Journal of Safety Research 94 (2025) 193–215

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

How do motorists' pre-crash behaviors contribute to the injury severity of police officers? Using interpretable machine learning to untangle the behavioral pathways in crashes involving police vehicles

Ningzhe Xu [a], Jun Liu [a,*], Zihe Zhang [b], Steven Jones [a,b,c]

[a] Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL 35487, United States
[b] Alabama Transportation Institute, The University of Alabama, Tuscaloosa, AL 35487, United States
[c] Transportation Policy Research Center, The University of Alabama, Tuscaloosa, AL 35487, United States

ARTICLE INFO

Keywords:
Police officers
Injury severity
Traffic crash
Pre-crash behavior

ABSTRACT

Introduction: Police officers are integral to enforcing traffic laws and providing assistance to motorists. While performing their duties on the road or roadside, they encounter significant hazards, many of which arise from the negligent or inappropriate behaviors of drivers. Despite the prevalence of these risks, there is a paucity of research specifically examining the outcomes of traffic crashes involving police officers or police vehicles, particularly in relation to the injury severity sustained by police officers in such incidents. The objective of this study is to explore the roles of motorist behaviors in police-involved crashes. *Method:* Specifically, using police-involved crashes data from 2017 to 2021 in Alabama, this study employs a path analysis framework integrated with interpretable machine learning to examine the behavioral pathways from contributing factors to motorists' pre-crash behaviors, ultimately leading to the injury severity of police officers. The path analysis allows the identification of various factors that directly contribute to injury severity and those that indirectly contribute to injury severity through their relationships with motorists' pre-crash behaviors. *Results:* The results indicate that several factors are directly associated with policy injury severities, including motorists' pre-crash behaviors (such as failure to yield, driving under the influence (DUI), and traffic signal violations) that are significantly correlated with increased injury severity for police officers involved in crashes. Further, these pre-crash behaviors are associated with contributing factors including vehicle speeds, lighting conditions, seatbelt usage, and weather conditions. Thus, these contributing factors serve as significant predictors of motorists' pre-crash behaviors, indirectly impacting police injury severity through their influence on unsafe driving practices. *Practical Applications:* This study provides valuable insights into the behavioral pathway leading to police injury severities. By understanding the direct and indirect contributors to police injury severity, policymakers can develop targeted interventions to address these key risk factors and improve overall road safety for police officers.

1. Introduction

Police officers, also known as law enforcement officers (LEOs), play a critical role in traffic safety, enforcing traffic laws, providing assistance to motorists, and responding to traffic incidents. While performing their duties on the road or roadside, they encounter significant hazards, many of which arise from the negligent or inappropriate behaviors of drivers they encounter. Between 2011 and 2020, motor vehicle-related incidents resulted in the deaths of 454 law enforcement officers, accounting for 33% of all line-of-duty deaths (LODDs), according to the National Institute for Occupational Safety and Health (NIOSH, 2022). Notably, 131 police officers were fatally struck by passing vehicles while responding to traffic incidents during the same period (National Law Enforcement Officer Memorial Fund, 2021). Moreover, there are numerous nonfatal police-involved traffic crashes that could lead to long-term injuries, increased healthcare costs, and psychological trauma (García-Altés & Pérez, 2007; Mayou et al., 1993), thereby reducing the operational effectiveness of law enforcement agencies. The cumulative effect of these incidents extends beyond individual officers, ultimately impacting overall traffic safety and negatively affecting all road users.

To mitigate these risks and safeguard police officers on roads, as well as to ensure the efficacy of traffic enforcement and incident

* Corresponding author.
E-mail addresses: nxu3@crimson.ua.edu (N. Xu), jliu@eng.ua.edu (J. Liu), zzhang124@crimson.ua.edu (Z. Zhang), steven.jones@ua.edu (S. Jones).

https://doi.org/10.1016/j.jsr.2025.06.023
Received 18 February 2025; Received in revised form 28 March 2025; Accepted 20 June 2025
Available online 27 June 2025

management, it is imperative to conduct research identifying the factors that contribute to the severity of injuries sustained by police officers involved in traffic crashes. Compared to general road users, police vehicles often operate under emergency conditions, such as pursuits, which introduce unique risk factors like high-speed driving and running traffic signals or stop signs. Additionally, police officers spend significantly more time on the road than the general public, increasing their crash exposure. Their job duties and working environments also contribute to heightened psychological pressure while driving. These factors may result in distinct crash patterns and injury severities in police-involved crashes compared to those involving the general public. This area of research is relatively underexplored (Hossain et al., 2023; Xu et al., 2024a) compared to the extensive studies on injury severity among general road users, including motorists (e.g., Kockelman & Kweon, 2002; Xie et al., 2012), motorcycle riders (e.g., Santos et al., 2024; Zahid et al., 2024), bicyclists (e.g., Eluru et al., 2008; Kim et al., 2007), and pedestrians (e.g., Eluru et al., 2008; Kim et al., 2010; Sze & Wong, 2007). To identify the factors contributing to injury severity, many studies have focused on the direct relationships between injury severity and various determinants, including drivers' demographic attributes, vehicle characteristics, crash dynamics (e.g., speed), as well as roadway and weather conditions (Behnood & Mannering, 2017; Chen et al., 2016; Hu et al., 2014; Kaplan et al., 2014; Kim et al., 2013; Zhang et al., 2024). A critical insight highlighted by researchers is that driver behaviors are shaped by these determinants (Behnood & Mannering, 2017; Fu, 2008; Kaplan et al., 2014; Kim et al., 2007; Porter & England, 2000), and the driver behaviors prior to the event of a crash, which is called pre-crash behaviors, are found to be significantly related to crash injury severity outcomes (Dong et al., 2018; Liu et al., 2020; Xu et al., 2024a). Behavioral pathways may exist in traffic crashes, linking contributing factors, pre-crash behaviors, and injury severity. By analyzing the behavioral pathways in traffic crashes, researchers can uncover how various factors directly contribute to injury severity and how the same or different factors indirectly influence injury severity through their relationships with pre-crash behaviors. While several studies have explored the behavioral pathways in traffic crashes involving general road users such as motorists (Arvin et al., 2019; Gargoum & El-Basyouny, 2016; Liu & Khattak, 2017), bicyclists (Liu et al., 2021; Lu et al., 2022), and pedestrians (Okafor et al., 2023), there is a notable absence of research focused on investigating these pathways in police-involved crashes (Chu, 2016; Liu & Donmez, 2011; von Kuenssberg Jehle et al., 2005; Xu et al., 2024a). Therefore, this study aims to fill the research gap by exploring the behavioral pathways in police-involved crashes. Specifically, this study focuses on examining the roles of motorists' pre-crash behaviors and addresses the following research questions:

- How do various factors contribute to unsafe motorists' pre-crash behaviors?
- How do motorists' pre-crash behaviors contribute to the severity of injuries sustained by police officers involved in traffic crashes?
- How do various factors directly contribute to police injury severity?
- How do factors indirectly influence police injury severity through their relationships with motorists' pre-crash behaviors?

Speaking of the modeling methods, traditionally, researchers have utilized statistical modeling and analysis to identify contributing factors for crash severity (Eluru et al., 2008; Kim et al., 2007, 2010; Kockelman & Kweon, 2002; Liu et al., 2019; Sze & Wong, 2007) or driver behaviors (Fu, 2008; Porter & England, 2000). The results of statistical models are easy to interpret and can provide valuable insights. However, there are some limitations associated with traditional statistical models. For example, these models often require specific assumptions regarding data distribution before model fitting, such as the Independence of Irrelevant Alternatives. Besides, traditional statistical models tend to have relatively lower predictive ability and may struggle to capture the nonlinear relationships between contributing factors and target variables. To overcome these limitations, in recent years, researchers have increasingly utilized machine learning methods to solve different kinds of transportation problems (Ijaz et al., 2021; Xu et al., 2021; Yang et al., 2023; Zhang et al., 2018; Zhang et al., 2024; Zhang et al., 2024). Zhao et al. (2020) indicated that machine learning methods could provide impact assessments of contributing factors comparable to those obtained from logit models, suggesting that the interpretations derived from machine learning methods are reliable. According to the aforementioned advantages of machine learning, this study employed machine learning methods to model and investigate the pre-crash behaviors and injury severity of police officers.

Using data on police-involved crashes from 2017 to 2021 in Alabama, this study employs a path analysis framework integrated with rigorous modeling techniques to explore the behavioral pathways from contributing factors to motorists' pre-crash behaviors, ultimately leading to the injury severity of police officers. Note that police vehicles could be either the vehicle responsible for the crash or the non-fault party. Different machine learning methods may introduce certain biases, leading to varying modeling outcomes, such as differences in the impact of input features on the outputs. To mitigate potential biases from a single model, this study employs multiple machine learning classifiers, including Random Forest (RF), Adaptive Boosting (AdaBoost), Naïve Bayes (NB), Artificial Neural Network (ANN), and Linear Discriminant Analysis (LDA). These classifiers are used to train models for motorists' pre-crash behaviors and the injury severities of police officers. To interpret the machine learning models and understand how individual factors are associated with motorists' pre-crash behaviors and police injury severities, marginal effects are calculated based on these models to quantify the relationships between factors. Marginal effects offer an intuitive way to interpret the impact of contributing factors on model outputs intuitively by measuring the change in the model outputs (pre-crash behavior or police injury severity) with a one-unit change in the input features (i.e., contributing factors), which is a suitable measurement of direct and indirect impact for the path analysis framework. Path analysis is then employed to disentangle the direct and indirect contributing factors for police injury severity. Indirect contributing factors are quantified through their relationships with motorists' pre-crash behaviors. This study is expected to offer valuable insights into the behavioral pathways leading to police injury severities. By understanding the direct and indirect contributors to police injury severity, policymakers can develop targeted interventions to address these key risk factors and improve overall road safety for police officers.

2. Literature review

2.1. Police-involved crashes injury severity

Few studies investigated factors related to the injury/crash severity of police-involved traffic crashes (Chu, 2016; Hossain et al., 2023; Liu & Donmez, 2011; Xu et al., 2024a; Abdelwanis, 2013). Regarding pre-crash behaviors, most police-involved crash/injury severity research mentioned that distracted driving (regardless of police or non-police driver) would increase the probability of severe injury (Abdelwanis, 2013; Hu et al., 2014; Liu & Donmez, 2011; Noh, 2011). However, Liu and Donmez (2011) found that loss of thought would decrease the severity of injuries caused by police-involved crashes. Their findings also indicated that in-vehicle distractions could cause more severe injury for both police officers and other motorists, specifically for police officers, which agreed with Abdelwanis (2013), Hu et al. (2014), and Noh (2011). Besides distracted driving, speeding or driving too fast for conditions were also identified as major pre-crash behaviors that significantly contribute to severe injuries. Noh (2011) identified high-speed pursuits, driving too fast for conditions, and speeding as key factors related to severe injuries. This finding is consistent with Chu (2016), who also found that the speeding behavior of police officers is more likely to result in severe injury. The risk of fatality for first responders in crashes significantly increases when speeding coincides with a lack of seatbelt use. Xu et al. (2024a) indicated that as speeds increase, police officers are more likely to suffer severe injuries. Xu et al. (2024a) also found that behaviors such as police officers' improper seatbelt usage, police officers' sign/signal violation, motorists' failure to yield, DUI, and sign/signal violations were significantly associated with more severe injury. Chu (2016) found that traffic violations during emergency responses or pursuits were related to more severe injuries of police officers.

Other numerous factors have been identified as contributors to the injury severity of police-involved incidents. According to Noh's findings (2011), failure to maintain the proper lane, or running off the road were identified as factors related to severe injuries. Chu (2016) found that driving during night hours (6p.m. to 12p.m.) and rainy weather situations were more likely to increase the probability of severe injury. Young police officers were also found to have a higher likelihood of being involved in crashes resulting in injury compared to middle-aged officers. Failure to yield was identified as a critical factor that can lead to severe injury. The findings of Xu et al. (2024a) indicated that variables such as police officers' gender, and emergency situations were highly related to more severe injury. Other factors including intersection, dawn, dusk, dark with continuous light, curve and level roads, head-on crashes, sideswipe crashes, and the dates after holidays were found to be positively associated with more severe injury for police officers.

2.2. Pre-crash behaviors

Previous studies investigated various dangerous pre-crash behaviors, such as aggressive driving (Shinar & Compton, 2004; Taubman-Ben-Ari & Yehiel, 2012; Yang et al., 2022), DUI (Oh et al., 2020; Smailović et al., 2023; Wyatt & Novotna, 2021), traffic violations (Cheng et al., 2023; González-Iglesias et al., 2012; Li et al., 2021; Penmetsa & Pulugurtha, 2017; Porter & Berry, 2001; Porter & England, 2000; Precht et al., 2017; Zhang et al., 2016), failure to yield (Hussain et al., 2021; Li et al., 2023; Medina et al., 2023; Pai & Saleh, 2008), and speeding-related behaviors (Chevalier, Coxon, Chevalier, et al., 2016; Chevalier, Coxon, Rogers, et al., 2016; De Pelsmacker & Janssens, 2007; Gheorghiu et al., 2015; Kong et al., 2020; Stephens et al., 2017). Penmetsa and Pulugurtha (2017) found that going the wrong way is the most harmful pre-crash behavior for the innocent party of drivers. Speeding threatens the innocent party of drivers more than a signal violation. Driving under the influence of alcohol (DUIA) produced the risk for the innocent party of drivers two times the risk generated by driving under the influence of drugs (DUID). In general, gender and age were found to be the common identifiers of the aforementioned behaviors. Males were more likely to DUI (Oh et al., 2020; Wyatt & Novotna, 2021), violate traffic rules (González-Iglesias et al., 2012), speed (Perez et al., 2021; Stephens et al., 2017), not yield (Li et al., 2023), and drive aggressively (Shinar & Compton, 2004; Taubman-Ben-Ari & Yehiel, 2012; Yang et al., 2022). Although some studies indicated that there is no significant relationship between gender and traffic violations (Zhang et al., 2016) and the gap in the number of DUI drivers between males and females is shrinking (Oh et al., 2020), males were still more likely to drive aggressively. Regarding age, young drivers have a higher likelihood of traffic violations (González-Iglesias et al., 2012; Zhang et al., 2016), speeding (Chevalier, Coxon, Chevalier, et al., 2016; Chevalier, Coxon, Rogers, et al., 2016; Perez et al., 2021), and aggressive driving (Shinar & Compton, 2004; Taubman-Ben-Ari & Yehiel, 2012; Yang et al., 2022). On the other hand, older drivers were more likely to drive under the influence (Oh et al., 2020; Wyatt & Novotna, 2021).

Except for gender and age, different factors were found to be related to various behaviors. Studies regarding DUI revealed that several socio-demographic factors were related to DUIA and DUID. Oh et al. (2020) found that white, high-income, and high-education level people were more likely to DUIA. A review paper done by Wyatt and Novotna (2021) has the opposite findings. Their review results indicated that there is no difference in DUI behavior among races, and people with a lower socio-economic status were less likely to have DUI. Moreover, they claimed that DUID is normal in all age groups. Those who use marijuana frequently, young and male drivers, were more likely to drive under the influence of marijuana. Another review paper found that other types of factors, such as speeding, nighttime, not using a seatbelt, having DUI experience, older vehicles, and urban two-lane highways, were highly associated with DUIA (Smailović et al., 2023).

Socio-demographic and traffic-related factors played an important role in identifying traffic violations. Zhang et al. (2016) revealed that darkness without streetlights was negatively related to traffic signal violations. Normal city roads were positively related to traffic signal violations. Cheng et al. (2023) indicated that AADT, average traffic speed, signal timing, and phasing were important in identifying traffic violations at road intersections. Findings from Li et al. (2021) agreed with Zhang and Cheng. They found that high AADT and local roads positively affect traffic violation behaviors. Also, midnight, weekends, residential areas, low wind speed, and temperature were found to be positively associated with traffic violations. Precht et al. (2017) indicated that rainy days, daylight, and emotions such as anger, frustration, surprise, and sadness were more likely to be related to traffic violations. Running red light was also found to be positively related to improper usage of seatbelts and non-white people (Porter & Berry, 2001).

Rare studies regarding yielding behaviors analyzed the failure to yield. Medina et al. (2023) examined the reason for failure to yield at multi-lane roundabouts. They found that volume magnitude, uniformity, and imbalances across lanes were more likely to be related to failure to yield. Other papers focused on the behavior of yield or not. Allocating detection-based strategy and drivers with less driving experience were more likely to yield (Hussain et al., 2021; Li et al., 2023).

Speeding-related behavior was affected by multiple factors. Gheorghiu et al. (2015) investigated the relationship between peer pressure and young drivers' speeding behavior. They found that high peer pressure was associated with a high intention of speeding. Kong et al. (2020) analyzed naturalistic driving data and summarized that driving on a low functional class roadway, experiencing congestion, and the presence of a median was related to a higher speeding pattern. Perez et al. (2021) indicated that lower speed limits were more likely to be related to speeding behaviors.

2.3. Path analysis in crash research

Path analysis is often used in traffic safety research (Ahmad et al., 2023; Arvin et al., 2019; Liu et al., 2015, 2016, 2021; Liu & Khattak, 2017, 2018; Lu et al., 2022; Okafor et al., 2023; Park et al., 2021; Yu et al., 2019). It is a subset of Structural Equation Modeling (Golob, 2003). This approach is valuable for researchers as it helps to clarify the direct and indirect connections between variables. Path analysis enables researchers to evaluate mediation effects, which occur when the influence of an independent variable on a dependent variable is conveyed through one or more mediating variables. Thus, it can help researchers gain a deeper understanding of the complex relationship in traffic safety research.

Path analysis is treated as a modeling framework rather than a specific model in previous studies. Typically, because most previous studies focused on classification problems, binary logit models (Liu & Khattak, 2017, 2018), multinomial logit models (Liu et al., 2015, 2016), ordered logit models (Liu et al., 2015, 2016), Tobit regression (Ahmad et al., 2023), and ordered probit regression (Ahmad et al., 2023) are nested into path analysis. The negative binomial model and linear regression model are also employed to investigate the pathway relationship in safety research (Park et al., 2021). Additionally, researchers have made some efforts to combine other modeling frameworks like Geographic Weighted Regression (Liu et al., 2016; Liu & Khattak, 2017), latent class models (Yu et al., 2019), and random parameter logit and probit models

(Arvin et al., 2019; J. Liu et al., 2021; Okafor et al., 2023) with path analysis to capture the mediation effects with unobserved heterogeneities. As another famous type of modeling method, machine learning methods have the huge potential to reveal the relationships within the path analysis framework. Lu et al. (2022) has quantified the behavioral pathway in bicycle-motor vehicle crashes by utilizing multiple machine-learning classifier-supported path analysis.

2.4. Summary

In conclusion, there are a limited number of studies examining the crash/injury severity of police- or responder-involved crashes. Moreover, the findings from these studies show inconsistencies in the factors associated with police-involved crash/injury severity. It is possible that heterogeneity exists among different populations, highlighting the need for further exploration of factors related to police-involved crash/injury severity. Additionally, literature has proved that there are several potential behavior pathways that exist between pre-crash behaviors and crash injury severity. It is necessary to investigate whether the behavior pathway can be captured under the background of police-involved injury severity research. Compared to traditional statistical models, most machine learning models can uncover relationships between contributing factors and outcomes without requiring specific assumptions, such as the Independence of Irrelevant Alternatives (IIA) in binary logistic regression or the proportional odds assumption in ordered logistic regression. This flexibility allows researchers to explore various types of outcomes more easily. Given the identified research gaps and the advantages of machine learning models, this study integrates multiple machine learning models within a path analysis framework to investigate the behavior pathway between pre-crash behaviors and the injury severity of police officers.

3. Methodology

3.1. Data preparation

This study extracted the records of police-involved crashes from the Alabama statewide crash database named CARE. The crashes that occurred during the period from 2017 to 2021 were employed to analyze the behavior pathways among variables in interest, pre-crash behaviors, and police injury severity.[1] Fig. 1 shows the overview of data filtering and re-organizing steps. To identify police-involved multi-vehicle crashes, crash-level, vehicle-level, and person-level records were merged using the crash identification number, which is unique to each recorded crash. For any given crash, all associated vehicles and individuals share the same crash identification number across datasets. Police vehicles and personal vehicles were identified using the vehicle usage indicator and were merged into a single crash record based on the shared crash ID. In cases where multiple police vehicles were involved in a single crash, the following selection rules were applied to retain only one police vehicle per crash: (1) If one of the police vehicles was at fault, that record was retained; (2) if no police vehicles were at fault, the record with the highest police driver injury severity was retained; and (3) if all police vehicles had similar injury severity levels, one police vehicle record was selected at random. After applying these rules, crash and driver records (including both police drivers and motorists) were merged using the same crash identification number. Final error checks (described below) were then performed to generate the final dataset of police-involved multi-vehicle crashes used in this study.

This study developed two sets of models: (1) models for the motorist's pre-crash behaviors, and (2) models for the police injury severity. There are various pre-crash behaviors reported in the crash records. According to the findings of Xu et al. (2024a), motorists' pre-crash behaviors that significantly affect police injury severity in Alabama include failure to yield, sign/signal violations, and driving under the influence (DUI). Therefore, this study focuses on these behaviors as the main categories of motorists' pre-crash behaviors. For the models assessing police injury severities, the original dependent variable (injury severity) has five levels. Due to the small proportion of fatal injuries, this study merged the two highest severity levels—fatal and severe injuries. Data were error-checked before modeling, such as removing observations with missing values in pre-crash behaviors and the injury severity of police officers, as well as any data entries that were clearly incorrect or unreliable. For instance, crashes with conflicting crash times and lighting conditions, or mismatched coded and textual crash location information, were excluded from the analysis. Eventually, a total of 4,667 police-involved crashes were used for modeling.

To better explain the modeling results, independent variables were classified as police officer characteristics, non-police officer (referred to as motorist) characteristics, crash information, and pre-crash behaviors. Before further processing of the data, the missing value issue was checked by the authors. Police, crash, and motorist characteristics existed varying degrees of the missing value issue. Specifically, the missing rate of all police characteristics was lower than 4.0%. The missing rate of motorist age, gender, and seatbelt usage were 20.1%, 19.7%, and 25.6%, respectively. The missing value issue of speed is the most serious one. The missing rate was 43.5%. In addition, for the lighting conditions, road alignment, and collision manner, the missing rates were 0.5%, 8.7%, and 0.2%, respectively. There are some effective ways to address the missing value issue, such as removing samples or variables with missing values and data imputation. However, removing samples or variables would remove a lot of useful information that would be necessary to accurately capture the relationships between independent and dependent variables. In this case, data imputation is a useful way to address the missing value issue. Imputing by mean or mode is the easiest method, but it decreases the variable variance and introduces bias into the analysis. Compared to the aforementioned methods, multiple imputations via chained equations (MICE) could perform better because the goal of MICE is to maintain the relationships between independent and dependent variables as those in the original dataset after imputation. To keep as much information as possible for modeling and explanation purposes, MICE was adopted to impute the missing values based on existing data (Van Buuren, 2014). All variables that need to be imputed are multi-class variables, except for gender. Thus, polytomous and binary logistic regressions were selected as imputation methods.

Table 1 presents the descriptive statistics of variables after data imputation. Notably, approximately 14.4% of police officers were reported injured in the crashes they were involved in, slightly higher than the injury rate for all crashes in Alabama during the same period, which stood at around 12.1%. Most police officers (37.0%) belonged to the age group of 26 to 35 years old. Fewer female police officers (9.4%) were involved in crashes than others. Notably, about three-quarters of police-involved crashes occurred when police officers were not responding to an emergency call. Interestingly, motorists shared a similar age distribution with police officers, with the majority falling into the 26–35 age group. Approximately 43.5% of motorists were identified as female. For police-involved crashes, the most prevalent contributing circumstance was motorists failing to yield, accounting for 11.2% of the cases. Furthermore, around 42.1% of police-involved crashes occurred at a speed of less than 11MPH for the causal unit involved. Regarding the conditions under which the crashes occurred, the majority of them took place on clear days (70.0%), during daylight hours (65.1%), and on straight and level roadways (75.6%). In terms of collision manners, the two primary types of police-involved collisions were rear-ended (35.3%) and sideswipe (25.9%).

[1] The police discussed in this article are all drivers of the police vehicle.

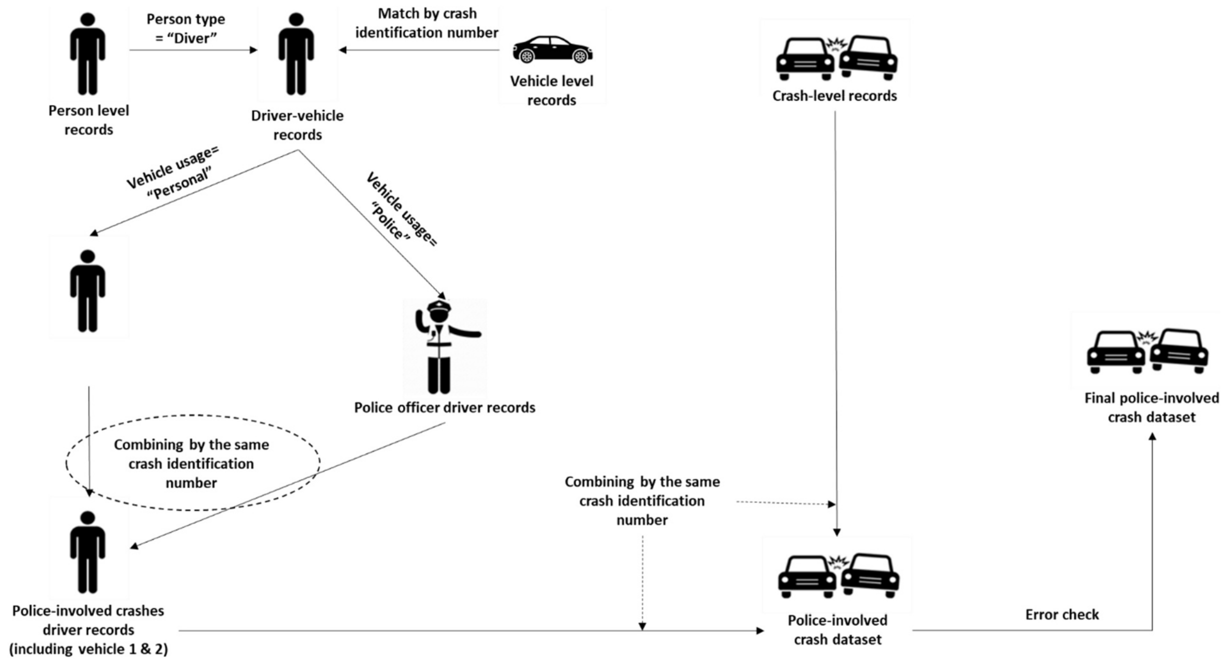

Fig. 1. Data filtering and re-organization.

3.2. Machine learning

This study employs machine learning to model the relationships between variables, from contributing factors to motorists' pre-crash behaviors and police injury severity. Given the variety of machine learning methods available, researchers often use different performance metrics, such as accuracy, F1-score, and recall, to compare models and determine which one performs best. However, it is important to note that a model may not exhibit the best performance across all performance metrics. As Box (1976) mentioned, "All models are wrong, but some are useful." This implies that while each model may introduce prediction bias, each can also provide valuable and unique insights that are not captured by other models (Bates & Granger, 1969; Makridakis & Winkler, 1983). To address the issue of bias and leverage the strengths of multiple models rather than relying on a single one, this study considers multiple machine learning methods to train models, including Random Forest (RF), Adaptive Boosting (AdaBoost), Naïve Bayes (NB), Linear Discriminant Analysis (LDA), and Artificial Neural Network (ANN). The marginal effects of the certain variable in different models were averaged to keep as much as possible information from each trained model. Hyperparameters of these classifiers were meticulously tuned using the "caret" package in R (Kuhn, 2008), utilizing repeated cross-validation based on predictive accuracy. The dataset was split into training and testing sets based on the ratio of 4:1, which means that taking 80% of the samples into the training set randomly, the rest of the samples were used to construct the testing sets. All models (if needed) were tuned hyperparameters by three-time 20-fold cross-validation using the training set based on the criteria of best validation accuracy. To ensure the model interpretation results are reliable, the multicollinearity issue should be excluded. The variable inflation factors (VIF) were checked, and all values were less than 5, which means that no multicollinearity issue existed. The confusion matrixes of all machine learning models can be found in Appendix A. Also, a comparison table of model performance metrics is also shown in Appendix A.

3.2.1. Random Forest

The Random Forest algorithm (RF) is an ensemble learning method that builds upon the bagging tree concept introduced by Breiman (2001). Its fundamental idea is to combine multiple weak classifiers, such as Decision Trees, into a robust and powerful classifier known as the Random Forest. To achieve this, the algorithm employs bootstrapping, which involves randomly selecting samples from the dataset with replacements to create new training sets for each tree in the forest. In the Random Forest, a specified number of trees (ntree) are constructed through bootstrapping. The data points that remain unselected during bootstrapping are called out-of-bag (OOB) observations. To reduce potential correlations between the trees, the algorithm randomly selects a subset of features (mtry) from the total pool of p features for each tree. The Random Forest algorithm's prediction process relies on each tree's individual predictions. To be precise, the prediction probability for each class is determined by the proportion of occurrences of that class in the predictions generated by all the trees in the forest. One of the advantages of the Random Forest is its robustness to skewed distributions, outliers, and missing values, which makes it suitable for various types of datasets. The "randomForest" package developed by Liaw and Wiener (2002) was utilized to train the Random Forest models in this study. The selected combinations of "mtry" and "ntree" for the pre-crash behavior and injury severity models were (5,150) and (6,150), respectively.

3.2.2. AdaBoost

Adaptive Boosting (AdaBoost) is an ensemble learning algorithm that effectively combines multiple weak classifiers, such as decision trees or logistic regression models. In this study, decision trees were chosen as the weak classifiers. Unlike RF, AdaBoost assigns weights to observations, which play a crucial role in the algorithm. All observations have equal weights initially, but if an observation is misclassified, its weight is boosted for the subsequent classifier iteration. This process results in the creation of weak classifiers with different weights. After convergence, each weak classifier receives a score, and the final classifier is determined by the linear combination of the weak classifiers from each stage (Zhu et al., 2009). AdaBoost tends to exhibit higher accuracy than RF, as it considers each classifier's weight. However, it can be computationally time-consuming and sensitive to outliers. For this study, the AdaBoost models were trained using the "Adabag" package in R (Alfaro et al., 2013). The combinations of "mfinal" (number of iterations) and "maxdepth" (the maximum depth of each tree) for each model were set to (50, 5) and (300, 9), respectively.

Table 1
Descriptive statistics after data imputation (N = 4,667).

Variable	Description	Frequency	Percent
Police injury severity	No injury (O)	4,087	87.6%
	Not visible but complains of pain (C)	248	5.3%
	Non-Incapacitating (B)	277	5.9%
	Incapacitating or fatal (K & A)	55	1.2%
Police age	18–25 years old	612	13.1%
	26–35 years old	1,727	37.0%
	36–45 years old	1,100	23.6%
	46–55 years old	897	19.2%
	56 + years old	331	7.1%
Police gender	Female	440	9.4%
	Not female	4,227	90.6%
Police vehicle emergency statuses	Not emergency	3,533	75.7%
	On an emergency call	1,134	24.3%
Police seatbelt	Use seatbelt	4,478	96.0%
	Not use seatbelt	170	3.6%
	Other	19	0.4%
Police vehicle type	Passenger car	1,741	37.3%
	SUV	2,395	51.3%
	Pickup	337	7.2%
	Other	194	4.2%
Motorist age	18–25 years old	1,011	21.7%
	26–35 years old	1,774	38.0%
	36–45 years old	590	12.6%
	46–55 years old	473	10.1%
	56 + years old	819	17.5%
Motorist gender	Female	2,030	43.5%
	Not female	2,637	56.5%
Motorist seatbelt	Use seatbelt	4,155	89.0%
	Not use seatbelt	339	7.3%
	Other	173	3.7%
Motorist vehicle type	Passenger car	2,115	45.3%
	SUV	915	19.6%
	Pickup	600	12.9%
	Other	1,037	22.2%
Primary pre-crash behavior (for the at-fault driver)	Motorist: Other	3,885	83.2%
	Motorist: DUI	145	3.1%
	Motorist: Fail to yield	523	11.2%
	Motorist: Sign/signal violation	114	2.4%
Estimated causal unit vehicle speed at impact	1–10 mph	1,964	42.1%
	11–20 mph	741	15.9%
	21–30 mph	481	10.3%
	31–40 mph	484	10.4%
	41–50 mph	382	8.2%
	51–60 mph	247	5.3%
	61–70 mph	139	3.0%
	70 + mph	128	2.7%
	Stationary	101	2.2%
Crash location	On road	4,124	88.4%
	Intersection	99	2.1%
	Median	24	0.5%
	Off road	53	1.1%
	Shoulder or roadside	152	3.3%
	Other	215	4.6%
Lighting condition	Day light	3,038	65.1%
	Dawn/dusk	202	4.3%
	Dark with continuous light	309	6.6%
	Dark with spot lights	651	13.9%
	Dark without light	408	8.7%
	Other	42	0.9%
Road alignment	Straight and level	3,529	75.6%
	Straight with down grade	444	9.5%
	Straight with up grade	326	7.0%
	Curve and level	162	3.5%
	Curve with grade	145	3.1%
	Other	61	1.3%
Collision manner	Rear-end	1,646	35.3%
	Causal unit backing	111	2.4%
	Head-on	218	4.7%
	Angle	351	7.5%
	Sideswipe	519	11.1%
	Side-impact	1,211	25.9%
	Other	611	13.1%
Weather	Clear	3,265	70.0%
	Cloudy	834	17.9%
	Rain	403	8.6%
	Mist/Fog	141	3.0%
	Other	24	0.5%

3.2.3. *Naïve Bayes classifier*

The naïve Bayes classifier (NB) is a probabilistic parametric classifier that relies on Bayes' theorem and adopts the maximum likelihood method for parameter estimation (Ng & Jordan, 2001). It assumes that all factors are independent of each other, known as the "naïve Bayes assumption," which may not hold in most real-world tasks. Despite this simplifying assumption, the naïve Bayes classifier still demonstrates good performance in practice (McCallum & Nigam, 1998). As aforementioned, to access potential multicollinearity among contributing factors, VIFs were calculated. All VIF values were below 5, indicating no severe multicollinearity exist. In other words, although strict independence does not exist among predictors, their interrelationships are not strong enough to substantially affect the model's assumptions or interpretation. It combines the naïve Bayes conditional probability model with the maximum a posteriori decision rule (Inkoom et al., 2019). For conducting the naïve Bayes classifier in this study, the R package "e1071″ was utilized (Meyer et al., 2019).

3.2.4. *Linear discriminant analysis*

Linear discriminant analysis (LDA) is a Bayesian classifier suitable for both binary and multiple classification tasks. However, compared to logistic regression, LDA necessitates more data assumptions (Pohar et al., 2004). Specifically, LDA assumes that the independent variables follow a normal distribution with equal covariance matrices and that each independent variable is approximately normally distributed within each class. LDA typically performs well when dealing with small sample-size datasets (James et al., 2013). While the mathematical expression of LDA bears similarities to logistic regression, the estimation of coefficients differs between the two. This study employed the LDA function in the R MASS package to implement the models.

3.2.5. *Artificial neural network*

The Artificial Neural Network (ANN) is a computational model designed to mimic the structure and functioning of biological nervous systems. It consists of three types of neuron layers: the input layer, the hidden layer, and the output layer. Data are input into the model through the input layer, and then it undergoes processing in the hidden layer. The resulting probabilities of each class are then output through the output layer. In this study, single-hidden-layer ANN models were trained using the "nnet" package in R (Ripley et al., 2016). The specific combinations for the number of units in the hidden layer and the weight decay constant for each model were (6, 0.6) and (2, 0.6).

3.3. *Marginal effects*

To interpret the machine learning models and quantitatively represent the relationships between factors, this study calculates the marginal effects for the input features/independent variables of interest. The definition of marginal effects is holding other variables as constant, the change of predictions (predicted value or probability) for the dependent variable when one unit value changes for a continuous variable or one category changes for a categorical variable (Williams, 2012). Marginal effects are widely used to measure the impact of independent variables on the dependent variable (Fu et al., 2022; Liu et al., 2021; Xu et al., 2025; Xu et al., 2023). Several ways of marginal effect calculation exist in previous literature, including the marginal effect at the mean (Silva Filho et al., 2021; Sun et al., 2020), the marginal effect at the representative value (Silva Filho et al., 2021), and marginal effect based on partial dependence (Molnar, 2020). For statistical models commonly used in injury severity research, such as logistic regression and ordered logit—which have closed-form mathematical structures—the marginal effects are calculated based on the estimated model parameters. However, for most machine learning models (e.g., RF, AdaBoost, and ANN), there is no closed-form structure from which to derive the marginal effect of each contributing factor. As such, marginal effects must be numerically approximated. Since the intuitive definition of marginal effects is consistent between closed-form and numerical estimation approaches, the interpretations remain comparable. This study adopted the definition of partial dependence to estimate the average marginal effects (AMEs). The AMEs represent the quantified relationships between the injury severities of responders, motorists' pre-crash behaviors, and associated independent variables.

All variables in this study are categorical or ordered variables. According to the definition, this study estimated the AMEs by calculating the difference in the predicted probability of the dependent variable between the target and base categories of certain independent variables. Thus, the AME of each category of a specific variable is based on the base category of this variable. The AME of the base category will be zero and will not be presented in the final results. The predicted probability was the calculated partial dependence of the target category of the specific independent variable. The partial dependence is estimated by the Monte Carlo method using training data (Molnar, 2020). The formulas of partial dependence and AME calculation are shown as follows:

$$\widehat{f}_S(x_S) = \frac{1}{n}\sum_{i=1}^{n}\widehat{f}\left(x_S, x_C^{(i)}\right) \tag{5}$$

$$AME = \widehat{f}_{Scategory} - \widehat{f}_{Sbase} \tag{6}$$

where, the $\widehat{f}_S$ is the partial function. Specifically, $\widehat{f}_{Scategory}$ and $\widehat{f}_{Sbase}$ are the calculated partial dependences of the target and base categories of the certain variable. The n is the number of observations in the training data. The x_S and $x_C^{(i)}$ are the set of independent variables. The x_S is the independent variable that needs to be calculated the partial dependence. The $x_C^{(i)}$ is the set of value or category of other independent variables at the i_{th} observations. Note that the partial function (Equation (5)) assumes that independent variables in the sets of independent variable C and S are not correlated with each other.

Given multiple models, multiple marginal effects will be calculated for the same input features, and these marginal effects are expected to vary across models. If the signs of the marginal effects from all or most models are consistent, it strengthens and validates the estimated relationships between factors. This study presents marginal effects calculated from all trained machine learning models. Additionally, the marginal effects from different models are averaged to represent the overall relationships between factors. This method aligns with the approach of combining models to enhance predictions. Researchers have designed multiple model combination methods, such as the simple average method and the weighted average combination method (Clemen, 1989). The simple average method was widely used to improve prediction ability because of the attributes of straightforward and robust (Fang, 2003; Palm & Zellner, 1992; Wong et al., 2007). Model combination or aggregation, known as ensemble learning, plays a crucial role in machine learning. Specifically, the concept of stacking is closely related to the simple average method. To mitigate the impact of extreme values, the maximum and minimum marginal values are removed prior to averaging the marginal effects from different models. This approach provides a more validated representation of the overall relationships between factors.

3.4. Path analysis

Path analysis is a subset of Structural Equation Modeling (Golob, 2003). It allows researchers to explore pathway relationships among contribution, intermediate, and dependent variables. Specifically, path analysis links models to reveal the direct and indirect relationships between variables (Xu et al., 2024b). This study measures direct and indirect relationships by average marginal effects (AMEs). The modeling framework is shown in Fig. 2. Two sets of models are trained in this study: (1) Motorists' pre-crash behavior models for identifying the correlates of motorists' pre-crash behavior in police officer-involved crashes, and (2) police officers' injury severity model for exploring the correlates of injury severity of police officers. The motorists' pre-crash behaviors and police officers' injury severity models can be written in simple forms as:

$$Y_1 = f_{precrash}(X_1, X_2) \tag{1}$$

$$Y_2 = f_{injury}(X_2, X_3, Y_1) \tag{2}$$

Equation (1) represented the motorists' pre-crash behavior models. Each contributing variable X_1 (motorists' characteristics) and X_2 (crash information) can affect the probability of motorist pre-crash behaviors Y_1 by the averaged AMEs from five machine learning models. Equation (2) is the simple form of police officers' injury severity models. Similar to Equation (1), contributing variables X_2, X_3 and Y_1 (crash information, police officer characteristics, and motorists' pre-crash behaviors) affect the probability of the injury severity level of police officers. All variables used in path analysis are extracted from the training set.

As mentioned, path analysis links models to reveal direct and indirect relationships by using AMEs. The AME between each independent variable and dependent variable is called the direct relationship. In this study, two types of direct relationships are captured and denoted as $DirectMEonY_1$ and $DirectMEonY_1$, which represent the direct relationships between independent variables and pre-crash behaviors and injury severity of police officers. According to calculated direct relationships, indirect relationships can be captured using the following equation:

$$IndirectME(ofVariableofInterestonY_2) = DirectMEonY_1 \times DirectMEonY_2 \tag{3}$$

$$TotalME\,(ofVariableofInterstonY_2) = DirectMEonY_2 + \sum_{i=1}^{K} IndirectME(ofVariableofInterestonY_2) \tag{4}$$

Each path has a unique indirect correlation. It means that the number of indirect correlations equals the number of paths (K). Based on direct and indirect relationships, the total relationship can be calculated by adding them together (Equation (4)).

This study hypothesized that behavioral pathway relationships exist among crash information, motorists' characteristics, motorists' pre-crash behaviors, and police injury severity of the police-involved crashes in Alabama. For example, motorists' characteristics (e.g., gender and age) and/or crash information (e.g., lighting conditions) may be related to motorists' pre-crash behaviors, such as failure to yield. Simultaneously, motorists' failure to yield is found to be associated with police injury severity. The motorists' characteristics and/or crash information will indirectly affect police injury severity through the motorists' pre-crash behavior of failure to yield.

4. Results

This section summarizes the results of the machine learning and path analysis, presenting the estimated AMEs of variables on motorists' pre-crash behaviors and the injury severity of police officers. According to the averaged estimated AMEs of variables on motorists' pre-crash behaviors and the injury severity of police officers. Based on the averaged

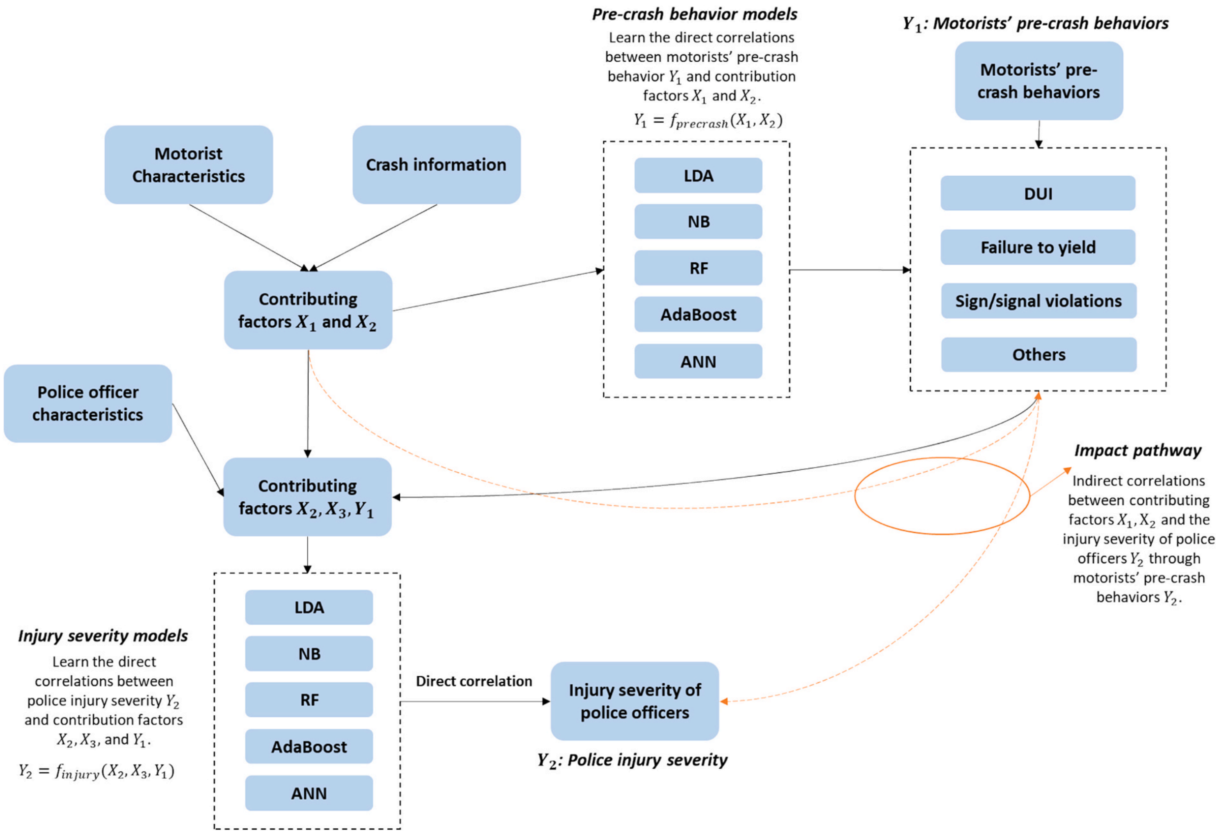

Fig. 2. Two-step path analysis modeling framework with multiple machine learning classifiers.

estimated AMEs of variables on both dependent variables, some variables exhibited a slight relationship (absolute averaged AMEs less than 1.0%) with the dependent variables, while others demonstrated a strong relationship (absolute averaged AMEs greater than 5%). To ensure clarity and conciseness in the results and discussion sections, we highlight variables with absolute averaged AMEs exceeding 2%. This threshold was chosen to balance interpretability and relevance, as a substantial share of key variables exceeded this level (about 40%) while smaller effects are still reported in the appendices. Note that this section shows the averaged AMF of motorists' pre-crash behavior and police injury severity models. The AMF for each machine learning model is shown in Appendixes B and C.

4.1. Motorists' pre-crash behaviors

The models of motorists' pre-crash behaviors included independent variables such as motorists' age, gender, seatbelt usage, lighting conditions, and causal unit speed in crashes. By reviewing the averaged AMEs under each category of pre-crash behaviors, considering the threshold of 2%, every independent variable can be found to have strong relationships with different kinds of pre-crash behavior except for the gender of the motorist. Note that the model results can only indicate whether the relationships between independent and dependent variables were strong or not. It is difficult to infer causation directly from the results. Also, since the variables were extracted from the crash database, the behaviors were recorded only when the crash happened. These behaviors can also lead to the happening of traffic conflicts or do not lead to anything. Thus, the results of the motorists' pre-crash behaviors model can only be used as a reference and be used to estimate the indirect relationships between those independent variables and police injury severity.

4.1.1. Failure to yield

Fig. 3 shows the independent variables that are strongly associated with the pre-crash behavior of motorists' failure to yield. It is obvious that motorists' age, seatbelt usage, lighting conditions, weather, road alignments, and the speed of the causal unit greatly impact the behavior of failure to yield in police-involved crashes.

Socio-demographic and other characteristics related to motorists were found to have a strong relationship with motorists' failure to yield, especially the age of motorists. Motorists who were 18–25, 36–45, and older than 55 years old when the crash happened were more likely to have the pre-crash behavior of failure to yield (the AMEs are 3.4%, 2.5%, and 5.2%) than 26–35 years old. In comparing motorists based on seatbelt usage, those who did not wear seatbelts when the crash happened were found to have a negative relationship with the pre-crash behavior of failure to yield; the AME is 2.2%.

All considered crash-related independent variables were able to affect the probability of motorists' failure to yield when the police-involved crashes happened. Compared to daylight, motorists were 5.0% more likely to fail to yield under the dawn or dusk lighting conditions. Mist or fog weather was also a vital variable that could increase the probability of motorists' failure to yield when police-involved crashes happened. The AMEs of road alignments showed that straight and level roads were more likely to occur in police-involved crashes caused by motorists' failure to yield. The probabilities of motorists' failure to yield for other alignments were from 0.8% to 4.3% lower. Another important variable is the speed of the causal unit. Compared to the speed of 0–10 miles per hour (MPH), except for 11–20 MPH, the probability of motorists' failure to yield when police-involved crashes happened decreased with the increasing speed in general, the range of probability decreasing was from 5.4% to 12.3%. The probability of motorists' failure to yield would be 5.5% higher when the speed of the causal unit is 11–20 MPH compared to the speed of 0–10 MPH. In addition, if the causal unit was stationary, compared to the speed of

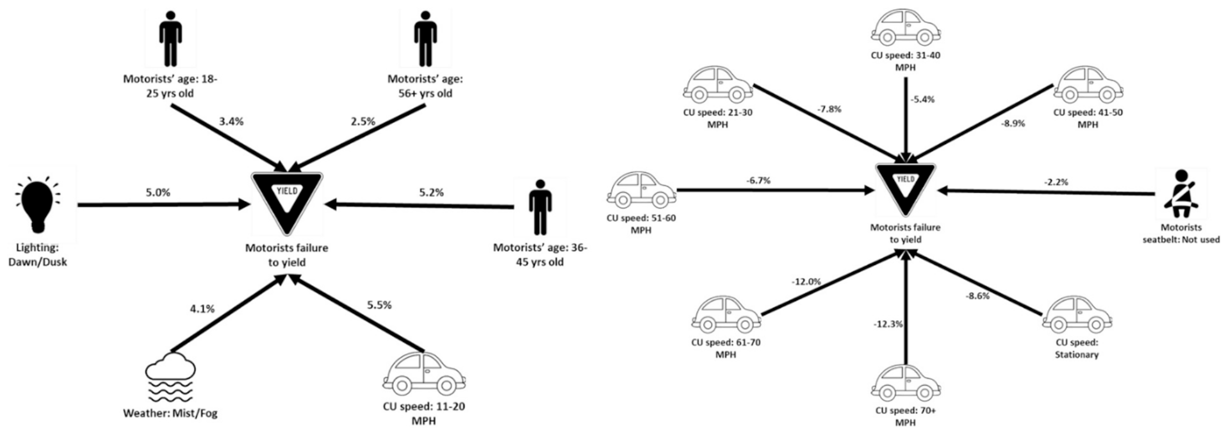

Fig. 3. Direct impact on motorists failure to yield (N = 4,667).

0–10 MPH, the probability of motorists' failure to yield when police-involved crashes happened would be 8.6% lower.

4.1.2. Sign/signal violations

Fig. 4 displays the direct impact of independent variables on the motorists' sign/signal violations when the police-involved crashes happened. Most of them showed a positive significant relationship with the behavior of motorist sign/signal violations when police-involved crashes happened. All of those independent variables are crash-related variables. Compared to daylight, motorists were 2.0% more likely to violate signs or signals when the police-involved crash happened under the lighting condition of dark with continuous light. Similar to the behavior of failure to yield, the probability of motorists violating the sign/signal when the police-involved crash happened on a mist or fog day was higher than on a clear day (the AME is 3.9%). Motorists were most likely to violate the signs or signals when the police-involved crashes happened with the 21–30 MPH causal unit speed compared to 0–10 MPH causal unit speed (the AME is 8.7%). Also, compared to the causal unit speed of 0–10 MPH, motorists were 2.0%, 2.0%, and 2.1% more likely to violate the traffic signs or signals with the speed of 11–20, 31–40, and 41–40 MPH. Curve and level roads were 2.0% less likely to relate to motorists' sign/signal violations than straight and level roads.

4.1.3. DUI

The results of the DUI model are shown in Fig. 4. Motorists' seatbelt usages were found to be significantly related to motorists' DUI when police-involved crashes happened. Motorists who did not wear seatbelts were 6.5% more likely to be identified as DUI than people who wore the seatbelt. Lighting conditions are another vital variable. Compared to daylight, dark (whether with light or not) conditions were found to be associated with motorists' DUI when the police-involved crashes happened, with the AMEs that were higher than 5.5%. In comparing motorists based on the causal unit speed, those who stopped were found to have a positive relationship with DUI when police-involved crashes happened. Motorists were 3.7% more likely to be predicted as a person who DUI when the causal unit speed is 51–60 MPH than the 0–10 MPH speed of the causal unit. With higher causal unit speed (i.e., 61–70 or higher than 70 MPH), the probability of motorists' DUI when police-involved crashes happened increased to 6.7% and 5.4% higher than the 0–10 MPH speed of the causal unit.

4.2. Injury severities of police officers

Most of the independent variables, except the motorists' vehicle types, were found to be meaningfully related to the injury severity of police officers. Unlike motorists' pre-crash behavior models, the relationships between independent and dependent variables for the police injury severity models are more representative. The police could be injured by traffic crashes only when the crash happened. The data extracted from the crash database can relatively completely reflect the relationships captured by police injury severity models.

Table 2 shows the direct relationships between independent

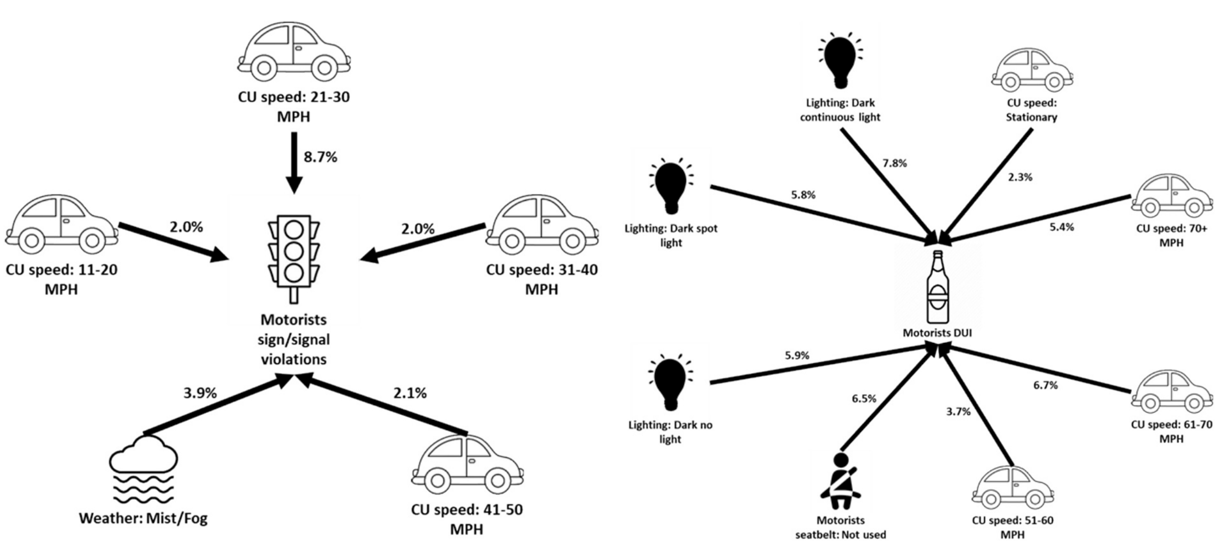

Fig. 4. Direct impacts on motorists sign/signal violations and DUI (N = 4,667).

Table 2
Averaged direct impacts on police injury severity (Number of samples in the training set = 3,734).

Y = Police injury severity Variable		PDO	C	B	K & A
Police age (base: 26–35 years old)	18–25 years old	−1.7%	0.1%	1.4%	0.4%
	36–45 years old	0.0%	0.1%	0.1%	0.1%
	46–55 years old	−2.7%	1.1%	0.9%	0.7%
	56 + years old	0.4%	0.3%	−1.1%	0.1%
Police gender (base: Not female)	Female	−2.9%	2.3%	1.1%	0.0%
Emergency (base: Non-emergency)	Emergency	−3.0%	1.4%	1.7%	0.4%
Police vehicle type (base: Passenger car)	SUV	1.7%	−1.4%	−0.1%	−0.1%
	Pickup	3.6%	−1.2%	−1.4%	−1.2%
	Other	0.9%	0.0%	−0.2%	−0.5%
Motorist vehicle type (base: Passenger car)	SUV	−1.4%	1.3%	0.3%	0.1%
	Pickup	−1.0%	1.5%	−0.2%	−0.2%
	Other	0.0%	0.6%	−0.4%	−0.3%
Motorists' pre-crash behaviors (base: Other behaviors)	Fail to yield	−12.9%	5.4%	6.6%	1.0%
	Sign/Signal violation	−17.0%	7.6%	7.7%	1.7%
	DUI	−15.1%	8.3%	3.5%	2.6%
Police Seatbelt (base: used)	Not used	−23.8%	6.7%	12.1%	5.4%
	Other	−18.8%	9.6%	6.8%	0.4%
Crash location (base: On road)	Intersection	−9.6%	−0.6%	4.0%	6.5%
	Median	−8.8%	4.4%	4.3%	−0.3%
	Off road	1.9%	−2.3%	1.2%	−0.5%
	Other	2.9%	−1.8%	−0.6%	−0.5%
	Shoulder/ Roadside	2.3%	−1.7%	0.0%	−0.3%
Causal unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−2.6%	2.3%	0.9%	0.1%
	21–30 MPH	−1.2%	1.1%	0.5%	−0.2%
	31–40 MPH	−5.6%	3.7%	1.1%	0.6%
	41–50 MPH	−12.4%	4.9%	6.5%	0.7%
	51–60 MPH	−10.7%	5.0%	3.9%	1.3%
	61–70 MPH	−16.4%	6.6%	8.3%	1.1%
	70 + MPH	−17.4%	2.8%	9.7%	3.6%
	Stationary	−1.1%	0.2%	2.1%	−0.4%
Lighting condition (base: Daylight)	Dawn/Dusk	−6.0%	4.4%	0.6%	−0.2%
	Dark continuous light	−3.4%	1.5%	1.3%	0.7%
	Dark spot light	−2.3%	−0.1%	1.8%	0.6%
	Dark no light	−3.6%	−0.7%	2.9%	1.8%
	Other	−1.6%	2.4%	−2.2%	−0.3%
Weather (base: Clear)	Cloudy	−1.2%	0.7%	0.1%	0.4%
	Rain	−0.7%	1.0%	0.1%	−0.3%
	Mist/Fog	1.8%	−2.0%	−0.4%	0.7%
	Other	5.4%	−2.2%	−2.6%	−0.5%
Road alignment (base: Straight and level)	Straight with down grade	−0.9%	0.4%	−0.3%	1.1%
	Straight with up grade	1.5%	0.0%	−1.6%	0.2%
	Curve and level	−2.1%	−2.6%	4.8%	0.2%
	Curve and grade	−3.5%	−1.0%	1.7%	2.8%
	Other	−2.8%	−0.3%	1.9%	1.0%
Collision manner (base: Rear-end)	Head on	−5.8%	−1.7%	1.8%	4.2%
	CU backing	5.3%	−2.4%	−3.1%	−0.7%
	Angle	−1.6%	1.7%	0.7%	0.1%
	Sideswipe (same direction)	4.9%	−1.9%	−2.5%	−0.3%
	Side impact	−5.3%	2.1%	2.6%	0.6%
	Other	0.8%	−1.8%	1.6%	−0.2%

variables and police injury severity. Compared to police officers who were 26–35 years old when the crash happened, 46–55-year-old police officers were 2.7% less likely to experience a property damage-only (PDO) crash. Their injury severity would be higher than PDO. Female police officers were found to be less likely to experience a PDO crash than males. They had a 2.3% higher probability of being reported as possible injuries. Emergency is another vital contribution variable to police injury severity. The modeling results show that if police officers were in an emergency situation, they were more likely to suffer from more severe injury than in a non-emergency situation when they meet traffic crashes. Compared to passenger cars, if police vehicles were pickups, the police officers (the driver of the pickup) were positively associated with PDO. Seatbelt usage is a critical issue for police officers. Police officers, compared to those who wore seatbelts, were found to be 6.7% more likely to suffer a possible injury, 12.1% higher probability of suffering a minor injury, and 5.4% more likely to be killed or suffer a severe injury when they did not wear the seatbelt.

Motorists' pre-crash behaviors were the key factor in the injury severity of police officers when police-involved crashes happen. Police officers were 5.4% or 6.6% more likely to suffer a possible or minor injury in the crash when the primary circumstance of the crash was motorists' failure to yield compared to other behaviors. Also, motorists' violation of signs or signals were 7.6% more likely to lead a possible injury for police officers. This behavior can also contribute to the high probability of minor injury to police officers. The most dangerous pre-crash behavior of motorists for police officers was DUI. Motorists' DUI was found to be 8.3% more likely to lead a possible injury, 3.5% more likely to lead a minor injury, and 2.6% more likely to lead a severe or fatal injury of police officers than other behaviors.

Intersections and medians are the top two dangerous crash locations for police officers. The result shows that, compared to the police-involved crashes that happened on the road, crashes that happened at intersections had 4.0% and 6.5% higher probabilities of leading to a minor injury and severe or fatal injuries for police officers. The harm of median-happened crashes was concentrated on the possible and minor injuries for police officers. They were 4.4% more likely to suffer a possible injury and 4.3% more likely to suffer a minor injury in the median-happened crashes than the on-road-happened crashes. The speed of the causal unit is one of the most important variables related to police injury severity. In general, the probability of police officers suffering from injuries increases with the increasing speed of the causal unit. Compared to the causal units, which drove under the speed of 0–10 MPH, causal units with 11–20 MPH and 31–40 MPH were 2.3% and 3.7% more likely to lead to the possible injury of police officers. When the speed of causal units was higher, the police officers started to have a higher probability of suffering possible or minor injuries. Causal units with a speed of faster than 70 MPH were found to be 9.7% and 3.6% more likely to lead to minor and severe or fatal injuries to police officers than causal units, which drove under the speed of 0–10 MPH. Compared to the causal units, which drove under the speed of 0–10 MPH, stationary causal units would be able to lead to a 2.1% higher probability that police officers suffered from a minor injury.

In comparing police injury severity based on lighting conditions, police officers who met the crash under dawn or dusk lighting conditions were found to have a positive relationship with possible injury. Also, if the police-involved crashes happened under dark lighting conditions, their injury chances were increased, especially under the no light dark condition (2.9% more likely to lead to a minor injury). Compared to clear weather, police-involved crashes that happened in mist or fog weather were positively associated with PDO crashes. Curve roads were found to be more risky than straight roads. Police officers were 4.8% more likely to suffer a minor injury on curved and level roads than on straight and level roads. If the curved roads are graded, police officers are 2.8% more likely to suffer a severe or fatal injury than on straight and level roads. Head-on and side-impact crashes were found to be more dangerous for police officers than rear-ended crashes. If police officers were involved in head-on crashes, the probability that they suffered a severe or fatal injury was 4.2% higher than police officers who were involved in rear-ended crashes. Side-impact crashes were more likely to lead to possible minor injuries for police officers than rear-ended crashes. Sideswipe crashes and crashes caused by the backing behavior of causal units were found to have 4.9% and 5.3% higher probabilities that lead to PDO crashes.

4.3. Path analysis

To clearly present and discuss the path analysis results, only the PDO path is shown in Table 3. The total impact on police injury severity is influenced by multiple paths, as revealed by the analysis. This underscores the complex nature of police injury severity and pre-crash behaviors.

The results demonstrate that the indirect impact of older motorists (older than 55 years old) on the PDO of police officers is −0.7%. Also, motorists who did not use seatbelts were found to be 0.8% less likely to associate with the outcome of PDO for police officers in the crash. For some variables that are related to both police injury severity and motorists' pre-crash behaviors, the total impact may be strengthened or eliminated by different behavioral pathways. For instance, the absolute value of the relationships between PDO crashes and speed levels 11–20 and 21–30 MPH increased after considering indirect impacts. Simultaneously, the total impacts of the higher speed level and stationary on

Table 3
Path analysis results of police injury severity – PDO (Number of sample in the training set = 3,734).

Y = PDO Variable		Direct impact Average	Indirect impact Fail to yield	 Sign or signal violation	 DUI	Total indirect impact	Total impact
Motorists' pre-crash behaviors (base: Other behaviors)	Fail to yield	−12.9%				0.0%	−12.9%
	Sign/Signal violation	−17.0%				0.0%	−17.0%
	DUI	−15.1%				0.0%	−15.1%
Police age (base: 26–35 years old)	18–25 years old	−1.7%				0.0%	−1.7%
	36–45 years old	0.0%				0.0%	0.0%
	46–55 years old	−2.7%				0.0%	−2.7%
	56 + years old	0.4%				0.0%	0.4%
Police gender (base: Not female)	Female	−2.9%				0.0%	−2.9%
Emergency (base: Non-emergency)	Emergency	−3.0%				0.0%	−3.0%
Police vehicle type (base: Passenger car)	SUV	1.7%				0.0%	1.7%
	Pickup	3.6%				0.0%	3.6%
	Other	0.9%				0.0%	0.9%
Motorist age (base: 26–35 years old)	18–25 years old		−0.5%	−0.1%	0.1%	−0.4%	−0.4%
	36–45 years old		−0.3%	0.3%	0.0%	−0.1%	−0.1%
	46–55 years old		0.1%	0.1%	−0.1%	0.0%	0.0%
	56 + years old		−0.7%	−0.2%	0.1%	−0.7%	−0.7%
Motorist gender (base: Not female)	Female		−0.1%	−0.1%	0.2%	0.0%	0.0%
Motorist vehicle type (base: Passenger car)	SUV	−1.4%	0.0%	0.1%	0.1%	0.2%	−1.2%
	Pickup	−1.0%	−0.1%	0.1%	−0.2%	−0.1%	−1.1%
	Other	0.0%	0.4%	0.1%	0.3%	0.8%	0.9%
Police Seatbelt (base: used)	Not used	−23.8%				0.0%	−23.8%
	Other	−18.8%				0.0%	−18.8%
Motorist Seatbelt (base: used)	Not used		0.2%	−0.1%	−0.8%	−0.8%	−0.8%
	Other		0.3%	0.2%	0.0%	0.4%	0.4%
Crash location (base: On road)	Intersection	−9.6%				0.0%	−9.6%
	Median	−8.8%				0.0%	−8.8%
	Off road	1.9%				0.0%	1.9%
	Other	2.9%				0.0%	2.9%
	Shoulder/Roadside	2.3%				0.0%	2.3%
Causal unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−2.6%	−0.7%	−0.3%	−0.1%	−1.2%	−3.7%
	21–30 MPH	−1.2%	0.8%	−1.0%	−0.1%	−0.3%	−1.5%
	31–40 MPH	−5.6%	0.7%	−0.4%	0.1%	0.4%	−5.3%
	41–50 MPH	−12.4%	1.2%	−0.4%	−0.3%	0.6%	−11.8%
	51–60 MPH	−10.7%	0.9%	0.1%	−0.5%	0.5%	−10.1%
	61–70 MPH	−16.4%	1.4%	0.1%	−0.8%	0.7%	−15.7%
	70 + MPH	−17.4%	1.4%	0.1%	−0.7%	0.8%	−16.6%
	Stationary	−1.1%	1.2%	0.1%	−0.3%	1.1%	0.0%
Lighting condition (base: Daylight)	Dawn/Dusk	−6.0%	−0.7%	0.1%	−0.1%	−0.7%	−6.7%
	Dark continuous light	−3.4%	0.2%	−0.4%	−0.9%	−1.1%	−4.5%
	Dark spot light	−2.3%	0.1%	−0.2%	−0.7%	−0.8%	−3.1%
	Dark no light	−3.6%	0.0%	0.2%	−0.8%	−0.5%	−4.2%
	Other	−1.6%	0.5%	0.1%	−0.3%	0.4%	−1.2%
Weather (base: Clear)	Cloudy	−1.2%	−0.1%	0.1%	0.1%	0.1%	−1.1%
	Rain	−0.7%	−0.1%	0.0%	0.2%	0.1%	−0.6%
	Mist/Fog	1.8%	−0.5%	−0.7%	0.2%	−1.0%	0.8%
	Other	5.4%	0.3%	−0.1%	0.2%	0.4%	5.8%
Road alignment (base: Straight and level)	Straight with down grade	−0.9%	0.3%	0.1%	0.2%	0.5%	−0.4%
	Straight with up grade	1.5%	0.1%	0.1%	0.0%	0.3%	1.8%
	Curve and level	−2.1%	0.3%	0.3%	0.0%	0.5%	−1.5%
	Curve and grade	−3.5%	0.5%	0.1%	−0.2%	0.4%	−3.1%
	Other	−2.8%	−1.5%	0.2%	−0.2%	−1.5%	−4.3%
Collision manner (base: Rear-end)	Head on	−5.8%				0.0%	−5.8%
	CU backing	5.3%				0.0%	5.3%
	Angle	−1.6%				0.0%	−1.6%
	Sideswipe (same direction)	4.9%				0.0%	4.9%
	Side impact	−5.3%				0.0%	−5.3%
	Other	0.8%				0.0%	0.8%

police officers who were suffering from injuries were eliminated by 0.4% to 1.1% compared to direct impact. The total impacts of different lighting conditions on PDO are lower than direct impacts, particularly for dark conditions with continuous light (−1.1% indirect impact). This means that the impact of lighting these lighting conditions on police officers suffering from injuries was enhanced. Except for the weather conditions of mist or fog, other weather conditions were not affected significantly by indirect impacts. The total impact of mist or fog on PDO crashes was decreased from 1.8% to 0.8%. Considering indirect impacts, the total impacts of road alignments on PDO increased by 0.4% on average compared to direct impacts, with consistent directional signs.

5. Discussion

The motorist pre-crash behavior models provided some valuable insights that related to motorist's pre-crash behaviors in police-involved crashes. Lighting and weather conditions were found to be significantly related to motorists' pre-crash behaviors. Specifically, under dawn or dusk lighting conditions and mist or fog weather conditions, motorists were more likely to fail to yield to police vehicles, leading to police-involved crashes. Mist or fog is also a key weather condition that is positively associated with motorists' likelihood of violating traffic signs and signals. DUI was highly related to dark lighting conditions, including dark with no light, dark with the spotlight, and dark with continuous light. The aforementioned findings indicated that, in police-involved crashes, visibility is a critical element. These findings suggest that better visibility conditions are associated with a lower likelihood of motorists engaging in risky pre-crash behaviors. In this case, to prevent these pre-crash behaviors from happening, increasing self-visibility for police officers is necessary and important. In low visibility situations, Dynamic Message Signs (DMS) and Variable Speed Limit (VSL) would be useful. Furthermore, drivers are encouraged to take online courses or training to develop appropriate driving behavior under low visibility conditions. Note that for DUI behavior, the association is not only related to low visibility. Reduced visibility at night may interact with the impairing effects of alcohol. Increased blood alcohol concentration can impair drivers' visual functions, coordination, and ability to track moving objects,[2] which may elevate the risk of crashes during nighttime or low-visibility conditions. In addition, nighttime often overlaps with social drinking hours, which may contribute to a higher frequency of DUI behavior during this period.

Previous literature suggested that male and young drivers were more likely to drive with some risky behaviors, such as DUI and different kinds of aggressive driving (Yang et al., 2022). However, the modeling results did not agree with previous findings. These differences may be influenced by the characteristics of the crash dataset and the modeling methods used. For example, the majority of DUI-involved drivers in this study were over 25 years old, and gender effects were not consistently captured across all models due to differing learning mechanisms. Also, it might be because the police vehicles can warn and make drivers calm down when they are driving. It is unknown if the correlations between variables and motorists' pre-crash behavior will be similar after removing the restraint about police-involved crashes only in the training and testing datasets.

According to the injury severity models and path analysis results, several variables were found to be significantly associated with more severe injuries for police officers. These variables include seatbelt usage, intersections, dark conditions without any lighting, curved and graded roads, and pre-crash behaviors. From the perspective of police officer injury prevention, it is essential to identify the underlying reasons for these associations and propose effective countermeasures.

Motorists' pre-crash behaviors, including failure to yield, sign or signal violations, and DUI, are strongly linked to more severe injuries for police officers, especially possible and minor injuries. Failure to yield increases the likelihood of officer injuries by 12.9%, likely due to a high non-compliance rate with Move-over Laws in Alabama. The efforts of guiding people to follow the Move-over Laws, which means that motorists will move over or slow down for emergency vehicles and people on the side of the road, will be helpful in mitigating behaviors like failure to yield. The modeling results of pre-crash behaviors also support this point of view. The AMEs of road alignments showed that straight and level roads were more likely to occur in police-involved crashes caused by motorists' failure to yield. This finding may be explained by drivers paying less attention or making mistakes on straight and level roads. These roads may lead to less careful scanning for other vehicles, poor judgment of how fast another vehicle is approaching, or distraction from the flashing lights of police vehicles. All of these factors can increase the chance of failing to yield. Also, increasing the perceptibility of police vehicles might be a useful way to help drivers notice the working police vehicle and then move over. Research conducted by Carrick and Washburn (2012) indicated that using blue and red emergency lights is an efficient way to increase compliance with the Move-over Law. However, they also found that the compliance rate of slowing down when not moving over was quite low. This would be another issue that needs to be addressed through education, publicity, and enforcement to protect police officers.

Simultaneously, motorists who violate traffic signs or signals raise the risk of officer injuries by 17.0%. This finding is highly associated with the relationship between intersection and police injury severity. Similar to the research findings of Penmetsa and Pulugurtha (2017), traffic signs or signal violations will increase the injury severity of the innocent party of drivers, even though the outcome of sign or signal violations is not that severe compared to other violations. To protect police officers, other drivers, and roadside responders/workers, it is essential to prevent traffic signs or signal violation behavior. According to the results of the pre-crash behavior model, it is clear that Alabamians did not take traffic signs or signal violations seriously at low and medium speeds. Enhancing the enforcement strength by applying some equipment and strategies, such as red-light cameras, high visibility enforcement programs, and automated enforcement systems (Cohn et al., 2020; Li et al., 2020; Pantangi et al., 2020), for the roads with low and medium speed limits (e.g., less or equal to 45 MPH) might be a potential solution.

Motorists' DUIs are another big issue. Police officers who were involved in crashes were 2.6% more likely to be killed or suffer from severe injuries when motorists DUI. DUI was a common problem across the United States. The best way to mitigate the loss caused by DUI is to reduce the number of DUI behaviors. To prevent DUI behaviors, public safety campaigns, legal sanctions, and education were widely adopted (Wyatt & Novotna, 2021). Alabama has a law to restrain DUI (Alabama Code Title 32. Motor Vehicles and Traffic § 32-5A-191 | FindLaw, n.d.). However, as Stringer (2019) mentioned, the effectiveness of laws and rules will be undermined if people think the certainty of punishment is low. Family-based prevention and intervention might generate slow but long-term positive results in mitigating DUI. Also, research indicated that safe-ride programs (e.g., designating drivers before and after drinking and providing alternative transportation options) would positively impact on reducing DUI rate (Gieck & Slagle, 2010).

6. Limitations

This study used Alabama crash data, which was extracted from crash reports. The accuracy of crash data people collect determines the quality of results. It is possible that the crash reporters were provided with inaccurate information by witnesses and the crash involved because the crash reporters might not have been at the scene of the crash when the collision happened. One another limitation of this study is the lack of detailed pre-crash scenario data, such as vehicle maneuvers, trajectory

[2] https://www.nhtsa.gov/risky-driving/drunk-driving#alcohol-abuse-and-cost-5091.

paths, or driver intent, which could further enhance understanding of police injury severity in traffic crashes. While we included contributing circumstances (i.e., motorist pre-crash behaviors) and manner of collision as proxy indicators, these do not fully capture the complexity of pre-crash interactions. Future research incorporating more granular scenario-level data, if available, could provide a deeper behavioral perspective on police-involved crashes. For the models of pre-crash behaviors, some of the estimated relationships cannot represent the causal relationship. Because the dataset only includes the pre-crash information of the crashes that happened. There are still lots of uncaptured samples of motorists' pre-crash behaviors without any crashes. Also, this study only considered the potential behavioral pathway among independent variables, motorists' pre-crash behaviors, and the injury severity of police officers. In addition, this study adopted multiple machine learning methods to identify variables of interest. The unbalanced issue of the dataset may have affected the performance of the models.

7. Conclusions

This study aimed to understand the roles of motorists' pre-crash behaviors in police-involved traffic crashes. Utilizing data on police-involved crashes from 2017 to 2021 in Alabama, this study employed a path analysis framework integrated with interpretable machine learning techniques to examine the behavioral pathways from contributing factors to motorists' pre-crash behaviors, ultimately leading to the injury severity of police officers. The findings of this study effectively address existing research gaps and provide valuable insights for law enforcement agencies and organizations. These insights can inform the development of targeted strategies to enhance the safety of law enforcement personnel on the road, thereby improving overall traffic safety and positively affecting all road users.

Key results from models for the motorists' pre-crash behaviors in police-involved crashes showed that 18–25 years old, 36–45 years old, and motorists who were older than 55 years old were more likely to have the pre-crash behavior of failure to yield. Those who did not wear seatbelts when the crash happened were found to have a negative relationship with the pre-crash behavior of failure to yield. Motorists under mist or fog, straight and level roads, dawn or dusk lighting conditions, and causal unit speed of 11–20 MPH were more likely to fail to yield to police vehicles. On the contrary, the causal unit speed was higher than 20 MPH, and the causal unit stationary was negatively related to motorists' failure to yield to police vehicles. Mist or fog was also found to be related to motorists who violated traffic signs and signals. Compared to 0–10 MPH, when the causal unit speed is between 11–50 MPH, motorists were more likely to violate traffic signs and signals. Motorists' DUI was highly related to lighting conditions, causal unit speed, and seatbelt usage of motorists. Dark conditions with no light, dark with a spotlight, and dark with continuous light conditions were positively related to motorists' DUI. With higher casual unit speed (i.e., 51 MPH or higher), the probability of motorists' DUI when police-involved crashes happened was higher than the 0–10 MPH speed of the causal unit. Motorists who did not wear the seatbelt were more likely to be identified as DUI than people who wore the seatbelt.

According to the police injury severity modeling results, this study found that police officers who did not wear seatbelts were more likely to suffer from more severe injuries. Motorists' pre-crash behaviors (i.e., failure to yield, traffic sign/signal violations, and DUI) were positively associated with more severe injuries. Compared to other locations, the intersection is the most dangerous crash location for police officers. A faster speed of causal unit shows a higher probability of suffering from more severe injuries. Dark no-light conditions and curves with grade roads were more likely to lead police officers to suffer from more severe injuries. Compared to rear-end crashes, head-on crashes were the riskiest collision manner for police officers. The path analysis results indicated that motorists who were older than 55 years old and motorists who did not wear seatbelts, mist or fog, speed, and road alignments were able to impact police officer injury severity indirectly through motorists' pre-crash behaviors.

This study provides insights into the behavioral pathways in police-involved traffic crashes. Key variables that were directly or indirectly related to increased injury severity of police officers are identified. The research results suggested that law enforcement agencies and organizations should focus on mitigating failure to yield, traffic signs or signal violations, and DUI to protect police officers on the roadway. Several strategies, such as enhancing the enforcement of the Move-over Law, enhancing the enforcement strength by applying some equipment and strategies for the roads with low and medium speed limits (e.g., less or equal to 45 MPH), and employing safe-ride programs, could be referenced by the law enforcement agencies and organizations. Moreover, increasing the visibility of police officers and vehicles during low visibility scenarios might be helpful for preventing motorists from failing to yield to police vehicles and violating traffic signs and signals. Future studies could adopt data with better quality and enlarge the research areas, such as focusing on different states in the United States, to acquire more generated and accurate results. Also, integrating this study with some other modeling technology, like geographically and temporally weighted regression, could provide the opportunity to investigate the localized behavioral pathway and propose more targeted analysis and suggestions.

Declaration of Generative AI in Scientific Writing

During the preparation of this work the authors used ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the published article.

CRediT authorship contribution statement

Ningzhe Xu: Writing – original draft, Formal analysis, Conceptualization. **Jun Liu:** Writing – review & editing, Formal analysis, Conceptualization. **Zihe Zhang:** Writing – original draft, Formal analysis. **Steven Jones:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Centers for Disease Control and Prevention (CDC) for their support through Grant #R01OH012695, and the Alabama Transportation Institute (ATI) at the University of Alabama for additional support. Special appreciation is extended to the Center for Advanced Public Safety (CAPS) at the University of Alabama for providing the data used in this analysis. ChatGPT was utilized solely for editing and grammar-checking during the preparation of this manuscript. The views expressed in this paper are those of the authors, who are solely responsible for the accuracy and interpretation of the information presented.

Appendix

Appendix A.1. Confusion matrixes for each model (Number of samples in the testing set = 933)

Models	Pre-crash behavior models					Police injury severity models				
	Actual values	Predicted values				Actual values	Predicted values			
		Other	Fail to yield	Sign/Signal Violation	DUI		O	C	B	K&A
RF	Other	749	46	2	3	O	811	3	5	0
	Fail to yield	87	6	0	0	C	50	1	2	0
	Sign/Signal Violation	27	1	0	0	B	47	1	2	0
	DUI	31	1	0	0	K&A	11	0	0	0
ANN	Other	777	3	0	0	O	816	3	0	0
	Fail to yield	93	0	0	0	C	53	0	0	0
	Sign/Signal Violation	28	0	0	0	B	50	0	0	0
	DUI	32	0	0	0	K&A	10	1	0	0
AdaBoost	Other	708	47	8	17	O	764	23	30	2
	Fail to yield	82	11	0	0	C	47	1	4	1
	Sign/Signal Violation	26	2	0	0	B	44	3	3	0
	DUI	28	1	0	3	K&A	8	1	2	0
NB	Other	777	1	0	2	O	806	1	10	2
	Fail to yield	92	0	0	1	C	51	1	1	0
	Sign/Signal Violation	28	0	0	0	B	45	1	4	0
	DUI	31	0	0	1	K&A	9	0	1	1
LDA	Other	772	2	2	4	O	793	2	6	18
	Fail to yield	92	0	0	1	C	51	1	1	0
	Sign/Signal Violation	27	1	0	0	B	42	3	4	1
	DUI	29	0	0	3	K&A	7	1	0	3

Appendix A.2. Comparison of performance metrics

Pre-crash behavior models	RF	ANN	AdaBoost	NB	LDA	Police officer injury severity models	RF	ANN	AdaBoost	NB	LDA
Accuracy	0.7922	0.8153	0.7576	0.8164	0.8132	Accuracy	0.8725	0.8746	0.8232	0.8703	0.8585
Precision (other)	0.8378	0.8355	0.8389	0.8373	0.8391	Precision (O)	0.8825	0.8784	0.8853	0.8847	0.8880
Recall (other)	0.9363	0.9962	0.9077	0.9962	0.9897	Recall (O)	0.9902	0.9963	0.9328	0.9841	0.9683
F1 (other)	0.8843	0.9088	0.8719	0.9098	0.9082	F1 (O)	0.9333	0.9336	0.9084	0.9318	0.9264
Precision (fail to yield)	0.1111	0.0000	0.1803	0.0000	0.0000	Precision (C)	0.2000	0.0000	0.0357	0.3333	0.1429
Recall (fail to yield)	0.0645	0.0000	0.1183	0.0000	0.0000	Recall (C)	0.0189	0.0000	0.0189	0.0189	0.0189
F1 (fail to yield)	0.0816	0.0000	0.1429	0.0000	0.0000	F1 (C)	0.0345	0.0000	0.0247	0.0357	0.0333
Precision (violation)	0.0000	0.0000	0.0000	0.0000	0.0000	Precision (B)	0.2222	0.0000	0.0769	0.2500	0.3636
Recall (violation)	0.0000	0.0000	0.0000	0.0000	0.0000	Recall (B)	0.0400	0.0000	0.0600	0.0800	0.0800
F1 (violation)	0.0000	0.0000	0.0000	0.0000	0.0000	F1 (B)	0.0678	0.0000	0.0674	0.1212	0.1311
Precision (DUI)	0.0000	0.0000	0.1500	0.2500	0.3750	Precision (K&A)	0.0000	0.0000	0.0000	0.3333	0.1364
Recall (DUI)	0.0000	0.0000	0.0938	0.0313	0.0938	Recall (K&A)	0.0000	0.0000	0.0000	0.0909	0.2727
F1 (DUI)	0.0000	0.0000	0.1154	0.0556	0.1500	F1 (K&A)	0.0000	0.0000	0.0000	0.1429	0.1818
Macro precision	0.2372	0.2089	0.2923	0.2718	0.3035	Macro precision	0.3262	0.2196	0.2495	0.4504	0.3827
Macro recall	0.2502	0.2490	0.2799	0.2569	0.2709	Macro recall	0.2623	0.2491	0.2529	0.2935	0.3350
Macro F1	0.2415	0.2272	0.2825	0.2413	0.2646	Macro F1	0.2589	0.2334	0.2501	0.3079	0.3182

Appendix B.1. Direct impacts on pre-crash behaviors according to the RF model (Number of sample in the training set = 3,734)

Y = Motorist pre-crash behavior		Average Marginal Effects			
Variable		Other precrash behavior	Fail to yield	Sign/signal violation	DUI
Motorist age (base: 26–35 years old)	18–25 years old	−2.1%	3.1%	0.5%	−1.5%
	36–45 years old	−3.0%	3.2%	−0.3%	0.0%
	46–55 years old	−2.9%	0.1%	1.1%	1.7%
	56 + years old	−7.3%	6.7%	0.8%	−0.3%
Motorist gender (base: Not female)	Female	−0.7%	1.1%	0.4%	−0.7%
Motorist vehicle type (base: Passenger car)	SUV	−2.0%	1.7%	0.5%	−0.3%
	Pickup	−7.7%	5.6%	0.7%	1.4%
	Other	−1.7%	3.6%	−0.5%	−1.4%
Motorist Seatbelt (base: used)	Not used	−12.9%	5.8%	1.3%	5.7%
	Other	−7.6%	3.1%	0.7%	3.8%
Lighting condition (base: Daylight)	Dawn/Dusk	−14.2%	10.7%	0.9%	2.6%
	Dark continuous light	−13.9%	4.3%	2.5%	7.1%
	Dark spot light	−9.7%	3.2%	1.1%	5.4%
	Dark no light	−12.9%	7.5%	0.1%	5.2%
	Other	−7.6%	3.5%	0.1%	4.0%
Weather (base: Clear)	Cloudy	−3.7%	3.2%	0.5%	0.0%
	Rain	−7.6%	7.3%	1.1%	−0.8%
	Mist/Fog	−14.9%	10.9%	4.6%	−0.6%

(continued on next page)

(*continued*)

Y = Motorist pre-crash behavior Variable		Average Marginal Effects Other precrash behavior	Fail to yield	Sign/signal violation	DUI
	Other	−8.1%	5.6%	3.2%	−0.6%
Road alignment (base: Straight and level)	Straight with down grade	−1.1%	0.9%	0.8%	−0.7%
	Straight with up grade	−5.6%	4.4%	0.6%	0.6%
	Curve and level	−1.6%	2.0%	−0.6%	0.2%
	Curve and grade	−3.7%	2.0%	0.3%	1.5%
	Other	−14.0%	12.3%	−0.6%	2.3%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−9.1%	5.9%	1.9%	1.3%
	21–30 MPH	−4.1%	−3.2%	6.0%	1.3%
	31–40 MPH	0.8%	−3.1%	2.1%	0.2%
	41–50 MPH	1.6%	−6.8%	2.3%	3.0%
	51–60 MPH	0.4%	−4.3%	0.0%	3.9%
	61–70 MPH	1.8%	−7.9%	−0.2%	6.3%
	70 + MPH	4.7%	−7.9%	−0.1%	3.4%
	Stationary	3.1%	−5.5%	−0.1%	2.5%

Appendix B.2. Direct impacts on pre-crash behaviors according to the NB model (Number of sample in the training set = 3,734)

Y = Motorist pre-crash behavior Variable		Average Marginal Effects Other precrash behavior	Fail to yield	Sign/signal violation	DUI
Motorist age (base: 26–35 years old)	18–25 years old	−8.0%	7.2%	0.9%	−0.1%
	36–45 years old	−4.3%	4.3%	−1.4%	1.3%
	46–55 years old	−3.2%	1.8%	0.0%	1.3%
	56 + years old	−7.3%	7.7%	0.8%	−1.3%
Motorist gender (base: Not female)	Female	−0.2%	1.8%	0.8%	−2.4%
Motorist vehicle type (base: Passenger car)	SUV	2.1%	−0.5%	−0.3%	−1.3%
	Pickup	−0.2%	−0.2%	−0.8%	1.2%
	Other	12.7%	−8.2%	−1.1%	−3.4%
Motorist Seatbelt (base: used)	Not used	−3.1%	−3.7%	0.7%	6.1%
	Other	15.6%	−11.2%	−2.4%	−2.0%
Lighting condition (base: Daylight)	Dawn/Dusk	−5.2%	4.3%	−0.2%	1.2%
	Dark continuous light	−7.8%	−1.6%	2.1%	7.3%
	Dark spot light	−5.0%	−2.3%	1.3%	6.1%
	Dark no light	−1.2%	−4.0%	−1.5%	6.8%
	Other	11.1%	−10.5%	−2.0%	1.4%
Weather (base: Clear)	Cloudy	0.9%	0.6%	−0.8%	−0.7%
	Rain	1.8%	0.6%	−0.1%	−2.3%
	Mist/Fog	−6.2%	4.0%	4.0%	−1.8%
	Other	10.1%	−5.2%	−1.9%	−3.0%
Road alignment (base: Straight and level)	Straight with down grade	7.7%	−5.5%	−0.6%	−1.5%
	Straight with up grade	3.7%	−2.6%	−1.1%	0.0%
	Curve and level	7.0%	−5.0%	−2.6%	0.6%
	Curve and grade	7.7%	−8.6%	−0.9%	1.8%
	Other	−3.8%	4.9%	−2.5%	1.4%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−8.8%	5.3%	2.4%	1.1%
	21–30 MPH	0.2%	−8.7%	6.8%	1.7%
	31–40 MPH	4.7%	−6.8%	2.4%	−0.3%
	41–50 MPH	5.9%	−11.9%	2.6%	3.4%
	51–60 MPH	5.8%	−9.1%	−1.0%	4.3%
	61–70 MPH	9.8%	−15.2%	−0.9%	6.3%
	70 + MPH	9.3%	−15.1%	−0.9%	6.7%
	Stationary	10.4%	−12.8%	−0.9%	3.2%

Appendix B.3. Direct impacts on pre-crash behaviors according to the AdaBoost model (Number of sample in the training set = 3,734)

Y = Motorist pre-crash behavior Variable		Average Marginal Effects Other precrash behavior	Fail to yield	Sign/signal violation	DUI
Motorist age (base: 26–35 years old)	18–25 years old	1.2%	2.9%	−1.1%	−3.1%
	36–45 years old	1.7%	2.5%	−8.2%	4.0%
	46–55 years old	1.5%	0.5%	−2.8%	0.8%
	56 + years old	0.5%	0.2%	0.8%	−1.5%
Motorist gender (base: Not female)	Female	0.5%	0.4%	2.4%	−3.3%
Motorist vehicle type (base: Passenger car)	SUV	0.4%	−0.5%	−4.8%	5.0%
	Pickup	0.5%	0.9%	−2.4%	1.0%
	Other	2.4%	−0.4%	0.6%	−2.6%
Motorist Seatbelt (base: used)	Not used	−1.1%	−3.8%	−2.9%	7.7%
	Other	0.1%	−1.6%	−1.2%	2.7%
Lighting condition (base: Daylight)	Dawn/Dusk	−0.6%	−2.0%	−3.6%	6.1%
	Dark continuous light	−3.6%	−6.8%	−0.7%	11.0%
	Dark spot light	−3.2%	−3.8%	−2.4%	9.4%

(*continued on next page*)

(*continued*)

Y = Motorist pre-crash behavior		**Average Marginal Effects**			
Variable		**Other precrash behavior**	**Fail to yield**	**Sign/signal violation**	**DUI**
	Dark no light	−0.5%	−3.0%	−9.6%	13.2%
	Other	3.7%	−5.6%	−2.5%	4.4%
Weather (base: Clear)	Cloudy	2.4%	4.3%	−2.7%	−4.1%
	Rain	2.7%	2.7%	−0.2%	−5.2%
	Mist/Fog	0.1%	2.3%	1.9%	−4.3%
	Other	1.5%	3.0%	−1.0%	−3.5%
Road alignment (base: Straight and level)	Straight with down grade	2.5%	0.5%	−1.5%	−1.5%
	Straight with up grade	1.6%	0.6%	−2.6%	0.4%
	Curve and level	2.7%	−0.1%	−4.8%	2.2%
	Curve and grade	1.5%	−1.7%	−0.3%	0.6%
	Other	1.9%	0.9%	−4.3%	1.5%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−3.3%	−1.2%	7.7%	−3.3%
	21–30 MPH	−3.2%	−8.5%	14.3%	−2.6%
	31–40 MPH	0.6%	−3.7%	10.0%	−6.8%
	41–50 MPH	−1.7%	−7.1%	8.3%	0.4%
	51–60 MPH	−0.8%	−4.5%	−8.5%	13.8%
	61–70 MPH	4.7%	−13.4%	−8.3%	17.0%
	70 + MPH	4.7%	−14.7%	−7.9%	17.9%
	Stationary	3.4%	−1.2%	−8.2%	6.0%

Appendix B.4. Direct impacts on pre-crash behaviors according to the ANN model (Number of sample in the training set = 3,734)

Y = Motorist pre-crash behavior		**Average Marginal Effects**			
Variable		**Other precrash behavior**	**Fail to yield**	**Sign/signal violation**	**DUI**
Motorist age (base: 26–35 years old)	18–25 years old	−1.3%	2.6%	−0.1%	−1.2%
	36–45 years old	−0.4%	1.8%	−1.4%	0.0%
	46–55 years old	1.3%	−2.1%	−0.3%	1.0%
	56 + years old	−5.9%	5.6%	1.2%	−0.9%
Motorist gender (base: Not female)	Female	1.4%	−0.2%	0.1%	−1.3%
Motorist vehicle type (base: Passenger car)	SUV	0.9%	−0.4%	0.0%	−0.4%
	Pickup	−1.6%	1.2%	−0.6%	0.9%
	Other	6.5%	−3.6%	−0.3%	−2.6%
Motorist Seatbelt (base: used)	Not used	−6.0%	0.8%	0.5%	4.8%
	Other	8.8%	−6.6%	−1.3%	−0.9%
Lighting condition (base: Daylight)	Dawn/Dusk	−4.1%	3.9%	−0.4%	0.6%
	Dark continuous light	−6.5%	−1.2%	1.6%	6.0%
	Dark spot light	−4.8%	−0.8%	1.3%	4.4%
	Dark no light	−3.7%	0.3%	−1.0%	4.4%
	Other	5.9%	−5.5%	−0.7%	0.3%
Weather (base: Clear)	Cloudy	1.5%	−0.2%	−0.7%	−0.7%
	Rain	0.5%	1.1%	−0.4%	−1.2%
	Mist/Fog	−5.8%	3.7%	3.2%	−1.0%
	Other	3.6%	−2.6%	−0.3%	−0.7%
Road alignment (base: Straight and level)	Straight with down grade	5.1%	−3.5%	−0.3%	−1.2%
	Straight with up grade	1.7%	−1.3%	−0.3%	0.0%
	Curve and level	4.5%	−3.2%	−1.6%	0.3%
	Curve and grade	5.8%	−6.2%	−0.6%	0.9%
	Other	−16.7%	15.5%	−0.1%	1.3%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−7.3%	5.5%	1.7%	0.2%
	21–30 MPH	1.4%	−7.1%	5.4%	0.3%
	31–40 MPH	5.8%	−5.8%	1.3%	−1.3%
	41–50 MPH	8.7%	−10.3%	1.1%	0.5%
	51–60 MPH	7.9%	−8.1%	−0.9%	1.2%
	61–70 MPH	10.7%	−12.0%	−0.8%	2.0%
	70 + MPH	10.5%	−11.8%	−0.8%	2.0%
	Stationary	11.2%	−10.5%	−0.9%	0.1%

Appendix B.5. Direct impacts on pre-crash behaviors according to the LDA model (Number of sample in the training set = 3,734)

Y = Motorist pre-crash behavior		**Average Marginal Effects**			
Variable		**Other precrash behavior**	**Fail to yield**	**Sign/signal violation**	**DUI**
Motorist age (base: 26–35 years old)	18–25 years old	−2.2%	4.2%	0.3%	−2.3%
	36–45 years old	1.3%	1.0%	−1.5%	−0.8%
	46–55 years old	2.1%	−1.7%	−0.6%	0.1%
	56 + years old	−2.1%	3.4%	0.6%	−2.0%
Motorist gender (base: Not female)	Female	0.7%	0.1%	0.5%	−1.3%
Motorist vehicle type (base: Passenger car)	SUV	1.9%	−0.6%	−0.4%	−0.8%
	Pickup	−0.9%	0.5%	−0.8%	1.1%
	Other	9.2%	−5.0%	−0.9%	−3.3%

(*continued on next page*)

(*continued*)

Y = Motorist pre-crash behavior		**Average Marginal Effects**			
Variable		**Other precrash behavior**	**Fail to yield**	**Sign/signal violation**	**DUI**
Motorist Seatbelt (base: used)	Not used	−11.0%	−4.2%	1.3%	9.6%
	Other	5.0%	0.1%	−1.0%	0.2%
Lighting condition (base: Daylight)	Dawn/Dusk	−6.9%	6.8%	−0.3%	0.4%
	Dark continuous light	−10.0%	−1.1%	2.1%	9.1%
	Dark spot light	−5.3%	−1.6%	1.1%	5.8%
	Dark no light	−4.7%	0.0%	−1.1%	5.7%
	Other	3.7%	−4.8%	−0.8%	1.9%
Weather (base: Clear)	Cloudy	0.5%	0.6%	−0.5%	−0.6%
	Rain	0.8%	1.3%	−0.2%	−1.9%
	Mist/Fog	−11.9%	4.7%	9.6%	−2.4%
	Other	6.9%	−5.1%	1.3%	−3.0%
Road alignment (base: Straight and level)	Straight with down grade	5.3%	−3.4%	−0.5%	−1.4%
	Straight with up grade	3.3%	−1.9%	−0.9%	−0.5%
	Curve and level	4.6%	−2.8%	−1.8%	0.0%
	Curve and grade	3.5%	−4.9%	−0.9%	2.2%
	Other	–23.0%	21.7%	−1.4%	2.7%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−8.5%	6.7%	1.4%	0.3%
	21–30 MPH	−5.8%	−7.8%	13.3%	0.3%
	31–40 MPH	6.0%	−6.6%	1.4%	−0.8%
	41–50 MPH	6.1%	−9.3%	1.5%	1.8%
	51–60 MPH	4.9%	−7.6%	−0.2%	2.9%
	61–70 MPH	3.3%	−10.5%	−0.3%	7.5%
	70 + MPH	4.6%	−10.4%	−0.2%	6.0%
	Stationary	8.8%	−9.8%	−0.3%	1.4%

Appendix C.1. Direct impacts on police injury severity according to the RF model (Number of sample in the training set = 3,734)

Y = Police injury severity		**Average Marginal Effects**			
Variable		**O**	**C**	**B**	**K&A**
Police age (base: 26–35 years old)	18–25 years old	−2.1%	0.4%	1.5%	0.3%
	36–45 years old	−0.5%	0.4%	0.0%	0.2%
	46–55 years old	−3.0%	1.4%	1.0%	0.6%
	56 + years old	−2.0%	1.4%	0.2%	0.4%
Police gender (base: Not female)	Female	−5.4%	2.9%	2.3%	0.2%
Emergency (base: Non-emergency)	Emergency	−3.8%	1.8%	1.9%	0.1%
Police vehicle type (base: Passenger car)	SUV	1.7%	−1.2%	−0.6%	0.1%
	Pickup	1.0%	0.0%	−0.9%	−0.1%
	Other	−0.7%	0.6%	0.0%	0.1%
Motorist vehicle type (base: Passenger car)	SUV	−2.6%	1.7%	0.6%	0.3%
	Pickup	−3.1%	1.7%	1.0%	0.4%
	Other	−2.6%	1.1%	1.4%	0.1%
Motorist pre-crash behavior (base: Other)	Fail to yield	−15.9%	6.5%	8.0%	1.4%
	Sign/Signal violation	−16.5%	8.1%	6.8%	1.6%
	DUI	−18.7%	10.2%	5.7%	2.8%
Police Seatbelt (base: used)	Not used	–23.9%	5.9%	12.6%	5.4%
	Other	−21.1%	8.0%	9.9%	3.2%
Crash location (base: On road)	Intersection	−10.1%	2.3%	4.5%	3.2%
	Median	−11.1%	5.6%	4.9%	0.6%
	Off road	−5.8%	0.1%	5.3%	0.3%
	Other	−3.6%	0.9%	2.4%	0.2%
	Shoulder/Roadside	−2.1%	0.2%	1.6%	0.3%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−3.3%	2.3%	0.8%	0.2%
	21–30 MPH	−2.5%	1.7%	0.9%	0.0%
	31–40 MPH	−7.4%	3.8%	2.7%	0.9%
	41–50 MPH	−14.3%	5.8%	7.6%	1.0%
	51–60 MPH	−11.8%	6.1%	4.6%	1.1%
	61–70 MPH	−16.7%	6.4%	8.5%	1.9%
	70 + MPH	−18.5%	6.4%	9.8%	2.4%
	Stationary	−5.9%	3.2%	2.5%	0.2%
Lighting condition (base: Daylight)	Dawn/Dusk	−6.3%	3.7%	2.3%	0.3%
	Dark continuous light	−4.0%	1.9%	1.6%	0.5%
	Dark spot light	−3.3%	0.7%	2.3%	0.3%
	Dark no light	−4.3%	0.2%	3.2%	0.9%
	Other	−5.0%	3.0%	1.6%	0.5%
Weather (base: Clear)	Cloudy	−2.7%	1.6%	0.8%	0.3%
	Rain	−2.4%	1.3%	0.9%	0.2%
	Mist/Fog	−2.4%	0.1%	1.9%	0.4%
	Other	−1.7%	0.0%	1.4%	0.3%
Road alignment (base: Straight and level)	Straight with down grade	−2.9%	1.2%	0.9%	0.8%
	Straight with up grade	−1.1%	0.7%	0.0%	0.3%
	Curve and level	−4.3%	−0.3%	4.2%	0.4%
	Curve and grade	−8.7%	0.9%	5.9%	1.9%

(*continued on next page*)

(*continued*)

Y = Police injury severity		Average Marginal Effects			
Variable		O	C	B	K&A
	Other	−6.5%	1.7%	3.4%	1.3%
Collision manner (base: Rear-end)	Head on	−6.0%	0.2%	3.3%	2.4%
	CU backing	4.7%	−2.5%	−1.8%	−0.4%
	Angle	−2.4%	1.0%	1.3%	0.2%
	Sideswipe (same direction)	5.0%	−2.3%	−2.3%	−0.5%
	Side impact	−5.0%	1.5%	3.1%	0.4%
	Other	0.2%	−1.3%	1.4%	−0.3%

Appendix C.2. Direct impacts on police injury severity according to the NB model (Number of sample in the training set = 3,734)

Y = Police injury severity		Average Marginal Effects			
Variable		O	C	B	K&A
Police age (base: 26–35 years old)	18–25 years old	−2.5%	0.3%	1.6%	0.6%
	36–45 years old	1.1%	−0.2%	−0.9%	0.0%
	46–55 years old	−1.3%	0.6%	0.1%	0.5%
	56 + years old	3.1%	0.0%	−2.9%	−0.2%
Police gender (base: Not female)	Female	−1.8%	1.9%	0.3%	−0.5%
Emergency (base: Non-emergency)	Emergency	−5.9%	1.9%	3.5%	0.6%
Police vehicle type (base: Passenger car)	SUV	1.9%	−2.0%	0.2%	−0.1%
	Pickup	5.9%	−2.0%	−2.1%	−1.7%
	Other	1.3%	0.0%	−0.4%	−0.8%
Motorist vehicle type (base: Passenger car)	SUV	−1.3%	1.4%	−0.2%	0.0%
	Pickup	−0.9%	1.5%	−0.2%	−0.3%
	Other	3.3%	−1.3%	−1.4%	−0.6%
Motorist pre-crash behavior (base: Other)	Fail to yield	−11.9%	5.3%	6.4%	0.2%
	Sign/Signal violation	−19.5%	9.2%	8.9%	1.4%
	DUI	−16.2%	8.7%	5.1%	2.5%
Police Seatbelt (base: used)	Not used	−26.5%	8.3%	12.6%	5.5%
	Other	−19.9%	13.7%	7.0%	−0.9%
Crash location (base: On road)	Intersection	−9.7%	0.1%	4.9%	4.6%
	Median	−6.9%	4.3%	3.9%	−1.2%
	Off road	6.6%	−5.2%	0.0%	−1.3%
	Other	9.3%	−4.4%	−3.4%	−1.5%
	Shoulder/Roadside	1.2%	−2.3%	1.5%	−0.4%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−5.9%	3.7%	1.5%	0.8%
	21–30 MPH	−2.8%	2.2%	0.7%	−0.1%
	31–40 MPH	−6.8%	4.5%	1.3%	1.0%
	41–50 MPH	−12.1%	4.7%	6.5%	0.9%
	51–60 MPH	−10.6%	4.9%	4.0%	1.8%
	61–70 MPH	−16.6%	6.7%	8.9%	1.0%
	70 + MPH	−17.3%	3.1%	9.5%	4.6%
	Stationary	−1.0%	−0.6%	2.3%	−0.7%
Lighting condition (base: Daylight)	Dawn/Dusk	−6.3%	5.7%	0.9%	−0.2%
	Dark continuous light	−5.9%	2.8%	1.9%	1.2%
	Dark spot light	−3.6%	0.1%	2.7%	0.7%
	Dark no light	−7.0%	−0.1%	4.8%	2.3%
	Other	5.2%	0.5%	−4.9%	−0.8%
Weather (base: Clear)	Cloudy	−1.2%	0.7%	0.0%	0.5%
	Rain	−0.7%	1.3%	0.0%	−0.6%
	Mist/Fog	2.7%	−2.6%	−0.9%	0.8%
	Other	10.7%	−4.4%	−5.3%	−1.0%
Road alignment (base: Straight and level)	Straight with down grade	−0.7%	0.1%	−0.6%	1.1%
	Straight with up grade	1.9%	−0.3%	−1.9%	0.3%
	Curve and level	−0.8%	−4.3%	4.6%	0.5%
	Curve and grade	−1.7%	−2.0%	1.0%	2.6%
	Other	0.3%	−1.5%	0.0%	1.2%
Collision manner (base: Rear-end)	Head on	−6.6%	−1.7%	1.9%	6.4%
	CU backing	10.2%	−4.2%	−5.1%	−0.9%
	Angle	−4.6%	3.3%	1.0%	0.4%
	Sideswipe (same direction)	4.2%	−1.6%	−2.9%	0.2%
	Side impact	−8.7%	3.3%	4.5%	1.0%
	Other	−0.4%	−2.1%	2.5%	−0.1%

Appendix C.3. Direct impacts on police injury severity according to the AdaBoost model (Number of sample in the training set = 3,734)

Y = Police injury severity		**Average Marginal Effects**			
Variable		**O**	**C**	**B**	**K&A**
Police age (base: 26–35 years old)	18–25 years old	1.0%	−1.4%	2.5%	−2.1%
	36–45 years old	1.2%	−0.6%	1.2%	−1.8%
	46–55 years old	−1.0%	−0.7%	0.3%	1.5%
	56 + years old	0.2%	−0.2%	−1.5%	1.5%
Police gender (base: Not female)	Female	−0.4%	−1.3%	1.3%	0.4%
Emergency (base: Non-emergency)	Emergency	−1.0%	−1.3%	1.4%	0.8%
Police vehicle type (base: Passenger car)	SUV	0.9%	−1.3%	1.5%	−1.1%
	Pickup	1.3%	1.6%	0.5%	−3.4%
	Other	1.0%	−0.1%	1.5%	−2.3%
Motorist vehicle type (base: Passenger car)	SUV	−0.8%	1.1%	0.5%	−0.8%
	Pickup	−0.7%	1.4%	−0.6%	−0.2%
	Other	−0.4%	1.1%	−0.3%	−0.4%
Motorist pre-crash behavior (base: Other)	Fail to yield	−2.8%	−0.1%	0.8%	2.1%
	Sign/Signal violation	−4.4%	1.2%	1.1%	2.1%
	DUI	−1.6%	0.9%	−4.9%	5.5%
Police Seatbelt (base: used)	Not used	−3.3%	−1.8%	−0.1%	5.2%
	Other	−2.0%	0.3%	1.7%	0.1%
Crash location (base: On road)	Intersection	−3.1%	−5.9%	−2.6%	11.5%
	Median	−1.6%	0.7%	1.3%	−0.5%
	Off road	1.0%	−3.6%	2.8%	−0.2%
	Other	2.9%	−3.2%	0.5%	−0.3%
	Shoulder/Roadside	1.3%	−0.6%	−0.5%	−0.1%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−1.4%	2.6%	2.6%	−3.9%
	21–30 MPH	0.1%	0.7%	3.6%	−4.3%
	31–40 MPH	−0.5%	2.9%	0.1%	−2.6%
	41–50 MPH	−4.2%	2.0%	5.1%	−2.9%
	51–60 MPH	−3.2%	0.1%	2.8%	0.3%
	61–70 MPH	−4.0%	3.0%	4.1%	−3.1%
	70 + MPH	−5.3%	−1.5%	3.7%	3.1%
	Stationary	−1.5%	1.0%	5.6%	−5.0%
Lighting condition (base: Daylight)	Dawn/Dusk	−0.1%	2.9%	−2.2%	−0.6%
	Dark continuous light	−0.7%	−0.3%	0.8%	0.2%
	Dark spot light	−0.3%	−2.6%	1.1%	1.8%
	Dark no light	−1.0%	−5.1%	1.3%	4.8%
	Other	−0.1%	3.6%	−3.4%	0.0%
Weather (base: Clear)	Cloudy	0.1%	−0.2%	−0.8%	0.9%
	Rain	1.5%	0.4%	−1.5%	−0.4%
	Mist/Fog	0.2%	−2.1%	0.5%	1.5%
	Other	1.0%	−0.1%	−1.1%	0.2%
Road alignment (base: Straight and level)	Straight with down grade	0.6%	−1.2%	−3.0%	3.6%
	Straight with up grade	0.5%	1.1%	−2.1%	0.4%
	Curve and level	0.2%	−6.5%	6.4%	−0.2%
	Curve and grade	0.6%	−4.5%	−0.1%	4.0%
	Other	0.3%	−2.6%	2.1%	0.2%
Collision manner (base: Rear-end)	Head on	−0.3%	−3.7%	0.1%	3.9%
	CU backing	3.3%	3.5%	−3.7%	−3.1%
	Angle	0.4%	1.5%	1.4%	−3.4%
	Sideswipe (same direction)	5.1%	0.8%	−1.8%	−4.1%
	Side impact	−1.3%	2.0%	2.0%	−2.7%
	Other	1.6%	−2.9%	5.9%	−4.6%

Appendix C.4. Direct impacts on police injury severity according to the ANN model (Number of sample in the training set = 3,734)

Y = Police injury severity		**Average Marginal Effects**			
Variable		**O**	**C**	**B**	**K&A**
Police age (base: 26–35 years old)	18–25 years old	−1.4%	0.1%	1.1%	0.3%
	36–45 years old	−0.6%	0.3%	0.2%	0.1%
	46–55 years old	−4.1%	1.7%	1.9%	0.5%
	56 + years old	0.8%	0.3%	−0.9%	−0.3%
Police gender (base: Not female)	Female	−3.3%	2.4%	0.8%	0.1%
Emergency (base: Non-emergency)	Emergency	−2.7%	1.0%	1.4%	0.3%
Police vehicle type (base: Passenger car)	SUV	1.9%	−1.0%	−0.7%	−0.2%
	Pickup	5.4%	−1.7%	−3.0%	−0.7%
	Other	1.3%	−0.2%	−0.9%	−0.2%
Motorist vehicle type (base: Passenger car)	SUV	−1.3%	1.1%	0.2%	0.0%
	Pickup	−0.9%	1.6%	−0.4%	−0.2%
	Other	0.5%	0.3%	−0.6%	−0.2%
Motorist pre-crash behavior (base: Other)	Fail to yield	−12.1%	5.1%	5.5%	1.4%
	Sign/Signal violation	−14.9%	5.4%	7.5%	2.0%
	DUI	−10.8%	6.1%	3.9%	0.8%
Police Seatbelt (base: used)	Not used	−20.9%	6.6%	11.1%	3.3%

(*continued on next page*)

(*continued*)

Y = Police injury severity Variable		Average Marginal Effects O	C	B	K&A
	Other	−15.4%	7.0%	6.7%	1.7%
Crash location (base: On road)	Intersection	−9.0%	0.4%	6.5%	2.1%
	Median	−8.6%	3.4%	4.1%	1.1%
	Off road	2.6%	−1.5%	−0.9%	−0.2%
	Other	4.4%	−1.8%	−2.1%	−0.5%
	Shoulder/Roadside	5.0%	−2.3%	−2.3%	−0.5%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−2.2%	1.6%	0.5%	0.1%
	21–30 MPH	−0.5%	0.5%	0.0%	0.0%
	31–40 MPH	−5.1%	3.4%	1.5%	0.2%
	41–50 MPH	−11.8%	5.1%	5.4%	1.3%
	51–60 MPH	−10.3%	4.5%	4.7%	1.1%
	61–70 MPH	−16.0%	6.7%	7.5%	1.9%
	70 + MPH	−16.4%	3.5%	9.9%	3.0%
	Stationary	−0.5%	−0.2%	0.5%	0.2%
Lighting condition (base: Daylight)	Dawn/Dusk	−5.2%	3.9%	1.2%	0.1%
	Dark continuous light	−3.3%	1.2%	1.7%	0.4%
	Dark spot light	−1.7%	−0.2%	1.5%	0.4%
	Dark no light	−2.9%	−1.1%	3.0%	1.0%
	Other	0.4%	0.7%	−0.8%	−0.3%
Weather (base: Clear)	Cloudy	−1.4%	0.7%	0.6%	0.1%
	Rain	−1.0%	0.8%	0.2%	0.0%
	Mist/Fog	2.8%	−1.7%	−0.9%	−0.2%
	Other	7.1%	−2.9%	−3.4%	−0.8%
Road alignment (base: Straight and level)	Straight with down grade	−1.0%	0.7%	0.3%	0.1%
	Straight with up grade	2.5%	−1.1%	−1.2%	−0.3%
	Curve and level	−2.3%	−1.3%	2.7%	0.9%
	Curve and grade	−4.6%	0.0%	3.4%	1.1%
	Other	−5.2%	2.3%	2.4%	0.6%
Collision manner (base: Rear-end)	Head on	−5.0%	−0.3%	4.0%	1.3%
	CU backing	7.9%	−3.3%	−3.8%	−0.8%
	Angle	−1.2%	1.4%	−0.1%	−0.1%
	Sideswipe (same direction)	5.4%	−2.3%	−2.5%	−0.6%
	Side impact	−5.7%	2.5%	2.6%	0.6%
	Other	1.2%	−1.5%	0.2%	0.1%

Appendix C.5. Direct impacts on police injury severity according to the LDA model (Number of sample in the training set = 3,734)

Y = Police injury severity Variable		Average Marginal Effects O	C	B	K&A
Police age (base: 26–35 years old)	18–25 years old	−1.5%	−0.1%	1.1%	0.5%
	36–45 years old	−0.8%	0.3%	0.1%	0.4%
	46–55 years old	−3.7%	1.3%	1.3%	1.0%
	56 + years old	0.2%	0.7%	−1.0%	0.1%
Police gender (base: Not female)	Female	−3.7%	2.7%	1.2%	−0.2%
Emergency (base: Non-emergency)	Emergency	−2.5%	1.4%	1.6%	−0.6%
Police vehicle type (base: Passenger car)	SUV	1.5%	−1.7%	0.1%	0.1%
	Pickup	4.0%	−1.8%	−1.2%	−1.0%
	Other	0.3%	0.1%	−0.1%	−0.3%
Motorist vehicle type (base: Passenger car)	SUV	−1.6%	1.3%	0.0%	0.3%
	Pickup	−1.1%	1.4%	−0.1%	−0.2%
	Other	0.0%	0.5%	−0.2%	−0.3%
Motorist pre-crash behavior (base: Other)	Fail to yield	−14.7%	5.7%	9.1%	−0.1%
	Sign/Signal violation	−26.1%	12.5%	12.3%	1.4%
	DUI	−18.4%	14.2%	1.6%	2.6%
Police Seatbelt (base: used)	Not used	−53.6%	7.7%	16.3%	29.7%
	Other	−46.6%	40.6%	6.6%	−0.5%
Crash location (base: On road)	Intersection	–22.6%	−2.3%	2.5%	22.5%
	Median	−16.1%	9.9%	7.2%	−1.0%
	Off road	2.0%	−1.7%	0.8%	−1.1%
	Other	1.4%	−0.4%	−0.2%	−0.8%
	Shoulder/Roadside	4.5%	−2.2%	−1.0%	−1.2%
Casual unit speed in the crash (base: 0–10 MPH)	11–20 MPH	−2.2%	2.0%	0.2%	−0.1%
	21–30 MPH	−0.5%	1.0%	0.0%	−0.5%
	31–40 MPH	−5.0%	3.8%	0.5%	0.7%
	41–50 MPH	−13.2%	4.8%	8.1%	0.4%
	51–60 MPH	−11.1%	5.7%	3.2%	2.2%
	61–70 MPH	−20.6%	8.1%	12.2%	0.3%
	70 + MPH	−24.3%	1.8%	11.3%	11.2%
	Stationary	−0.7%	−0.3%	1.6%	−0.6%
Lighting condition (base: Daylight)	Dawn/Dusk	−6.5%	7.1%	−0.2%	−0.4%
	Dark continuous light	−2.9%	1.3%	0.5%	1.1%
	Dark spot light	−1.9%	−0.4%	1.7%	0.6%

(*continued on next page*)

(*continued*)

Y = Police injury severity		**Average Marginal Effects**			
Variable		**O**	**C**	**B**	**K&A**
	Dark no light	−3.7%	−1.0%	2.6%	2.1%
	Other	−7.9%	11.0%	−2.5%	−0.7%
Weather (base: Clear)	Cloudy	−1.0%	0.6%	−0.1%	0.5%
	Rain	−0.5%	1.0%	0.1%	−0.5%
	Mist/Fog	2.5%	−2.1%	−1.3%	0.9%
	Other	8.1%	−3.7%	−3.2%	−1.2%
Road alignment (base: Straight and level)	Straight with down grade	−1.0%	0.5%	−0.8%	1.3%
	Straight with up grade	2.1%	−0.3%	−1.8%	0.0%
	Curve and level	−3.1%	−2.3%	5.7%	−0.3%
	Curve and grade	−4.2%	−1.2%	0.7%	4.7%
	Other	−3.5%	−1.1%	1.2%	3.4%
Collision manner (base: Rear-end)	Head on	−24.3%	−3.1%	−2.3%	29.7%
	CU backing	2.9%	−1.4%	−2.0%	0.4%
	Angle	−1.0%	2.0%	−1.1%	0.1%
	Sideswipe (same direction)	4.4%	−1.8%	−2.7%	0.1%
	Side impact	−5.0%	1.9%	2.2%	0.9%
	Other	1.2%	−1.9%	0.9%	−0.2%

References

Abdelwanis, N. (2013). *Characteristics and contributing factors of emergency vehicle crashes* [Dissertation or Thesis, Clemson University]. https://www.proquest.com/dissertations-theses/characteristics-contributing-factors-emergency/docview/1499080049/se-2?accountid=14472.

Ahmad, N., Arvin, R., & Khattak, A. J. (2023). Exploring pathways from driving errors and violations to crashes: The role of instability in driving. *Accident Analysis and Prevention, 179*. https://doi.org/10.1016/j.aap.2022.106876

Alabama Code Title 32. Motor Vehicles and Traffic § 32-5A-191 | FindLaw. (n.d.). Retrieved March 2, 2023, from https://codes.findlaw.com/al/title-32-motor-vehicles-and-traffic/al-code-sect-32-5a-191.html.

Alfaro, E., Gamez, M., & Garcia, N. (2013). adabag: An R package for classification with boosting and bagging. *Journal of Statistical Software, 54*, 1–35. https://doi.org/10.18637/jss.v054.i02

Arvin, R., Kamrani, M., & Khattak, A. J. (2019). The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data. *Accident Analysis and Prevention, 132*. https://doi.org/10.1016/j.aap.2019.07.002

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society, 20*(4), 451–468. https://doi.org/10.1057/jors.1969.103

Behnood, A., & Mannering, F. (2017). Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research, 16*, 35–47. https://doi.org/10.1016/j.amar.2017.08.001

Box, G. E. P. (1976). Science and statistics. *Source: Journal of the American Statistical Association, 71*(356), 56. https://doi.org/10.2307/2286841

Breiman, L. (2001). *Random Forests, 45*. https://doi.org/10.1023/A:1010933404324

Carrick, G., & Washburn, S. (2012). The move over law. *Transportation Research Record, 2281*, 1–7. https://doi.org/10.3141/2281-01

Chen, C., Zhang, G., Qian, Z., Tarefder, R. A., & Tian, Z. (2016). Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accident Analysis and Prevention, 90*, 128–139. https://doi.org/10.1016/j.aap.2016.02.011

Cheng, Z., Zhang, L., Zhang, Y., Wang, S., & Huang, W. (2023). A systematic approach for evaluating spatiotemporal characteristics of traffic violations and crashes at road intersections: An empirical study. *Transportmetrica A: Transport Science, 19*(3). https://doi.org/10.1080/23249935.2022.2060368

Chevalier, A., Coxon, K., Chevalier, A. J., Wall, J., Brown, J., Clarke, E., Ivers, R., & Keay, L. (2016). Exploration of older drivers' speeding behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour, 42*, 532–543. https://doi.org/10.1016/j.trf.2016.01.012

Chevalier, A., Coxon, K., Rogers, K., Chevalier, A. J., Wall, J., Brown, J., Clarke, E., Ivers, R., & Keay, L. (2016). A longitudinal investigation of the predictors of older drivers' speeding behaviour. *Accident Analysis and Prevention, 93*, 41–47. https://doi.org/10.1016/j.aap.2016.04.006

Chu, H. C. (2016). Risk factors for the severity of injury incurred in crashes involving on-duty police cars. *Traffic Injury Prevention, 17*(5), 495–501. https://doi.org/10.1080/15389588.2015.1109082

Clemen, R. T. (1989). Combining forecasts: A review and annotated. *International Journal of Forecasting, 5*, 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Cohn, E. G., Kakar, S., Perkins, C., Steinbach, R., & Edwards, P. (2020). Red light camera interventions for reducing traffic violations and traffic crashes: A systematic review. In *Campbell Systematic Reviews* (Vol. 16, Issue 2). Wiley Blackwell. doi: 10.1002/cl2.1091.

De Pelsmacker, P., & Janssens, W. (2007). The effect of norms, attitudes and habits on speeding behavior: Scale development and model building and estimation. *Accident Analysis and Prevention, 39*(1), 6–15. https://doi.org/10.1016/j.aap.2006.05.011

Dong, B., Ma, X., & Chen, F. (2018). Analyzing the injury severity sustained by non-motorists at mid-blocks considering non-motorists' pre-crash behavior. *Transportation Research Record, 2672*(38), 138–148. https://doi.org/10.1177/0361198118777354

Eluru, N., Bhat, C. R., & Hensher, D. A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention, 40*(3), 1033–1054. https://doi.org/10.1016/j.aap.2007.11.010

Fang, Y. (2003). Forecasting combination and encompassing tests. *International Journal of Forecasting, 19*, 87–94. https://doi.org/10.1016/S0169-2070(01)00121-2

Fu, H. (2008). Identifying repeat DUI crash factors using state crash records. *Accident Analysis and Prevention, 40*(6), 2037–2042. https://doi.org/10.1016/j.aap.2008.08.020

Fu, X., Nie, Q., Li, X., Liu, J., Nambisan, S., & Jones, S. (2022). The role of the built environment in emergency medical services delays in responding to traffic crashes. *Journal of Transportation Engineering, Part A: Systems, 148*(10). https://doi.org/10.1061/jtepbs.0000726

García-Altés, A., & Pérez, K. (2007). The economic cost of road traffic crashes in an urban setting. *Injury Prevention, 13*(1), 65. https://doi.org/10.1136/ip.2006.012732

Gargoum, S. A., & El-Basyouny, K. (2016). Exploring the association between speed and safety: A path analysis approach. *Accident Analysis and Prevention, 93*, 32–40. https://doi.org/10.1016/j.aap.2016.04.029

Gheorghiu, A., Delhomme, P., & Felonneau, M. L. (2015). Peer pressure and risk taking in young drivers' speeding behavior. *Transportation Research Part F: Traffic Psychology and Behaviour, 35*, 101–111. https://doi.org/10.1016/j.trf.2015.10.014

Gieck, D. J., & Slagle, D. M. (2010). Examination of a university-affiliated safe ride program. *Journal of Alcohol and Drug Education, 54*(1), 37–55. https://www.jstor.org/stable/45128197.

Golob, T. F. (2003). *Structural equation modeling for travel behavior research*. 1–25. doi: 10.1016/S0191-2615(01)00046-7.

González-Iglesias, B., Gómez-Fraguela, J. A., & Luengo-Martín, M.Á. (2012). Driving anger and traffic violations: Gender differences. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*(4), 404–412. https://doi.org/10.1016/j.trf.2012.03.002

Hossain, M. M., Zhou, H., & Das, S. (2023). Data mining approach to explore emergency vehicle crash patterns: A comparative study of crash severity in emergency and non-emergency response modes. *Accident Analysis and Prevention, 191*. https://doi.org/10.1016/j.aap.2023.107217

Hu, F., Lv, D., Zhu, J., & Fang, J. (2014). Related risk factors for injury severity of e-bike and bicycle crashes in Hefei. *Traffic Injury Prevention, 15*(3), 319–323. https://doi.org/10.1080/15389588.2013.817669

Hussain, Q., Alhajyaseen, W. K. M., Pirdavani, A., Brijs, K., Shaaban, K., & Brijs, T. (2021). Do detection-based warning strategies improve vehicle yielding behavior at uncontrolled midblock crosswalks? *Accident Analysis and Prevention, 157*. https://doi.org/10.1016/j.aap.2021.106166

Ijaz, M., Lan, L., Zahid, M., & Jamal, A. (2021). A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw. *Accident Analysis and Prevention, 154*. https://doi.org/10.1016/j.aap.2021.106094

Inkoom, S., Sobanjo, J., Barbu, A., & Niu, X. (2019). Pavement crack rating using machine learning frameworks: partitioning, bootstrap forest, boosted trees, naïve bayes, and K -nearest neighbors. *Journal of Transportation Engineering, Part B: Pavements, 145*(3), Article 04019031. https://doi.org/10.1061/jpeodx.0000126

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer. doi: 10.1007/978-1-0716-1418-1_1.

Kaplan, S., Vavatsoulas, K., & Prato, C. G. (2014). Aggravating and mitigating factors associated with cyclist injury severity in Denmark. *Journal of Safety Research, 50*, 75–82. https://doi.org/10.1016/j.jsr.2014.03.012

Kim, J. K., Kim, S., Ulfarsson, G. F., & Porrello, L. A. (2007). Bicyclist injury severities in bicycle-motor vehicle accidents. *Accident Analysis and Prevention, 39*(2), 238–251. https://doi.org/10.1016/j.aap.2006.07.002

Kim, J. K., Ulfarssom, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis and Prevention, 42*(6), 1751–1758. https://doi.org/10.1016/j.aap.2010.04.016

Kim, J. K., Ulfarsson, G. F., Kim, S., & Shankar, V. N. (2013). Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis and Prevention, 50*, 1073–1081. https://doi.org/10.1016/j.aap.2012.08.011

Kockelman, K. M., & Kweon, Y.-J. (2002). Driver injury severity: an application of ordered probit models. In *Accident Analysis and Prevention* (Vol. 34). www.elsevier.com/locate/aap.

Kong, X., Das, S., Jha, K., & Zhang, Y. (2020). Understanding speeding behavior from naturalistic driving data: Applying classification based association rule mining. *Accident Analysis and Prevention, 144*. https://doi.org/10.1016/j.aap.2020.105620

Kuhn, M. (2008). Journal of Statistical Software Building Predictive Models in R Using the caret Package. http://www.jstatsoft.org/.

Li, X., Oviedo-Trespalacios, O., Pooyan Afghari, A., Kaye, S. A., & Yan, X. (2023). Yield or not to yield? an inquiry into drivers' behaviour when a fully automated vehicle indicates a lane-changing intention. *Transportation Research Part F: Traffic Psychology and Behaviour, 95*, 405–417. https://doi.org/10.1016/j.trf.2023.05.012

Li, Y., Abdel-Aty, M., Yuan, J., Cheng, Z., & Lu, J. (2020). Analyzing traffic violation behavior at urban intersections: A spatio-temporal kernel density estimation approach using automated enforcement system data. *Accident Analysis and Prevention, 141*. https://doi.org/10.1016/j.aap.2020.105509

Li, Y., Li, M., Yuan, J., Lu, J., & Abdel-Aty, M. (2021). Analysis and prediction of intersection traffic violations using automated enforcement system data. *Accident Analysis and Prevention, 162*. https://doi.org/10.1016/j.aap.2021.106422

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2* (3), 18–22.

Liu, J., Jones, S., Adanu, E. K., & Li, X. (2021). Behavioral pathways in bicycle-motor vehicle crashes: From contributing factors, pre-crash actions, to injury severities. *Journal of Safety Research, 77*, 229–240. https://doi.org/10.1016/j.jsr.2021.02.015

Liu, J., & Khattak, A. J. (2017). Gate-violation behavior at highway-rail grade crossings and the consequences: Using geo-Spatial modeling integrated with path analysis. *Accident Analysis and Prevention, 109*, 99–112. https://doi.org/10.1016/j.aap.2017.10.010

Liu, J., & Khattak, A. J. (2018). Are gates at rail grade crossings always safe? Examining motorist gate-violation behaviors using path analysis. *Transportation Research Part F: Traffic Psychology and Behaviour, 55*, 314–324. https://doi.org/10.1016/j.trf.2018.03.014

Liu, J., Khattak, A. J., Li, X., & Fu, X. (2019). A spatial analysis of the ownership of alternative fuel and hybrid vehicles. *Transportation Research Part D: Transport and Environment, 77*, 106–119. https://doi.org/10.1016/j.trd.2019.10.018

Liu, J., Khattak, A. J., Richards, S. H., & Nambisan, S. (2015). What are the differences in driver injury outcomes at highway-rail grade crossings? Untangling the role of pre-crash behaviors. *Accident Analysis and Prevention, 85*, 157–169. https://doi.org/10.1016/j.aap.2015.09.004

Liu, J., Li, X., & Khattak, A. J. (2020). An integrated spatio-temporal approach to examine the consequences of driving under the influence (DUI) in crashes. *Accident Analysis and Prevention, 146*. https://doi.org/10.1016/j.aap.2020.105742

Liu, J., Wang, X., Khattak, A. J., Hu, J., Cui, J. X., & Ma, J. (2016). How big data serves for freight safety management at highway-rail grade crossings? a spatial approach fused with path analysis. *Neurocomputing, 181*, 38–52. https://doi.org/10.1016/j.neucom.2015.08.098

Liu, Z., & Donmez, B. (2011). Effects of distractions on injury severity in police-involved crashes. *Proceedings of the transportation research board 90th annual meeting, Washington, DC. 01333498*.

Lu, W., Liu, J., Fu, X., Yang, J., & Jones, S. (2022). Integrating machine learning into path analysis for quantifying behavioral pathways in bicycle-motor vehicle crashes. *Accident Analysis & Prevention, 168*, Article 106622. https://doi.org/10.1016/j.aap.2022.106622

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science, 29*(9), 987–996. https://doi.org/10.1287/mnsc.29.9.987

Mayou, R., Bryant, B., & Duthie, R. (1993). Psychiatric consequences of road traffic accidents. *British Medical Journal, 307*(6905), 647–651. https://doi.org/10.1136/bmj.307.6905.647

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. *AAAI-98 Workshop on Learning for Text Categorization, 752*(1), 41–48.

Medina, A., Bansen, J., Williams, B., Pochowski, A., Rodegerdts, L., Markosian, J., Li, Y., & Gibbons, R. (2023). *Reasons for drivers failing to yield at multi-lane roundabout exits: transportation pooled fund study final report.* https://rosap.ntl.bts.gov/view/dot/66498.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., Lin, C.-C., & Meyer, M. D. (2019). Package 'e1071.'. *The R Journal.*

Molnar, C. (2020). *Interpretable machine learning.* Lulu. com.

National Law Enforcement Officer Memorial Fund. (2021). *Causes of law enforcement deaths*. National law enforcement officer memorial fund. https://nleomf.org/wp-content/uploads/2021/07/causes-le-deaths-updated-July-19-2021.pdf.

Ng, A., & Jordan, M. (2001). On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. In T. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems* (Vol. 14). MIT Press. https://proceedings.neurips.cc/paper_files/paper/2001/file/7b7a53e239400a13bd6be6c91c4f6c4e-Paper.pdf.

NIOSH. (2022, August 22). *Law enforcement officer motor vehicle safety*. https://www.cdc.gov/niosh/topics/leo/default.html.

Noh, E. Y. (2011). Characteristics of law enforcement officers' fatalities in motor crashes 13. Type of report and period covered NHTSA Technical Report. www.ntis.gov.

Oh, S., Vaughn, M. G., Salas-Wright, C. P., AbiNader, M. A., & Sanchez, M. (2020). Driving under the influence of Alcohol: Findings from the NSDUH, 2002–2017. *Addictive Behaviors, 108*. https://doi.org/10.1016/j.addbeh.2020.106439

Okafor, S., Liu, J., Kofi Adanu, E., & Jones, S. (2023). Behavioral pathway analysis of pedestrian injury severity in pedestrian-motor vehicle crashes. *Transportation Research Interdisciplinary Perspectives, 18*. https://doi.org/10.1016/j.trip.2023.100777

Pai, C. W., & Saleh, W. (2008). Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: Focusing on the effects of driver's failure to yield and junction control measures. *Accident Analysis and Prevention, 40*(2), 479–486. https://doi.org/10.1016/j.aap.2007.08.003

Palm, F. C., & Zellner, A. (1992). To combine or not to combine? issues of combining forecasts. *Journal of Forecasting, 11*(8), 687–701. https://doi.org/10.1002/for.3980110806

Pantangi, S. S., Fountas, G., Anastasopoulos, P. C., Pierowicz, J., Majka, K., & Blatt, A. (2020). Do high visibility enforcement programs affect aggressive driving behavior? an empirical analysis using naturalistic driving study data. *Accident Analysis and Prevention, 138*. https://doi.org/10.1016/j.aap.2019.105361

Park, E. S., Fitzpatrick, K., Das, S., & Avelar, R. (2021). Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis. *Accident Analysis and Prevention, 150*. https://doi.org/10.1016/j.aap.2020.105896

Penmetsa, P., & Pulugurtha, S. S. (2017). Risk drivers pose to themselves and other drivers by violating traffic rules. *Traffic Injury Prevention, 18*(1), 63–69. https://doi.org/10.1080/15389588.2016.1177637

Perez, M. A., Sears, E., Valente, J. T., Huang, W., & Sudweeks, J. (2021). Factors modifying the likelihood of speeding behaviors based on naturalistic driving data. *Accident Analysis and Prevention, 159*. https://doi.org/10.1016/j.aap.2021.106267

Pohar, M., Blas, M., & Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis: A simulation study. *Metodoloski Zvezki, 1*(1), 143.

Porter, B. E., & Berry, T. D. (2001). A nationwide survey of self-reported red light running: measuring prevalence, predictors, and perceived consequences. In *Accident Analysis and Prevention* (Vol. 33). doi: 10.1016/S0001-4575(00)00087-7.

Porter, B. E., & England, K. J. (2000). Predicting red-light running behavior: a traffic safety study in three urban settings. *Journal of Safety Research* (Vol. 31, Issue 1).

Precht, L., Keinath, A., & Krems, J. F. (2017). Identifying the main factors contributing to driving errors and traffic violations – results from naturalistic driving data. *Transportation Research Part F: Traffic Psychology and Behaviour, 49*, 49–92. https://doi.org/10.1016/j.trf.2017.06.002

Ripley, B., Venables, W., & Ripley, M. B. (2016). Package 'nnet.'. *R Package Version, 7* (3–12), 700.

Santos, K., Firme, B., Dias, J. P., & Amado, C. (2024). Analysis of motorcycle accident injury severity and performance comparison of machine learning algorithms. *Transportation Research Record, 2678*(1), 736–748. https://doi.org/10.1177/03611981231172507

Shinar, D., & Compton, R. (2004). Aggressive driving: An observational study of driver, vehicle, and situational variables. *Accident Analysis and Prevention, 36*(3), 429–437. https://doi.org/10.1016/S0001-4575(03)00037-X

Silva Filho, R. L. C., Adeodato, P. J. L., & dos Santos Brito, K. (2021). Interpreting classification models using feature importance based on marginal local effects. *Brazilian Conference on Intelligent Systems, 484–497*. https://doi.org/10.1007/978-3-030-91702-9_32

Smailović, E., Pešić, D., Antić, B., & Marković, N. (2023). A review of factors associated with driving under the influence of alcohol. *Transportation Research Procedia, 69*, 281–288. https://doi.org/10.1016/j.trpro.2023.02.173

Stephens, A. N., Nieuwesteeg, M., Page-Smith, J., & Fitzharris, M. (2017). Self-reported speed compliance and attitudes towards speeding in a representative sample of drivers in Australia. *Accident Analysis and Prevention, 103*, 56–64. https://doi.org/10.1016/j.aap.2017.03.020

Stringer, R. J. (2019). Policing the drunk driving problem: a longitudinal examination of DUI Enforcement and Alcohol Related Crashes in the U.S. (1985–2015). *American Journal of Criminal Justice, 44*(3), 474–498. doi: 10.1007/s12103-018-9464-4.

Sun, F., Liu, M., Wang, Y., Wang, H., & Che, Y. (2020). The effects of 3D architectural patterns on the urban surface temperature at a neighborhood scale: Relative contributions and marginal effects. *Journal of Cleaner Production, 258*. https://doi.org/10.1016/j.jclepro.2020.120706

Sze, N. N., & Wong, S. C. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis and Prevention, 39*(6), 1267–1278. https://doi.org/10.1016/j.aap.2007.03.017

Taubman-Ben-Ari, O., & Yehiel, D. (2012). Driving styles and their associations with personality and motivation. *Accident Analysis and Prevention, 45*, 416–422. https://doi.org/10.1016/j.aap.2011.08.007

Van Buuren, S. (2014). *Package "mice" Title Multivariate Imputation by Chained Equations*. doi: 10.18637/jss.v045.i03.

von Kuenssberg Jehle, D., Wagner, D. G., Mayrose, J., & Hashmi, U. (2005). Seat belt use by police: should they click it? *Journal of Trauma and Acute Care Surgery, 58*(1). https://journals.lww.com/jtrauma/fulltext/2005/01000/seat_belt_use_by_police_should_they_click_it_.20.aspx.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal, 12*(2), 308–331. https://doi.org/10.1177/1536867X1201200209

Wong, K. K. F., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management, 28*(4), 1068–1078. https://doi.org/10.1016/j.tourman.2006.08.003

Wyatt, B., & Novotna, G. (2021). Driving under the influence of alcohol and drugs: a scoping review. *Journal of Social Work Practice in the Addictions, 21*(2), 119–138. https://doi.org/10.1080/1533256X.2021.1893952

Xie, Y., Zhao, K., & Huynh, N. (2012). Analysis of driver injury severity in rural single-vehicle crashes. *Accident Analysis and Prevention, 47*, 36–44. https://doi.org/10.1016/j.aap.2011.12.012

Xu, N., Liu, J., Zhang, Z., & Jones, S. (2024a). Injury severity of police officers involved in traffic crashes: A spatial analysis of Alabama. *Safety Science, 172*. https://doi.org/10.1016/j.ssci.2023.106406

Xu, N., Nie, Q., Liu, J., & Jones, S. (2023). Post-pandemic shared mobility and active travel in Alabama: A machine learning analysis of COVID-19 survey data. *Travel Behaviour and Society, 32*. https://doi.org/10.1016/j.tbs.2023.100584

Xu, N., Nie, Q., Liu, J., & Jones, S. (2024b). Linking short-and long-term impacts of the COVID-19 pandemic on travel behavior and travel preferences in Alabama: A machine learning-supported path analysis. *Transport policy, 151*, 46–62. https://doi.org/10.1016/j.tranpol.2024.04.002

Xu, N., Pena-Bastidas, J., Yang, C., Liu, J., Hockstad, T., & Jones, S. (2025). Urban and regional Air Mobility (URAM) and relocation decisions in the United States: Insights from a machine learning-supported path analysis. *Transport Policy*. https://doi.org/10.1016/j.tranpol.2025.05.021

Xu, Y., Yan, X., Liu, X., & Zhao, X. (2021). Identifying key factors associated with ridesplitting adoption rate and modeling their nonlinear relationships. *Transportation Research Part A: Policy and Practice, 144*, 170–188. https://doi.org/10.1016/j.tra.2020.12.005

Yang, K., Yang, H., & Du, L. (2023). A data-driven traffic shockwave speed detection approach based on vehicle trajectories data. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*. https://doi.org/10.1080/15472450.2023.2270415

Yang, L., Li, X., Guan, W., & Jiang, S. (2022). Assessing the relationship between driving skill, driving behavior and driving aggressiveness. *Journal of Transportation Safety and Security, 14*(5), 737–753. https://doi.org/10.1080/19439962.2020.1812785

Yu, R., Zheng, Y., Abdel-Aty, M., & Gao, Z. (2019). Exploring crash mechanisms with microscopic traffic flow variables: A hybrid approach with latent class logit and path analysis models. *Accident Analysis and Prevention, 125*, 70–78. https://doi.org/10.1016/j.aap.2019.01.022

Zahid, M., Habib, M. F., Ijaz, M., Ameer, I., Ullah, I., Ahmed, T., & He, Z. (2024). Factors affecting injury severity in motorcycle crashes: Different age groups analysis using Catboost and SHAP techniques. *Traffic Injury Prevention, 25*(3), 472–481. https://doi.org/10.1080/15389588.2023.2297168

Zhang, G., Tan, Y., & Jou, R. C. (2016). Factors influencing traffic signal violations by car drivers, cyclists, and pedestrians: A case study from Guangdong, China. *Transportation Research Part F: Traffic Psychology and Behaviour, 42*, 205–216. https://doi.org/10.1016/j.trf.2016.08.001

Zhang, J., Li, Z., Pu, Z., & Xu, C. (2018). Comparing prediction performance for crash injury severity among various machine learning and statistical methods. *IEEE Access, 6*, 60079–60087. https://doi.org/10.1109/ACCESS.2018.2874979

Zhang, X., Zhou, Z., Xu, Y., & Zhao, X. (2024). Analyzing spatial heterogeneity of ridesourcing usage determinants using explainable machine learning. *Journal of Transport Geography, 114*. https://doi.org/10.1016/j.jtrangeo.2023.103782

Zhang, Z., Nie, Q., Liu, J., Hainen, A., Islam, N., & Yang, C. (2024). Machine learning based real-time prediction of freeway crash risk using crowdsourced probe vehicle data. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 28*(1), 84–102. https://doi.org/10.1080/15472450.2022.2106564

Zhang, Z., Xu, N., Liu, J., & Jones, S. (2024). Exploring spatial heterogeneity in factors associated with injury severity in speeding-related crashes: An integrated machine learning and spatial modeling approach. *Accident Analysis and Prevention, 206*. https://doi.org/10.1016/j.aap.2024.107697

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society, 20*, 22–35. https://doi.org/10.1016/j.tbs.2020.02.003

Zhu, J., Zou, H., Rosset, S., & Hastie, T. (2009). Multi-class AdaBoost *. In *Statistics and Its Interface* (Vol. 2). doi: 10.4310/SII.2009.v2.n3.a8.

Ningzhe Xu is a Ph.D. candidate of the Department of Civil, Construction and Environmental Engineering at the University of Alabama. His research interests include travel behavior, evacuation behavior, and roadside responder safety.

Jun Liu, Ph.D. is an Associate Professor of Civil Engineering and the Director of NextGen Transportation Lab at the University of Alabama. His research interests include shared mobility and planning, road safety and responder safety, ITS/CAVs, and sustainable transportation. Dr. Liu has 200+ publications including five book chapters, 60+ journal articles, 120+ conference papers, and 20+ technical reports. Dr. Liu has served as the PI or co-PI in a broad range of grants/projects sponsored by NSF, US DOT, AAA Foundation, and State DOTs (Alabama and Georgia). Dr. Liu is currently an Assistant Editor for the Journal of Intelligent Transportation Systems and a Handling Editor for the journal of Transportation Research Record. He is also an Editorial Board Member for the journal of Accident Analysis & Prevention and the Journal of Safety Research.

Zihe Zhang, Ph.D.is a Postdoctoral Research Associate at the Alabama Transportation Institute (ATI). Her research interests include transportation safety, intelligent transportation systems, shared mobility, micro-mobility, automation, and machine learning applications in transportation research. She received her B.S. and M.S. degrees from Chang'an University in China and her doctoral degree from the University of Alabama.

Steven Jones has more than 20 years experience in transportation engineering and planning. He has participated in projects in the United States, Europe, Africa, and Asia. Jones has served as principal investigator on more than $10 million in externally-sponsored projects. He has authored or co-authored more than 150 journal articles, conferences papers, design manuals, and project reports on a range of transportation topics. He directs the Transportation and Human Development Lab at the University of Alabama.