



Article info

Type of article:

Original research paper

DOI:

<https://doi.org/10.58845/jstt.utt.2026.en.6.2.339-353>

***Corresponding author:**

Email address:

hhz0001@auburn.edu

Received: 09/02/2026

Received in Revised Form:

08/04/2026

Accepted: 29/05/2026

Machine learning analysis of wrong-way driving crash severity factors: evidence from California Highway Patrol data

Tonghui Li¹, Md Roknuzzaman¹, Zijie Zhao², Fangjian Yang³, Huaguo Zhou^{1,*}

¹Department of Civil and Environmental Engineering, Auburn University, Auburn, AL, USA; tzi0065@auburn.edu, mzr0177@auburn.edu

²Beijing Urban Construction Design & Development Group Co. Limited, Beijing, China; zhaozijie@bjucd.com

³WSP U.S., Atlanta, GA, USA; Fang.Yang@wsp.com

Abstract: Wrong-way driving (WWD) crashes remain a major highway safety concern in the United States, yet many existing datasets do not capture key behavioral details, such as the distance traveled in the wrong direction before a crash. This study uses a comprehensive five-year California Highway Patrol dataset (2016–2020) to examine WWD crash severity using tree-based ensemble learning models, with Random Forest (RF) as a baseline model and XGBoost selected as the final model for interpretation. The dataset includes unique variables such as WWD distance, driver demographics, BAC level, safety equipment use, and crash context. Among the tested models, XGBoost achieved the best overall performance, with an accuracy of 56.35%, and showed good classification capability across fatal, injury, and property-damage-only (PDO) crashes. Variable importance results identified driver age, WWD distance, BAC, time of day, number of vehicles involved, safety equipment use, and driver sex as the most influential predictors. Partial dependence analysis revealed strong non-linear effects: younger and older drivers, higher BAC levels, longer WWD distances, nighttime conditions, and multi vehicle involvement were associated with more severe crash outcomes, while seatbelt use consistently reduced severity. In addition, the Severity Index analysis, where higher values represent greater economic loss, showed that age, WWD distance, number of vehicles, and safety equipment were also key determinants of crash-related economic burden. The findings support targeted countermeasures, including education for high-risk drivers, rapid correction strategies near WWD entry points, and stronger enforcement of seatbelt and impaired-driving laws.

Keywords: Wrong-Way driving, Machine learning model, Crash severity, California highway patrol, Contributing factor.

1. Introduction

Wrong-Way driving (WWD) has been recognized for decades as a significant safety concern on highways, freeways, and access ramps

[1], [2]. Although WWD crashes account for only about 1.3% of total roadway crashes, they are disproportionately fatal, causing an estimated 360 to 507 deaths annually in the United States from

2016 to 2020 [3]. The elevated severity of WWD crashes stems from their frequent association with high-speed, head-on collisions or sideswipes involving vehicles traveling in the opposite direction. These scenarios increase the relative velocities between colliding vehicles, resulting in higher kinetic energy transfer upon impact. The consequence is a notably elevated fatality rate. Based on the Fatality Analysis Reporting System (FARS) data from 2004 to 2022, an average of about 383 fatal WWD crashes occur each year, with a fatality rate of 1.35, approximately 22% higher than the 1.10 rate observed in other crash types [4, 5]. As such, a comprehensive understanding of the factors contributing to WWD crash severity is needed to inform effective prevention strategies and mitigate both the human and economic costs associated with these events.

Prior research has explored a wide range of determinants influencing WWD crash occurrence and their severity through various modeling approaches. Studies have implemented random parameter ordered probit models and multiple correspondence analysis to identify key risk factors, consistently finding strong associations between greater crash severity and poor lighting conditions, lack of seatbelt use, specific vehicle types (notably pick-ups or SUVs), older vehicles, and adverse weather conditions [6, 7]. Additional analyses using ordered logit, generalized ordered logit, and partial proportional odds models, drawing on data from states such as Illinois [2] and Alabama have revealed elevated WWD severity linked to driver status, rural settings, dark lighting conditions, and head-on collisions. Meanwhile, factors such as older driver age, afternoon driving, and wet runway surfaces have shown correlations with lower injury severity [8].

Binomial logistic models have also been utilized, particularly with Florida crash data, to uncover predictive factors for WWD fatal crash severity. These models have identified older and male drivers, high blood alcohol concentration

(BAC), out-of-state drivers, various health impairments (such as poor eyesight or fatigue), non-use of seatbelts, high average annual daily traffic (AADT), arterial roadways, dark or nighttime conditions, rural environments, weekends, and nighttime driving as significant contributors [9, 10]. In Louisiana, the Equivalence Class Clustering and Bottom-Up Lattice Traversal (ECLAT) algorithm further highlighted driver impairment as a pivotal factor in WWD severity [11]. Recent applications of decision tree models have also underscored factors such as front-to-front collisions, speed, day of the week, light conditions, driver age, and impairment as influential contributors to WWD crash outcomes [12]. Large truck-involved WWD crashes present additional risk, with random parameter ordered logit models showing increased severity for incidents involving higher speeds, seatbelt non-use, and drug involvement [13].

Despite these extensive efforts, much of the existing literature tends to rely on conventional statistical models and standardized variable sets, constraining the depth of analysis. The widespread use of Firth's penalized likelihood model ordered and generalized ordered logit models, random parameter ordered probit models, binomial logistic models, and the ECLAT algorithm [2, 6, 7, 11] reflects these methodological limitations. While these approaches have deepened understanding of WWD crash severity, they are often restricted by underlying assumptions and rigid input structures. By contrast, machine learning models provide enhanced flexibility and stronger predictive performance for crash severity research. Some recent studies have successfully used machine learning techniques for more accurate prediction of traffic accident severity and for identifying relevant contributing factors or travel demand selection [14, 15].

Additionally, investigations of WWD crashes that employ expanded variable sets, and diverse geographic contexts remain rare. To address this gap, the present study utilizes detailed crash data

from the CHP and compares RF and XGBoost models to examine contributing factors to WWD crash severity. With the CHP dataset, the largest repository of traffic crash cases in California [16], this study takes advantage of enriched variables, such as WWD distance before the crash, specific BAC values, and driver age. Variables like the WWD distance are critical, as they may reflect geometric or design features at entry points; previous research has indicated correlations between WWD crashes and parclo interchange terminals [4, 17]. These insights enable identification of interchange types at Wrong-Way entry locations, thereby informing interventions aimed at preventing future WWD incidents. What's more, this study addresses two related but distinct outcomes of WWD crashes. First, an injury-severity classification model was developed to examine factors associated with Fatal, Injury, and PDO crash outcomes. Second, a Severity Index analysis was conducted to evaluate the economic burden of WWD crashes. Accordingly, variables associated with higher fatality risk should not be interpreted as identical to those associated with greater crash-related economic loss.

2. Methodology

2.1. Data Collection

The CHP dataset used for this study covers the period from 2016 to 2020 and consists of 1,006 crash records with 44 variables. These variables include detailed information related to crash circumstances (e.g., city, latitude and longitude, point of entry, WWD distance before crash, crash date and time, number of fatalities, number of injuries, and number of vehicles involved), driver characteristics (e.g., drug use, blood alcohol concentration-BAC, age, sex, and license condition), and environmental factors (e.g., road condition and road surface type).

2.2. Crash Severity Classification

The crash severity levels were categorized into Fatal, Injury, and Property Damage Only (PDO) based on the number of deaths and injuries

recorded:

- Fatal: At least one death occurred.
- Injury: At least one injury occurred but no deaths.
- PDO: No injuries or deaths occurred.

2.3. Variable Selection

Based on the literature review, thirteen variables potentially related to crash severity were selected for analysis:

- WWD Distance Traveled Prior to the Crash
- Time of Day
- Day of the Week
- Entry from Off-Ramp
- Number of Vehicles
- Weather
- Lighting
- Safety Belt Used
- Drug Usage (Driver)
- BAC (Driver)
- Driver's Age
- Road Surface Condition
- Sex (Driver)

2.4. Data Preparation and Cleaning

Data mining and preprocessing were executed in R Studio (tidyverse and mice package) to ensure a reproducible pipeline. The initial raw dataset contained 1006 crash records, which were subjected to a rigorous cleaning protocol to address missingness and ensure statistical integrity.

Continuous variables included WWD distance traveled prior to the crash, driver's age, and the number of vehicles involved. While previous literature often discretizes age, this study retained it as a continuous variable. This allowed for the detection of non-linear breakpoints in crash severity that categorical grouping might obscure.

Categorical variables included Safety Belt Used, Drug, Sex, Off-Ramp Entry, Road Surface, and Lighting. These were primarily binary variables, representing whether a seatbelt was used, whether the driver was under drug influence,

the driver's sex, whether entry occurred from an off-ramp, road surface condition (dry or wet), and lighting condition (daylight or dark).

The Day of the Week variable represented all seven days but was also grouped into binary categories (weekday vs. weekend). The Time-of-Day variable was segmented into 24 one-hour intervals, representing distinct periods throughout the day. The Weather variable included multiple conditions such as clear, cloudy, foggy, rainy, and misty. The BAC variable was treated as a categorical variable with four levels based on prior research [18–20]:

- None: 0.00%
- Legal: 0.01%–0.07%
- Inebriated: 0.08% - 0.14%
- Highly Intoxicated: > 0.14%

To ensure data integrity and minimize potential selection bias, a systematic cleaning protocol was applied to the initial dataset. This process involved excluding records that lacked critical dependent variables or contained logical inconsistencies resulting in the removal of 172 cases. For records with minor data gaps, continuous variables were handled via median imputation to preserve sample size without introducing distributional bias, while missing categorical entries were empty designation to maintain the record's inclusion. Following this rigorous filtering and optimization process, the final dataset comprised 843 high-fidelity crash records for analysis.

2.5. Random Forest (RF) Analysis

The RF method is an advanced ensemble learning technique that constructs a multitude of decision trees to improve predictive performance and robustness [21]. Unlike traditional decision tree algorithms that evaluate all predictors at each split, the RF algorithm randomly selects a subset of predictors for consideration. Mathematically, if p represents the total number of predictors in the dataset, the number of predictors considered at each split (m) is typically defined as the square root

of p ($m = \sqrt{p}$). At each decision node, a distinct subset of predictors is randomly chosen from the full set of available variables. From this subset, one predictor is selected to perform the split based on impurity measures such as Gini impurity or entropy.

This randomization process reduces the correlation between individual trees, thereby decreasing the overall variance of the ensemble without increasing bias. As a result, the RF algorithm often achieves lower test error compared to a single decision tree. The model operates by constructing multiple decision trees using different bootstrap samples of the dataset. Predictions from all trees are then aggregated by averaging for regression problems or majority voting for classification tasks to produce the final output.

This ensemble approach enhances predictive accuracy and stability, as it mitigates the overfitting issues commonly associated with single decision trees. Moreover, Random Forest inherently handles missing data, provides resilience against outliers, and efficiently models complex variable interactions and nonlinear relationships. Given these strengths, the RF method was adopted in this study for crash severity classification due to its proven ability to capture intricate patterns within heterogeneous crash data [22].

2.6. Extreme Gradient Boosting (XGBoost) Algorithm

To further enhance the predictive accuracy and address the limitations of the Random Forest model, this study also implemented the XGBoost algorithm. As a scalable and highly efficient realization of gradient-boosted decision trees (GBDT), XGBoost operates on the principle of boosting, where multiple "weak" learners (decision trees) are constructed sequentially rather than in parallel [23]. Unlike Random Forest, which aims to reduce variance through independent tree averaging, XGBoost minimizes a regularized objective function that combines a loss function (measuring the difference between predicted and

actual outcomes) and a penalty term for model complexity to prevent overfitting.

Mathematically, in each iteration, the algorithm adds a new tree that best predicts the errors of the previous ensemble, effectively "learning" from the mistakes of prior trees. The objective function at iteration can be expressed as:

$$L^{(t)} = \sum_{i=1}^n l(y_i, y_i^{t-1} + f_t(x_i)) + \Omega(f_t) \quad (1)$$

Where,

L is the differentiable loss function and Ω denotes the regularization term.

To optimize this objective efficiently, XGBoost employs a second-order Taylor expansion and a sparsity-aware split-finding algorithm, which enables it to handle missing data and complex variable interactions common in crash datasets. By iteratively refining the decision boundaries, XGBoost often achieves superior precision and recall in imbalanced classification tasks, such as crash severity prediction, making it a robust benchmark for evaluating the critical factors influencing WWD fatalities [24].

2.7. Performance Metric

To determine the optimal predictive framework for crash severity classification, the performance of the RF and XGBoost models was evaluated using a comprehensive suite of metrics: accuracy, precision, recall (sensitivity), specificity, and the F1-score. While Accuracy provides a general measure of correctly classified instances, it can be misleading in the context of imbalanced crash datasets. Therefore, Recall and Specificity were utilized to assess the models' ability to correctly identify positive instances and avoid false alarms, respectively. Precision was employed to measure the reliability of positive predictions, while the F1-score, the harmonic mean of precision and recall, offered a balanced assessment of model completeness. To visualize these results, a Confusion Matrix was generated to map True Positives (TP), False Positives (FP), False

Negatives (FN), and True Negatives (TN). Because the study focused on distinguishing Fatal, Injury, and PDO crashes, model performance was primarily evaluated using class-specific precision, recall, and F1-score, together with overall accuracy and confusion matrices.

3. Results and Discussion

3.1. Descriptive Statistics

Table 1 presents an overview of 843 wrong-way driving (WWD) crashes that occurred in California between 2016 and 2020. Among these, 137 were fatal crashes, 428 resulted in injuries, and 278 were property damage only (PDO) crashes. These findings align with previous research indicating that WWD crashes tend to be more severe than other crash types and are more likely to occur during nighttime conditions [3].

Table 2 summarizes the descriptive statistics for continuous variables. The results indicate that the average distance traveled by WWD vehicles before a crash was longest in fatal crashes, followed by injury and PDO crashes.

Although the median values for these continuous variables did not differ substantially across severity levels, the mean values displayed reasonable distinctions. Specifically, fatal crashes were associated with the greatest average WWD distance and the largest number of vehicles involved, consistent with the higher impact and complexity of such crashes.

Table 3 summarizes the comparative performance of the XGBoost and RF models in predicting crash severity levels.

Following empirical tuning and a rigorous train-test split, the XGBoost model demonstrated superior predictive capability, achieving an overall accuracy of 56.35%, compared to 51.98% for the RF model. While these accuracies reflect the inherent stochasticity and complexity of WWD crashes, both models perform significantly better than a random baseline, providing a realistic assessment of crash severity risk.

A detailed examination of the category-wise metrics reveals that the XGBoost model offers a

more balanced trade-off between precision and recall across the severity classes.

XGBoost achieved the highest F1-score (0.641) in predicting Injury crashes, demonstrating strong sensitivity (0.703) for this category. For the critical Fatal class, predicting outcomes remains challenging due to the extreme class imbalance and the highly random nature of fatal collisions; however, XGBoost yielded a precision of 0.476, indicating that nearly half of the crashes it flagged as fatal were indeed fatal. While the RF model achieved a slightly higher recall for the Fatal class (0.366 compared to XGBoost's 0.244), XGBoost

provided greater overall stability and precision across the remaining categories.

RF and XGBoost were compared; XGBoost performed better and was therefore selected for subsequent interpretation analyses. In this case, further study was based on the results from XGBoost. To further demystify the internal decision mechanisms of these ensemble models, two ML interpretability techniques: Variable Importance and Partial Dependence Plots (PDPs) were employed. Additionally, the Severity Index was computed to assess the associated economic impact of these crash severity outcomes.

Table 1. Overview of cleaned dataset

Variable	Description	Category	Fatal	Injury	PDO
SEVERITY	Crash Severity	PDO (2)			278
		Injury (1)		428	
		Fatal (0)	137		
OFFRAMP	Whether enter from Off Ramp	Yes (1)	56	191	135
		No (0)	81	237	143
DOWNO	Day of Week	Weekday (0)	71	245	178
		Weekends (1)	66	183	100
SEXNO	Driver's Sex	Female (1)	47	151	83
		Male (2)	90	277	195
SAFENO	Safety Belt Used (Driver)	YES (1)	88	383	245
		NO (0)	49	45	33
RDSURFNO	Road Surface	Dry (1)	129	400	251
		Wet (0)	8	28	27
LIGHTNO	Lighting condition	Dark (1)	123	349	218
		Daylight (0)	14	79	60
DRUG	Drug used (Driver)	Yes (1)	42	52	25
		No (0)	95	376	253
WEATHER	Weather	Clear (1)	115	377	241
		Cloudy (2)	16	27	17
		Fog (3)	4	9	4
		Rain (4)	2	15	16
BACRECODE	BAC (Driver)	None (0)	53	158	136
		Legal (1)	3	10	5
		Inebriated (2)	22	78	27
		Highly intoxicated (3)	59	185	110

Table 2. Descriptive Statistics for continuous variables

Variable	Description	Fatal	Injury	PDO
DISTANCETRV	WWD Distance Traveled Before the Crash (MILE)	Mean = 1.61,	Mean = 1.05,	Mean = 0.77,
		Median = 0.42,	Median = 0.42,	Median = 0.42,
		Min = 0.04,	Min = 0.01,	Min = 0.01,
		Max = 22.20	Max = 22.00	Max = 11.90
NOVEHICLES	Number of Vehicles Involved	Mean = 2.37,	Means = 2.13,	Mean = 1.86,
		Median = 2.00,	Median = 2.00,	Median = 2.00,
		Min = 1.00,	Min = 1.00,	Min = 1.00,
		Max = 10.00	Max = 7.00	Max = 5.00
AGE	Age	Mean = 38.66,	Mean = 36.91,	Mean = 40.05,
		Median = 33.00,	Median = 32.00,	Median = 33.00,
		Min = 17.00,	Min = 17.00,	Min = 17.00,
		Max = 93.00	Max = 92.00	Max = 97.00

Table 3. Model Performance

Model	Severity Class	Precision	Recall	F1-Score	Overall Accuracy
XGBOOST	Fatal	0.476	0.244	0.323	56.35%,
	Injury	0.588	0.703	0.641	
	PDO	0.538	0.506	0.522	
RF	Fatal	0.357	0.366	0.361	51.98%
	Injury	0.607	0.555	0.58	
	PDO	0.484	0.542	0.511	

3.2. Model Training and Empirical Tuning

The cleaned dataset consisted of 843 WWD crash records and was divided into training and testing subsets, with 591 observations used for model development and 252 reserved for out-of-sample evaluation, corresponding to an approximate 70/30 split. To reduce overfitting and improve robustness, both Random Forest (RF) and XGBoost models were empirically tuned using grid search combined with five-fold cross-validation. For the RF model, the tuning process evaluated multiple values of the number of variables randomly selected at each split (*mtry* = 2, 3, 5, and 8), and the optimal value was identified as *mtry* = 2. For the XGBoost model, the grid search selected *max_depth* = 4 and *eta* = 0.1 as the best-performing hyperparameters, and the final model was trained with *nrounds* = 100.

Because the crash severity data were

imbalanced, different imbalance-handling strategies were applied for the two ensemble models. For XGBoost, class imbalance was addressed through inverse-frequency sample weighting, thereby imposing a larger penalty when minority-class fatal crashes were misclassified. For RF, balanced sampling was used during tree construction so that minority-class observations were more consistently represented in the bootstrap samples. Model performance was evaluated on the independent test set using overall accuracy, class-specific precision, recall, and F1-score. The final XGBoost model achieved an accuracy of 56.35%, whereas the final RF model achieved an accuracy of 51.98%. XGBoost showed better overall predictive performance, while RF yielded slightly higher recall for the fatal-crash class. Therefore, XGBoost was retained for subsequent interpretation analyses.

3.3. Variable Importance

Fig. 1 presents the results of a variable

importance analysis, quantified based on the "Total Gain" metric within an XGBoost model.

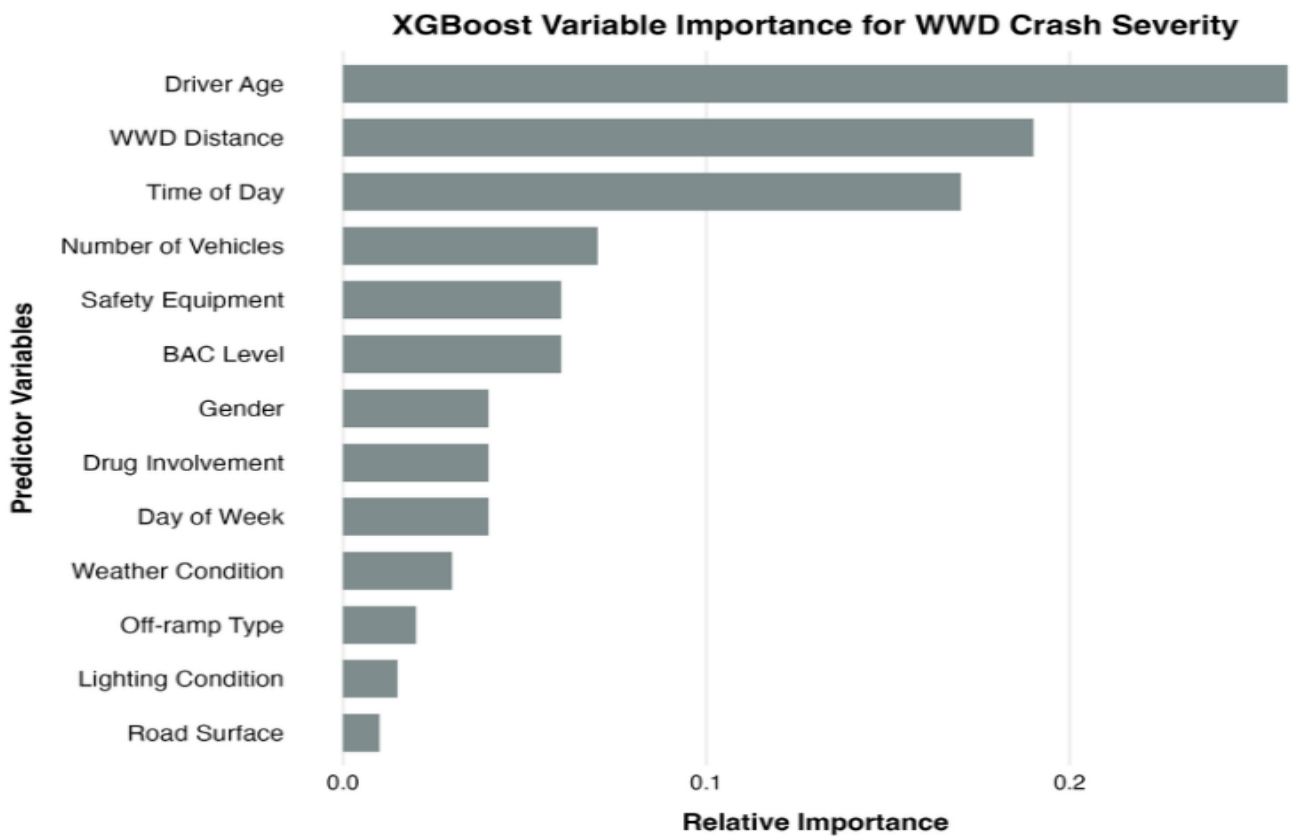


Fig. 1. Variable Importance Plot

The analysis reveals that Driver Age is the most influential predictor of accident severity, registering a score exceeding 0.25; it is closely followed by WWD Distance and Time of Day, both of which constitute major trip-level influencing factors. Although variables such as Number of Vehicles Involved, Safety Equipment Usage, BAC Level, Driver Gender, Drug Involvement, and Day of the Week also demonstrate moderate importance, environmental and infrastructure-related factors such as Weather, Lighting, and Road Surface Conditions contribute the least to the model's predictive gain. Ultimately, these findings indicate that when determining the severity of Wrong-Way driving accidents, human centric demographic and temporal characteristics play a far more critical role than external physical environmental conditions. This conclusion underscores the necessity of implementing safety interventions targeted at drivers, rather than

limiting efforts solely to infrastructure-level modifications.

3.4. Partial Dependence Plots

The PDPS were employed to visualize the marginal effect of key predictors on the probability of fatal WWD crashes while accounting for the average influence of all other variables. Following the variable importance rankings identified in Fig. 1, the nine most influential predictors were selected for detailed PDP analysis. This method allows for a granular investigation into the functional relationships between these high-impact variables and crash severity, revealing specific thresholds and non-linear trends that are critical for understanding the underlying mechanisms of WWD fatalities.

Fig. 2 shows the top nine XGBoost PDPs which indicate that WWD crash severity is shaped by several strong non-linear risk factors.

Substance impairment stands out as one of

the most important predictors: both high BAC levels and drug involvement are associated with a sharp shift toward fatal outcomes. Spatially, crashes occurring after a driver travel beyond the initial short correction window become much more severe, suggesting that failure to recover early greatly increases the chance of a catastrophic collision. Time of day also matters, with the highest fatality risk concentrated in the early morning hours, especially around 2:00–4:00 AM, when fatigue and low visibility likely intensify crash consequences.

The plots also show that crash severity

worsens substantially when multiple vehicles are involved, reflecting the greater violence of head-on or high-energy impacts, while safety equipment clearly reduces lethality and improves outcomes. In addition, driver age exhibits a multi-peak vulnerability pattern, with especially high fatality risk among very young and very old drivers, and a smaller peak around middle age. Gender and day of week contribute more modest but still meaningful differences, indicating that demographic and travel-pattern factors help refine the distinction between fatal and non-fatal WWD crashes.

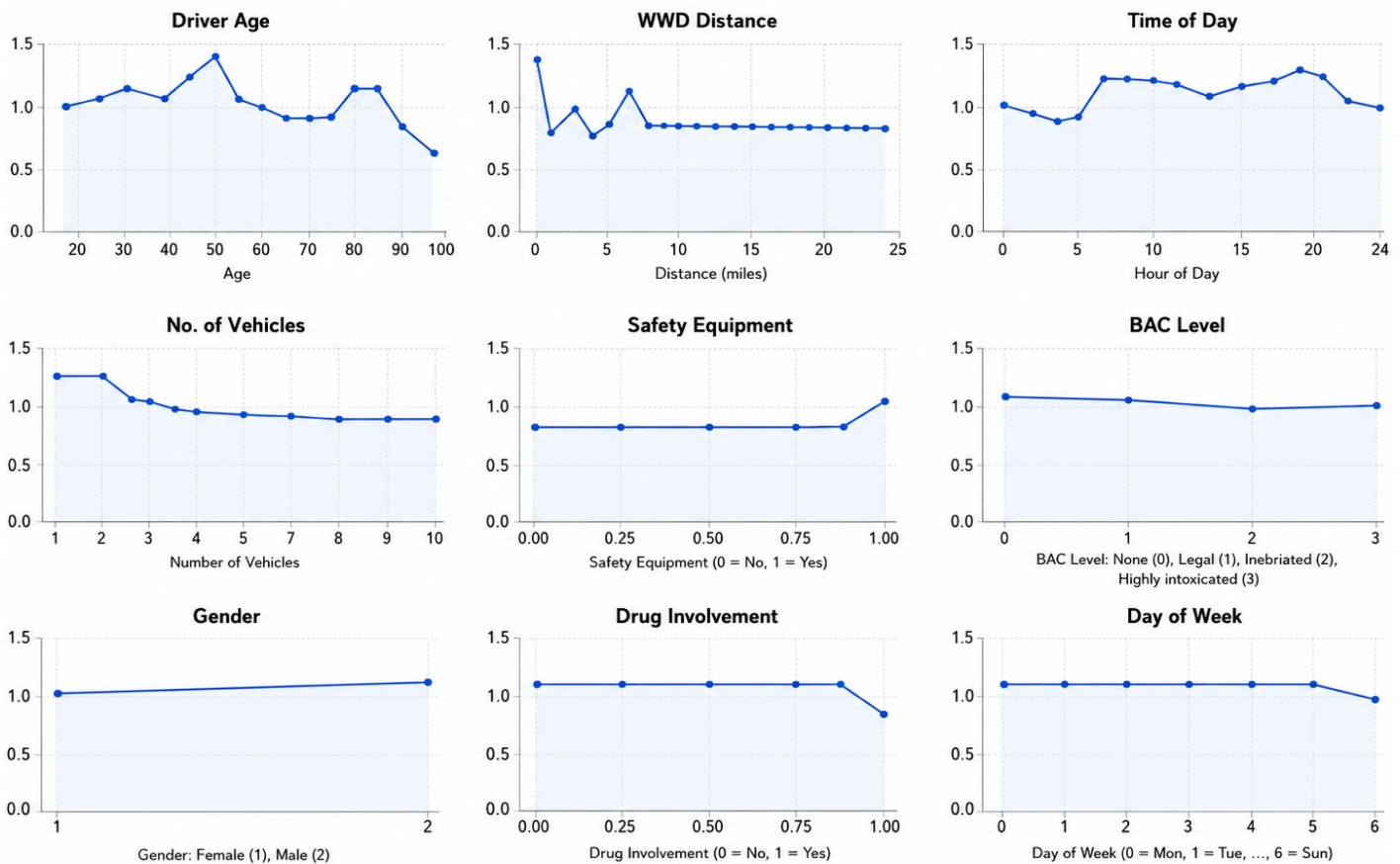


Fig. 2. Top 9 – Partial Dependence Plots for Fatal Crash

3.5. Severity Index

To assess the economic impact of crash severity, a Severity Index was computed using crash cost equivalencies from the Highway Safety Manual [25]. In this framework, fatal crashes are weighted 541.7 times more than PDO crashes, while injury crashes are weighted 11.2 times higher. The index was calculated using the

following formula:

$$\begin{aligned} \text{Severity Index} = & (541.7 \times \% \text{ of fatal crashes} \\ & + 11.2 \times \% \text{ of injury crashes} \\ & + \% \text{ of PDO crashes}) \end{aligned} \tag{2}$$

Fig. 3 presents the Partial Dependence Plot results for the top nine variables associated with the WWD crash Severity Index, where a higher y

value indicates greater economic loss. Under this definition, the figure should be interpreted as showing how each factor changes the expected economic burden of a WWD crash rather than its fatality risk alone. Overall, the plots reveal that the relationships are largely non-linear, with some variables producing clear threshold effects and others showing only modest marginal influence.

Driver age exhibits one of the most irregular and dynamic patterns in the figure. Rather than changing monotonically, the Severity Index fluctuates across age groups, suggesting that the

economic consequences of WWD crashes vary substantially by driver life stage. Several middle-aged and older age groups correspond to relatively high Severity Index values, while the curve declines noticeably at the oldest end of the distribution. This pattern indicates that age is not simply a linear proxy for crash cost; instead, different age cohorts may be associated with different combinations of exposure, driving behavior, vehicle type, and crash configuration, all of which can shape the final economic loss. [26, 27].

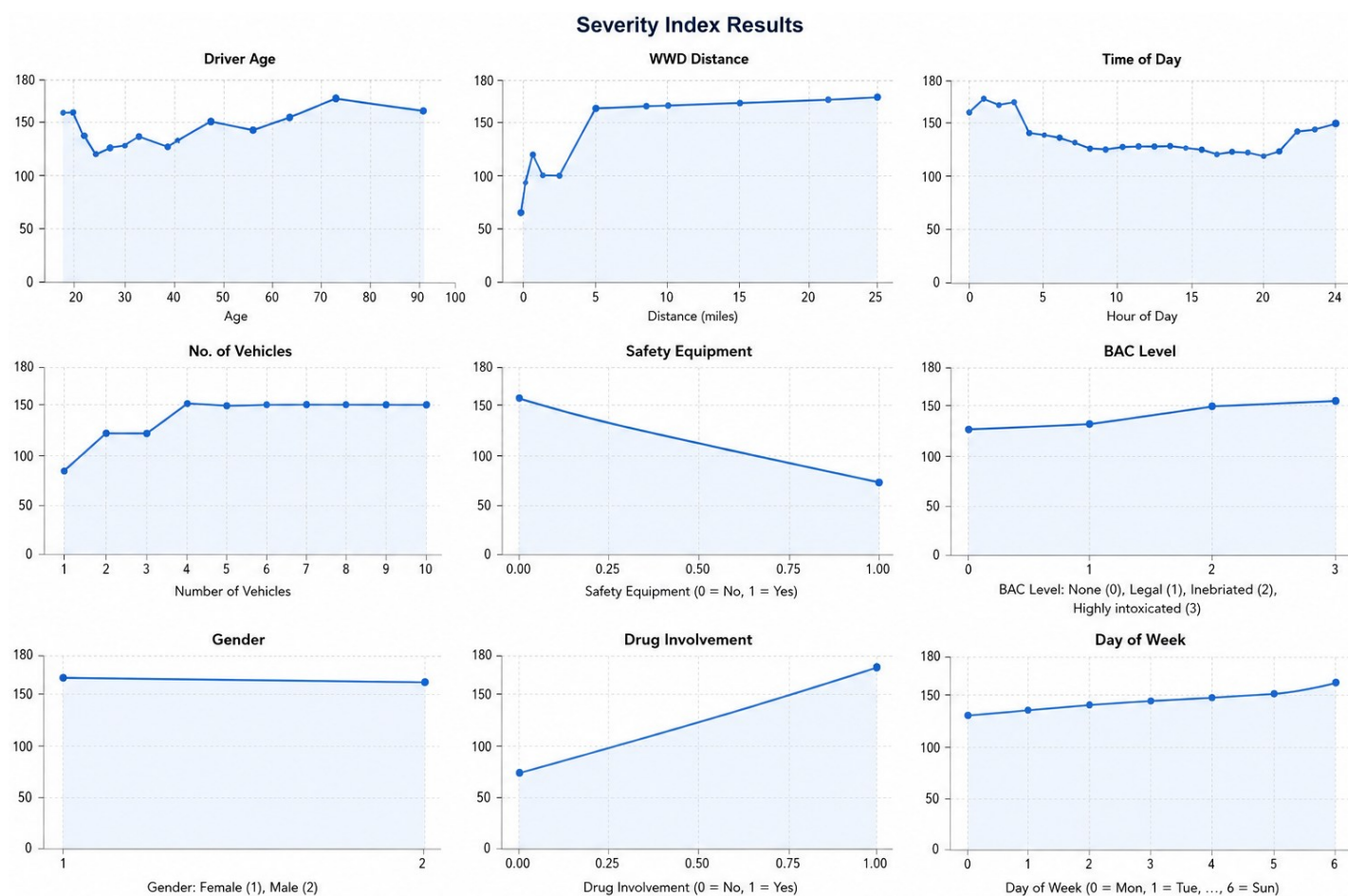


Fig. 3. Severity Index Results for the Top 9 Variables

The analysis also reveals a strong positive relationship between WWD distance and crash severity. Severity increases sharply with greater wrong-way distances, particularly beyond one mile, underscoring the importance of enabling early driver correction mechanisms to mitigate risk. Time of day further influences crash outcomes, with nighttime crashes generally showing higher

severity levels. Nonetheless, most severity index values during late-night hours remain below 100, suggesting that while nighttime conditions contribute to risk, their overall economic impact is moderate compared to other factors.

WWD distance shows a much clearer threshold-type relationship. The Severity Index is highest when the Wrong-Way travel distance is

very short, then drops sharply as distance increases, followed by a relatively stable plateau with only limited fluctuation. This result suggests that the greatest economic losses may occur in crashes that happen shortly after the onset of wrong-way entry, possibly because these events take place in high-conflict areas such as ramps, gore zones, or immediate merge locations. After that initial decline, the relatively flat curve implies that once the vehicle continues beyond the early stage of Wrong-Way movement, the additional distance travel does not dramatically increase the expected economic loss.

Time of day also has a meaningful effect on the Severity Index. The plot shows comparatively lower values during some early-hour periods, followed by a clear upward shift and a sustained higher level across much of the daytime and afternoon range, before declining somewhat again later in the cycle. This indicates that the economic burden of WWD crashes may be greater during busier daytime periods, likely because higher traffic interaction, denser traffic streams, and more multi-party involvement can increase total property damage and associated costs. In this sense, the time-of-day effect in Fig. 3 appears to reflect not only crash occurrence conditions but also the broader traffic environment in which the crash takes place.

Day of week displays a gradual downward trend. The Severity Index is relatively high at the beginning of the weekly cycle, then decreases step by step and eventually levels off. This pattern suggests that WWD crashes occurring earlier in the week may impose somewhat greater economic losses than those occurring later. Although the variation is not as dramatic as that observed for age or distance, the trend still indicates that weekly travel patterns, traffic demand composition, or trip purpose may modestly influence the financial consequences of these crashes.

Drug involvement appears to have a distinct binary effect. For most of the variable range, the

curve remains essentially flat, but it rises sharply at the final category, indicating that crashes involving drugs are associated with higher expected economic losses than crashes without drug involvement. This result suggests that drug-related WWD crashes may be more disruptive, more complex, or more damaging once they occur, even if the marginal effect is not expressed gradually across the full range of the predictor. The abrupt jump in the PDP also implies that drug involvement functions more as a categorical risk indicator than as a continuously varying influence [28-29].

BAC level shows a modest but noticeable stepwise pattern. The Severity Index is highest at lower BAC categories, then declines slightly as BAC level increases, with only a minor rebound at the highest end. Compared with several other variables, the BAC effect is relatively moderate, and the overall slope is downward rather than upward. This means that, in terms of economic loss captured by the Severity Index, higher BAC levels in this figure are not associated with greater monetary consequences. Instead, BAC may be more strongly related to other dimensions of crash seriousness, such as injury or fatality severity, than to the direct economic-loss measure used here [30-33].

The number of vehicles involved has a comparatively stable PDP, but the curve still indicates a slight upward shift at the highest category. This suggests that the economic burden tends to increase when more vehicles are involved, although the marginal effect is not as strong as for age or WWD distance. This finding is reasonable, since crashes with more vehicles typically generate greater cumulative property damage, more claims, and more complicated post-crash consequences, even when the change in the Severity Index is not dramatic across the full variable range [34].

Safety equipment status is one of the clearest categorical predictors in the figure. The Severity Index remains high across most of the range and then drops sharply for one category,

indicating a substantial difference in economic loss between crashes with different safety-equipment conditions. Interpreted practically, this result supports the conclusion that safety protection meaningfully reduces the economic consequences of WWD crashes. The sharp contrast also suggests that this variable has a strong differentiating role in the model, even though its effect is expressed through category change rather than gradual variation.

Finally, sex of driver shows only a very small marginal effect. The PDP remains nearly flat for most of the range, with only a slight decline at the final category. This indicates that driver sex contributes relatively little to the overall variation in economic loss compared with the other predictors. Although the variable may still improve model fit, its practical influence appears limited, and any sex-based difference in WWD crash economic burden is much smaller than the differences associated with age, WWD distance, safety equipment, or drug involvement.

4. Conclusion

This study investigated the factors contributing to the severity of WWD crashes in California using an XGBoost model. The model achieved an overall accuracy of 56.35% and provided a useful basis for examining patterns across crash severity categories. Model interpretation techniques, including Variable Importance analysis and PDPs, offered further insights into the relative influence and marginal effects of key predictors.

The analysis identified driver age, WWD distance prior to the crash, time of day, day of week, BAC level, drug involvement, number of vehicles involved, safety equipment status, and driver sex as the principal factors associated with the Severity Index of WWD crashes. Interpreted as an indicator of expected economic loss, the results showed that these variables influenced crash outcomes in clearly non-linear ways. Among them, driver age and WWD distance exhibited the most

pronounced patterns. Age-related effects were not monotonic, suggesting that the economic burden of WWD crashes differs across driver groups rather than increasing or decreasing uniformly. WWD distance showed a threshold-like relationship, with the highest Severity Index occurring at very short WWD distances, followed by a rapid decline and then a relatively stable pattern. This suggests that the greatest economic losses may occur shortly after Wrong-Way entry, when conflicts are likely to arise in high-risk areas such as ramps or merge zones.

The results also indicated that crash severity, in economic terms, varied by traffic context and crash composition. Time of day showed higher Severity Index values during daytime and afternoon periods than during overnight hours, implying that denser traffic conditions may amplify the total economic consequences of WWD crashes. Day of week displayed a mild downward trend, suggesting somewhat greater losses earlier in the week. Drug involvement was associated with a noticeable increase in expected economic loss, while BAC showed a more moderate and non-monotonic effect, indicating that alcohol impairment may not translate directly into larger economic losses in the same way it affects injury or fatality risk. In addition, crashes involving more vehicles tended to generate somewhat higher Severity Index values, reflecting the greater damage and complexity of multi-vehicle events.

Safety equipment remained one of the most important protective factors in the model. The partial dependence results showed a clear difference between safety-equipment categories, indicating that protective restraint use substantially reduces the economic burden of WWD crashes. By contrast, driver sex contributed only a limited marginal effect compared with the other predictors. Overall, the Severity Index analysis suggests that age, WWD distance, traffic environment, crash composition, and safety protection are central to understanding the economic consequences of

WWD crashes.

It should be noted that the Severity Index analysis represents a different outcome from the crash severity classification model. The former reflects expected economic burden, whereas the latter focuses on the likelihood of Fatal, Injury, or PDO outcomes. Therefore, similarities or differences across these two sets of results should be interpreted with respect to their different analytical objectives.

From a policy perspective, these findings support several practical interventions. First, early detection and rapid correction of WWD movements near entry points may be especially important, since the highest economic losses appear to occur during the initial segment of WWD. Second, continued enforcement and promotion of seatbelts and other safety equipment use remain essential for reducing crash costs. Third, targeted prevention strategies for driver groups showing greater economic-loss sensitivity, along with measures addressing impaired driving and multi-vehicle exposure, may further reduce the burden of WWD crashes. Future research should incorporate additional behavioral, roadway, and environmental variables, as well as broader multi-state datasets, to better capture regional differences in traffic conditions, roadway design, and driver behavior. Applying more advanced interpretability methods may also improve understanding of the mechanisms through which these factors influence WWD crash losses.

Acknowledgments

This study was collaboratively developed by the authors. Conceptualization was led by Huaguo Zhou and Tonghui Li, with methodology developed by Tonghui Li and Huaguo Zhou. Tonghui Li conducted the formal analysis, curated the data, and prepared the visualizations. Investigation was carried out by Tonghui Li together with Md Roknuzzaman, Fangjian Yang, and Zijie Zhao. The original draft was prepared by Huaguo Zhou, Tonghui Li, and Md Roknuzzaman, and the

manuscript was reviewed and edited by Tonghui Li, Md Roknuzzaman, Huaguo Zhou, Fangjian Yang, and Zijie Zhao. Huaguo Zhou provided supervision, managed project administration, and led funding acquisition. This research received no external funding. Institutional Review Board approval was not applicable, and informed consent was obtained from all subjects involved in the study. Data were obtained from CHP and are subject to access restrictions; they may be made available by the corresponding author with permission from CHP. ChatGPT was used for grammar checking and writing improvement. The authors declare no conflicts of interest.

References

- [1]. H. Zhou, Q. Chang, Y. Song, M. Jalayer, P. Hosseini, P.-S. Lin, C. Chen, P.A. Tobias. (2023). *Wrong-Way Driving Solutions Handbook*. National Cooperative Highway Research Program. *Transportation Research Board, and National Academies of Sciences, Engineering, and Medicine*. DOI: <https://doi.org/10.17226/27199>
- [2]. H. Zhou, J. Zhao, M. Pour-Rouholamin, P.A. Tobias. (2015). Statistical Characteristics of Wrong-Way Driving Crashes on Illinois Freeways. *Traffic Injury Prevention*, 16(8), 760–767. <https://doi.org/10.1080/15389588.2015.1020421>
- [3]. U.S. Department of Transportation Federal Highway Administration. *About Intersection Safety 2024*.
- [4]. Q. Chang, H. Zhou, Y. Song, M. Roknuzzaman. (2025). Impact of Traditional Traffic Control Devices on Wrong-Way Driving Incidents at Interchange off-ramps. *IATSS Research*, 49(4), 448–458. <https://doi.org/10.1016/j.iatssr.2025.09.008>
- [5]. National Highway Traffic Safety Administration (NHTSA). *Fatality Analysis Reporting System (FARS)*. <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>

- (accessed September 25, 2025)
- [6]. M. Jalayer, R. Shabanpour, M. Pour-Rouholamin, N. Golshani, H. Zhou. (2018). Wrong-Way Driving Crashes: A Random-Parameters Ordered Probit Analysis of Injury Severity. *Accident Analysis & Prevention*, 117, 128–135. <https://doi.org/10.1016/j.aap.2018.04.019>
- [7]. M. Jalayer, M. Pour-Rouholamin, H. Zhou. (2018). Wrong-Way Driving Crashes: A Multiple Correspondence Approach to Identify Contributing Factors. *Traffic Injury Prevention*, 19(1), 35–41. <https://doi.org/10.1080/15389588.2017.1347260>
- [8]. H. Zhou, M. Pour-Rouholamin, B. Zhang, J. Wang, R. Turochy. (2018). A Study of Wrong-Way Driving Crashes in Alabama Volume 1: Freeways. Publication ALDOT-Belt-SP07(906)-Wrong Way. *Auburn University*.
- [9]. R.V. Ponnaluri. (2018). Modeling Wrong-Way Crashes and Fatalities on Arterials and Freeways. *IATSS Research*, 42(1), 8–17. <https://doi.org/10.1016/j.iatssr.2017.04.001>
- [10]. C. Kadeha, H. Haule, M.S. Ali, P. Alluri, R. Ponnaluri. (2021). Modeling Wrong-Way Driving (WWD) Crash Severity on Arterials in Florida. *Accident Analysis & Prevention*, 151, 105963. <https://doi.org/10.1016/j.aap.2020.105963>
- [11]. S. Das, A. Dutta, M. Jalayer, A. Bibeka, L. Wu. (2018). Factors Influencing the Patterns of Wrong-Way Driving Crashes on Freeway Exit Ramps and Median Crossovers: Exploration Using ‘Eclat’ Association Rules to Promote Safety. *International Journal of Transportation Science and Technology*, 7(2), 114–123. <https://doi.org/10.1016/j.ijst.2018.02.001>
- [12]. S.R. Nafis, P. Alluri, W. Wu, B.M.G. Kibria. (2022). Wrong-Way Driving Crash Injury Analysis on Arterial Road Networks Using Non-Parametric Data Mining Techniques. *Journal of Transportation Safety & Security*, 14(10), 1702–1730. <https://doi.org/10.1080/19439962.2021.1960660>
- [13]. S. Anam, G. Azimi, A. Rahimi, X. Jin. (2022). Severity Analysis of Large-Truck Wrong-Way Driving Crashes in the State of Florida. *Vehicles*, 4(3), 766–779. <https://doi.org/10.3390/vehicles4030043>
- [14]. X. Zhang, X. Zhao. (2022). Machine Learning Approach for Spatial Modeling of Ridesourcing Demand. *Journal of Transport Geography*, 100, 103310. <https://doi.org/10.1016/j.jtrangeo.2022.103310>
- [15]. M. Yan, Y. Shen. (2022). Traffic Accident Severity Prediction Based on Random Forest. *Sustainability*, 14(3), 1729. <https://doi.org/10.3390/su14031729>
- [16]. California State Transportation Agency. (2023). 2023 Crash Data on California State Highways (Road Miles, Travel, Crashes, Crash Rates). Department of Transportation. Division of Research, Innovation, and System Information, Sacramento, California.
- [17]. Q. Chang. (2022). Effects of Geometric Design Features and Traditional Traffic Control Devices on Wrong-Way Driving Incident at Partial Cloverleaf Interchange Terminal: A Machine-Learning Approach. *Doctoral Dissertation. Auburn University*.
- [18]. R. Hingson, M. Winter. (2003). Epidemiology and Consequences of Drinking and Driving. *Alcohol Research & Health*, 27(1), 63–78.
- [19]. A.C. Wagenaar, M.M. Maldonado-Molina, L. Ma, A.L. Tobler, K.A. Komro. (2007). Effects of Legal BAC Limits on Fatal Crash Involvement: Analyses of 28 States from 1976 through 2002. *Journal of Safety Research*, 38(5), 493–499. <https://doi.org/10.1016/j.jsr.2007.06.001>
- [20]. A.C. Wagenaar, M.M. Maldonado-Molina. (2007). Effects of Drivers’ License Suspension Policies on Alcohol-Related Crash

- Involvement: Long-Term Follow-Up in Forty-Six States. *Alcoholism: Clinical and Experimental Research*, 31(8), 1399–1406. <https://doi.org/10.1111/j.1530-0277.2007.00441.x>
- [21]. L. Breiman. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- [22]. X.Wang, S.H. Kim. (2019). Prediction and Factor Identification for Crash Severity: Comparison of Discrete Choice and Tree-Based Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(9), 640–653. <https://doi.org/10.1177/0361198119844456>
- [23]. S. Fouladvand, M. Noshad, M.K. Goldstein, V.J. Periyakoil, J.H. Chen. (2023). Mild cognitive impairment: data-driven prediction, risk factors, and workup. *AMIA Joint Summits on Translational Science Proceedings*, 2023, 167-175.
- [24]. A.B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, A.K. Mohammadian. (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, 136, 105405.
- [25]. AASHTO. (2011). Highway Safety Manual. American Association of State Highway and Transportation Officials.
- [26]. A. Adebisi, J. Ma, J. Masaki, J. Sobanjo. (2019). Age-Related Differences in Motor-Vehicle Crash Severity in California. *Safety*, 5(3), 48. <https://doi.org/10.3390/safety5030048>
- [27]. G.A. Ryan, M. Legge, D. Rosman. (1998). Age Related Changes in Drivers' Crash Risk and Crash Type. *Accident Analysis & Prevention*, 30(3), 379–387. [https://doi.org/10.1016/S0001-4575\(97\)00098-5](https://doi.org/10.1016/S0001-4575(97)00098-5)
- [28]. C. Se, T. Champahom, S. Jomnonkwao, V. Ratanavaraha. (2024). Examining Factors Affecting Driver Injury Severity in Speeding-Related Crashes: A Comparative Study across Driver Age Groups. *International Journal of Injury Control and Safety Promotion*, 31(2), 234–255. <https://doi.org/10.1080/17457300.2023.2300458>
- [29]. P.L. Zador, S.A. Krawchuk, R.B. Voas, and Westat. (2000). Relative Risk of Fatal Crash Involvement by BAC, Age, and Gender. *National Highway Traffic Safety Administration. US Department of Transportation*. <https://doi.org/10.21949/1525465>
- [30]. T.L. Traynor. (2005). The Impact of Driver Alcohol Use on Crash Severity: A Crash Specific Analysis. *Transportation Research Part E: Logistics and Transportation Review*, 41(5), 421–437. <https://doi.org/10.1016/j.tre.2005.03.005>
- [31]. N.H. Mounce, O.J. Pendleton. (1992). The Relationship between Blood Alcohol Concentration and Crash Responsibility for Fatally Injured Drivers. *Accident Analysis & Prevention*, 24(2), 201–210. [https://doi.org/10.1016/0001-4575\(92\)90038-K](https://doi.org/10.1016/0001-4575(92)90038-K)
- [32]. K. Shyhalla. (2014). Alcohol Involvement and Other Risky Driver Behaviors: Effects on Crash Initiation and Crash Severity. *Traffic Injury Prevention*, 15(4), 325–334. <https://doi.org/10.1080/15389588.2013.822491>
- [33]. Q. Wu, F. Chen, G. Zhang, X.C. Liu, H. Wang, S.M. Bogus. (2014). Mixed Logit Model-Based Driver Injury Severity Investigations in Single- and Multi-Vehicle Crashes on Rural Two-Lane Highways. *Accident Analysis & Prevention*, 72, 105–115. <https://doi.org/10.1016/j.aap.2014.06.014>
- [34]. K.A. Abay, R. Paleti, C.R. Bhat. (2013). The Joint Analysis of Injury Severity of Drivers in Two-Vehicle Crashes Accommodating Seat Belt Use Endogeneity. *Transportation Research Part B: Methodological*, 50, 74–89. <https://doi.org/10.1016/j.trb.2013.01.007>