In-Depth Understanding of Pedestrian- Vehicle Near-Crash Events at Signalized Intersections: An Interpretable Machine Learning Approach

Article *in* Transportation Research Record Journal of the Transportation Research Board · November 2022

DOI: 10.1177/03611981221136138

Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–13

DOI: 10.1177/03611981221136138
journals.sagepub.com/home/trr
SAGE

In-Depth Understanding of Pedestrian–Vehicle Near-Crash Events at Signalized Intersections: An Interpretable Machine Learning Approach

Xiaoqiang Kong[1], Subasish Das[2], Yunlong Zhang[1], Zihang Wei[1], and Chi-hao Yuan[1]

Abstract

This study used a pedestrian-involved near-crash database and adopted an interpretable machine learning framework using SHapley Additive exPlanations (SHAP) to understand the factors associated with critical pedestrian-involved near-crash events. The results indicate that pedestrians with a relatively higher walking speed are more likely to be involved in critical near-crash events. Furthermore, critical pedestrian-involved near-crash events are highly associated with vehicles with driving speeds of less than 10 mph. A higher pedestrian volume is highly associated with critical near-crash events with left-turn vehicles. It is possible that a higher pedestrian volume increases the occurrence of jaywalking behavior or encourages more pedestrians to step into the crosswalk when they should not. By contrast, a higher pedestrian volume is highly associated with non-critical near-crash events with right-turn vehicles. Right-turn vehicles often expect that there will be pedestrians crossing, and a higher volume of pedestrian traffic increases a driver's awareness and caution while turning. The study also found that a longer signal cycle is highly associated with critical near-crash events when the pedestrian volume is low, while a relatively short signal cycle length is highly associated with critical near-crash events when the pedestrian volume is high. During non-peak hours, pedestrians have less tolerance for a relatively longer signal cycle. Moreover, a relatively shorter signal cycle length at peak hours will limit the number of pedestrians that can cross during a cycle and encourage the possibility of pedestrians jaywalking or stepping onto the crosswalk when they should not.

Keywords

operations, traffic control devices, pedestrians, bicycles, human factors, safety, motorcycles and mopeds, pedestrian and bicyclist safety

Pedestrian safety has been an important research topic for many years. Pedestrians are the most vulnerable road users, and the consequences of any type of crash involving pedestrians could be severe. Despite many years during which transportation researchers, engineers, and agencies have promoted pedestrian safety, the number of pedestrians injured or killed in crashes is still concerning. In the United States, there were 6,205 pedestrians killed and more than 76,000 injured in crashes in 2019 (*1*). Based on the data provided by the U.S. Federal Highway Administration, from 2015 to 2018 about one-third of pedestrian fatalities occurred at intersections, and nearly half of them occurred at signalized intersections (*2*). Signalized intersections are often built to facilitate the mobility and safety of vehicles and pedestrians. With the pedestrian signal, the signalized intersection is intended to provide a safe walking environment. However, the number of fatal pedestrian crashes at signalized intersections is not trivial. Although many conflicting trajectories are separated by signal phasing, potential conflicts between pedestrians and vehicles are still unavoidable given the

[1]Zachry Department of Civil & Environmental Engineering, Texas A&M University, College Station, TX
[2]Texas State University, San Marcos, TX

Corresponding Author:
Subasish Das, subasish@txstate.edu

complexity of pedestrian and vehicle movements at intersections, and pedestrian safety at signalized intersections still requires attention and more in-depth study.

Many previous studies have focused on finding factors that correlate with pedestrian-involved crashes at signalized intersections (*3–7*). The majority of these factors can be classified into the following categories: intersection characteristics, built environment (*5*, *8–12*), pedestrian-related factors such as pedestrian volume or speed (*13–17*), and other factors such as traffic volume (*15*). For intersection characteristics, factors such as the number of right-turn-only lanes, the number of nonresidential driveways within 50ft of the intersection, and road geometrics are associated with pedestrian-related crashes (*5*, *8*, *11*). Lee and Abdel-Aty (*8*) used four years of intersection pedestrian–vehicle crash data in Florida to study the factors that influence intersection pedestrian-related crash frequency and severity. The results show that driver demographics, road geometrics, traffic, and environmental factors are the main contributors to pedestrian-related crash frequency and severity. Schneider et al. (*11*) analyzed intersection pedestrian-related crash risk levels by comparing intersection pedestrian crash rates with intersection characteristics. Miranda-Moreno et al. (*5*) investigated the relationship between the built environment, pedestrian activities, and pedestrian-vehicle crash occurrence. The results show that the built environment has a strong association with pedestrians while it is not directly related to pedestrian-vehicle crash occurrences. Driver-related factors such as driver demographics have been examined in multiple studies. Iryo-Asano and Alhajyaseen (*18*) indicated that unexpected speed changes of pedestrians might lead to dangerous pedestrian–vehicle conflicts. This study analyzed the sudden speed changes made by pedestrians in a quantitative way and developed a model to represent pedestrian travel time by considering this factor.

However, only a few studies have focused on the nonlinear relationship between these factors and pedestrian-involved crashes. A recent study published by Ding et al. (*19*) stated that many factors present clear nonlinear effects on pedestrian–vehicle collisions, which challenges the linearity assumptions commonly used by conventional statistical models. With the prevalence of machine learning in recent years, machine learning algorithms have been applied to many studies to overcome the nonlinearity constraints and achieve better model performance. The common concern about machine learning algorithms is interpretability. Many methods, such as neural networks, could achieve decent accuracy by tuning many hyperparameters. As a result, the model often has remarkable predictive power but poor interpretability.

Within the above-mentioned published studies, the common data sources for pedestrian-involved crashes are police-reported data, typically collected after the crash events (*9*, *15*, *20*). This type of crash data are commonly managed and maintained by transportation agencies and are widely used in transportation safety studies. In addition, some studies used simulated data (*21–23*). The majority of published studies have used pedestrian–vehicle crash data, such as FARS, to analyze the intersection pedestrian safety factors. However, most pedestrian–vehicle conflicts are without actual physical interaction, and these conflicts are unlikely to be recorded in the database. Even with a considerable amount occurring each year, vehicle–vehicle collisions are still considered sparse events (*24*, *25*). Therefore, the surrogate data set, such as the near-crash data, has gradually become a popular supplement source for safety analysis. However, these near-crash data are often only considered vehicle–vehicle events (*26–29*). Compared with vehicle–vehicle collisions, the occurrence of pedestrian–vehicle collision events is sparser. However, pedestrian–vehicle near-crash data are not widely available to the public. In recent years, a few studies have used the pedestrian-involved near-crash data collected through naturalistic driving studies (*30–32*). To effectively reduce the pedestrian crash risk, there is a need to investigate the safety factors of surrogate crash data, such as near-crash data, more comprehensively because they are also likely to result in severe pedestrian injuries.

This study used a pedestrian near-crash database developed from video footage from Bellevue, Washington. The post-encroachment time (PET) was adopted as a criterion for capturing the near-crash event through image recognition techniques. PET is defined as the time between when the first road user leaves the conflict point and the second road user arrives at the conflict point. All events less than 10 s were collected as a potential conflict between the pedestrian and the vehicle. The pedestrian–vehicle conflict events were categorized into critical conflict (0–2 s), minor conflict (2–3 s), potential conflict (3–5 s), and interactions (larger than 5 s) based on the PET intervals. In addition, other information such as vehicle speed, pedestrian speed, the traffic volume of each movement direction (through, right-turn, and left-turn), pedestrian volume on each crosswalk of the intersection, traffic signal information, speed limit, and historical crash information are all available for the analysis. The volume of vehicles and pedestrians is collected at a 15-min interval.

This analysis is conducted in two steps. First, the study adopted the XGBoost algorithm to train a predictive classification model to classify the critical and noncritical near-crash events. The critical event considers any near-crash events with less than 3 s PET, and the noncritical event considers any near-crash events with more than 5 s PET. Then, an interpretable machine learning framework was established on this trained predictive model to interpret the nonlinear relationships. The main contributions of this study are listed below:

- use pedestrian–vehicle near-crash data for the study;
- establish an interpretable machine learning framework to understand the impacts of selected factors on the pedestrian safety;
- investigate the nonlinear relationships between the near-crash events and other factors; and
- explore the effect of the signal cycle and its interaction effects with the pedestrian volume on pedestrian safety

Methodology

Data Description

The data set analyzed in this study is from the city of Bellevue, Washington. The data set was collected from the city's existing 360 HD traffic camera network. Unlike traditional crash data sets in which crashes can only be recorded after they happen, this novel data set contains near-miss conflict records identified based on video footage.

Currently,data from six intersections have been made available online by the Institute of Transportation Engineers (ITE) as part of the ITE Vision Zero Design Sandbox Competition (*33*). The raw data contain three parts: (1) intersection geometric and signal information; (2) intersection traffic flow information based on all movements; and (3) safety-related event information which includes all near-missed events identified from the video footage.

These six intersections from the city of Bellevue, Washington, are:

- 124th Avenue NE and NE 8th Street;
- 116th Avenue NE and Northup Way;
- 148th Avenue SE and SE 22nd Street;
- 112th Avenue NE and NE 8th Street;
- 100th Avenue and Main Street;
- Bellevue Way NE and NE 8th Street.

To prepare the data set for analysis, three steps were performed on the raw data set. First, the data used in this study only need to include vehicle–pedestrian-related safety events. Because each safety-related event includes two road users, records were kept only if at least one of the road users was a pedestrian. As for the second step, each safety event was classified as: a critical conflict with PET between 0 and 2 s; a minor conflict with PET between 2 and 3 s; a potential conflict with PET between 3 and 5 s; and an interaction with PET greater than 5 s (*34*). All critical conflicts and minor conflicts are considered critical events, and all interactions are considered non-critical events. The third step was to assign traffic volume, intersection geometrics, and traffic signal information to each safety event record. Based on the vehicle and pedestrian movements in each safety event, traffic-related variables such as volume, speed, and signal information were selected corresponding to the road users' movements in each record. For example, if a vehicle's movement was southbound through, only the traffic-related variables corresponding to southbound through movement were included.

Tables 1 and 2 present the descriptive statistics for the continuous and categorical variables of the data set used in the XGBoost model. It provides general information about the data in two PET groups. For some features, the values from the two groups are different, such as vehicle volume, while some features are similar, such as pedestrian volume. For critical events, the vehicle volume in the involved-vehicle direction is smaller than the vehicle volume of non-critical events. Another interesting observation is that the intersections with more total crash events or pedestrian crash events are more likely to have critical events. For categorical features, the majority of critical events occurred to the right-turning vehicles. For the right-turning or through vehicles, the left signal feature is labeled as Not Applicable.

Critical and Non-Critical Near-Crash Events Classification Using XGBoost

Building a classification model with decent accuracy is critical for applying the SHapley Additive exPlanations (SHAP) interpretable machine learning framework. The problem is formulated as a binary classification task: critical and non-critical near-crash events. The assumption for this classification is that the selected features contribute to the near-crash events. The basic concept of the XGBoost algorithm is the collective power of a group of weak learners:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F} \quad (1)$$

$\hat{y}_i$ represents the classification results critical or non-critical (the mathematical representations are 0 and 1 during the training process). K is the total number of trees, while f_k is a function for the k th tree. $\mathcal{F}$ is a functional space for trees. Each tree f_k is a weak learner.

The objective function minimizes the total loss:

$$\min \mathcal{L}^{(t)} = min\left(\sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{T} \Omega(f_i)\right) \quad (2)$$

Table 1. Descriptive Statistics of Continuous Variables

		PET							
		Critical				Non-Critical			
Variable names	Definition	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
veh_vol	Traffic volume in the involved-vehicle direction	41.2	30.2	1	431	47.8	44.1	1	481
ped_vol	Pedestrian volume in the involved-pedestrian direction	25.8	17.1	1	76	24.2	16.7	1	84
total_vol	Total traffic volume of the intersection	655.6	222.7	26	1,102	644.0	238.3	35	1,130
total_crash	Total crash count from 2015 to 2019c	47.1	16.1	3	73	45.3	16.5	3	73
total_ped_crash	Total pedestrian-involved crash count from 2015 to 2019	1.8	0.6	0	2	1.7	0.7	0	2
ped_signal	Pedestrian signal length (s)	6.5	0.9	4	7	6.3	1.0	4	7
ped_signal_fla	Pedestrian “Don’t walk” flashing signal length (s)	16.1	2.8	7	20	15.8	2.9	7	20
min_cl	Minimum signal cycle (s)	96.5	7.2	75	100	95.7	7.6	75	100
max_cl	Maximum signal cycle (s)	144.2	13.6	100	150	143.1	14.4	100	150
veh_speed	Involved-vehicle speed (mph)	10.9	4.6	1	48	13.4	6.6	1	73
ped_speed	Involved-pedestrian speed (mph)	4.7	2.5	1	22	3.9	1.9	1	30

Note: PET = post-encroachment time; SD = standard deviation; Min. = minimum; Max. = maximum.

Table 2. Count of Events for the Categorical Variables

Variable names	Definition	Values	Critical	Non-critical
veh_mov	Movement of the involved vehicle	Left	98	4,280
		Right	715	8,366
		Through	98	3,008
left_turn_sig	Left-turn signal type of the direction of the involved vehicle	Not applicable	813	11,374
		Permissive	11	380
		Pro-Per	11	228
		Protected	67	3,559
		Split	9	113
wkd	Weekday/weekends	Weekdays	751	12,667
		Weekends	160	2,987
peakhr	Peak hour/off-peak hour	Off-peak	670	11,428
		Peak	241	4,226
speeding	Is the involved vehicle speeding?	No	909	15,184
		Yes	2	470
num_lane	Number of lanes in the direction of the involved vehicle. For right-turn vehicle, if there is no exclusive right-turn lane, the number of lanes is zero	0	350	5,032
		1	447	6,518
		2	111	3,864
		3	3	240
speed_limit	Speed limit	25	38	1,136
		30	829	13,605
		35	44	913

where $l()$ is the chosen loss function and $\Omega(f_i)$ is the regularization term to control the complexity of the tree model and reduce the chance of overfitting.

The regularization function is written as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \qquad (3)$$

where γ is the parameter of the penalty for the complexity and λ is the degree of the regularization of function f. T stands for the number of leaves of the tree model and w_j means the weight of the jth leaf of the tree model. The following steps are: (1) to use Taylor expansion to rewrite the expression of the objective function; and (2) to obtain the optimal weight of the jth leaf. As the iteration of this process goes on, the summation of the weak learners leads to the steepest gradient and reaches the optimal results (*35*).

Interpreting Near-Crash Classification Using SHAP

As machine learning algorithms are becoming more and more popular, people are often amazed by their remarkable predictive accuracy, which is empowered by numerous hyperparameters inside the black box. High accuracy and poor interpretability often become a trade-off for researchers. For many non-theoretical studies, interpretability is the core of the study and grants power to researchers to understand the data. This study takes advantage of an interpretable machine learning algorithm (SHAP) to use the power of a tree boosting algorithm (XGBoost) to understand factors affecting pedestrian safety at signalized intersections. SHAP calculates the contribution of each feature from a coalitional game theory perspective (*36*). The main function can be written as:

$$\varnothing_i(f,x) = \sum\nolimits_{S \subseteq S_{all/\{i\}}} \frac{|S|!(M-|S|-1)!}{M!}[f_x(S \cup \{i\}) - f_x(S)] \tag{4}$$

where

M is the set that contains all features,

i is the feature I,

S is one of the subsets of M, and

$f_x(S)$ refers to the trained model with the feature.

SHAP interaction value is another advantage of this interpretable machine learning framework. It stands for the effect of a pair feature, which is similar to the interaction effects in the regression models. The interaction value is calculated based on the following function:

$$\phi_{i,j} = \sum\nolimits_{S \subseteq \setminus \{i,j\}} \frac{|S|!(M-|S|-2)!}{2(M-1)!} \delta_{ij}(S) \tag{5}$$

where $i \neq j$ and

$$\delta_{ij}(S) = f_x(S \cup \{i,j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S)$$

For a feature, the summation of all its SHAP interaction values with other features equals its SHAP value like the following equation:

$$\sum \phi_{i,j} = \phi_i \tag{6}$$

Model Performance on Imbalanced Data

Training a model with acceptable performance is a critical step for this interpretable framework. A well-trained model indicates the model and trained hyperparameters could capture these hidden associations among features and the target. Thus, the interpretable machine learning framework could uncover the hidden relationships learned by using these hyperparameters. Table 3 shows the general performance of the XGBoost model and the performance scores of five other commonly adopted classification models: LightGBM, AdaBoost, Random Forest, GBDT, and logistic regression, which are also reported as comparison methods. Except for the performance score of the models, this research requires extra attention to these scores since the data set is heavily imbalanced (critical/positive events: 911; non-critical/negative events: 15,654). Heavily imbalanced data could provide misleading results. For example, with imbalanced data consisting of 10,000 negative cases and 100 positive cases, if the model predicts all cases negative, the accuracy rate is still 99%. However, this 99% accuracy rate is not very meaningful, especially when the positive cases are the focus of the question.

Table 3. Model Performance on Testing Data (25% of the Whole Data Set)

	Accuracy	Precision (weighted)	F1 (weighted)	AUC
XGBoost	0.80	0.93	0.85	0.70
LightGBM	0.94	0.93	0.92	0.50
AdaBoost	0.93	0.91	0.92	0.55
Random forest	0.95	0.93	0.92	0.51
GBDT	0.95	0.91	0.92	0.50
Logistic regression	0.94	0.90	0.92	0.51

Note: AUC = area under the curve.

Table 3 shows that all models have a decent performance from the accuracy, precision, and F1 values. However, with a heavily imbalanced data set, it is important to check the weighted confusion matrix (Table 4) and confirm the real performance of the two models.

Performance measures like true positive (TP) and false positive (FP) are instances of correct and incorrect classifications per original class, respectively. True negative (TN) and false negative (FN) are cases of correct and incorrect rejections per original class, respectively. Some of the common performance measures are:

- $Precision = \frac{TP}{TP+FP}$ = class agreement of the data labels with the positive labels
- $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ = overall accuracy
- $F1\ or\ F-score = \frac{2 \times Precision \times Recall}{Precision + Recall}$ = weighted average of the recall and precision
- $Area\ Under\ the\ Curve\ (AUC)$ = The Receiver Operating Characteristic curve where TP is plotted against FP.

Table 4. Confusion Matrix of the Testing Data (25% of the Whole Data Set)

	XGBoost		LightGBM	
	Predicted 0	Predicted 1	Predicted 0	Predicted 1
Actual 0	3,227	693	3,916	4
Actual 1	98	124	220	2
	AdaBoost		Random forest	
	Predicted 0	Predicted 1	Predicted 0	Predicted 1
Actual 0	3,845	75	3,917	3
Actual 1	197	25	218	4
	GBDT		Logistic regression	
	Predicted 0	Predicted 1	Predicted 0	Predicted 1
Actual 0	3,227	693	3,917	3
Actual 1	98	124	222	0

Not surprisingly, except for the XGBoost model, other models barely captured the positive cases, and the weights of these coefficients heavily lean toward classifying all cases as negative/non-critical events. Therefore, the XGBoost model was selected to perform the classification task and prepare the trained model for the SHAP framework.

Results and Discussions

Explore the Importance of Individual Features

Global Feature Importance. Figure 1 shows the global feature importance of all features selected for training the predictive model. The length of the bar indicates the importance of a feature on the classification tasks. A longer bar means that the feature has a higher overall contribution to the model to classify the critical and non-critical near-crash events correctly. The vehicle movement feature has the highest contribution on this classification task, followed by pedestrian speed, vehicle speed, left-turn signal type, the total volume of the intersection, total pedestrian volume, signal cycle length, and so forth. Most of the associations between the factors and pedestrian-involved near-crashes have been mentioned in previous studies (*8*, *16*, *37–39*). This finding approves the validity of using near-crash data as surrogate data for studying pedestrian safety. For the most dominant feature (vehicle movement), the reason behind this dominance is explained in the descriptive statistics. Out of 911 critical near-crash events, 719 of them involved vehicles turning right. This echoes a study from Yue et al. (*20*), which stated that drivers often pay attention to the upcoming vehicles on the left side while making the right-turn, especially when turning right on red, and that

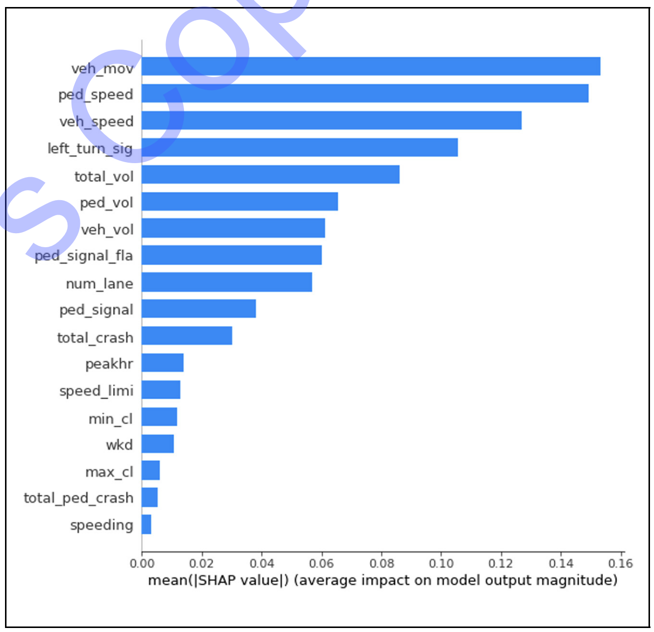

Figure 1. Global feature importance.
Note: veh_mov = movement of the involved vehicle; ped_speed = involved-pedestrian speed (mph); veh_speed = involved-vehicle speed (mph); left_turn_sig = left-turn signal type of the direction of the involved vehicle; total_vol = total traffic volume of the intersection; ped_vol = pedestrian volume in the involved-pedestrian direction; veh_vol = traffic volume in the involved-vehicle direction; ped_signal_fla = pedestrian "Don't walk" flashing signal length (s); num_lane = number of lanes in the direction of the involved vehicle; ped_signal = pedestrian signal length (s); total_crash = Total crash count from 2015 to 2019; peakhr = peak hour/off-peak hour; speed_limi = speed limit; min_cl = minimum signal cycle (s); total_ped_crash = total pedestrian-involved crash count from 2015 to 2019.

is where the most conflicts occur. It is worth mentioning that the minimum and maximum traffic signal cycle of an intersection pose influences the occurrence of critical

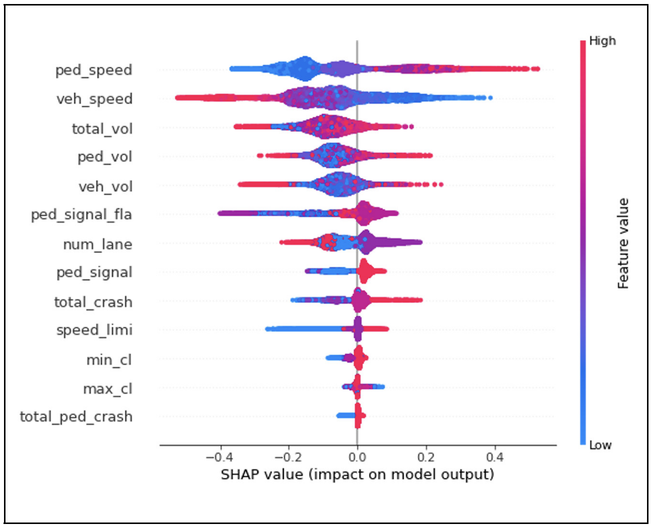

Figure 2. Local explanation summary for continuous variables.
Note: ped_speed = involved-pedestrian speed (mph); veh_speed = involved-vehicle speed (mph); total_vol = total traffic volume of the intersection; ped_vol = pedestrian volume in the involved-pedestrian direction; veh_vol = traffic volume in the involved-vehicle direction; ped_signal_fla = pedestrian "Don't walk" flashing signal length (s); num_lane = number of lanes in the direction of the involved vehicle; ped_signal = pedestrian signal length (s); total_crash = total crash count from 2015 to 2019; speed_limi = speed limit; min_cl = minimum signal cycle (s); total_ped_crash = total pedestrian-involved crash count from 2015 to 2019.

pedestrian–vehicle safety events. The length of the signal cycle is related to the length of traffic time phase duration and pedestrian walk time. The improper minimum cycle or maximum cycle could increase the chance of violations by the pedestrians caused by impatience.

Local Explanation Summary. While Figure 1 provides the average importance of features on the model output, Figure 2 offers a more detailed feature contribution to the classification task for continuous features. The colored bars in Figure 2 consist of points, which are observations of this data set. The feature value of each observation is colored from blue to red gradually. Lower values are bluer and higher values are redder. The X-axis position of each point represents its SHAP value, which is considered as the localized marginal effect of a feature on the response feature (critical or non-critical safety events in this case) in a specific observation. A more positive SHAP value of a feature represents that feature having a more decisive influence on that observation being classified as a critical safety event. Similarly, a more negative feature indicates that a feature has a more significant influence on that observation being classified as a non-critical event. Moreover, the thickness of the bar in Figure 2 means the number of observations at a certain feature value.

In Figure 1, out of all continuous variables, pedestrian speed is the most dominant feature associated with the critical/non-critical pedestrian-vehicle safety events. Figure 2 shows that a higher pedestrian speed leads to a higher occurrence of critical safety events. Typically, there are two scenarios for these critical safety events with pedestrians walking at a relatively higher speed. The first scenario is that a pedestrian who rushes into the crosswalk is aware of the environment, and another scenario is that the distracted pedestrian is unaware of the surroundings. For the first scenario, the pedestrian crosses the crosswalk with a great amount of assertion, knowing the time gap with another vehicle is very short. The second scenario is that the pedestrian crosses the crosswalk recklessly without checking the environment before crossing. Both scenarios are concerning, but the second one is more severe. The second feature that affects the occurrence of critical events is vehicle speed. The safety events are more likely to be classified as non-critical events when the involved vehicle's speed is high. This may suggest that, in general, the pedestrian would avoid the possible conflict when the vehicle speed is high (indicating aggressive driving behavior of the drivers), and the pedestrian could be more aggressive when the vehicle speed is low. The slow-moving vehicle may boost the confidence of both pedestrian and vehicle driver given the perception that their safety is under control; drivers think they can immediately stop the vehicle, and the pedestrian thinks they can avoid the collision if anything happens. For drivers, similar issues can arise. Drivers are somewhat more confident about stopping a vehicle rather than when the vehicle is operated in high speed. The behavioral adaptation related issues among drivers/pedestrians and yielding/crossing patterns require additional analysis, which is outside the scope of this study. The traffic volume of the whole intersection, pedestrian volume, and vehicle volume in the conflict directions are important features associated with critical near-crash events. The total traffic volume generally describes how busy that intersection was when the near-crash event happened. The three-traffic volume-related features show nonlinear associations with the occurrences of critical/non-critical events. More investigation on these nonlinear relationships will be presented in a later discussion. This plot also shows how the rest of the continuous variables contribute to the classification tasks. The length of the pedestrian phase and the "Don't walk" flashing signal affects the probability of the occurrence of critical near-crash events. Longer pedestrian signals (walk indication and "Don't walk" flashing signals) increase the chance of critical events. For the number of lanes in the conflict vehicle direction, no exclusive lane or many exclusive lanes lower the probability of critical near-crash events. In this analysis, no exclusive lane is only for a turning movement, and two or three lanes are often for through traffic. No exclusive lane for a turning movement may require more of a driver's attention while turning. For the through traffic, which often has more lanes than turning traffic, the pedestrian traffic light must be red when the through traffic is allowed. The odds of having critical near-crash events between the pedestrian and through traffic are relatively low compared with the turning traffic. The plot also points out that the total crashes in the past five years at each intersection are positively associated with the occurrence of the critical near-crash event. Near-crash events at the intersection with higher historical crash counts are more likely to be classified as critical near-crash events.

Partial Dependence Plot (PDP). The PDP presents the SHAP values (marginal effects) of individual features across all the predictions. The x-axis is the value of the feature, the y-axis on the left side is the SHAP value, and the y-axis on the right side is the value of another feature. Similar to Figure 2, a higher and positive SHAP value supports the critical event classification, and a negative and lower SHAP value supports the non-critical event classification. The individual observations are colored based on the value of the feature on the right side of the y-axis, which grants more power to the interpretation. Out of the 18 features selected for training the classification model, four features and one intersection effect with the most obvious and insightful patterns are discussed in this section, based on the importance and meaningfulness of the feature impact. The four features are pedestrian speed, vehicle speed, vehicle volume, and pedestrian volume, and the interaction effect is between the pedestrian volume and signal cycle.

Pedestrian speed: Figure 3 shows the general trend of the impact of pedestrian speed on critical near-crash events. The color in the figure indicates the vehicle volume at the moment of that near-crash event. Increasing pedestrian speed contributes to more critical near-crash events. When the pedestrian rushed through the crosswalk, it is possible that the pedestrian tried to cross the street during the pedestrian red indication ("Don't walk") or at the end of the pedestrian flashing indication, and that is where the most critical near-crash events occurred. According to the Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD), the recommended average speed for calculating the pedestrian clearance intervals for traffic signals is 4.0 ft/s (about 3 mph) (*40*). The Design and Safety of Pedestrian Facilities published by ITE (*41*) suggests the walking speed to be up to 8 ft/s (5.4 mph). Thus, this study considers any walking speed less than 3 mph to be slow walking, and any walking speed higher than 5 mph to be a fast-walking or running pace. There were also a few pedestrians with really high speeds over 10 mph. They were either running fast or riding scooters/other tools. In Figure 4, when the speed of slow walking pedestrians increases from 0 to 5 mph, the near-crash event is more likely to be a non-critical event. However, when the speed of the fast-walking pedestrian increases from 5 to 10 mph, the near-crash event is more likely to be a critical event. Moreover, the increasing pedestrian speed in this range would increase the likelihood of turning a near-crash event into a critical event. There are also some observations with higher than 10 mph, which could

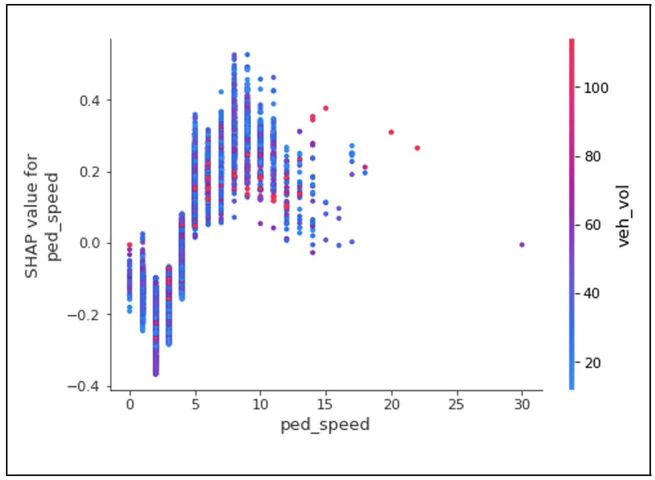

Figure 3. Partial dependence plot of pedestrian speed and vehicle volume.
Note: SHAP = SHapley Additive exPlanations; ped_speed = involved-pedestrian speed (mph); veh_vol = traffic volume in the involved-vehicle direction.

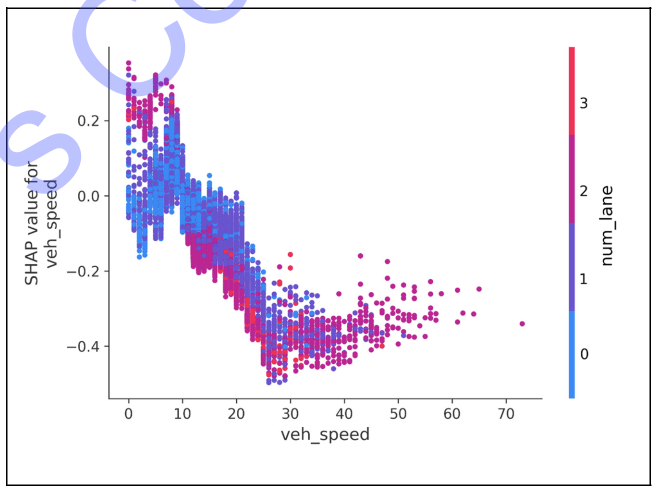

Figure 4. Partial dependence plot (PDP) of conflict vehicle speed and number of lanes.
Note: SHAP = SHapley Additive exPlanations; veh_speed = involved-vehicle speed (mph); num_lane = number of lanes in the direction of the involved vehicle.

be people running fast or using scooters or other tools. The image recognition algorithm mistakenly classified them as pedestrians during the data collection process through video footage at these intersections. However, a trend can still be observed. When the speed of an object is more than 10 mph, the impact of the increasing speed slightly decreases but still positively contributes to the critical event. In short, for pedestrians walking at a normal speed while crossing the crosswalk, the near-crash event is more likely to be a non-critical event; for pedestrians walking at a relatively high speed, the increasing speed is positively associated with critical events.

Vehicle speed: Figure 3 is the PDP of the speed of conflict vehicles. In general, the increasing vehicle speed lowers the chance of critical pedestrian-vehicle near-crash events. There is a noticeable difference between vehicle speeds less than and more than 10 mph. A near-crash event involving a conflicting vehicle with a less than 10 mph speed is more likely to be classified as a critical event. By contrast, if the involved vehicle's speed was over 10 mph, that critical event is more likely to be a non-critical event. The dominance of this feature in classifying the near-crash events as non-critical events increases as the speed is over 10 mph. There are several possible explanations for this critical speed (10 mph) and the trend in the plot. First, while the vehicle speed is above the critical speed (10 mph), this might be the incoming vehicle's speed threshold for pedestrians who want to cross the crosswalk safely. Whenever the incoming vehicle is above this critical speed, the pedestrian may take extra caution and choose not to be anywhere close to that vehicle's trajectory, such as by stepping onto the crosswalk. Pedestrians may feel more comfortable crossing the street when the incoming vehicles are at a lower speed. Another possible explanation is that these critical events are classified based on the time gap between the pedestrian and the vehicle to reach the conflict point. Critical events have a shorter PET, which means the two objects were really close. In that scenario, the vehicles may reduce their speed to avoid a possible collision. When the vehicle reduces speed to a lower level, especially for these turning vehicles, this may boost the pedestrian's confidence and cause them to be more aggressive.

Another finding is that for involved vehicles with speeds less than 10 mph, the near-crash event is more likely to be a critical event if there is more than one lane in the direction of that vehicle. Interestingly, for vehicles with speeds over 10 mph, more than one lane of the direction of that vehicle becomes a factor that decreases the chance of a critical near-crash event. Left-turn and through are two movement types, possibly with more than one exclusive lane. Normally, the left-turn and through movements should not be conflicted with any crosswalk. Traffic signal engineers would set red lights

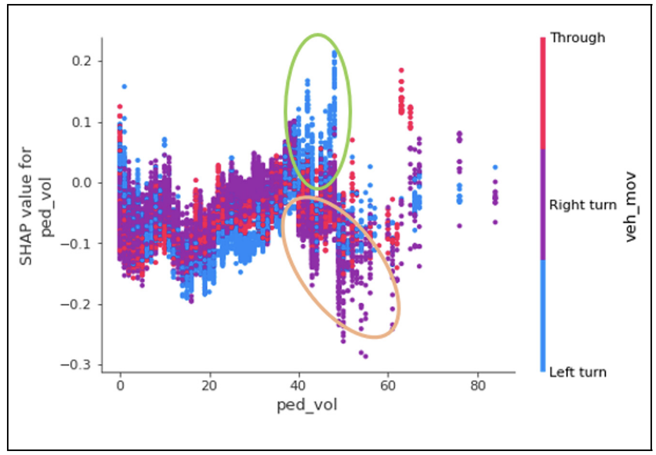

Figure 5. Partial dependence plot (PDP) of the pedestrian volume and vehicle movement.
Note: SHAP = SHapley Additive exPlanations; veh_speed = involved-vehicle speed (mph); num_lane = number of lanes in the direction of the involved vehicle.

for pedestrians when the conflict left-turn signal indication is protected, and the vehicle should yield for the pedestrian when the left-turn signal is permissive and the pedestrian indication is green, to avoid possible conflicts. Thus, in this scenario, either the vehicle or the pedestrian may violate the traffic light or be very aggressive. For the movements with more lanes (left-turn and through), lower vehicle speed, aggressive behavior, and possible violations of the traffic signal lead to more critical near-crash events.

Pedestrian volume: The pedestrian volume affects the odds of near-crash events insignificantly when the volume of pedestrians is less than 40 per 15 min. However, in the green interval in Figure 5, there is a group of conflicts showing the strong association between a relatively high pedestrian volume and critical near-crash incidents with left-turn vehicles. As mentioned above, the pedestrian indication should be red while the conflict left-turn signal indication is protective and the left-turn indication should be permissive when the conflict pedestrian indication is green. A relatively high pedestrian volume, around 40 per 15 min (i.e., 160 pedestrians/h), could increase the odds of critical near-crash occurrences. One possible explanation is that a high pedestrian volume increases the jaywalking rate. The conflicts between these unlawful crossing pedestrians and left-turn vehicles are more likely classified as critical conflict. Another possible reason could be the pedestrian allow-to-walk interval is not sufficient for a large crowd of pedestrians, and some of them step onto the crosswalk while the "Don't walk" indication is on. In this scenario, the pedestrians may still be on the crosswalk walking while the vehicle's left-turn signal turns to protective or permissive, which leads to

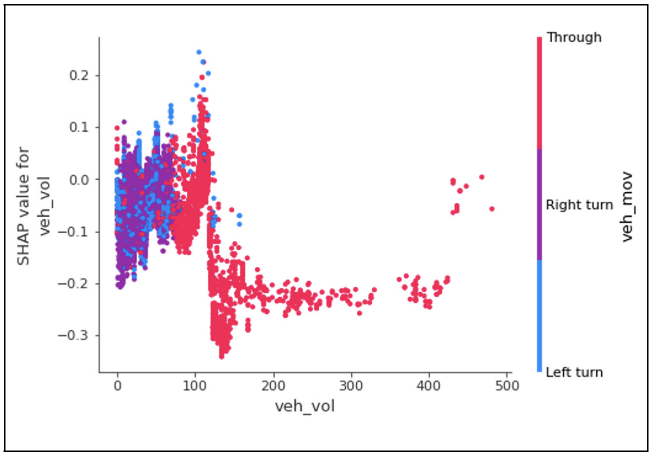

Figure 6. Partial dependence plot (PDP) of the vehicle volume and vehicle movement.
Note: SHAP = SHapley Additive exPlanations; veh_vol = traffic volume in the involved-vehicle direction; veh_mov = movement of the involved vehicle.

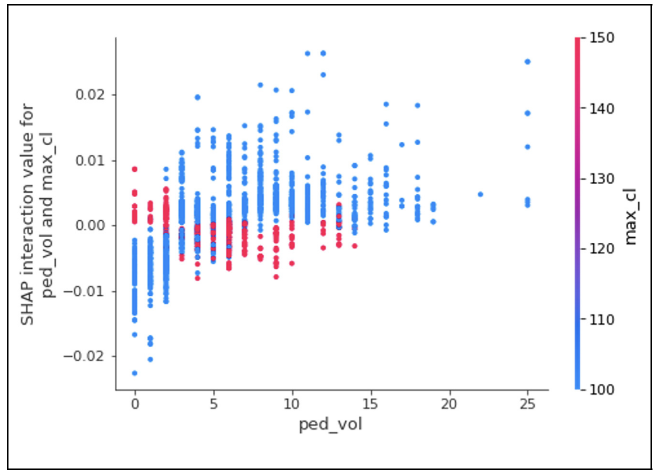

Figure 7. Interaction effect of pedestrian volume and maximum signal cycle.
Note: SHAP = SHapley Additive exPlanations; ped_vol = pedestrian volume in the involved-pedestrian direction; max_cl = maximum signal cycle (s).

these critical near-crash events. When the pedestrian volume is continuously going up, the probability of critical near-crash events occurring goes down. Additionally, for these near-crash events, while the pedestrian volume is over 40 per 15 mins, the conflicts involved with right-turn vehicles (see the light brown ellipse in Figure 5) are the least associated with the critical events. The possible explanation for this is that the right-turn vehicles become more cautious when many pedestrians are trying to cross the street. On the other hand, the pedestrians walking on the conflict crossing are usually unexpected for the left-turn or through vehicles, which leads to more critical near-crashes in that scenario.

Vehicle volume: Figure 6 describes the impact of vehicle volume on the moving direction of the conflicting vehicle on the classification task. For right-turn or left-turn vehicles, the 15 min maximum volume is less than 100, and this volume factor, in general, has no obvious association between the increased traffic volume and the odds of critical near-crash events. However, this association becomes obvious for the through movement, especially when the through traffic volume is over 100 vehicles in 15 min (i.e., 400 vehicles per hour [vph]). The critical near-crash event is less likely to occur for the through movement when the through traffic is more than 100 per 15 min (i.e., 400 vph). As discussed above, in most conflicts of pedestrians and through traffic, the pedestrians are more likely to be the ones who violated the traffic lights rule and tried to find a gap to cross the street without a green light. This kind of jaywalking behavior has less chance of occurring when the through traffic is heavy.

Interaction effect between the pedestrian volume and signal cycle: The interaction effect indicates the interactive impact of two features on the classification task, except for the impacts directly from individual features (see Figure 7). When the pedestrian volume is relatively low, less than 4 pedestrians per 15 min (i.e., 16 pedestrians/h), and the maximum signal cycle length is relatively long, higher than 130 s, the near-crash events are more likely to be classified as critical events. Moreover, a relatively short signal cycle length, about 100 s, would increase the chance of critical near-crash events when the pedestrian volume is higher than 5 per 15 min (i.e., 20 pedestrians/h). While it is only a few pedestrians, and probably during off-peak hours, the pedestrian has less patience with the long signal cycle length. However, long cycle lengths mitigate the critical near-crash frequency when the pedestrian volume increases because the long signal cycle length often also provides a longer crossing time for pedestrians.

The SHAP interpretable machine learning framework provides a novel perspective for investigating factors associated with pedestrian near-crash events. The key findings of this research are listed as follows.

- For critical and non-critical pedestrian–vehicle near-crashes, the majority of these events involved right-turn vehicles.
- The critical pedestrian near-crash has more likelihood of happening with a pedestrian who walks faster than normal walking speed (3–5 mph). There are multiple reasons that may cause pedestrians to walk faster than normal. One is the walking nature of a pedestrian, as some pedestrians may just walk faster than others. However, machine learning leans toward the overall average

trend rather than focusing on a small group of pedestrians. Therefore, a fast-walking pedestrian may suggest that the pedestrian violates the traffic signal, stepping onto the crosswalk while the "Don't walk" indication is on and trying to cross the street as fast as they can, or they might be trying to avoid the oncoming traffic which might violate the traffic rule.

- A vehicle with a driving speed under 10 mph is more likely to be involved in a critical near-crash event. High-speed vehicles are highly associated with non-critical near-crash events. Pedestrians may tend to avoid relatively fast vehicles. Lower-speed vehicles may boost the confidence of both the drivers and pedestrians with the perception that they can react to an emergency if anything occurs.
- A relatively high pedestrian volume (higher than 160 pedestrians/h) is highly associated with critical near-crash events with left-turn vehicles. Interestingly, a relatively high pedestrian volume is highly associated with non-critical near-crash events with right-turn vehicles. The conflict between pedestrians and left-turn vehicles is less likely to happen given the common design of traffic signal timing plans. The probability of having jaywalking pedestrians is higher when the pedestrian volume is high, which creates more dangerous near-crash events. High pedestrian volume increases visibility for the right-turn vehicles, and it increases driver awareness and caution about crossing pedestrians. Thus, the likelihood of the occurrence of the critical near-crash event drops.
- For through traffic, the critical pedestrian near-crash event is most likely to occur when the volume of through traffic is less than 400 vph. The odds significantly decrease when the through traffic volume is more than 400 vph. The conflicts between the through traffic and pedestrian are most likely caused by the pedestrian violating the traffic signal. Pedestrians may try to find a gap between oncoming vehicles to pass the street instead of waiting for the long signal cycle.
- While pedestrian volume is low, a long signal cycle increases the odds of a critical near-crash event. For intersections with very few pedestrians, pedestrians may lose their patience while waiting for the long signal cycle. When the pedestrian volume is high, the short signal cycle is highly associated with the occurrence of critical near-crash events.

Conclusions

This study establishes an interpretable machine learning framework by using a pedestrian–vehicle near-crash data set at signalized intersections to understand the associations between the selected factors and the critical near-crash events. These results reveal the nonlinear relationship between factors and critical near-crash events. Moreover, the results show great insights by incorporating traffic signal information into the analysis.

The findings shed new light on pedestrian safety analysis. First, surrogate data for pedestrian-involved crashes has not been widely accessible yet. This study established the relationship between the valuable surrogate data set and the pedestrian-involved crash data set. Many factors, which are correlated with pedestrian-involved crashes, were found to be associated with pedestrian-involved near-crash events. Second, the nonlinear effects of these factors on pedestrian-involved crashes are stated in previous studies but have not been thoroughly studied. The interpretable performance measure PDPs provide a detailed understanding of how the changing value of these factors could affect the odds of critical near-crash events which involve pedestrians. Third, not many studies have considered the signal cycle as a factor. The results prove the impact of the signal cycle on the critical near-crash event. Moreover, the interaction effect of the pedestrian volume and signal cycle is investigated and offers meaningful insights.

This analysis could be an exemplary case for studying pedestrian safety at signalized intersections. With these insights provided by this interpretable machine learning framework, countermeasures could be selected or designed. For example, a high pedestrian volume increases the odds of critical conflict with the left-turn vehicle and decreases the chance of critical conflict with a right-turn vehicle. The conflict between the left-turn vehicle and the pedestrian often suggests a violation of the traffic signal on one side. For example, it is possible that pedestrians violate the traffic signal because they are waiting too long for the signal cycle. A signal cycle length adjustment can be beneficial in reducing these events. Further investigation is needed to confirm this speculation and adjust the signal cycle to better accommodate these needs. The results also show that low pedestrian volume and long signal cycle increase the probability of critical near-crash events. Very few pedestrians can be associated with the off-peak hours and low traffic volume. If that is the case, shortening the signal cycle may decrease the chance of the occurrence of critical near-crash events. However, given the limitation of data resources, the historical pedestrian waiting time for each near-crash event cannot be retrieved. Further investigation is needed to confirm our speculations.

The current study also has other limitations. This is a case study with detailed information from only six signalized intersections. However, the data set is robust and provides many additional details, which is much harder to attain in many studies. Some factors, such as total crashes at the intersection, may not have enough variation across the data to become factors with dominant power of prediction. These factors are possible with small variation across the data to have great prediction power and dominant associations with the critical near-crash events. In recent years, several studies have applied interpretable machine learning in safety analysis (*42–45*). There is a need for additional research focus on the application of machine learning related explainability for wide adoption of machine learning based models.

Acknowledgments

The authors would like to thank the BriskLUMINA, BriskVANTAGE, and ITE Vision Zero Sandbox Design Competition for offering this valuable near-crash data set.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: X. Kong; data collection: X. Kong; analysis and interpretation of results: X. Kong, S. Das; draft manuscript preparation: X. Kong, S. Das, Y. Zhang, Z. Wei, C. Yuan; All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Xiaoqiang Kong https://orcid.org/0000-0002-8120-0754
Subasish Das https://orcid.org/0000-0002-1671-2753
Zihang Wei https://orcid.org/0000-0002-1790-022X

References

1. NHTSA. *Traffic Safety Facts - Overview of Motor Vehicle Crashes in 2019*. Publication DOT HS 813 060. National. Highway Traffic Safety Administration. NHTSA's National Center for Statistics and Analysis, Washington, D.C., 2020.
2. FHWA. About | Intersection Safety - Safety | Federal Highway Administration. https://safety.fhwa.dot.gov/intersection/about/. Accessed May 19, 2021.
3. Kröyer, H. R. Pedestrian and Bicyclist Flows in Accident Modelling at Intersections. Influence of the Length of Observational Period. *Safety Science*, Vol. 82, 2016, pp. 315–324.
4. Lee, J., M. Abdel-Aty, and Q. Cai. Intersection Crash Prediction Modeling With Macro-Level Data From Various Geographic Units. *Accident Analysis & Prevention*, Vol. 102, 2017, pp. 213–226.
5. Miranda-Moreno, L. F., P. Morency, and A. M. El-Geneidy. The Link Between Built Environment, Pedestrian Activity and Pedestrian–Vehicle Collision Occurrence at Signalized Intersections. *Accident Analysis & Prevention*, Vol. 43, No. 5, 2011, pp. 1624–1634.
6. Strauss, J., L. F. Miranda-Moreno, and P. Morency. Multimodal Injury Risk Analysis of Road Users at Signalized and Non-Signalized Intersections. *Accident Analysis & Prevention*, Vol. 71, 2014, pp. 201–209.
7. Wang, J., H. Huang, and Q. Zeng. The Effect of Zonal Factors in Estimating Crash Risks by Transportation Modes: Motor Vehicle, Bicycle and Pedestrian. *Accident Analysis & Prevention*, Vol. 98, 2017, pp. 223–231.
8. Lee, C., and M. Abdel-Aty. Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida. *Accident Analysis & Prevention*, Vol. 37, No. 4, 2005, pp. 775–786.
9. Lee, J., M. Abdel-Aty, and I. Shah. Evaluation of Surrogate Measures for Pedestrian Trips at Intersections and Crash Modeling. *Accident Analysis & Prevention*, Vol. 130, 2019, pp. 91–98.
10. Rankavat, S., and G. Tiwari. Association Between Built Environment and Pedestrian Fatal Crash Risk in Delhi, India. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2519: 61–66.
11. Schneider, R. J., M. C. Diogenes, L. S. Arnold, V. Attaset, J. Griswold, and D. R. Ragland. Association Between Roadway Intersection Characteristics and Pedestrian Crash Risk in Alameda County, California. *Transportation Research Record: Journal of the Transportation Research Board*, 2010. 2198: 41–51.
12. Ukkusuri, S., L. F. Miranda-Moreno, G. Ramadurai, and J. Isa-Tavarez. The Role of Built Environment on Pedestrian Crash Frequency. *Safety Science*, Vol. 50, No. 4, 2012, pp. 1141–1151.
13. Chimba, D., A. Musinguzi, and E. Kidando. Associating Pedestrian Crashes With Demographic and Socioeconomic Factors. *Case studies on Transport Policy*, Vol. 6, No. 1, 2018, pp. 11–16.
14. Esmaili, A., K. Aghabayk, N. Parishad, and A. N. Stephens. Investigating the Interaction Between Pedestrian Behaviors and Crashes Through Validation of a Pedestrian Behavior Questionnaire (PBQ). *Accident Analysis & Prevention*, Vol. 153, 2021, p. 106050.
15. Munira, S., I. N. Sener, and B. Dai. A Bayesian Spatial Poisson-Lognormal Model to Examine Pedestrian Crash Severity at Signalized Intersections. *Accident Analysis & Prevention*, Vol. 144, 2020, p. 105679.
16. Russo, B. J., E. James, C. Y. Aguilar, and E. J. Smaglik. Pedestrian Behavior at Signalized Intersection Crosswalks: Observational Study of Factors Associated With Distracted Walking, Pedestrian Violations, and Walking Speed. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 1–12.
17. Stammen, J. A., S. Ko, D. A. Guenther, and G. Heydinger. A Demographic Analysis and Reconstruction of Selected Cases From the Pedestrian Crash Data Study. *SAE Transactions*, 2002, SAE Technical Paper 2002-01-0560, pp. 840–853.

18. Iryo-Asano, M., and W. K. Alhajyaseen. Modeling Pedestrian Crossing Speed Profiles Considering Speed Change Behavior for the Safety Assessment of Signalized Intersections. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 332–342.
19. Ding, C., P. Chen, and J. Jiao. Non-Linear Effects of the Built Environment on Automobile-Involved Pedestrian Crash Frequency: A Machine Learning Approach. *Accident Analysis & Prevention*, Vol. 112, 2018, pp. 116–126.
20. Yue, L., M. Abdel-Aty, Y. Wu, O. Zheng, and J. Yuan. In-Depth Approach for Identifying Crash Causation Patterns and its Implications for Pedestrian Crash Prevention. *Journal of Safety Research*, Vol. 73, 2020, pp. 119–132.
21. Agarwal, N. K. Estimation of Pedestrian Safety at Intersections Using Simulation and Surrogate Safety Measures. PhD dissertation. *University of Kentucky, Lexington*, 2011.
22. Chen, P., W. Zeng, and G. Yu. Assessing Right-Turning Vehicle-Pedestrian Conflicts at Intersections Using an Integrated Microscopic Simulation Model. *Accident Analysis & Prevention*, Vol. 129, 2019, pp. 211–224.
23. Chrysler, S. T., O. Ahmad, and C. W. Schwarz. Creating Pedestrian Crash Scenarios in a Driving Simulator Environment. *Traffic Injury Prevention*, Vol. 16, No. sup1, 2015, pp. S12–S17.
24. Guo, F., S. G. Klauer, J. M. Hankey, and T. A. Dingus. Near Crashes as Crash Surrogate for Naturalistic Driving Studies. *Transportation Research Record: Journal of the Transportation Research Board*, 2010. 2147: 66–74.
25. Zhang, Y., and T. Cheng. Graph Deep Learning Model for Network-Based Predictive Hotspot Mapping of Sparse Spatio-Temporal Events. *Computers, Environment and Urban Systems*, Vol. 79, 2020, p. 101403.
26. Huisingh, C., E. B. Levitan, M. R. Irvin, P. MacLennan, V. Wadley, and C. Owsley. Visual Sensory and Visual-Cognitive Function and Rate of Crash and Near-Crash Involvement Among Older Drivers Using Naturalistic Driving Data. *Investigative Ophthalmology & Visual Science*, Vol. 58, No. 7, 2017, pp. 2959–2967.
27. Kong, X., S. Das, and Y. Zhang. Mining Patterns of Near-Crash Events With and Without Secondary Tasks. *Accident Analysis & Prevention*, Vol. 157, 2021, p. 106162.
28. Perez, M. A., J. D. Sudweeks, E. Sears, J. Antin, S. Lee, J. M. Hankey, and T. A. Dingus. Performance of Basic Kinematic Thresholds in the Identification of Crash and Near-Crash Events Within Naturalistic Driving Data. *Accident Analysis & Prevention*, Vol. 103, 2017, pp. 10–19.
29. Swain, T. A., G. McGwin, J. M. Wood, J. F. Antin, and C. Owsley. Naturalistic Driving Techniques and Association of Visual Risk Factors With At-Fault Crashes and Near Crashes by Older Drivers With Vision Impairment. *JAMA Ophthalmology*, Vol. 139, No. 6, 2021, pp. 639–645.
30. Su, J., J. Chen, H. Wang, W. Chen, and K. Wang. Establishment and Analysis on Typical Road Traffic Near-Crash Scenarios Related to Pedestrian in China. *Traffic and Transportation*, 2017, pp. 209–214.
31. Wu, J., H. Xu, Y. Zheng, and Z. Tian. A Novel Method of Vehicle-Pedestrian Near-Crash Identification With Roadside LiDAR Data. *Accident Analysis & Prevention*, Vol. 121, 2018, pp. 238–249.
32. Wu, J., H. Xu, Y. Zhang, and R. Sun. An Improved Vehicle-Pedestrian Near-Crash Identification Method With a Roadside LiDAR Sensor. *Journal of Safety Research*, Vol. 73, 2020, pp. 211–224.
33. ITE. Vision Zero Sandbox Competition. Institute of Transportation Engineers. https://www.ite.org/membership/sandboxcompetition/. Accessed May 25, 2021.
34. Samara, L., P. St-Aubin, F. Loewenherz, N. Budnick, and L. Miranda-Moreno. Video-Based Network-Wide Surrogate Safety Analysis to Support a Proactive Network Screening Using Connected Cameras: Case Study in the City of Bellevue (WA) United States. Presented at 100th Annual Meeting of the Transportation Research Board, Washington, D.C., 2020.
35. Chen, T., and C. Guestrin. Xgboost: A Scalable Tree Boosting System. *Proc., 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, San Francisco, CA, Association for Computing Machinery, New York, August 2016, pp. 785–794.
36. Lundberg, S. M., and S.-I. Lee. A Unified Approach to Interpreting Model Predictions. *arXiv Preprint arXiv:1705.07874*, 2017.
37. Hoxha, G., A. Shala, and R. Likaj. Pedestrian Crash Model for Vehicle Speed Calculation at Road Accident. *International Journal of Civil Engineering and Technology*, Vol. 8, No. 9, 2017, pp. 1093–1099.
38. Mukherjee, D., and S. Mitra. A Comparative Study of Safe and Unsafe Signalized Intersections From the View Point of Pedestrian Behavior and Perception. *Accident Analysis & Prevention*, Vol. 132, 2019, p. 105218.
39. Xie, S. Q., N. Dong, S. C. Wong, H. Huang, and P. Xu. Bayesian Approach to Model Pedestrian Crashes at Signalized Intersections With Measurement Errors in Exposure. *Accident Analysis & Prevention*, Vol. 121, 2018, pp. 285–294.
40. FHWA. *Manual on Uniform Traffic Control Devices*. FHWA, Washington, D.C., 2009.
41. ITE. *Safety of Pedestrian Facilities*. Institute of Transportation Engineers Technical Council Committee 5A-5, Washington, D.C., 1994.
42. Wei, Z., S. Das, and Y. Zhang. Apply Explainable Machine Learning Techniques in Daily Crash Occurrence and Severity Modeling for Rural Interstates. *Transportation Research Record: Journal of the Transportation Research Board*, 2022.
43. Wei, Z., S. Das, and Y. Zhang. Short Duration Crash Prediction for Rural Two-Lane Roadways: Applying Explainable Artificial Intelligence. *Transportation Research Record: Journal of the Transportation Research Board*, 2022.
44. Das, S., and I. Tsapakis. Interpretable Machine Learning Approach in Estimating Traffic Volume on Low-Volume Roadways. *International Journal of Transportation Science and Technology*, Vol. 9, No. 1, 2020, pp. 76–88.
45. Das, S., S. Datta, H. Zubaidi, and I. Obaid. Applying Interpretable Machine Learning to Classify Tree and Utility Pole Related Crash Injury Types. *IATSS Research*, Vol. 45, No. 3, 2019, pp. 310–316.