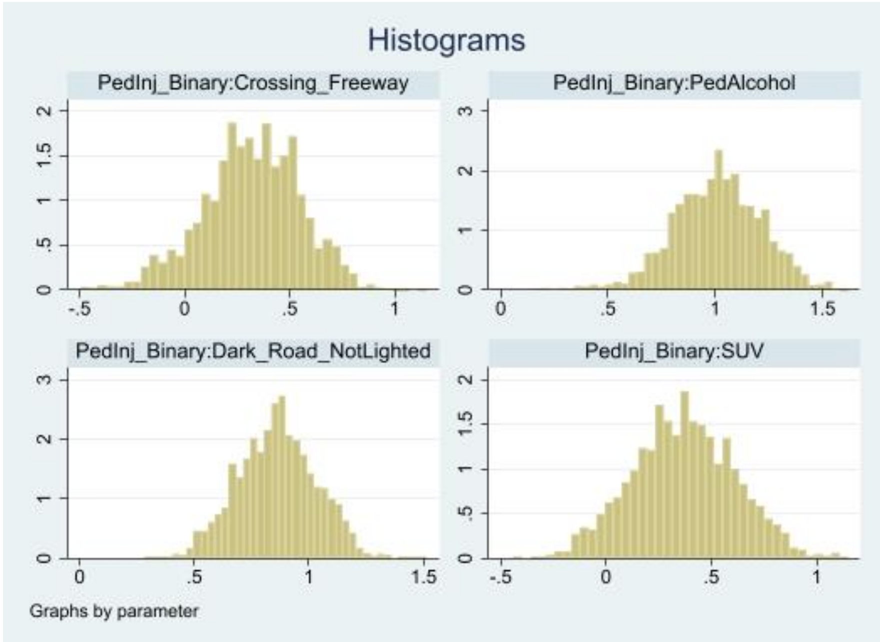
The log marginal likelihood values are -439.42 (Model 1), -483.58 (Model 2), and -461.05 (Model 3). These values indicate the overall fit of the models, with Model 1 having the highest log marginal likelihood. The Deviance Information Criteria (DIC) values are 883.99 (Model 1), 883.29 (Model 2), and 881.19 (Model 3). Lower DIC values indicate a better model fit. Model 3 has the lowest DIC, indicative of the best-fitting model among the three. The posterior distributions of key variables in each of the Bayesian models are presented in Figure 3 (a-c).


Histograms
PedInj_Binary:Crossing_Freeway
PedInj_Binary:PedAlcohol
PedInj_Binary:Dark_Road_NotLighted
PedInj_Binary:SUV
Graphs by parameter

(a)


Histograms
PedInj_Binary:Crossing_Freeway
PedInj_Binary:PedAlcohol
PedInj_Binary:Dark_Road_NotLighted
PedInj_Binary:SUV
Graphs by parameter

(b)



(c)

Figure 3 (a-c). Posterior Distributions of Key Variables in Bayesian Binary Logit Models with a) Flat priors b) Weak normal priors ~N (0,1000) c) Strong normal priors ~N(0, 10)

TABLE 3. Results of Bayesian Binary Logit Models with Different Prior Specifications

Variables	Model 1			Model 2			Model 3		
	Bayesian Model with Flat Priors			**Bayesian Model with Weak Normal Prior ~N(0, 1000)**			**Bayesian Model with Strong Normal Prior ~N(0, 10)**		
	Mean	MCSE	Median	Mean	MCSE	Median	Mean	MCSE	Median
Constant	-0.001	0.06	-0.03	-0.04	0.04	-0.04	-0.11	0.05	-0.10
Icy Road	1.15	0.07	1.14	1.06	0.07	1.03	1.10	0.06	1.09
Crossing Freeway	0.37	0.02	0.37	0.36	0.04	0.36	0.32	0.02	0.32
Pedestrian Alcohol	0.99	0.02	0.98	0.96	0.02	0.95	1.01	0.035	1.013
Walking along the Road in Traffic Direction	0.53	0.02	0.53	0.53	0.02	0.54	0.56	0.01	0.56
Disabled Vehicle Related	0.60	0.03	0.59	0.62	0.02	0.62	0.57	0.02	0.57
Rural Freeway	0.47	0.01	0.47	0.47	0.01	0.47	0.45	0.01	0.45
Male Driver	0.67	0.04	0.66	0.62	0.03	0.62	0.61	0.03	0.61
Dark Road Not Lighted	0.89	0.01	0.88	0.88	0.03	0.88	0.86	0.01	0.86
SUV	0.37	0.04	0.38	0.29	0.02	0.30	0.36	0.026	0.37
Standing in Roadway	0.89	0.05	0.90	0.90	0.07	0.91	0.83	0.06	0.81
Summary Statistics									
Log Marginal Likelihood	-446.42			-494.45			-468.27		
DIC	894.64			894.88			893.61		

Note. MCSE: Monte Carlo Standard Error

6.2 Results of Random Forest

In addition to the frequentist and Bayesian statistical parametric models, a supervised machine learning algorithm random forest classifier was also applied. Like the other models applied in the study, the random forest was trained on 80% of the data (N_{train} = 715) and tested on the remaining 20% (N_{test} = 167). An extended grid search was performed to determine the optimal hyperparameters of the model with 10-fold cross-validation. The 10-fold cross-validation process involves dividing the available data into ten equal subsets. Nine of these subsets are used for training the model, while one subset is kept for testing to evaluate the model's predictive accuracy. This process is carried out recursively ten times, with each subset serving as the testing data once. The final estimation is obtained by averaging the results from all ten iterations. Hyperparameters of the random forest include the number of decision trees, the number of variables considered for splitting the trees' nodes "m_{try}", minimum node size (also called tree depth), and a fraction of the sample used for fitting the model. A range of parameter values for the number of trees (200, 400, 500, 600), "m_{try}" (3, 6, 8), minimum node size (1, 5, 10), and sample fraction (0.6, 0.7, 0.8), resulting in 108 different combinations, were considered in the grid search. The optimal hyperparameters obtained through the grid search consist of 200 trees, 6 variables for splitting the node, a tree depth of 5, and a sample fraction of 0.6, resulting in a minimum cross-validation error of 0.580. Figure 4 shows the results of the top ten combinations of hyperparameters ranked according to increasing cross-validation error.

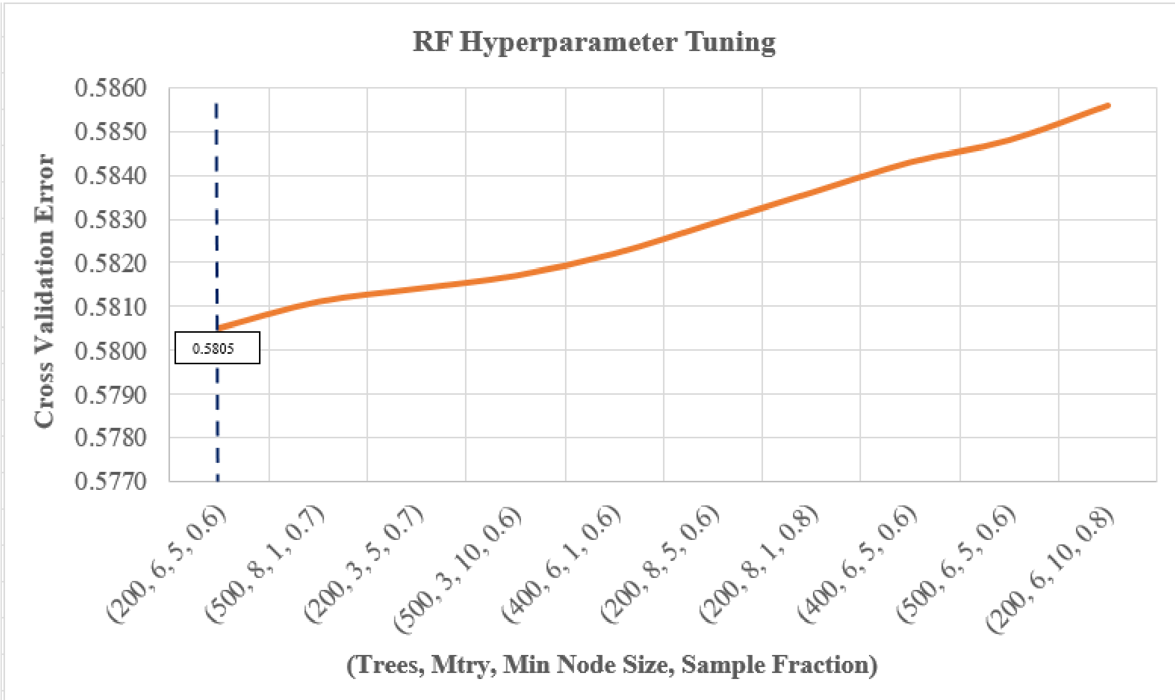


Figure 4. Selection of Optimal Hyperparameters for Random Forest Classifier

6.3 Prediction Results

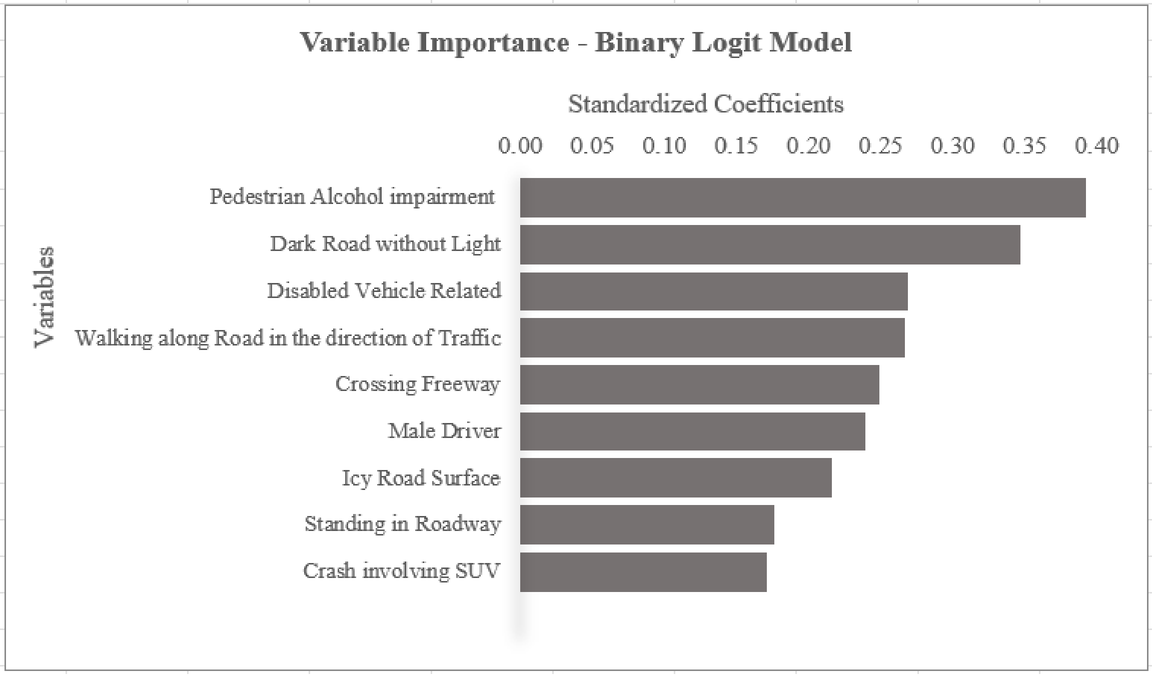
Predictions on pedestrian crash injury severity using the test data (Ntest = 167) were obtained from the estimated models on the training dataset. Confusion matrices, widely used for evaluating the models' predictive performance, were developed. The frequentist binary logit model and the Bayesian model with flat prior (Model 1) yielded a similar prediction accuracy of 62.87% for the test dataset. Models 2 and 3 yielded slightly better predictive accuracy of 64.67%. The random forest model yielded an out-of-sample prediction accuracy of 81.43% for the test data. The model achieved the highest balanced accuracy of 81.25% as opposed to the highest balanced accuracy (80.35%) achieved by the Bayesian model with a strong normal prior. The enhanced prediction and balanced accuracy for the random forest model provide evidence of the superior predictive performance of ML models to statistical parametric models. Table 4 provides the predictive performances of all models on the test dataset.

TABLE 4. Confusion Matrices of the Models for Test Dataset

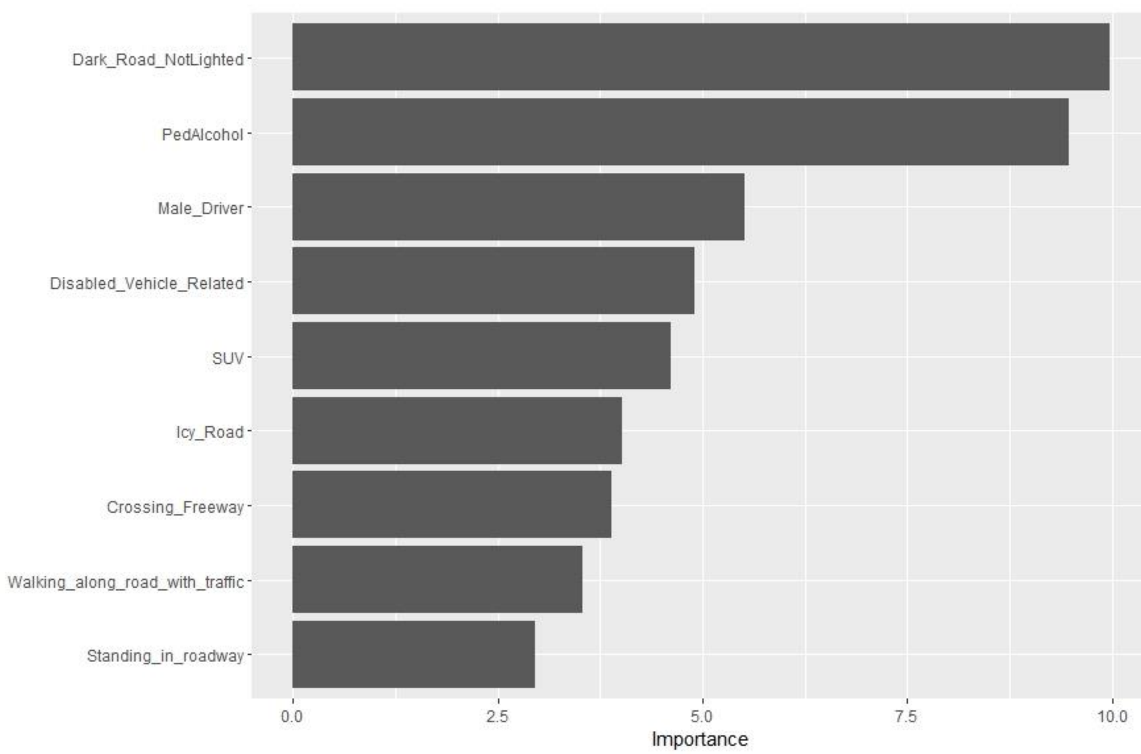
	Frequentist Binary Logit Model		
Observed Outcomes	**Predicted Outcomes**		
	Fatal & Severe Injury	Minor Injury	**Total**
Fatal & Severe Injury	57	32	89
Minor Injury	30	48	78
Total	87	80	167
Performance Measures			
Accuracy	62.87%		
Sensitivity	0.6404		
Specificity	0.6154		
Balanced Accuracy	62.79%		
	Bayesian Model with Flat Prior		
	Fatal & Severe Injury	Minor Injury	**Total**
Fatal & Severe Injury	59	30	89
Minor Injury	32	46	78
Total	91	76	167
Performance Measures			
Accuracy	62.87%		
Sensitivity	0.6629		
Specificity	0.5897		
Balanced Accuracy	62.63%		
	Bayesian Model with Weak Normal Prior ~N(0, 1000)		
	Fatal & Severe Injury	Minor Injury	**Total**
Fatal & Severe Injury	65	24	89
Minor Injury	35	43	78
Total	100	67	167
Performance Measures			
Accuracy	64.67%		
Sensitivity	0.7303		
Specificity	0.5513		
Balanced Accuracy	64.08%		
	Bayesian Model with Strong Normal Prior ~N(0, 10)		
	Fatal & Severe Injury	Minor Injury	**Total**
Fatal & Severe Injury	62	27	89
Minor Injury	32	46	78
Total	94	73	167
Performance Measures			
Accuracy	64.67%		
Sensitivity	0.6966		

Specificity		0.5897	
Balanced Accuracy		64.31%	
		Random Forest Classifier	
	Fatal & Severe Injury	Minor Injury	**Total**
Fatal & Severe Injury	75	14	89
Minor Injury	17	61	78
Total	92	75	167
Performance Measures			
Accuracy		81.43%	
Sensitivity		0.843	
Specificity		0.782	
Balanced Accuracy		81.25%	

The variable importance plots presented in Figure 5 (a, b) rank the most important predictors of pedestrian crash injury severity in freeway crashes yielded by the frequentist binary logit and the random forest model. The variable importance for the binary logit model was determined by obtaining the standardized coefficients of the predictors. The standardized coefficients represent the impact on the probability of the response variable when the explanatory variables change one standard deviation. The variable importance plot for the binary logit model shows that pedestrian alcohol impairment was the most important variable with the highest standardized coefficient of 0.393 followed by freeway pedestrian crashes in dark conditions without road lights. For the variable importance plot of random forest, the indicator of pedestrian crashes in dark conditions without road lights ("Dark_Road_NotLighted") was found the most important predictor followed by pedestrian alcohol impairment ("PedAlcohol"), and crashes involving male drivers ("Male_Driver"). Among the pedestrian actions, the indicators of pedestrians crossing the roadway, walking along the road in the direction of traffic, and standing in the roadway were the most important variables. The findings of the ML model highly complemented the results of the statistical model as all the important variables in the RF model were also found statistically significant in the frequentist and Bayesian binary logit models.



(a)



(b)

Figure 5 (a-b). Variable Importance Plots for a) Frequentist Binary Logit Model b) Random Forest

7. DISCUSSION

The analysis focused on the role of key correlates of pedestrian crash injury severity on interstates including pedestrian actions, roadway conditions, and vehicle characteristics, within the context of the safe systems approach. This approach emphasizes designing a transportation system that accommodates human error and minimizes crash severity. Understanding why pedestrians engage in risky behaviors and identifying where these behaviors occur is essential to developing effective safety interventions. Two research questions were formulated and highlighted in the earlier sections.

To find answers to the first research question which dealt with the influence of pedestrian actions and behaviors on crash injury severity on interstates, the study found that pedestrians crossing freeways, walking along the road in the direction of traffic, standing in the roadway, and alcohol-impaired face a significantly higher risk of fatal and severe injuries, given a crash. This finding aligns with the inherent dangers of crossing high-speed roadways where drivers do not expect pedestrian presence (*4*) and reflects the dangers of walking close to high-speed traffic without a barrier (*39*). Regarding standing on the roadway and pedestrian alcohol impairment, previous studies show that stationary pedestrians are highly vulnerable to being struck by vehicles (*23*) while pedestrians' judgment, coordination, and reaction times are impaired on the consumption of alcohol, leading to poor decision-making, misjudging vehicle speeds, and difficulty in avoiding oncoming traffic (*40*). Countermeasures such as enhancing pedestrian fencing to prevent freeway crossings, providing safe waiting areas off the road, and increasing public awareness about the dangers of walking on the road while intoxicated can be effective. Crashes involving disabled vehicles were associated with higher pedestrian injury severity. These findings are consistent with studies indicating that stranded pedestrians on freeways due to disabled vehicles are at high risk due to their unexpected presence on the roadway (*23*). Providing emergency services and roadside assistance more quickly and efficiently, along with better signage and lighting, can potentially reduce these risks.

As an answer to the second research question, the study found that pedestrian crashes involving SUVs and on icy road surfaces and dark unlit freeway sections were significantly associated with higher pedestrian crash injury severity. These conditions contribute to reduced visibility, longer stopping distances, and more severe impacts during collisions. The finding related to crashes on dark roads aligns with findings from previous studies which show that improved roadway lighting enhances pedestrian visibility and reduces the risk of severe injuries (*41–44*). Installing more roadway lights, enhancing vehicle headlights, encouraging the use of high beam assist technology, and promoting the use of reflective clothing for pedestrians can improve pedestrian visibility and safety (*23*).

8. LIMITATIONS

Several limitations should be acknowledged. The findings of this study are based on data from pedestrian crashes on interstates in North Carolina, which may limit the generalizability of the results. Different regions may exhibit considerable differences in driver and pedestrian behaviors, roadway conditions, traffic regulations, and enforcement practices. As such, the results may not be directly applicable to other states or countries. Further research is needed to verify if similar patterns hold in different geographic and cultural contexts. Furthermore, the results of the Bayesian models depend on the choice of prior distributions. While different prior specifications were used to assess the robustness of the findings, the results may still be sensitive to the chosen priors. Other modeling approaches, such as multi-level or mixed-effects models, could be explored in future research to account for potential hierarchical structures and unobserved heterogeneity in the data.

9. CONCLUSIONS

The main objective of this study was to investigate the factors influencing pedestrian injury severity in crashes occurring on interstates in North Carolina, with a focus on pedestrian actions, roadway conditions, vehicle characteristics, and spatial context of these crashes. The analysis was based on police-reported pedestrian crash data from North Carolina, spanning from 2007 to 2022, which was meticulously coded using the Pedestrian and Bicycle Crash Analysis Tool (PBCAT). The dataset included 882 pedestrian crash observations on interstates. Frequentist and Bayesian binary logit models with various prior specifications and a robust machine learning algorithm, Random Forest, were estimated to analyze the factors affecting pedestrian crash injury severity on interstates. The study was conducted to find answers to two research questions. The first research question dealt with the influence of pedestrian actions and behaviors on pedestrian crash injury severity in crashes on interstates. The main findings indicate that unsafe pedestrian actions such as crossing the freeway, walking along the road in the direction of traffic, standing in the roadway, and alcohol impairment significantly increase the likelihood of fatal and severe injuries. The second research question was based on assessing the role of roadway and vehicle characteristics in pedestrian crash injury severity on freeways. The study found that roadway conditions such as icy surfaces and dark unlit roads, along with the involvement of SUVs, are key factors associated with higher injury severity in pedestrian crashes. The study's findings have significant planning and policy implications for improving pedestrian safety on interstates. Enhancing roadway infrastructure, enforcing traffic regulations, and promoting public awareness are crucial steps toward reducing pedestrian fatalities and severe injuries on high-speed roadways. The findings can guide policymakers and practitioners in developing effective strategies to protect pedestrians on high-speed roadways, ultimately contributing to the reduction of pedestrian fatalities and severe injuries.

This study contributes to the existing body of knowledge by providing a comprehensive account of the critical unsafe pedestrian actions, roadway conditions, and vehicle characteristics that influence pedestrian safety on interstates. In addition to the conventional frequentist methods, the application of Bayesian models with different prior specifications offers robust insights into the factors affecting pedestrian crash injury severity on interstates. Future research should aim to explore more sophisticated modeling approaches, investigate the effectiveness of the suggested pedestrian safety countermeasures, and explore innovative solutions to enhance pedestrian safety on interstates.

ACKNOWLEDGMENT

The authors thank the USDOT-sponsored Center for Pedestrian and Bicycle Safety (CPBS) for providing financial support for this paper's research, authorship, and/or presentation/publication.

AUTHOR CONTRIBUTIONS

The authors confirm their contribution to the paper as follows:

Study Conception and Design: Sheikh M. Usman, Asad J. Khattak; ***Data Collection and Preliminary Analysis:*** Sheikh M. Usman; ***Analysis and Interpretation of Results:*** Sheikh M. Usman, Asad J. Khattak; ***Draft Manuscript Preparation:*** Sheikh M. Usman, Asad J. Khattak. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Tennessee Department of Transportation. Tennessee Pedestrian Laws. *https://www.tn.gov/tdot/multimodal-transportation-resources/bicycle-and-pedestrian-program/resources11/tennesse-pedestrian-laws.html*.
2. NHTSA. Traffic Safety Facts: Pedestrian 2018 Data. *https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812850*.
3. NHTSA. FARS Encyclopedia: People-Pedestrians. *Retrieved from https://www-fars.nhtsa.dot.gov/People/PeoplePedestrians.aspx*.
4. Johnson, C. Pedestrian Fatalities on Interstate Highways: Characteristics and Countermeasures. *Transportation Research Record*, Vol. 1578, No. 1, 1997, pp. 23–29.
5. Tefft, B. Impact Speed and a Pedestrian's Risk of Severe Injury or Death. *Accident Analysis and Prevention*, Vol. 50, 2013, pp. 871–878.
6. Hossain, A., X. Sun, R. Thapa, and J. Codjoe. Applying Association Rules Mining to Investigate Pedestrian Fatal and Injury Crash Patterns under Different Lighting Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2676, No. 6, 2022, pp. 659–672.
7. AAA Foundation for Traffic Safety. Pedestrian Fatalities on Interstate Highways, United States, 1993–2012. *Retrieved from https://aaafoundation.org/wp-content/uploads/2017/12/PedestrianFatalitiesonInterstatesReport.pdf*, 2014.
8. Iragavarapu, V., S. H. Khazraee, D. Lord, and K. Fitzpatrick. Pedestrian Fatal Crashes on Freeways in Texas. *Paper presented at the 94th Annual Meeting of the Transportation Research Board, Washington, DC.*, 2015.
9. Kim, J. K., G. F. Ulfarsson, V. N. Shankar, and S. Kim. Age and Pedestrian Injury Severity in Motor-Vehicle Crashes: A Heteroskedastic Logit Analysis. *Accident Analysis and Prevention*, Vol. 40, No. 5, 2008, pp. 1695–1702. https://doi.org/10.1016/j.aap.2008.06.005.
10. Haleem, K., P. Alluri, and A. Gan. Analyzing Pedestrian Crash Injury Severity at Signalized and Non-Signalized Locations. *Accident Analysis and Prevention*, Vol. 81, 2015, pp. 14–23. https://doi.org/10.1016/j.aap.2015.04.025.
11. Usman, S., A. Khattak, and L. Patwary. Nighttime Pedestrian Safety in Different Communities: Application of Artificial Intelligence Techniques. *Journal of Safety Research*, Vol. 94, 2025, pp. 340–361. https://doi.org/10.1016/j.jsr.2025.06.028.
12. Usman, S., A. Khattak, and L. Patwary. Nighttime Pedestrian Safety in Burdened Communities: Application of Artificial Intelligence Techniques. *Faculty Publications and Other Works -- Civil & Environmental Engineering. https://trace.tennessee.edu/utk_civipubs/36/*, 2025, pp. 1–27.

13. Ferenchak, N., R. Gutierrez, and P. Singleton. Shedding Light on the Pedestrian Safety Crisis: An Analysis across the Injury Severity Spectrum by Lighting Condition. *Traffic Injury Prevention*, Vol. 23, No. 7, 2022, pp. 434–439.

14. Usman, S., and A. Khattak. Applying Artificial Intelligence to Examine Nighttime Pedestrian Crash Injury Severity at Intersections. *Faculty Publications and Other Works -- Civil & Environmental Engineering, https://trace.tennessee.edu/utk_civipubs/37* , 2025, pp. 1–25.

15. Usman, S., and A. Khattak. Applying Artificial Intelligence Techniques to Examine Nighttime Pedestrian Crash Injury Severity at Intersections. *Journal of Transportation Safety & Security*, 2025, pp. 1–47. https://doi.org/10.1080/19439962.2025.2548807.

16. Fridman, L., R. Ling, L. Rothman, M. Cloutier, C. Macarthur, B. Hagel, and A. Howard. Effect of Reducing the Posted Speed Limit to 30 Km per Hour on Pedestrian Motor Vehicle Collisions in Toronto, Canada—A Quasi Experimental, Pre-Post Study . *BMC Public Health*, Vol. 20, No. 1, 2020.

17. Amoh-Gyimah, R., E. Aidoo, M. Akaateba, and S. Appiah. The Effect of Natural and Built Environmental Characteristics on Pedestrian-Vehicle Crash Severity in Ghana. *International Journal of Injury Control and Safety Promotion*, Vol. 24, No. 4, 2017, pp. 459–468.

18. Blazquez, C., and M. Celis. A Spatial and Temporal Analysis of Child Pedestrian Crashes in Santiago, Chile. *Accident Analysis and Prevention*, Vol. 50, 2013, pp. 304–311.

19. Kemnitzer, C., C. Pope, A. Nwosu, S. Zhao, L. Wei, and M. Zhu. An Investigation of Driver, Pedestrian, and Environmental Characteristics and Resulting Pedestrian Injury. *Traffic Injury Prevention*, Vol. 20, No. 5, 2019, pp. 510–514.

20. Islam, S., and S. L. Jones. Pedestrian At-Fault Crashes on Rural and Urban Roadways in Alabama. *Accident Analysis and Prevention*, Vol. 72, 2014, pp. 267–276. https://doi.org/10.1016/j.aap.2014.07.003.

21. Mansfield, T., D. Peck, D. Morgan, B. McCann, and P. Teicher. The Effects of Roadway and Built Environment Characteristics on Pedestrian Fatality Risk: A National Assessment at the Neighborhood Scale. *Accident Analysis and Prevention*, Vol. 121, 2018, pp. 166–176.

22. Dong, C., A. Khattak, C. Shao, and K. Xie. Exploring the Factors Contribute to the Injury Severities of Vulnerable Roadway User Involved Crashes. *International Journal of Injury Control and Safety Promotion*, Vol. 26, No. 3, 2019, pp. 302–314.

23. Wang, J., and J. Cicchino. Fatal Pedestrian Crashes on Interstates and Other Freeways in the United States. *Journal of Safety Research*, Vol. 74, 2020, pp. 1–7.

24. Hossain, A., X. Sun, N. Zafri, and J. Codjoe. Investigating Pedestrian Crash Patterns at High-Speed Intersection and Road Segments: Findings from the Unsupervised Learning

Algorithm. *International Journal of Transportation Science and Technology*, Vol. 14, 2023, pp. 186–201.

25. USDOT. What Is a Safe System Approach? *Retrieved from https://www.transportation.gov/NRSS/SafeSystem*, 2022.

26. NCDOT. NCPedBikeCrashes. *Retrieved from https://ncdot.maps.arcgis.com/home/item.html?id=d00e97d23146472cb5da78fc3dbc0274*.

27. Wali, B., A. Khattak, H. Bozdogan, and M. Kamrani. How Is Driving Volatility Related to Intersection Safety? A Bayesian Heterogeneity-Based Analysis of Instrumented Vehicles Data . *Transportation Research Part C: Emerging Technologies*, Vol. 92, 2018, pp. 504–524. https://doi.org/10.1016/j.trc.2018.05.017.

28. Usman, S., and M. Adeel. Exploring University Campus Community's Perceptions and Barriers toward Biking: A Case Study of the University of Tennessee Knoxville. *Journal of Transportation Technologies*, Vol. 14, No. 2, 2024, pp. 161–178. https://doi.org/10.4236/jtts.2024.142010.

29. Ntzoufras, I. *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, 2011.

30. Usman, S., A. Khattak, and A. L. Patwary. Are There Differences in Public Interest toward Automated Vehicles' Ownership in Burdened and Non-Burdened Communities? *Transportation Planning and Technology*, 2025, pp. 1–28. https://doi.org/10.1080/03081060.2025.2546096.

31. Usman, S., A. Khattak, S. Chakraborty, I. Mahdinia, and R. Tavassoli. Detection of Distracted Driving through the Analysis of Real-Time Driver, Vehicle, and Roadway Volatilities. *Journal of Transportation Safety & Security*, Vol. 16, No. 12, 2024, pp. 1426–1447. https://doi.org/10.1080/19439962.2024.2341393.

32. Khattak, A., S. Chakraborty, I. Mahdinia, and S. Usman. Safety Enhancement by Detecting Driver Impairment Through Analysis of Real-Time Volatilities. *Collaborative Sciences Center for Road Safety*, 2023, pp. 1–75.

33. Adeel, M., A. Khattak, S. Mishra, and D. Thapa. Enhancing Work Zone Crash Severity Analysis: The Role of Synthetic Minority Oversampling Technique in Balancing Minority Categories. *Accident Analysis & Prevention*, Vol. 208, 107794, 2024. https://doi.org/10.1016/j.aap.2024.107794.

34. Usman, S., A. Khattak, S. Chakraborty, I. Mahdinia, and R. Tavassoli. Distracted Driving Detection through the Analysis of Real-Time Driver, Vehicle, and Roadway Volatilities. *Faculty Publications and Other Works -- Civil & Environmental Engineering, https://trace.tennessee.edu/utk_civipubs/38*, 2025, pp. 1–25.

35. Ijaz, M., L. Lan, S. M. Usman, M. Zahid, and A. Jamal. Investigation of Factors Influencing Motorcyclist Injury Severity Using Random Parameters Logit Model with

Heterogeneity in Means and Variances. *International Journal of Crashworthiness*, Vol. 27, No. 5, 2021, pp. 1412–1422. https://doi.org/10.1080/13588265.2021.1959153.

36. Ijaz, M., L. Liu, Y. Almarhabi, A. Jamal, S. M. Usman, and M. Zahid. 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, 2022. https://doi.org/10.3390/ijerph191710526.

37. Rashid, H., S. Shah, and S. Usman. Factors Influencing Motorists' Injury Severities: An Empirical Assessment of Crashes in District Peshawar, Pakistan. *Global Social Sciences Review, VIII*, 2023, pp. 211–224. https://doi.org/10.31703/gssr.2023(VIII-II).20.

38. Usman, S., and A. Khattak. Beyond the Conventional: Exploring Pedestrian Safety on Interstates with Bayesian and Machine Learning Models. *Journal of Safety Research*, Vol. 94, 2025, pp. 141–156. https://doi.org/10.1016/j.jsr.2025.06.018.

39. Stoker, P., A. Garfinkel-Castro, M. Khayesi, W. Odero, M. Mwangi, M. Peden, and R. Ewing. Pedestrian Safety and the Built Environment: A Review of the Risk Factors. *Journal of Planning Literature*, Vol. 30, No. 4, 2015, pp. 377–392.

40. Hossain, A., X. Sun, S. Das, M. Jafari, and A. Rahman. Investigating Pedestrian-Vehicle Crashes on Interstate Highways: Applying Random Parameter Binary Logit Model with Heterogeneity in Means. *Accident Analysis and Prevention*, Vol. 199, No. 107503, 2024.

41. Wanvik, P. O. Effects of Road Lighting: An Analysis Based on Dutch Accident Statistics 1987–2006. *Accident Analysis & Prevention*, Vol. 41, 2009, pp. 123–128.

42. Khattak, A., M. King, M. Adeel, S. Usman, and I. Mahdinia. Advancing Crash Investigation With Connected and Automated Vehicle Data–Phase 2. *Collaborative Sciences Center for Road Safety*, Vol. CSCRS R-42, 2024.

43. Adeel, M., M. King, S. Usman, and A. Khattak. Advancing Crash Investigation with Connected and Automated Vehicle Data: Insights from a Survey of Law Enforcement. *Journal of Transportation Safety & Security*, 2025, pp. 1–30. https://doi.org/10.1080/19439962.2025.2462803.

44. King, M., S. Usman, M. Adeel, and A. Khattak. Advancing Crash Investigation with Connected and Automated Vehicle Data: Insights from a Survey of Law Enforcement. *Faculty Publications and Other Works -- Civil & Environmental Engineering. https://trace.tennessee.edu/utk_civipubs/32*, 2024, pp. 1–20.