Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 10, No. 06, June, 2024

XGBoost-SHAP and Unobserved Heterogeneity Modelling of Temporal Multivehicle Truck-Involved Crash Severity Patterns

Wimon Laphrom [1], Chamroeun Se [1], Thanapong Champahom [2], Sajjakaj Jomnonkwao [3], Warit Wipulanusat [4], Thaned Satiennam [5], Vatanavongs Ratanavaraha [3*]

[1] *Institute of Research and Development, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.*

[2] *Department of Management, Faculty of Business Administration, Rajamangala University of Technology Isan, Nakhon Ratchasima 30000, Thailand.*

[3] *School of Transportation Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.*

[4] *Department of Civil Engineering, Thammasat School of Engineering, Thammasat University, Pathumthani 12120, Thailand.*

[5] *Department of Civil Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand.*

Received 14 February 2024; Revised 03 May 2024; Accepted 09 May 2024; Published 01 June 2024

Abstract

This paper aims to address the critical issue of multivehicle truck crashes in developing regions, with a focus on Thailand, by analyzing the factors that influence injury severity and comparing the effectiveness of predictive models. Utilizing advanced random parameters and the XGBoost machine learning algorithm, we conducted a comprehensive analysis of injury severity factors in multivehicle truck-involved accidents, contrasting weekdays and weekends. Our findings reveal that the XGBoost model significantly outperforms the heterogeneous logit model in predicting crash severity outcomes, demonstrating superior accuracy, sensitivity, specificity, precision, F1 score, and area under the curve (AUC) in both model training and testing phases. Key risk factors identified include motorcycle involvement, head-on collisions, and crashes occurring during late night/early morning hours, with environmental elements like road lane numbers and weekend hours also playing a significant role. The study introduces XGBoost as a novel and improved method for truck safety analysis, capable of capturing the complex interactions within multivehicle crash data and offering actionable insights for targeted interventions to reduce crash severity. By highlighting specific risk factors and the effectiveness of XGBoost, this research contributes to the development of data-driven strategies for enhancing truck safety in developing countries.

Keywords: Truck-Involved Crashes; Injury Severities; Random Parameters; Machine Learning; eXtreme Gradient Boosting; SHAP.

1. Introduction

Trucks play a crucial role in powering Thailand's economy and transporting goods across the nation. However, they also pose a significant risk on the roads, being involved in a disproportionately high number of fatal crashes. Despite constituting only 20% of the vehicle population, trucks are implicated in 23% of fatal accidents, a stark contrast to the 15% attributed to non-truck vehicles [1-3]. This scenario underscores a pressing need for a balance between economic benefits and public safety, emphasizing the role of data-driven insights in crafting effective interventions.

Previous studies have extensively analyzed truck crash severity across various dimensions, such as rural versus urban settings, time of day, and specific road conditions [4–19]. These investigations highlight the multifaceted nature of crash severity factors but often focus on high-income countries or specific sub-groups of accidents. Crucially, the

* Corresponding author: vatanavongs@g.sut.ac.th

http://dx.doi.org/10.28991/CEJ-2024-010-06-011

differentiation between weekday and weekend crash severity has not been rigorously explored, particularly in the context of developing nations like Thailand. This gap in the literature points to an overlooked area of significant importance, given the unique socio-economic and infrastructural challenges faced by such countries.

Addressing this gap, our study delves into the severity of multivehicle truck crashes in Thailand, comparing weekday and weekend incidents through a dual-methodological approach that combines traditional statistical models and advanced machine learning techniques, including XGBoost. This novel comparison not only highlights the methodological advancements in injury severity prediction but also emphasizes the local nuances in Thailand, a context significantly underrepresented in current literature.

Most existing research leans towards developed countries, primarily the United States, with limited focus on the unique dynamics of truck crashes in developing contexts (see review of literature in Table 1). Furthermore, the potential of machine learning models, particularly XGBoost, to outperform traditional models in predicting injury severity has only recently been recognized, with little application in the nuanced settings of countries like Thailand [20–23].

By integrating these advanced methodologies to investigate a critical yet underexplored domain, our study aims to fill a significant knowledge gap. We propose not only to enhance the understanding of multivehicle truck crash severity in Thailand but also to contribute methodologically to the broader field of transportation safety research. Through this dual focus, we seek to offer actionable insights for reducing crash severity and, by extension, the associated human and economic tolls in developing regions reliant on truck transportation.

2. Literature Review

Table 1 reviews injury severity findings across truck crash studies over the past decade, highlighting key details on statistical methods, authorship, geographies, and insights. Most research centers on US data, with just one recent study in a developing nation (Iran) [24]. This definitively spotlights the need for further investigating developing contexts like Thailand, where truck crashes inflict immense economic and social costs. By extending severity analysis to the pressing yet overlooked Thai setting, this study helps address a critical knowledge gap regarding the truck safety landscapes of non-US developing countries. In an era when trucks persistently endanger lives across the developing world, related research and policy interventions must become more inclusive. This work marks an important step toward understanding and enhancing truck safety amidst the unique challenges in Thailand and similarly situated nations.

Table 1. Review of studies on injury severity in truck-related crashes over the past decade

Statistical approaches	Authors	Regions	Key findings
Random parameters ordered probit model	Islam & Hernandez [25]	Texas, US	Increased injury severity is associated with excessive speed, involvement of multiple vehicles, curved road conditions, reduced visibility in darkness, and lane changes.
Classification and Regression Tree (CART) model	Chang & Chien [21]	Taiwan	Driving under the influence of alcohol, the use of seatbelts, and the type of vehicle emerged as the primary factors significantly impacting the severity of injuries in truck accidents.
Random parameters logit model	Islam et al. [5]	Alabama, US	The effects of different factors on the severity of injuries differed depending on whether the accidents occurred in urban or rural areas.
Hierarchical random intercept model	Chen et al. [4]	New Mexico	Inclined roadways, isolated rural truck accidents involving a single vehicle, youthfulness, and the influence of alcohol or drugs are contributing factors to severe crashes.
Multinomial logit model	Dong et al. [26]	Tennessee, US	A significant percentage of trucks, unorthodox road layouts, adverse weather circumstances, and the driver's condition were more likely associated with fatal or serious injuries, moderate injuries, possible injuries, and damage-only accidents, respectively.
Random parameters logit model	Pahukula et al. [6]	Texas, US	The severity of injuries from incidents with large trucks was observed to vary based on factors such as traffic flow, illumination conditions, the state of the road surface, changes in season, and the percentage of trucks involved during various time intervals.
Random parameters logit and ordered logit model	Naik et al. [27]	Nebraska, US	The severity of injuries has associations with variables such as wind speed, precipitation, humidity, and air temperature.
Generalized ordered response logit model.	Osman et al. [28]	Minnesota, US	Elements contributing to a higher probability of serious accidents in work zones encompass incidents occurring during the day, lack of access restrictions, increased speed limits, and crashes on rural main roads.
Ordered probit models	Hao et al. [8]	US	Enhanced speed management for trucks can substantially reduce the severity of driver injuries, and fatigue-related driving was a notable predictor of increased injury severity.
Multinomial logit model	Dong et al. [29]	Tennessee, US	The likelihood of injury severity is reduced when truck drivers use seat belts and when accidents occur during daylight. However, the likelihood of severe injuries increases in rural road crashes.
Random parameters logit model	Uddin & Huynh [11]	Ohio, US	The severity of injuries varied based on average annual daily traffic (AADT), speed, and weather conditions.
Random parameters ordered probit model	Al-Bdairi & Hernandez [30]	Oregon, US	Fatigue was linked to an increased likelihood of less severe crashes.
Spatial generalized ordered probit model	Zou et al. [7]	New York, US	In the afternoon and evening, single-vehicle accidents tended to result in less severe injuries, whereas multi-vehicle accidents exhibited more severe outcomes during these time frames.
Random parameters logit model	Khan & Khattak [9]	US	Truck trailers crossing railway tracks crashes involving train collisions, and senior drivers were associated with more severe crashes.
Random parameters logit model	Al-Bdairi et al. [10]	Oregon, US	Crucial elements influencing injury severity include the lack of seatbelt wear, curved roads, driver exhaustion, accidents involving vehicle rollovers, loss of control over the vehicle, and the sobriety of the driver.

Gradient boosting data mining model	Zheng et al. [22]	North Dakota and Colorado, US	Crash severity tends to rise in the presence of wet road conditions, side winds, large vehicles, impacts with approaching traffic, and the involvement of either young or senior drivers.
Bayesian logistic regression	Ahmed et al. [12]	Wyoming, US	The probability of a severe accident increases to 2.3 times when heavy trucks are part of the incident.
Partial proportional odds model	Wang & Prato [31]	China	Factors linked to heightened injury severity include slopes greater than three percent, erratic driving patterns, vehicle overloading, conditions at night, and unfavorable climates.
Random parameters with heterogeneity in means and variances model	Behnood & Mannering [13]	Los Angeles, US	The ethnicity of the driver in crashes related to backing, sideswiping, hitting objects, collisions with parked vehicles, crashes involving fixed objects, and cases where the truck driver was at fault consistently had an impact on the likelihood of injury severity.
Random threshold random parameters hierarchical ordered probit model	Rahimi et al. [24]	Iran	Severe crashes were linked to roads that are curved and steep, and elevated driving speeds.
Random parameters logit models	Uddin & Huynh [15]	Ohio, US	Under normal conditions, poor visibility (whether it's dark or well-lit) and morning rush hours led to higher injury severity. When it rained, speeding intensified the severity of injuries, while injury severity levels were also increased by snowy weather, roads with curves, and collisions occurring in the afternoon.
Random parameters ordered probit models	Shao et al. [16]	US	In car-strike-truck accidents, both drivers under 25 and those over 64 years old had a significant impact, whereas in truck-strike-car accidents, only the age group of 55 to 64 showed significance.
Hierarchical Bayesian random intercept model	Haq et al. [32]	Wyoming, US	Various factors, including the presence of junctions, road downgrades, curves, and weather conditions, were found to have different effects on the severity of injuries sustained in different types of vehicle-truck collisions.
Binary logistic regression with the Bayesian random intercept model	Tahmidul Haq et al. [17]	Wyoming, US	The severity of injuries sustained by truck drivers in crashes involving different driving actions was notably influenced by factors such as the total weight of the vehicle, the age and sex of the driver, the time of day, the conditions of illumination, and the existence of intersections.
Hierarchical Bayesian random intercept model	Haq et al. [18]	Wyoming, US	The severity of driver injuries varied based on factors such as the presence of junctions, incidents involving vehicles running off the road, the type of median, driver residency, restrictions on licenses, construction areas, and license restrictions.
Random parameters logit with heterogeneity in means and variances model	Wang et al. [19]	Beijing-Shanghai and Changchun-Shenzhen	Certain factors exhibited relative spatial or temporal consistency, such as the length of the horizontal curve, AADT, early morning, and overcast weather.
Correlated random parameter logit with heterogeneity in means model	Wen et al. [33]	China	A curve with a moderate slope and a medium radius notably increases the probability of medium severity in comparison to a curve with a larger radius and a flat slope.
Random parameters bivariate probit model with heterogeneity in means and variances	Song et al. [14]	UK	Male, young, and older individuals operating trucks, along with inappropriate actions taken by truck drivers, elevate the likelihood of injuries among truck drivers. For car drivers, older age and encounters with unsignalized crossings increase the risk of injury.

Four key methodological approaches exist for analyzing transportation safety data, each with distinct trade-offs [34]. Traditional statistical models offer interpretability but struggle to establish causality amidst data limitations. Endogeneity/heterogeneity modeling helps address omitted variables and dynamics but encounters complexity constraints [35]. Data-driven machine learning delivers excellent predictive accuracy through detecting intricate patterns, yet causal insights may be lacking. Causal inference frameworks directly tackle causality but face implementation challenges. Ultimately, methodology selection necessitates navigating the inherent tension between predictive precision and explanatory power for uncovering causal mechanisms in transportation safety phenomena. This research harnesses the strengths of both traditional and data-driven methods to generate novel insights into the factors influencing truck crash severity outcomes.

As shown in Table 1, numerous truck crash studies have adopted advanced modeling to address unobserved heterogeneity. Such techniques include random parameter ordered models [25], random parameters logit [5], random threshold hierarchical ordered probit [24], Bayesian random intercept models [28], and heterogeneous mean/variance random parameters logit [13, 14]. Extensions like correlated random parameters and heterogeneity in means/variances provide superior flexibility in capturing unseen effects [8, 36, 37]. By leveraging state-of-the-art discrete choice and machine learning methods, this research pushes truck safety analysis forward, not just improving predictive accuracy but extracting deeper insights into the latent and complex factors influencing real-world crash outcomes.

Machine learning crash analysis is advancing with computational progress [20], offering model-free prediction without requiring deep process knowledge. While complementary to statistics, just two truck crash studies in Table 1 use data-driven techniques: CART [21] and gradient boosting [22]. From a comprehensive review study, random forest leads for injury severity prediction, followed by SVM, decision trees, and KNN [20]. However, XGBoost has shown immense promise recently, achieving 97% accuracy—far beyond 73.8% for the logit model [38]. Zhang et al. [23] also found XGBoost-dominated classifiers in accuracy, kappa, F1, and AUC. For single and multi-vehicle crashes, respectively, XGBoost achieved 82%/80% accuracy, 0.61/0.57 kappa, 0.78/0.75 F1, and 0.88/0.86 AUC. Moreover, Jamal et al. [39] reported 93% XGBoost accuracy, surpassing decision trees (88%), random forests (84%), and logit (63%). In injury severity analysis, XGBoost thus consistently excels across performance metrics, meriting focused application here. By comparing discrete choice and potent new data science methods like XGBoost, this study strengthens truck crash scholarship methodologically and substantively.

2.1. Research Gap and Contributions

In addressing the exigencies of transportation safety research, our study seeks to fill three pivotal gaps: the scarcity of multivehicle truck crash injury severity analysis in developing nations, the unexplored differences in crash severity between weekends and weekdays, and the comparative efficacy of XGBoost models versus sophisticated unobserved heterogeneity models like the mixed logit. The limited focus on developing countries, such as Thailand, underscores a critical void in understanding the unique dynamics influencing truck crashes outside high-income contexts. Furthermore, the nuanced distinction between weekend and weekday crash severities remains an uncharted domain, hinting at potentially different contributing factors that have yet to be empirically substantiated. Our investigation pioneers a comparative analysis between cutting-edge XGBoost models and mixed logit models—known for their adeptness at accounting for unobserved heterogeneity—specifically within the context of Thai multivehicle truck crash severity. This approach not only underscores an innovative methodological contribution but also addresses a doubly understudied area by spotlighting the uniquely vulnerable Thai context. Ultimately, this work aims to enhance the methodological foundations for transportation injury analysis while offering targeted insights for improving truck safety in developing regions.

3. Research Methodology

3.1. Discrete Choice Models with Random Parameters

Qualitative choice models, also known as discrete choice models, are designed to characterize, elucidate, and forecast decisions among discrete options or outcomes [40]. Recent progress in regression models for limited dependent variables has been propelled by advancements in computing power and the emergence of models based on simulation. The focus of these advancements has largely been on developing behavioral models that more accurately account for variations in parameters at the individual level. To address this variability, the modeling approach assumes that coefficients randomly differ among individuals, following a continuous distribution denoted as $g(\beta_i|\theta)$. To illustrate and justify the concept of random parameter models, consider the subsequent latent process [41]:

$$Y_{ij}^{*} = X_{ij}\beta_i + \varepsilon_{ij}, \tag{1}$$

where, $\beta_i \sim g(\beta_i|\theta)$, in this equation, Y_{it}^{*} represents an unobserved process for crash i that corresponds to injury severity j, β_i denotes a vector of estimable parameters of injury severity j. X_{ij} is a vector containing explanatory variable covariates, and ε_i is the normally distributed error term. For the binary outcome model, the conditional probability $f(Y_i^{*}|X_i,\beta_i)$ is defined as [41]:

$$f(Y_i^{*}|X_i,\beta_i) = [F(X_i\beta_i)^{Y_{it}}\,[1 - F(X_i\beta_i)]^{1-Y_i} \tag{2}$$

Here, $F(.)$ denotes CDF of the disturbance term, where $F(\varepsilon) = \exp(\varepsilon)\,/1 + \exp\,(\varepsilon)$ for the logit model. In the structural model described by Equation 1, we allow the vector coefficient β_i to vary for each individual involved in the crashes. However, the exact variations in these parameters across individuals are unknown. What is known is that they follow the population probability density function $g(\beta_i|\theta)$, where θ is the distribution moments including variance and mean. A fully parametric model is established when $g(\beta_i|\theta)$ and the distribution of ε are specified. he unconditional probability density function (PDF) can be described as the weighted average of the conditional probability, which is calculated over all potential values of β. This depends on the parameters characterizing the distribution of β_i [42]:

$$P_i(\theta) = \int f(Y_i^{*}|X_i,\beta_i)\, g(\beta_i)d\beta_i = \int \frac{\exp\,(\beta_i)}{1-\exp\,(\beta_i)}\, g(\beta_i)d\beta_i \tag{3}$$

where, $P_i(\theta)$ represents the weighted probability outcome, and $g(\beta_i) = f(\beta_i|\varphi)$ is the probability density function of β with the distribution parameter φ, which can take various forms like normal, triangular, uniform, log-normal, etc.

A significant and straightforward extension of the random parameter model involves permitting the coefficients to be correlated. In this scenario, Γ represents a lower triangular matrix, often referred to as the Cholesky matrix. The observed heterogeneity can be accounted for by allowing parameter heterogeneity to be partially systematic concerning observed variables. This means it allows observed factors to influence the mean value of unobserved factors. Formally, the parameter vector can be expressed as [26, 30, 43]:

$$\beta_i = \beta + \eta Z_i + \Gamma\omega_i, \tag{4}$$

Here, β_i denotes a parameter vector for crash i, β stands for the constant term for the random parameter, Z_i is a vector of explanatory variables that capture heterogeneity in the mean of random parameters, η represents the vector of estimable parameters corresponding to Z_i , ηZ_i represents the heterogeneous term that addresses unobserved heterogeneity arising from interactions (among explanatory variables) inducing variation in the parametric function of random parameters (i.e., their means are determined by the variables in Z_i, ω_i is an unobservable $K \times 1$ (K is the number of random parameters) latent random term with a mean value of zero, resulting in the mean and variance-covariance matrix of random parameters to be $E(\beta_i|\omega_i) = \beta + \eta Z_i$ and $Var(\beta_i|\omega_i) = \Gamma\Gamma^{T}$), respectively.

Marginal effects represent outcome probability changes given a one-unit increase in an independent variable, with other predictors held constant. By translating statistical significance into tangible meaning, they enhance model interpretability and policy relevance. This study computed average marginal effects (AME) for both heterogeneous logit models. The AME equations for the two models are [42]:

$$AME_{X_m} = \frac{1}{c}\sum_{q=1}^{c}[(P_m(X_{mq} = 1) - \left(P_m(X_{mq} = 0)\right)] \tag{5}$$

where X_m denotes the input variable m (e.g., attributes from the roadway, crash characteristics, etc.), c is the total number of observations that change their value from 0 to 1, and X_{mq} denotes the qth crash of the input X_m.

3.2. eXtreme Gradient Boosting (XGBoost)

XGBoost, an extension of gradient-boosting algorithms, was introduced by Chen & Guestrin [44] as an efficient and scalable approach to gradient-boosting trees. It is an ensemble method that builds upon Friedman's original Gradient Boosting approach [45], enhancing classification accuracy by recursively adding models that predict residuals from prior models. XGBoost is capable of utilizing maximum memory and hardware resources for data-intensive models, incorporating sparsity-aware data handling, and employing a weighted quantile sketch for approximate learning. These optimizations, achieved through insights on cache access patterns, data compression, and sharding, enable the training of billions of samples even in resource-constrained environments [31]. In the context of XGBoost modeling, we begin by describing the regularized learning objective. In mathematical terms, given a dataset *D* with *k* samples and *m* features, denoted as $D = \{(X_i, Y_i)\}$ $(\lceil D \rceil = k, X_i \in \mathbb{R}^m, Y_i \in \mathbb{R}$, tree ensemble model employs *N* additive functions to predict the output [44]:

$$\hat{Y}_i = \phi(X_i) = \sum_{n=1}^{N} f_n(X_i), f_n \in F, \tag{6}$$

Here, $F = \left\{f(x) = \omega_{q(x)}\right\}(q : \mathbb{R}^m \to T, \omega \in \mathbb{R}^T)$ denotes the space of regression trees (specifically, CART), *N* is the number of trees, *F* represents the space encompassing all trees, *q* is a tree structure, *T* denotes the number of tree leaves, f_n is an independent tree structure q, and leaf weight ω, and $q(x)$ corresponds to the input data. To learn the set of functions employed in the model, a regularized objective function can be minimized as follows:

$$\Lambda(\theta) = \sum_i l\left(\hat{Y}_i, Y_i\right) + \sum_n \Omega(f_n) \tag{7}$$

where $\Omega(f_n) = \gamma\mathrm{T} + \frac{1}{2}\lambda\|\omega\|^2$

where l denotes a differentiable convex loss function measuring the prediction $\hat{Y}_i$ and the target Y_i. The second term Ω penalizes the model's complexity, which consists of the regression tree functions. γ and λ are the regularization parameters. This additional regularization term helps smooth the final learned weights to prevent overfitting. In essence, the regularized objective tends to favor a model that employs simple yet predictive functions. However, optimizing the tree ensemble model is not feasible using traditional Euclidean space optimization methods. Instead, the model is trained in an additive manner, following this loss function:

$$\Lambda^{(t)} = \sum_{i=1}^{k} l(Y_i, \hat{Y}_i^{t-1} + f_t(X_i)) + \Omega(f_t) \tag{8}$$

This equation incrementally adds the function f_t that most improves the model according to Equation 6.

3.3. SHapley Additive exPlanations (SHAP)

XGBoost feature importance scores quantify variables' contribution to predictions, based on decision tree usage and impurity reductions. While insightful for variable selection and highlighting influential predictors, they do not convey relationship directions like discrete outcome model coefficients.

Compared to traditional methods, SHAP values not only identify the impact of each feature on a prediction but also reveal complex interactions between features. This approach aids in ranking feature importance, explaining model behavior globally and locally, and serving as a valuable tool for model improvement. Therefore, this study adopted this approach for machine learning model interpretation [31, 46]. The SHAP value of a variable is determined using the following equation [47]:

$$\emptyset_i = \sum_{S \subseteq \mathrm{X}\setminus[\mathrm{i}]} \frac{|S|!(|X| - |S| - 1)!}{|X|!}[f_{S\cup[i]}\left(x_{S\cup[i]}\right) - f_S(x_S)] \tag{9}$$

where $\emptyset_i$ represents the SHAP value or the marginal contribution of a variable. *X* denotes all variables, with S being a subset of these, and x_S indicating the values of the variables within *S*. To analyze the impact of a particular feature, one model, $f_{S\cup[i]}$ is created including the feature in question, while another model, f_S is trained without the variable of interest. The outcomes of these models are then compared to the existing output, shown as $f_{S\cup[i]}\left(x_{S\cup[i]}\right) - f_S(x_S)$. As the feature of interest is influenced by other variables in the model, this comparison is performed across every possible subset to calculate the differences.

3.4. Research Process

The research methodology, encapsulated in a detailed flow chart in Figure 1, commences with sourcing truck-related crash data for the years 2016-2020 from the Highways Accident Information Management System (HAIMS) Database, provided by the Department of Highways (DOH). This comprehensive dataset allows us to focus specifically on multivehicle truck-involved crashes, further categorizing incidents into weekend and weekday crashes. A preliminary analysis using a Likelihood Ratio Test confirms the differences in crash characteristics between these two categories, establishing a foundation for our comparative study. Subsequently, the data is partitioned into training and testing sets to facilitate the development and evaluation of our classification models. This phase includes a rigorous process of model training, hyperparameter tuning through an iterative refinement process, and evaluation to ascertain the models' effectiveness, leveraging metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The culmination of this process identifies the best model with optimized hyperparameters, which undergoes an overfitting assessment to ensure its reliability and accuracy. To enhance the interpretability of our findings, we apply SHapley Additive exPlanations (SHAP) for a comprehensive result interpretation. This methodological approach not only enables a direct comparison of model performances—specifically between the XGBoost Model and Discrete Choice Models with Random Parameters—but also facilitates a deeper understanding of the key findings and informs our recommendations for enhancing truck safety measures.

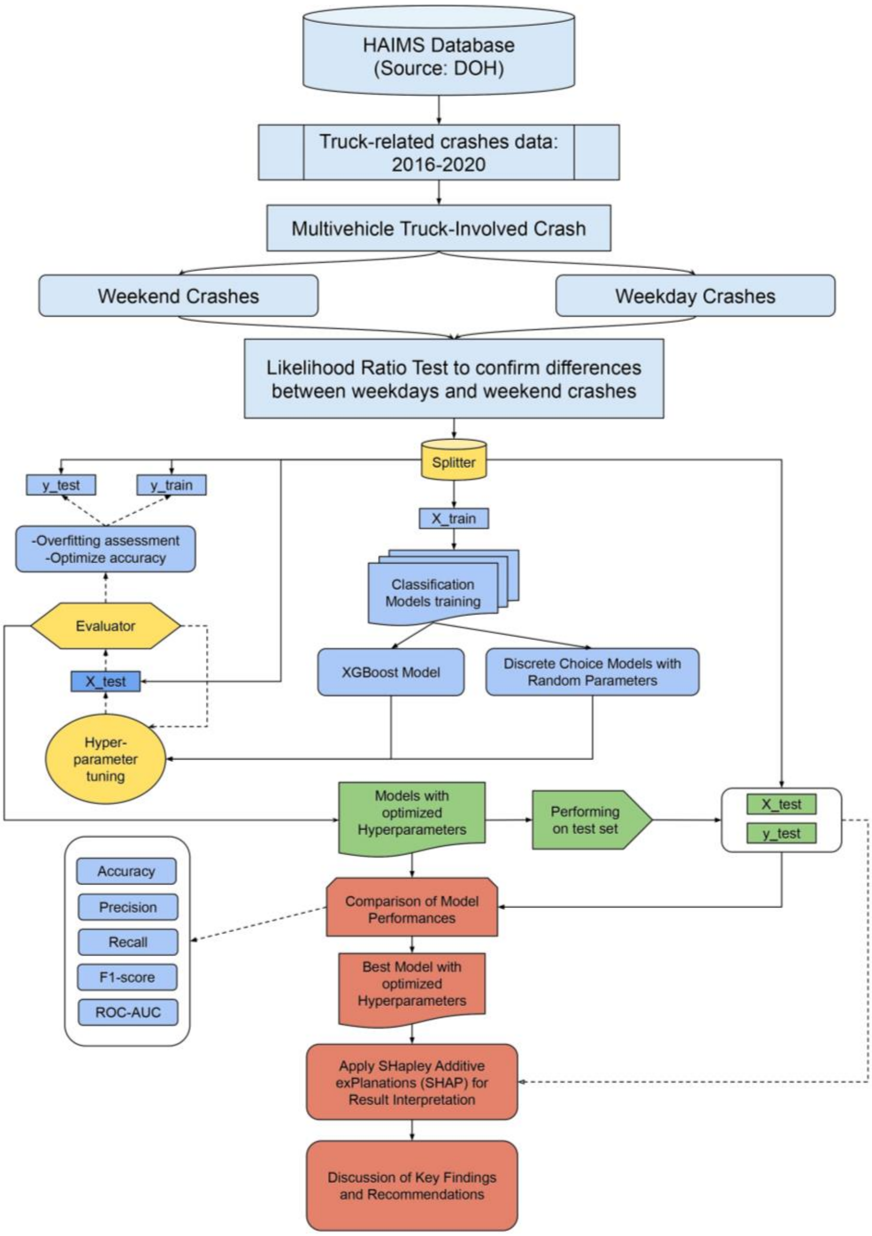

Figure 1. Research process flowchart

4. Data Collection

This study utilized Thailand's Department of Highways (DOH) data for 8,380 multivehicle truck crashes from 2016-2020, with weekend (24.5%) and weekday (75.5%) subsets exhibiting 689 and 2,160 severe/fatal crashes, respectively. Figure 2 maps the geographical distribution. The accident data offered detailed insights into a multitude of factors, such as characteristics and conditions of the road, conditions of driving, specifics of the accident, types of collisions, types of other vehicles involved, attributes related to the size of the truck, and chronological aspects (like the time of day). Table 2 presents descriptive statistics of the explanatory variables.

Figure 2. Geographical distribution of the multivehicle truck-involved crashes in Thailand from 2016-2020

Best practices mandate splitting data into mutually exclusive training (85% here) and test (15%) sets [48]. Key rationales include enabling unbiased performance evaluation on new data, detecting overfitting, parameter tuning, assessing generalizability, analyzing bias-variance trade-offs, model comparison, and simulating real-world conditions. By reserving an out-of-sample test set, models can be objectively evaluated on their ability to accurately predict unseen crash outcomes in applied settings. Both weekday and weekend crash severity analyses adopted this rigorous framework.

To justify our decision to separately estimate the models for weekday and weekend multivehicle truck-involved crashes, we conducted a transferability test using the likelihood ratio test (LRT) (i.e., $\chi^2 = -2[LL(\beta_{Total}) - LL(\beta_{Weekday}) - LL(\beta_{Weekend})]$. The resulting χ^2 value of 163.1, distributed with 46 degrees of freedom, indicates that estimating both weekday and weekend multivehicle truck-involved crashes separately is statistically warranted at a confidence level of 99.99%.

5. Model Evaluation

This study utilizes key evaluation metrics to benchmark model performance [49, 50]:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{10}$$

$$\text{Recall (Sensitivity)} = \frac{\text{TP}}{\text{TP+FN}} \tag{11}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{13}$$

$$\text{F} - \text{Score} = \frac{\text{2TP}}{\text{2TP+FN+FP}} \tag{14}$$

$$\text{AUC} = \frac{1}{2}(\text{Recall} + \text{Specificity}) \tag{15}$$

Table 2. Input variables description and descriptive statistic

Variable	Weekday		Weekend	
	Mean	**SD**	**Mean**	**SD**
Roadway Characteristics				
Main Lane	0.105	0.307	0.093	0.290
Frontage Lane	0.041	0.198	0.042	0.200
Lane = 4	0.404	0.491	0.388	0.487
Lane = 6/8	0.293	0.455	0.300	0.458
Lane > 8	0.079	0.270	0.084	0.277
Flush Median	0.065	0.246	0.054	0.225
Raised Median	0.192	0.394	0.175	0.380
Depressed Median	0.318	0.466	0.317	0.466
Barrier Median	0.209	0.407	0.228	0.420
Concrete Pavement	0.168	0.374	0.164	0.370
Curve road	0.097	0.296	0.101	0.301
Slope road	0.042	0.200	0.045	0.208
Intersection 4-leg	0.049	0.216	0.055	0.228
Intersection 3-leg	0.053	0.224	0.050	0.217
U-Turn	0.058	0.233	0.052	0.222
Wet road	0.101	0.301	0.108	0.310
Nighttime-lit Road	0.318	0.466	0.339	0.474
Nighttime-unlit Road	0.098	0.298	0.115	0.319
Driving conditions				
Over Speed Limit	0.732	0.443	0.733	0.442
Unexpected Crossing	0.113	0.316	0.101	0.301
Fatigue	0.047	0.212	0.055	0.229
Crash characteristics				
Read-end	0.572	0.495	0.586	0.493
Sideswipe	0.200	0.400	0.184	0.388
Head-on	0.076	0.265	0.080	0.272
Second party's vehicle types				
Motorcycle	0.219	0.414	0.205	0.404
Car	0.256	0.437	0.296	0.457
Van	0.038	0.192	0.041	0.199
Pickup truck	0.322	0.467	0.318	0.466
Bus	0.023	0.151	0.024	0.154
Truck sizes				
Truck 6W	0.278	0.448	0.247	0.431
Truck 8/10W	0.269	0.444	0.266	0.442
Truck >10W	0.550	0.498	0.572	0.495
Temporal attributes				
Morning peak [7:00-8:59]	0.094	0.292	0.091	0.288
Nighttime off-peak [19:30-23:59]	0.169	0.375	0.171	0.376
Early morning [0:00-6:59]	0.217	0.412	0.239	0.427
Afternoon peak [17:00-19:30]	0.113	0.316	0.120	0.325

These metrics are computed based on values from the confusion matrix, which includes: a) True positive (TP) rate: The proportion of actual positives that are correctly predicted. b) True negative (TN) rate: The proportion of actual negatives that are correctly predicted. c) False positive (FP) rate: The proportion of actual negatives that are incorrectly predicted as positives. d) False negative (FN) rate: The proportion of actual positives that are incorrectly predicted as negatives. Together, these comprehensive metrics enable holistic assessment across critical dimensions like correctness, true positive/negative rates, predictive value, imbalance handling, and discrimination. By leveraging multi-pronged evaluation grounded in established principles, this study achieves rigorous and transparent benchmarking essential for methodological advancements. The workflow and reporting standards further bolster reliability and framing for safety applications.

Grid search optimization of key XGBoost hyperparameters (listed in Table 3) minimizes overfitting, using log loss comparisons between training and test sets [51, 52]. Based on the model performance comparison Table 4, XGBoost consistently and considerably outperforms RPBLHM for predicting severity in both weekday and weekend multivehicle truck crashes. XGBoost achieves superior performance in all metrics across both training and testing data sets.

Table 3. Hyper-parameter values extracted with the use of grid search algorithms for XGBoost Models

Parameter	Description	Weekday crash model	Weekend crash model
eta	Learning rate: [0,1]	0.1	0.3
max_depth	Maximum depth of a tree: [0,20]	3	5
gamma	Min_split_loss: [0,10]	8	2
alpha	L1 regularization term on weights: [0,10]	6	2
lambda	L2 regularization term on weights: [0,10]	5	2
objective	Objective function	"binary:logistic"	"binary:logistic"
eval_metric	Evaluation metric	"logloss"	"logloss"

Table 4. XGBoost and RPBLHM's performance comparison

Model	Data	Accuracy	Sensitivity	Specificity	Precision	F-score	AUC
Weekday multivehicle truck-involved crash models							
XGBoost	Training	0.740	0.429	0.899	0.687	0.528	0.664
	Testing	0.732	0.428	0.896	0.691	0.529	0.662
RPBLHM	Training	0.719	0.424	0.871	0.628	0.507	0.648
	Testing	0.707	0.371	0.890	0.646	0.471	0.630
Weekend multivehicle truck-involved crash models							
XGBoost	Training	0.767	0.493	0.910	0.740	0.591	0.701
	Testing	0.764	0.545	0.856	0.615	0.578	0.701
RPBLHM	Training	0.718	0.479	0.842	0.611	0.537	0.661
	Testing	0.720	0.455	0.832	0.533	0.491	0.643

As shown in Table 4, the comparative analysis reveals distinct advantages in predictive performance. For weekday crash data, the XGBoost model achieves higher accuracy rates in both training (0.740) and testing phases (0.732) compared to the RPBLHM model (training: 0.719; testing: 0.707). This indicates a more precise overall classification of crash severities by the XGBoost model.

Notably, XGBoost's performance superiority is further underscored by its enhanced sensitivity, precision, and F-score. These metrics are critical as they reflect the model's ability to accurately identify severe crashes (sensitivity), correctly predict a higher proportion of severe crash outcomes as such (precision), and effectively balance precision and sensitivity (F-score). For example, in testing, XGBoost achieves a sensitivity of 0.428 and a precision of 0.691, outperforming the RPBLHM model's sensitivity of 0.371 and precision of 0.646. Such improvements are crucial for formulating precise safety interventions, as they ensure that severe crashes are identified more reliably.

The trend continues with weekend crash data, where XGBoost uniformly surpasses RPBLHM across all evaluative metrics, achieving a testing accuracy of 0.764 compared to RPBLHM's 0.720. The higher sensitivity, precision, and F-scores of XGBoost indicate a superior capability in identifying severe crashes while minimizing incorrect severe crash predictions.

XGBoost's exceptional performance can be attributed to its advanced machine-learning techniques, including ensemble learning and differential evolution mechanisms. These techniques allow XGBoost to adeptly navigate the complexities of multivehicle truck crash data, accommodating intricate variable interactions that traditional models like RPBLHM struggle with. As such, XGBoost's methodological sophistication not only enhances the accuracy of severe crash predictions but also reduces the likelihood of false alarms, thereby providing a robust framework for deriving safety-critical insights from complex data landscapes.

In summary, our analysis underscores the methodological advantages of the XGBoost model in predicting injury severity outcomes in multivehicle truck crashes, offering a significant improvement over traditional models in terms of accuracy, sensitivity, precision, and F-scores. This comprehensive performance benchmarking establishes XGBoost as the superior approach for analyzing and interpreting nuanced crash severity data, laying the groundwork for more effective safety interventions based on precise, data-driven insights.

6. Result and Discussion

Tables 5 and 6 present the results of the fixed effect model (FEM), random parameters model (RPM), and random parameters model with heterogeneity in means model for weekday and weekend truck-involved crashes, respectively. Likelihood ratio tests justify the superiority of RPBLHM over RPBL ($p<0.05$), and RPBL over the binary logit model, for weekday and weekend crashes models. The RPBLHM weekday and weekend R^2 values reach 0.173 and 0.183, respectively - indicating acceptable binary classification injury severity predictive performance [30, 53]. Figures 3 to 6 show the results of the SHAP analysis including the feature importance and impact of the input features on the model output predictions for weekday and weekend crashes. The following subsections focus on variable impact rankings to simplify cross-model comparisons.

Table 5. Weekday multivehicle truck-involved crash result using discrete outcome models [coefficient (Z-value)]

Variable	FEM	RPM	RPBHM	AME
Constant	-2.106(-10.24)	-2.24(-10.09)	-2.249(-10.04)	
Lane = 4	0.355(2.15)	0.451(2.53)	0.452(2.54)	0.0727
Raised Median	-0.298(-1.68)	-0.378(-2)	-0.38(-2)	-0.0597
Barrier Median	-0.744(-3.96)	-0.862(-4.27)	-0.867(-4.27)	-0.1407
Mean. Concrete pavement	-0.141(-1.46)	-0.283(-1.78)	-0.216(-1.29)	-0.0119
SD. Concrete pavement		0.92(2.26)	1.179(2.98)	
Distributional split. Concrete pavement			PDO/Minor = 57.27% Severe/Fatal = 42.73%	
Concrete pavement: Car			-0.621(-2.13)	
Mean. Curve road	-0.062(-0.5)	-0.319(-1.53)	-0.452(-1.98)	-0.0406
SD. Curve road		1.526(3.12)	1.356(2.71)	
Distributional split. Curve road			PDO/Minor = 63.06% Severe/Fatal = 36.94%	
Curve road: Pickup Truck			0.464(1.63)	
Intersection 3-leg	0.245(1.71)	0.266(1.76)	0.271(1.78)	0.0457
U-Turn	0.292(2.12)	0.304(2.1)	0.301(2.06)	0.0508
Nighttime-lit Road	0.479(4.71)	0.527(4.88)	0.537(4.94)	0.0910
Nighttime-unlit Road	0.512(3.99)	0.551(4.02)	0.554(4.04)	0.0953
Over Speed Limit	0.313(2.73)	0.35(2.85)	0.352(2.87)	0.0604
Unexpected Crossing	0.288(1.97)	0.31(2.01)	0.317(2.04)	0.0537
Fatigue	0.635(3.58)	0.72(3.81)	0.732(3.86)	0.1278
Read-end	0.388(3.63)	0.393(3.44)	0.398(3.46)	0.0648
Head-on	1.439(9.54)	1.586(9.4)	1.574(9.29)	0.2862
Motorcycle	1.84(19.47)	1.956(18.4)	1.972(18.46)	0.3803
Mean. Bus	0.245(1.14)	-0.675(-0.72)	-0.658(-0.73)	0.0087
SD. Bus		3.475(1.64)	3.457(1.7)	
Distributional split. Bus			PDO/Minor = 57.55% Severe/Fatal = 42.45%	
Early morning	0.217(1.98)	0.215(1.86)	0.219(1.88)	0.0361
Model statistic				
P	37	40	42	
LL(Contant)	-3445	-3445	-3445	
LL(B)	-2861.471	-2854.286	-2850.442	
R^2	0.169	0.171	0.173	
AIC	5796.943	5788.571	5784.884	
Model comparison				
Comparison pair		FE vs. RPBL	RPBL vs. RPBLHM	
Degree of freedom		3	2	
Chi-square		14.371	7.6869	
Significant level		0.002	0.02	

Note: Random parameters were estimated based on normal distribution ω_i~N(0,1) and 1000 Halton Draw. Correlations between random parameters were not significant.

Table 6. Weekend multivehicle truck-involved crash result using discrete outcome models [coefficient (Z-value)]

Variable	BL	RPBL	RPBLHM	AME
Constant	-1.087(-2.98)	-1.341(-3.12)	-1.415(-3.29)	
Depressed Median	-0.555(-2.02)	-0.474(-1.51)	-0.458(-1.46)	-0.0649
Barrier Median	-0.871(-2.91)	-0.821(-2.38)	-0.793(-2.29)	-0.1159
Mean. Concrete pavement	0.173(1.08)	-0.111(-0.36)	-0.119(-0.39)	0.0213
SD. Concrete pavement		1.674(2.27)	1.682(2.26)	
Distributional split. Concrete pavement			PDO/Minor = 52.82% Severe/Fatal = 47.18%	
Mean. Curve road	-0.159(-0.71)	-0.513(-1.32)	0.043(0.1)	0.0514
SD. Curve road		2.107(2.36)	2.128(2.38)	
Distributional split. Curve road			PDO/Minor = 49.19% Severe/Fatal = 50.81%	
Curve road: Lane = 4			-1.544(-2.1)	
Nighttime-unlit Road	0.741(3.51)	0.886(3.67)	0.906(3.75)	0.1436
Fatigue	0.541(1.82)	0.821(2.34)	0.849(2.43)	0.1333
Mean. Sideswipe	-0.072(-0.34)	-0.446(-1.25)	-0.434(-1.22)	-0.0285
SD. Sideswipe		1.64(2.66)	1.633(2.65)	
Distributional split. Sideswipe			PDO/Minor = 60.48% Severe/Fatal = 39.52%	
Head-on	1.117(4.3)	1.206(4)	1.19(3.94)	0.1943
Motorcycle	1.794(10.47)	2.155(9.57)	2.157(9.59)	0.3794
Nighttime off-peak	0.533(2.54)	0.625(2.59)	0.612(2.54)	0.0946
Early morning	0.417(2.26)	0.474(2.22)	0.462(2.17)	0.0695
Model statistic				
P	37	40	41	
LL(0)	-1130	-1130	-1130	
LL(B)	-931.074	-925.829	-923.008	
R^2	0.176	0.181	0.183	
AIC	1936.149	1931.657	1928.016	
Model comparison				
Comparison pair		FE vs. RPBL	RPBL vs. RPBLHM	
Degree of freedom		3	1	
Chi-square		10.492	5.6417	
Significant level		0.014	0.017	

Note: Random parameters were estimated based on normal distribution ω_i~N(0,1) and 1000 Halton Draw. Correlations between random parameters were not significant.

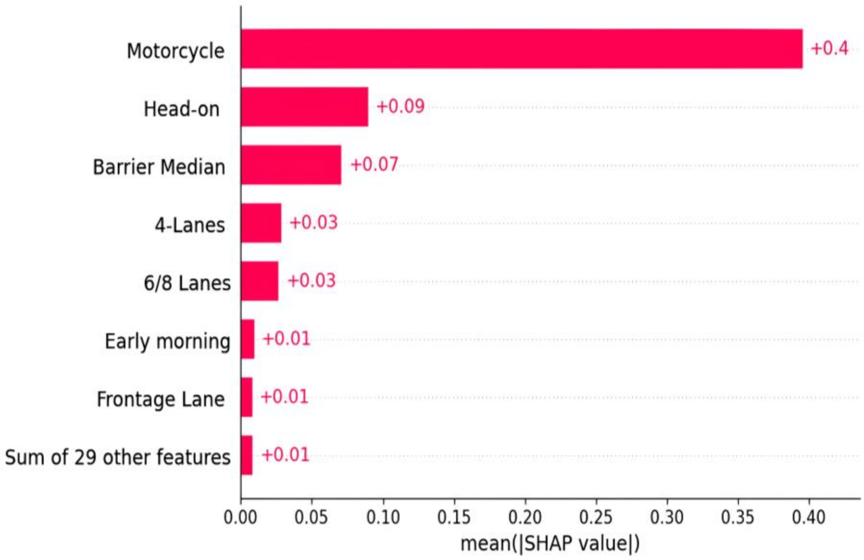

Figure 3. Feature importance of weekday crashes using SHAP

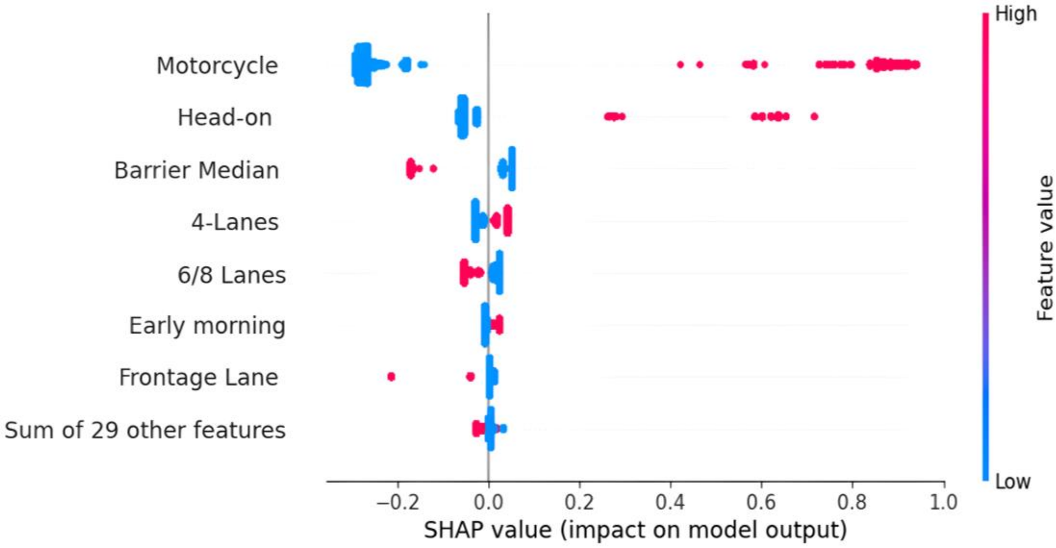

Figure 4. Impact of the input feature on the output prediction for weekday crashes using SHAP

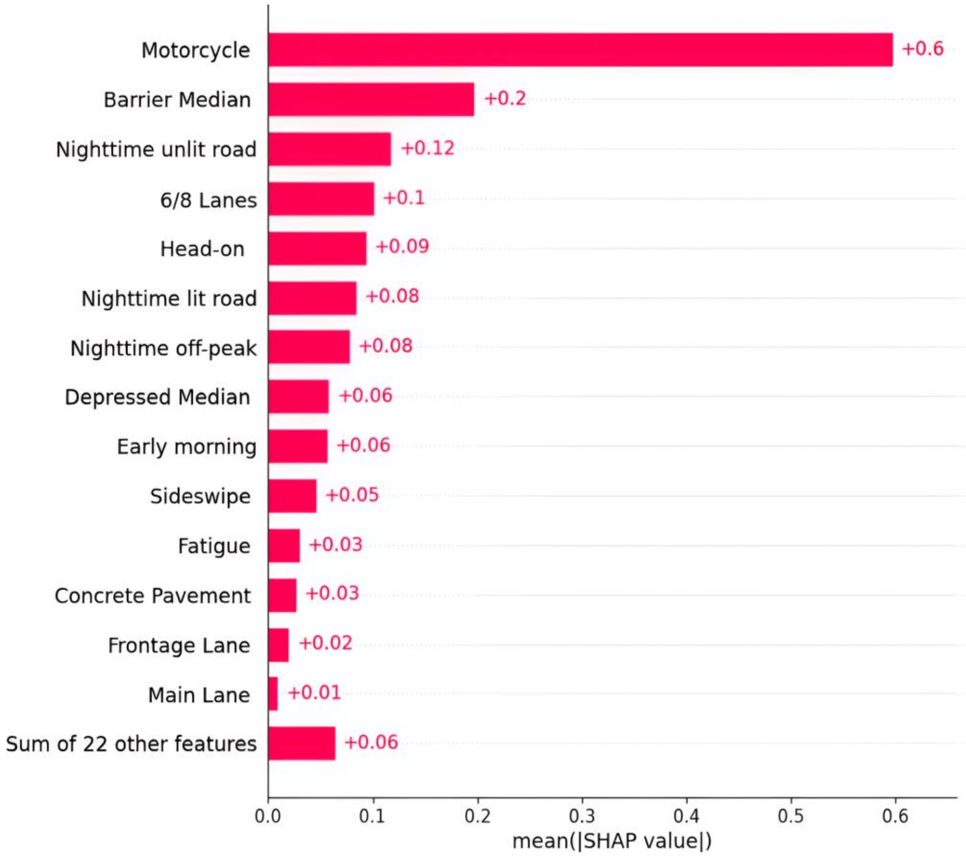

Figure 5. Feature importance of weekend crashes using SHAP

6.1. Weekday Crashes Model Result

According to the results from the XGBoost model, seven factors were pinpointed as the primary predictors of crash severity, with the involvement of a motorcycle identified as the most critical factor (as illustrated in Figures 3 and 4). This finding is consistent across both the XGBoost and RPBLHM models, underscoring motorcycle involvement as the paramount predictor of truck crash severity. In the RPBLHM model, the marginal effect of motorcycle involvement was quantified at 0.3803, signifying its substantial influence on crash outcomes. This result is both intuitive and reflective of the inherent risks associated with motorcycles. This finding is consistent with the previous studies [54-57]. Motorcycles lack the protective barriers that enclosed vehicles provide, exposing riders to greater harm in the event of a collision. Their relatively small size, when compared to large trucks, exacerbates the severity of impacts, markedly increasing the likelihood of serious or fatal injuries. The disparity in size and weight between motorcycles and trucks leads to more forceful collisions, further amplifying the risk of severe outcomes. Additionally, the absence of safety features that are standard in enclosed vehicles leaves motorcycle riders particularly vulnerable in crashes. Thus, when a motorcycle is implicated in a truck crash, the likelihood of severe or fatal outcomes escalates dramatically, a fact that is empirically supported by our model's identification of motorcycle involvement as the foremost predictor of increased crash severity.

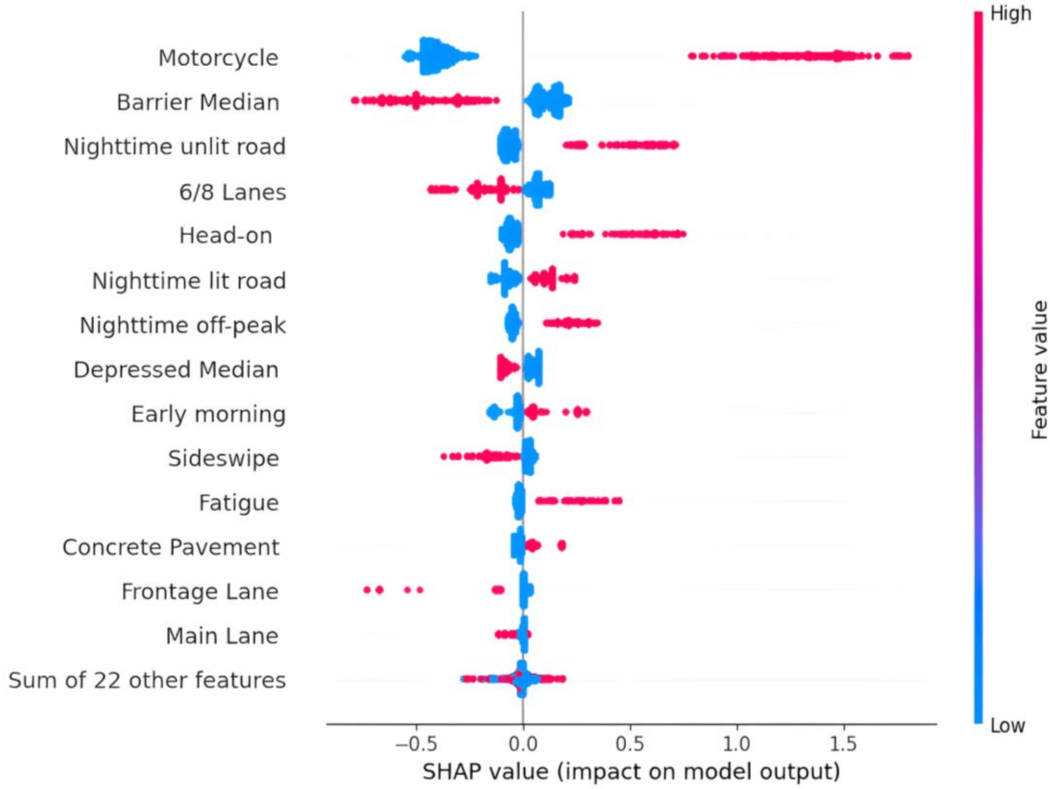

Figure 6. Impact of the input feature on the output prediction for weekend crashes using SHAP

The analysis identifies the head-on crash type as the next most significant factor influencing the severity of truck crashes, markedly increasing the likelihood of severe and fatal injuries, as depicted in Figure 4. This finding is corroborated by both the XGBoost and RPBLHM models, which consistently rank head-on collisions as the second most critical predictor of crash severity, evidenced by an Average Marginal Effect (AME) of 0.2862 in the RPBLHM model. The effect of head-on crashes found in this study is also in line with recent studies [14, 55, 58]. Head-on collisions are particularly devastating due to the sum of the velocities of the colliding vehicles, resulting in a significantly high impact force that can cause catastrophic damage [58]. The inherent size and weight disparity between large trucks and smaller passenger vehicles exacerbates the danger, with the greater mass of trucks contributing to higher kinetic energy at the moment of collision, thereby intensifying the impact [55]. Despite advancements in crash protection for trucks, the significant weight difference remains a critical factor that disadvantages them in head-on collisions with smaller vehicles, leading to a heightened risk of severe damage and occupant compartment intrusion. As such, the force exerted in head-on truck crashes significantly raises the likelihood of severe and fatal injuries. This effect is unequivocally demonstrated through the truck crash severity models, affirming head-on collisions as a paramount concern for road safety.

Within the findings of the XGBoost model, crashes occurring on roads equipped with barrier medians emerge as the third most influential factor, notably indicating that truck-involved crashes on these roads during weekdays are associated with lower levels of injury severity, as illustrated in Figure 4. This observation aligns with the outcomes presented by the RPBLHM model, suggesting a consistent understanding across modeling approaches regarding the protective impact of barrier medians. The rationale behind the reduced severity of crashes on roads with barrier medians stems from several key functions these structures serve. Firstly, median barriers effectively separate opposing lanes of traffic, significantly reducing the possibility of devastating head-on collisions and secondary impacts that often result in higher injury severities, consistent with Alrejjal, et al. [59]. Additionally, these barriers provide a crucial buffer zone, granting drivers increased reaction time to either avert accidents altogether or diminish the force of impact by reducing speed [60]. The enhanced visibility along roads with median barriers further aids in preventing hazardous lane departures, a benefit that is particularly significant for large vehicles such as trucks. Given their size and momentum, trucks find it especially challenging to cross tall concrete median barriers, thus maintaining them within their designated lanes during instances of control loss. Consequently, the presence of barrier medians acts as a preventive measure against severe truck crashes by mitigating the occurrence of head-on collisions and facilitating better control over the vehicle in emergent situations. The protective effect of barrier medians, as quantified by our model results, underscores their importance in promoting safer outcomes in truck-involved crashes on roads equipped with these structures.

The analysis conducted using the XGBoost model highlights a notable correlation between the number of lanes on a roadway and the severity of injuries resulting from truck crashes. Specifically, incidents occurring on 4-lane roads during weekdays are more likely to culminate in severe and fatal outcomes compared to those on roads with six or more lanes. Interestingly, multi-vehicle collisions on four-lane highways consistently lead to more serious injuries than similar incidents on two-lane roads or highways with six or more lanes. A similar finding was also revealed in a recent study [55]. A plausible explanation for this pattern is the unique role of four-lane highways as principal arteries connecting urban centers, often traversing rural expanses or extending over long interstate stretches. These roads are characterized by high-speed limits and heavy traffic flows, conditions that heighten the risk of high-impact rear-end collisions involving trucks [53]. The aerodynamics and momentum of trucks on such highways can exacerbate the severity of crashes, particularly in instances where vehicles crash into the rear of trucks, leading to underride accidents. These accidents often result in the intrusion of the passenger compartment, significantly increasing the likelihood of occupant ejection or fatality. Moreover, the limited space available in medians on four-lane roads may elevate the risk of secondary collisions following unsafe evasive maneuvers in response to impending crashes. This evidence underscores the heightened danger of severe injury outcomes on 4-lane roads, marking them as particularly hazardous for truck-involved incidents.

The XGBoost model has delineated crashes occurring during the early morning hours as a significant predictor of increased severity in truck crash outcomes. This period's association with severe and fatal crashes can be attributed to several factors that converge to elevate risk. Several studies have also discovered similar findings [15, 19, 27]. During these hours, traffic density typically decreases, leading to an increase in danger. With roads less congested, drivers are more inclined to accelerate, reaching higher speeds that can dramatically escalate the severity of any crash. High-speed incidents, even with minor initial loss of control, can result in significantly more violent impacts. Compounded by the early morning darkness, visibility diminishes, which crucially hampers drivers' ability to detect obstacles and react timely. The darkness of these hours reduces drivers' reaction times and complicates obstacle detection, further increasing the likelihood of severe accidents. Motorcyclists, who are already at a disadvantage due to the lack of protective enclosures and stability systems, find themselves in particularly perilous conditions. At high speeds, any collision can lead to catastrophic outcomes, including ejections or uncontrolled slides, heightening the risk of severe injuries or fatalities. These factors—reduced visibility, increased speeds, and the inherent vulnerability of motorcyclists—combine to create a perilous traffic environment during early morning hours. This explains the model’s finding that this time frame is associated with a higher likelihood of truck crashes resulting in severe and fatal injuries, underlining the need for targeted safety interventions during these hours.

6.2. Weekend Crashes Model Result

Analysis of the weekend crash data, as illustrated in Figures 5 and 6, reveals a broader array of variables significantly influencing the severity of truck-involved crashes, in comparison to weekday incidents. Echoing the patterns observed in weekday crashes, the involvement of a motorcycle in truck crashes emerges as the predominant predictor of increased injury severities, including severe and fatal outcomes. The elevated risk associated with motorcycle involvement can be attributed to a constellation of factors: the stark size disparity between motorcycles and trucks, the motorcycles' inherent instability, the absence of protective enclosures, challenges in visibility, and the pronounced likelihood of motorcycles becoming ensnared beneath trucks [56].

Parallel to findings from weekday incidents, truck crashes occurring on weekends that take place on roads with barrier medians or those spanning 6/8 lanes are less likely to result in severe or fatal outcomes, as opposed to incidents on roads with other types of medians or fewer lanes. This suggests that the presence of barrier medians and wider roadways contributes to mitigating crash severity, likely due to enhanced separation of traffic flows and increased space for maneuvering and crash avoidance.

Furthermore, consistent with weekday analysis, head-on collisions stand out as critical predictors of severe and fatal injuries during the weekend, surpassing other collision types such as sideswipe incidents. This underscores the particularly destructive nature of head-on impacts, which combine the vehicles' velocities into a singular, formidable force, dramatically increasing the potential for catastrophic damage and severe injuries.

The analysis distinctly identifies nighttime off-peak hours (19:30-23:59) as key predictors of crash severity during weekends, setting this period apart from typical weekday patterns. Despite not reporting day-of-week, the findings of previous studies support our result [7, 38]. This variance is attributed to the unique nature of weekend nighttime traffic, which often comprises a higher proportion of leisure travelers, alongside an increase in fatigued drivers and those impaired by alcohol [61]—factors less prevalent during the weekday. The likelihood of encountering intoxicated motorists escalates on weekends, particularly as late-night social activities wind down, contributing to heightened risks of severe crashes.

Additionally, the effects of fatigue, common during peak hours, extend into these weekend late-night periods. Truck drivers, at the end of long shifts and navigating in reduced visibility, face compounded risks. This combination of factors—darkness, unpredictable traffic flows, the presence of impaired road users, and driver fatigue—crafts a perilous traffic environment. The reduced reaction times and the likelihood of unsafe driving maneuvers that characterize this period amplify the potential for high-speed impacts and rollovers.

Moreover, crash severity involving impaired occupants is often exacerbated by lower seatbelt usage rates and diminished efficacy of front airbags. Thus, the confluence of these late-night factors during weekends significantly elevates the probability of severe truck crash outcomes, underscoring the distinct risks posed by this time frame in contrast to other periods.

The XGBoost model compellingly highlights that early morning hours (00:00-06:59) on weekends are significantly associated with an increased likelihood of severe truck crashes. This time frame is particularly perilous due to a combination of factors: diminished visibility under cover of darkness, heightened driver fatigue, less rigorous traffic management and enforcement, slower emergency response times, and the rigorous demands of trucking schedules. Intriguingly, the severity risks during these weekend early morning hours surpass those of weekday incidents, as evidenced by consistently higher AMEs, indicating that weekends carry distinct risks not as prevalent during the week. This observation is in line with established research findings [19, 62].

Additionally, the model underscores driver fatigue as a critical factor contributing to the severity of weekend crashes. Fatigue significantly undermines essential physical and cognitive functions, notably slowing reaction times, diminishing attention, impairing judgment, and reducing vehicle control efficacy. Previous studies have shown that fatigue can increase the likelihood of engaging in incapacitating or fatal crashes by up to eight times, especially in the context of multivehicle versus single-truck incidents [63]. Drowsy truck drivers are particularly vulnerable to failing to respond promptly to the movements of leading vehicles, drifting out of their lanes, and executing fewer defensive manoeuvres in the event of control loss. The interplay of fatigue, the veil of darkness, and the pressures of weekend trucking schedules constitutes a critical risk period in the early morning hours for truck operations.

6.3. Unobserved Heterogeneity

Despite involving more than two random parameters, neither the weekday nor weekend RPBLHM models exhibited significant random parameter correlations. Variables that were found to have varied effects on the weekday crash outcomes are concrete pavement, curve road, and crashes involving buses; whereas variables with varied effects on the weekend crash outcomes are concrete pavement, curve road, and sideswipe crash type (Table 5 and 6). However, both groups featured noteworthy heterogeneity in means. Weekday model insights include car involvement decreasing concrete pavement means, reducing severity likelihoods; and pickup trucks increasing curved road means, escalating severe/fatal odds. Meanwhile, for weekends, four-lane roads decreased curved road means, heightening PDO/minor probabilities. While random parameter correlations proved negligible, influential heterogeneity subtleties emerged - with vehicle types and road features interactively impacting day-specific risks through accident site means effects.

7. Summary and Conclusion

Multivehicle truck crashes persistently endanger lives and economies across developing regions, including middle-income Thailand. This study tackles this urgent problem through a comprehensive analysis of injury severity factors in Thai multivehicle truck-involved accidents, contrasting weekdays and weekends via advanced random parameters and XGBoost models.

The results reveal XGBoost as a breakthrough methodological asset for truck safety analysis. Across both weekday and weekend scenarios, XGBoost consistently and considerably surpasses heterogeneous logit at predicting truck crash severity outcomes. Specifically, XGBoost demonstrates unambiguously superior performance in accuracy, sensitivity, specificity, precision, F1 score, and AUC in both model training and testing. This evinces its excellence at correctly classifying severe crash outcomes while minimizing false alarms - achieving the elusive balance that defines state-of-the-art predictive power. Fundamentally, the dominance of XGBoost stems from its ensemble approach being uniquely equipped to handle the intricate data relationships inherent in truck safety datasets. The complex interplay of spatiotemporal factors, vehicle types, road features, and crash dynamics defies simplification. Yet XGBoost algorithmic flexibility allows it to capture subtle data nuances that evade traditional techniques like regression. This research thereby pioneers and validates XGBoost as an ideal injury severity prediction tool for illuminating the most influential risk factors from multifaceted multivehicle truck crash data. The actionable insights unlocked can thus guide transportation agencies towards targeted, optimized interventions to curtail truck crash severity more effectively than previously possible. Overall, the capability demonstrated by XGBoost represents an invaluable asset for enabling pivotal, lifesaving advances in data-driven truck safety.

This analysis provides new insight into factors influencing the heightened severity of truck crashes. Consistently across models, motorcycle involvement, head-on collisions, and late-night/early AM crashes emerged as predominant risk factors escalating truck crash severity. Furthermore, certain environmental factors, including barriers, road lane numbers, and weekend hours, were shown to disproportionately modulate severity likelihoods. Crucially, significant day-specific heterogeneity and interacting effects also transpire through diverse vehicle-road variable permutations. These findings spotlight the outsized threats posed by interactions between vulnerable road users and complex spatiotemporal conditions in exacerbating truck crash outcomes. Accordingly, infrastructure refinements and targeted enforcement should concentrate on locations with 4/fewer lanes undergoing periodic weekday volume surges. Meanwhile, trucking oversight prioritization should emphasize addressing schedule pressures encouraging fatigued, early AM, and late-night weekend driving, especially among operators exhibiting dangerous or intoxicated behaviors. Ultimately, this work highlights the reality of residual truck crash severity risks that will likely require multi-pronged solutions beyond vehicle-based countermeasures alone.

The findings offer actionable insights that can significantly inform and refine interventions and safety measures aimed at reducing the severity of multiple truck crashes. Based on the analysis results, several targeted interventions are proposed: (1) Enhanced Motorcycle Awareness: Given the high risk associated with motorcycle involvement in truck crashes, increasing awareness among truck drivers through training programs about motorcycles' vulnerabilities and promoting the use of advanced detection technologies could mitigate risks. Additionally, campaigns aimed at motorcyclists to maintain safe distances and visibility near trucks can be beneficial. (2) Infrastructure Improvements for Head-On Collision Prevention: The study highlighted head-on collisions as a major contributor to crash severity. Implementing physical barriers and median separations on highways, especially on routes identified with higher occurrences of such crashes, could prevent these high-impact events. Moreover, investing in road design that reduces the likelihood of head-on collisions, such as clearer lane demarcations and rumble strips, is crucial. (3) Regulating and Monitoring Driving Hours: Our findings regarding the timing of crashes suggest the need for stricter enforcement of regulations concerning driving hours, especially for truck drivers, to combat fatigue. Implementing electronic logging devices (ELDs) to accurately monitor and enforce rest periods for truck drivers can reduce crashes due to driver fatigue, particularly during early morning and late-night hours. (4) Alcohol and Substance Abuse Interventions: Considering the elevated risk of crashes during nighttime and early morning hours on weekends, interventions to reduce alcohol and substance abuse among all drivers, including increased DUI checkpoints and educational campaigns about the dangers of driving under the influence, are recommended. (5) Speed Management Strategies: Implementing dynamic speed limits based on the time of day and traffic conditions, along with stricter enforcement of existing speed regulations, can address the risks associated with high-speed collisions during less congested periods. These interventions and safety measures can significantly reduce the occurrence and impact of such incidents. Collaborative efforts among policymakers, transportation agencies, and the community are essential to implementing these strategies effectively and ensuring safer roadways for all users.

Our study uncovered surprising insights, particularly the pronounced influence of nighttime off-peak hours on weekends as a significant predictor of crash severity, underscoring behavioral or environmental differences that merit further exploration. The substantial protective effect of barrier median roads on reducing the severity of head-on collisions and rollovers was unexpected, highlighting roadway infrastructure's crucial role in crash mitigation. Additionally, the pronounced impact of motorcycle involvement across both weekdays and weekends as a key severity predictor points to the need for targeted safety measures for trucks and motorcycles. These findings challenge existing assumptions, suggesting new research avenues and multidisciplinary intervention strategies that include vehicle technology, road design, and traffic management to address the complex dynamics of truck crash severity.

Lastly, assessing the real-world impact of these interventions on reducing the severity of multivehicle-truck crashes and enhancing overall road safety is a critical next step. This would not only validate the theoretical contributions of the current research but also provide empirical evidence to guide policymakers and transportation authorities in implementing the most effective measures. This approach contributes further to the development of evidence-based strategies for improving road safety and minimizing the human and economic costs associated with multivehicle truck crashes.

8. Declarations

8.1. Author Contributions

Conceptualization, W.L. and C.S.; methodology, W.L. and C.S.; software, W.L.; validation, W.L., T.C., S.J., W.W., and T.S.; formal analysis, W.L. and C.S.; investigation, S.J. and V.R.; resources, S.J. and V.R.; data curation, W.L. and C.S.; writing—original draft preparation, W.L. and C.S.; writing—review and editing, W.L., C.S., and T.C.; visualization, C.S.; supervision, S.J. and V.R.; project administration, V.R.; funding acquisition, V.R. All authors have read and agreed to the published version of the manuscript.

8.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

8.3. Funding

This research has received funding support from (i) Suranaree University of Technology (SUT), (ii) the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation (PMU-B) [grant number B13F660067] and (iii) Thailand Science Research and Innovation (TSRI).

8.4. Acknowledgements

The authors express their gratitude to the Department of Highways of Thailand for providing access to the comprehensive nationwide road traffic crash.

8.5. Conflicts of Interest

The authors declare no conflict of interest.

9. References

[1] Boonpanya, T., & Masui, T. (2021). Assessing the economic and environmental impact of freight transport sectors in Thailand using computable general equilibrium model. Journal of Cleaner Production, 280, 124271. doi:10.1016/j.jclepro.2020.124271.

[2] NESDB. (2012). Structure of the GDP. National Economic & Social Development Board (NESDB), Bangkok, Thailand.

[3] DOH. (2024). Thailand traffic accident on national highways. Thailand Department of Highways (DOH), Bangkok, Thailand.

[4] Chen, C., Zhang, G., Tian, Z., Bogus, S. M., & Yang, Y. (2015). Hierarchical Bayesian random intercept model-based cross-level interaction decomposition for truck driver injury severity investigations. Accident Analysis and Prevention, 85, 186–198. doi:10.1016/j.aap.2015.09.005.

[5] Islam, S., Jones, S. L., & Dye, D. (2014). Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama. Accident Analysis and Prevention, 67, 148–158. doi:10.1016/j.aap.2014.02.014.

[6] Pahukula, J., Hernandez, S., & Unnikrishnan, A. (2015). A time of day analysis of crashes involving large trucks in urban areas. Accident Analysis and Prevention, 75, 155–163. doi:10.1016/j.aap.2014.11.021.

[7] Zou, W., Wang, X., & Zhang, D. (2017). Truck crash severity in New York city: An investigation of the spatial and the time of day effects. Accident Analysis and Prevention, 99, 249–261. doi:10.1016/j.aap.2016.11.024.

[8] Hao, W., Kamga, C., Yang, X., Ma, J. Q., Thorson, E., Zhong, M., & Wu, C. (2016). Driver injury severity study for truck involved accidents at highway-rail grade crossings in the United States. Transportation Research Part F: Traffic Psychology and Behaviour, 43, 379–386. doi:10.1016/j.trf.2016.09.001.

[9] Khan, W. A., & Khattak, A. J. (2018). Injury Severity of Truck Drivers in Crashes at Highway-Rail Grade Crossings in the United States. Transportation Research Record, 2672(10), 38–47. doi:10.1177/0361198118781183.

[10] Al-Bdairi, N. S. S., Hernandez, S., & Anderson, J. (2018). Contributing Factors to Run-Off-Road Crashes Involving Large Trucks under Lighted and Dark Conditions. Journal of Transportation Engineering, Part A: Systems, 144. doi:10.1061/jtepbs.0000104.

[11] Uddin, M., & Huynh, N. (2017). Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. Accident Analysis and Prevention, 108, 44–55. doi:10.1016/j.aap.2017.08.009.

[12] Ahmed, M. M., Franke, R., Ksaibati, K., & Shinstine, D. S. (2018). Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models. Accident Analysis and Prevention, 117, 106–113. doi:10.1016/j.aap.2018.04.011.

[13] Behnood, A., & Mannering, F. (2019). Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes. Analytic Methods in Accident Research, 23, 100102. doi:10.1016/j.amar.2019.100102.

[14] Song, D., Yang, X., Yang, Y., Cui, P., & Zhu, G. (2023). Bivariate joint analysis of injury severity of drivers in truck-car crashes accommodating multilayer unobserved heterogeneity. Accident Analysis and Prevention, 190, 107175. doi:10.1016/j.aap.2023.107175.

[15] Uddin, M., & Huynh, N. (2020). Injury severity analysis of truck-involved crashes under different weather conditions. Accident Analysis and Prevention, 141, 105529. doi:10.1016/j.aap.2020.105529.

[16] Shao, X., Ma, X., Chen, F., Song, M., Pan, X., & You, K. (2020). A random parameter ordered probit analysis of injury severity in truck involved rear-end collisions. International Journal of Environmental Research and Public Health, 17(2), 395. doi:10.3390/ijerph17020395.

[17] Tahmidul Haq, M., Zlatkovic, M., & Ksaibati, K. (2021). Assessment of Commercial Truck Driver Injury Severity as a Result of Driving Actions. Transportation Research Record: Journal of the Transportation Research Board, 2675(9), 1707–1719. doi:10.1177/03611981211009880.

[18] Haq, M. T., Zlatkovic, M., & Ksaibati, K. (2021). Assessment of commercial truck driver injury severity based on truck configuration along a mountainous roadway using hierarchical Bayesian random intercept approach. Accident Analysis & Prevention, 162, 106392. doi:10.1016/j.aap.2021.106392.

[19] Wang, C., Chen, F., Zhang, Y., & Cheng, J. (2022). Spatiotemporal instability analysis of injury severities in truck-involved and non-truck-involved crashes. Analytic Methods in Accident Research, 34, 100214. doi:10.1016/j.amar.2022.100214.

[20] Santos, K., Dias, J. P., & Amado, C. (2022). A literature review of machine learning algorithms for crash injury severity prediction. Journal of Safety Research, 80, 254–269. doi:10.1016/j.jsr.2021.12.007.

[21] Chang, L. Y., & Chien, J. T. (2013). Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Safety Science, 51(1), 17–22. doi:10.1016/j.ssci.2012.06.017.

[22] Zheng, Z., Lu, P., & Lantz, B. (2018). Commercial truck crash injury severity analysis using gradient boosting data mining model. Journal of Safety Research, 65, 115–124. doi:10.1016/j.jsr.2018.03.002.

[23] Zhang, S., Khattak, A., Matara, C. M., Hussain, A., & Farooq, A. (2022). Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents. PLoS ONE, 17(2), 262941. doi:10.1371/journal.pone.0262941.

[24] Rahimi, E., Shamshiripour, A., Samimi, A., & Mohammadian, A. (Kouros). (2020). Investigating the injury severity of single-vehicle truck crashes in a developing country. Accident Analysis and Prevention, 137, 105444. doi:10.1016/j.aap.2020.105444.

[25] Islam, M., & Hernandez, S. (2013). Large truck-involved crashes: Exploratory injury severity analysis. Journal of Transportation Engineering, 139(6), 596–604. doi:10.1061/(ASCE)TE.1943-5436.0000539.

[26] Dong, C., Richards, S. H., Huang, B., & Jiang, X. (2013). Identifying the factors contributing to the severity of truck-involved crashes. International Journal of Injury Control and Safety Promotion, 22(2), 116–126. doi:10.1080/17457300.2013.844713.

[27] Naik, B., Tung, L. W., Zhao, S., & Khattak, A. J. (2016). Weather impacts on single-vehicle truck crash injury severity. Journal of Safety Research, 58, 57–65. doi:10.1016/j.jsr.2016.06.005.

[28] Osman, M., Paleti, R., Mishra, S., & Golias, M. M. (2016). Analysis of injury severity of large truck crashes in work zones. Accident Analysis & Prevention, 97, 261–273. doi:10.1016/j.aap.2016.10.020.

[29] Dong, C., Dong, Q., Huang, B., Hu, W., & Nambisan, S. S. (2017). Estimating Factors Contributing to Frequency and Severity of Large Truck–Involved Crashes. Journal of Transportation Engineering, Part A: Systems, 143. doi:10.1061/jtepbs.0000060.

[30] Al-Bdairi, N. S. S., & Hernandez, S. (2017). An empirical analysis of run-off-road injury severity crashes involving large trucks. Accident Analysis and Prevention, 102, 93–100. doi:10.1016/j.aap.2017.02.024.

[31] Wang, Y., & Prato, C. G. (2019). Determinants of injury severity for truck crashes on mountain expressways in China: A case-study with a partial proportional odds model. Safety Science, 117, 100–107. doi:10.1016/j.ssci.2019.04.011.

[32] Haq, M. T., Zlatkovic, M., & Ksaibati, K. (2020). Investigating occupant injury severity of truck-involved crashes based on vehicle types on a mountainous freeway: A hierarchical Bayesian random intercept approach. Accident Analysis & Prevention, 144, 105654. doi:10.1016/j.aap.2020.105654.

[33] Wen, H., Ma, Z., Chen, Z., & Luo, C. (2023). Analyzing the impact of curve and slope on multi-vehicle truck crash severity on mountainous freeways. Accident Analysis & Prevention, 181, 106951. doi:10.1016/j.aap.2022.106951.

[34] Mannering, F., Bhat, C. R., Shankar, V., & Abdel-Aty, M. (2020). Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Analytic Methods in Accident Research, 25, 100113. doi:10.1016/j.amar.2020.100113.

[35] Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research, 11, 1–16. doi:10.1016/j.amar.2016.04.001.

[36] Hou, Q., Huo, X., Leng, J., & Mannering, F. (2022). A note on out-of-sample prediction, marginal effects computations, and temporal testing with random parameters crash-injury severity models. Analytic Methods in Accident Research, 33, 100191. doi:10.1016/j.amar.2021.100191.

[37] Se, C., Champahom, T., Jomnonkwao, S., Karoonsoontawong, A., & Ratanavaraha, V. (2021). Temporal stability of factors influencing driver-injury severities in single-vehicle crashes: A correlated random parameters with heterogeneity in means and variances approach. Analytic Methods in Accident Research, 32, 100179. doi:10.1016/j.amar.2021.100179.

[38] Goswamy, A., Abdel-Aty, M., & Islam, Z. (2023). Factors affecting injury severity at pedestrian crossing locations with Rectangular RAPID Flashing Beacons (RRFB) using XGBoost and random parameters discrete outcome models. Accident Analysis and Prevention, 181, 106937. doi:10.1016/j.aap.2022.106937.

[39] Jamal, A., Zahid, M., Tauhidur Rahman, M., Al-Ahmadi, H. M., Almoshaogeh, M., Farooq, D., & Ahmad, M. (2021). Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study. International Journal of Injury Control and Safety Promotion, 28(4), 408–427. doi:10.1080/17457300.2021.1928233.

[40] Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press, Cambridge, United Kingdom.

[41] Sarrias, M. (2016). Discrete choice models with random parameters in R: The Rchoice package. Journal of Statistical Software, 74(10), 1–31. doi:10.18637/jss.v074.i10.

[42] Washington, S., Karlaftis, M. G., Mannering, F., & Anastasopoulos, P. (2020). Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC, Boca Raton, United States.

[43] Greene, W. H., & Hensher, D. A. (2010). Modeling ordered choices: A primer. Cambridge University Press, Cambridge, United Kingdom.

[44] Chen, T., & Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. doi:10.1145/2939672.2939785.

[45] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232. doi:10.1214/aos/1013203451.

[46] Dong, S., Khattak, A., Ullah, I., Zhou, J., & Hussain, A. (2022). Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. International Journal of Environmental Research and Public Health, 19(5), 2925. doi:10.3390/ijerph19052925.

[47] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

[48] Yu, H., Li, Z., Zhang, G., Liu, P., & Ma, T. (2021). Fusion convolutional neural network-based interpretation of unobserved heterogeneous factors in driver injury severity outcomes in single-vehicle crashes. Analytic Methods in Accident Research, 30, 100157. doi:10.1016/j.amar.2021.100157.

[49] Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2020). Factors affecting injury severity in vehicle-pedestrian crashes: A day-of-week analysis using random parameter ordered response models and Artificial Neural Networks. International Journal of Transportation Science and Technology, 9(2), 100–115. doi:10.1016/j.ijtst.2020.01.001.

[50] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437. doi:10.1016/j.ipm.2009.03.002.

[51] Ryu, S. E., Shin, D. H., & Chung, K. (2020). Prediction model of dementia risk based on XGBoost using derived variable extraction and hyper parameter optimization. IEEE Access, 8, 177708–177719. doi:10.1109/ACCESS.2020.3025553.

[52] Wu, S., Yuan, Q., Yan, Z., & Xu, Q. (2021). Analyzing Accident Injury Severity via an eXtreme Gradient Boosting (XGBoost) Model. Journal of Advanced Transportation, 2021, 1–11. doi:10.1155/2021/3771640.

[53] Champahom, T., Jomnonkwao, S., Watthanaklang, D., Karoonsoontawong, A., Chatpattananan, V., & Ratanavaraha, V. (2020). Applying hierarchical logistic models to compare urban and rural roadway modeling of severity of rear-end vehicular crashes. Accident Analysis & Prevention, 141, 105537. doi:10.1016/j.aap.2020.105537.

[54] Rifaat, S. M., Tay, R., & De Barros, A. (2012). Severity of motorcycle crashes in Calgary. Accident Analysis & Prevention, 49, 44–49. doi:10.1016/j.aap.2011.02.025.

[55] Se, C., Champahom, T., Jomnonkwao, S., Chonsalasin, D., & Ratanavaraha, V. (2024). Modeling of single-vehicle and multi-vehicle truck-involved crashes injury severities: A comparative and temporal analysis in a developing country. Accident Analysis & Prevention, 197, 107452. doi:10.1016/j.aap.2023.107452.

[56] Se, C., Champahom, T., Jomnonkwao, S., Kronprasert, N., & Ratanavaraha, V. (2022). The impact of weekday, weekend, and holiday crashes on motorcyclist injury severities: Accounting for temporal influence with unobserved effect and insights from out-of-sample prediction. Analytic Methods in Accident Research, 36, 100240. doi:10.1016/j.amar.2022.100240.

[57] Se, C., Champahom, T., Jomnonkwao, S., Chaimuang, P., & Ratanavaraha, V. (2021). Empirical comparison of the effects of urban and rural crashes on motorcyclist injury severities: A correlated random parameters ordered probit approach with heterogeneity in means. Accident Analysis & Prevention, 161, 106352. doi:10.1016/j.aap.2021.106352.

[58] Zhu, X., & Srinivasan, S. (2011). A comprehensive analysis of factors influencing the injury severity of large-truck crashes. Accident Analysis & Prevention, 43(1), 49–57. doi:10.1016/j.aap.2010.07.007.

[59] Alrejjal, A., Farid, A., & Ksaibati, K. (2021). A correlated random parameters approach to investigate large truck rollover crashes on mountainous interstates. Accident Analysis & Prevention, 159, 106233. doi:10.1016/j.aap.2021.106233.

[60] Mehrara Molan, A., Rezapour, M., & Ksaibati, K. (2020). Investigating the relationship between crash severity, traffic barrier type, and vehicle type in crashes involving traffic barrier. Journal of Traffic and Transportation Engineering (English Edition), 7(1), 125–136. doi:10.1016/j.jtte.2019.03.004.

[61] Hao, W., Kamga, C., & Wan, D. (2016). The effect of time of day on driver's injury severity at highway-rail grade crossings in the United States. Journal of Traffic and Transportation Engineering, 3(1), 37–50. doi:10.1016/j.jtte.2015.10.006.

[62] Mohamad, I., Jomnonkwao, S., & Ratanavaraha, V. (2022). Using a decision tree to compare rural versus highway motorcycle fatalities in Thailand. Case Studies on Transport Policy, 10(4), 2165–2174. doi:10.1016/j.cstp.2022.09.016.

[63] Chen, F., & Chen, S. (2011). Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. Accident Analysis & Prevention, 43(5), 1677–1688. doi:10.1016/j.aap.2011.03.026.