

Article

Modeling Crash Injury Severity for Vulnerable Road Users Using CatBoost and SHAP: Uncovering Complex Risk Interactions

Mousa Abushattal ¹, Mohammad Nour Al-Marafi ², Rasha Al-Shamaseen ¹, Fadi Alhomaiddat ¹, Fareh Abudawaba ¹ and Ahmed Jaber ^{3,4,*}

¹ Civil Engineering Department, Al Hussein Bin Talal University, Ma'an 71111, Jordan; mousa.abushattal@ahu.edu.jo (M.A.); rashashamaseen@gmail.com (R.A.-S.); fadi.alhomaiddat@ahu.edu.jo (F.A.); fareh.abudawaba@ahu.edu.jo (F.A.)

² Civil and Construction Engineering, Tafila Technical University, Tafila 66110, Jordan; mohammad.almarafi@ttu.edu.jo

³ CEDARZ Center for Development and Research, Ramallah P622, Palestine

⁴ Faculty of Graduate Studies, An-Najah National University, Nablus P.O. Box 7, Palestine

* Correspondence: a.jaber@cedarz.co

Abstract

Rapid urbanization and increasing traffic density have increased the crash risk of vulnerable road users (VRUs), particularly pedestrians and cyclists. Meanwhile, the conventional statistical models have difficulties in capturing the nonlinear and complex nature of crash data, limiting their safety analysis effectiveness. This study utilized advanced Gradient Boosting machine learning and integrated it with SHapley Additive exPlanations (SHAP) using five years of crash data from Michigan, USA, employing a two-tiered modeling design consisting of a 4-class joint structure and binary subset frameworks. Rigorously evaluated using 10-fold stratified cross-validation to predict crash severity for VRUs, the CatBoost model had better predictive performance (AUC = 0.917) than LightGBM, Random Forest and the traditional Logistic Regression models. The analysis further indicated that prior crash actions, particularly risky crossing behaviors, are the most significant determinants of injury severity for both user groups. However, the pedestrian crash severity is strongly associated with lighting conditions and speed limits, while cyclist crash severity is more heavily influenced by intersection involvement and roadway geometry. Moreover, SHAP interaction analysis showed that the speed effect on severity significantly increases when it interacts with hazardous actions or poor visibility. The findings provide a critical insight into the implementation of effective measures and infrastructure improvements to increase the safety of VRUs.

Keywords: vulnerable road users (VRUs); crash injury severity; CatBoost; SHAP analysis; explainable artificial intelligence (XAI); traffic safety; pedestrian crashes; bicycle crashes; machine learning; crash severity prediction

Academic Editor: Anna Grana

Received: 9 June 2026

Revised: 21 July 2026

Accepted: 22 July 2026

Published: 27 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Urbanization and rapid population mobility intensely reshaped travel behaviors and road usage patterns in modern cities. Rapid urban development is changing residential, employment, and commuting structures while expanding road networks and increasing

traffic density. As a result, interactions and conflicts among road users have increased, raising exposure to crashes for non-motorized users. Road crashes involving vulnerable road users (VRUs), which include pedestrians, cyclist riders, and other unprotected users, represent significant challenges to global public health and safe transportation [1].

Globally, traffic crashes remain a leading cause of deaths and severe injury. The World Health Organization (WHO) reports that although there was recently a decline in fatal crashes, they still present consistent challenges for vulnerable users, particularly in complex urban circumstances that involve motorized and non-motorized interactions [2]. VRUs are more exposed and vulnerable than vehicle users and are therefore more at risk of injury. Additionally, according to the WHO, during the years 2010–2021, there were increases in fatal injuries to both pedestrians and bicyclists, 3% and 20%, respectively. Similarly in the United States, VRUs represented approximately 21.6% of all crash fatalities during 2023. More specifically, nearly 7000 pedestrians and over 1000 bicycle riders lost their lives due to road crashes [3]. These figures showed the continued existence of substantial and growing safety concerns for non-motorized roadway users. As such, the safety of VRUs represents a significant priority area in terms of traffic safety goals and objectives as well as being supportive of promoting active transportation and mitigating congestion. Developing effective countermeasures to reduce the risks faced by VRUs needs a comprehensive understanding of crash mechanisms and crash contributing factors. Previous studies on VRU safety have identified a wide range of these factors, including human-related factors, geometry factors, vehicle characteristics, traffic composition and conditions, and the surrounding environment [4]. However, the sophisticated interaction and fusion of these factors form a serious challenge for the conventional analytical approaches. As a result, we need more robust analytical frameworks that are capable of capturing the impact of contributing factors on VRU severity; this is a critical priority. This improvement can significantly enhance traffic management and decision-making.

To date, most of the conducted studies on VRU crash severity rely on conventional statistical approaches, order probit, multinomial logit, and mixed-effects models [5–8]. Even though these models provide interpretative ability and are capable of measuring the marginal effects of contributing predictors on severity class outcomes, they may be limited in terms of accuracy and explanation due to their commitment to restrictive assumptions regarding data linearity and distributional properties. Moreover, as the nature of crash data can be described with high-dimensional characteristics, nonlinear patterns, and complex variable interdependency, conventional statistical approaches may struggle to offer an understandable explanation for VRU crashes, specifically with large-scale datasets [6].

In response to these challenges, machine learning (ML) techniques such as Random Forests (RFs), Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Gradient Boosting, which have attracted growing attention recently, provide robust and superior performance compared to conventional approaches, particularly in crash severity prediction [9–11]. Compared to conventional approaches, these analytical tools allow us to identify nonlinear correlations and complex interactions in the data. Moreover, Gradient Boosting is a very effective method since it uses ensemble learning to increase model predictability, thereby achieving both high accuracy and robustness in complex scenarios such as crash analyses [10].

Despite their predictive advantages, ML models often face criticism for their limited transparency and interpretability, commonly referred to as “black-box” approaches [12]. The ML model is usually limited in terms of capturing individual variable influence on prediction, which restricts its usefulness in traffic safety applications. On the other hand, decision makers and infrastructure authorities require interpretable evidence to justify infrastructure investment, traffic regulations, and policy implementations. Recently, explainable artificial intelligence (XAI) techniques were introduced to provide complex ML

models with the power of interpretation while keeping a high level of predictive accuracy. This valuable additive to ML provides a promising trade-off between accuracy and interpretation, especially in gradient boosting models. Among these techniques, SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee [13], provides a chance to determine the predictors' importance and capture the marginal effect as a visualization in an intuitive manner. This additive technique forms a robust framework that allows the researchers to obtain a deeper understanding of individual risk factors impacting the crash severity and capture their variation under different circumstances.

While most existing studies investigate the safety of VRUs separately, pedestrians and cyclists commonly share the same traffic environment under comparable infrastructure and regulatory conditions. Analyzing them separately may overlook shared exposure patterns, common risk factors, and interrelated behavioral responses. Therefore, adopting a separate approach may lead to fragmented decisions and inconsistent recommendations. In contrast, a joint model approach offers a chance to integrate contributing factors across VRUs and provide a comprehensive understanding of crash severity mechanisms. The joint multiclass modeling framework enables the simultaneous analysis of multiple crash severity levels and road user groups. The similarities and differences can be captured between user groups (pedestrians and cyclists) by combining the classification problem in one analytical approach, which involves comparing the processes across user groups and capturing the decision-making-related factors, as well as identifying the factors associated with fatalities and serious injuries.

However, despite these advancements, several primary research gaps have been identified. Most previous studies of pedestrians and cyclists are either separated or focused on a single group. A joint analytical framework is needed to distinguish shared risk factors from user-specific mechanisms, thus adopting coordinated safety decisions. Moreover, although ML and SHAP can determine the importance of individual variables (marginal contributions), there are limited applications to features interactions such as how speed actually has a different effect under poor lighting or hazardous behavior, which is crucial for crash causation.

2. Literature Review

Previous safety studies indicate that injury severity of VRUs is the result of the interaction between multiple factors at different levels [4,14–17]. Crash outcomes are a result of a combination of vehicle attributes, human characteristics, and environmental and roadway conditions. In addition, the spatial and temporal context of where the crash takes place is also significant. These factors do not function alone but are interdependent to either reduce or increase risk based on the specific circumstances of the crash. Therefore, these results indicate that the severity of crashes for VRUs resulted from complex interactions rather than from a single factor. However, the majority of existing studies treat these contributing factors as independent predictors, largely failing to quantify how the simultaneous interaction of these variables amplifies crash risk.

Factors such as demographics, vehicle type, traffic conditions, and driver behavior all exist within an interaction context, which points towards the importance of targeting safety initiatives to high-risk zones, especially for VRUs [4,17]. Findings from previous studies also indicate that human- and traffic-related factors have a major impact on determining crash injury outcomes. Specifically, driver education and the behavior of VRUs can reduce risk, whereas high vehicle speeds, presence of heavy vehicles, driving under the influence, and crashes on major roads significantly increase the likelihood of severe crashes. In addition, environmental and road features, which include lighting, pavement quality, road alignment, and protective barriers, directly impact the severity of injuries for VRUs [15]. Taken together, these results indicate that the severity of crashes for VRUs

resulted from complex interactions rather than from a single, isolated factor. Recent studies on crash injury severity also showed the importance of accounting for heterogeneous and interactive effects [18,19]. A study performed by ref. [18] has employed a random parameters logit model with interaction terms to analyze roadway crashes on mountainous freeways with high proportion of heavy vehicles. Vehicle characteristics, traffic volume, roadway geometry, crash timing, and environment factors were identified as significant contributors to injury severity, and several interaction effects were found to improve the explanatory ability of the model. The study also showed that taking into account heterogeneity and interactions between risk factors gives a better understanding of crash mechanisms and helps to design more effective safety interventions.

Several safety studies have highlighted the impact of various factors on VRU injury outcomes to reduce serious injuries [4,19–21]. A study in Saint Petersburg, Russia, employed ordered probit model analysis using data collected from 13,888 pedestrian–vehicle crashes. They identified poor lighting, adverse weather, inadequate infrastructure, and number of people involved in a crash as contributing to higher probabilities of severe or fatal injuries. The study further indicated that there are gender-related effects associated with the degree of injury experienced by pedestrians and drivers; specifically, males tend to experience greater degrees of injury than females [19]. Similarly, Ma et al. [22] reviewed data from 2614 intersection crashes in Cook County, IL, USA. They found that pedestrian age, vehicle type, first point of contact, and weather conditions significantly influence crash severity. Their findings also revealed that certain age groups were more vulnerable to injuries depending on factors such as the number of vehicles involved, the type of traffic, the presence of road medians, and the behavior of drivers. On the other hand, the study by Khaki and Mohammadnazar [23] found that the behavior of pedestrians at uncontrolled mid-block crossings can independently increase or decrease the risks, regardless of conventional environmental or demographic factors. Specifically, factors traditionally assumed to affect injury severity, such as vehicle speed, vehicle type, and pedestrian age and gender, were found to have little influence on pedestrians' crossing decisions. However, pedestrian studies remain heavily reliant on traditional statistical models that struggle to capture the nonlinear relationship where behavioral risks suddenly fall into severe outcomes under specific environmental conditions

Cyclist crashes also represent a significant safety concern, which emphasizes the importance of identifying the related factors influencing crash severity when developing effective preventive measures. One topic that has achieved major attention in the literature is the potential behavioral effect associated with using a bicycle helmet, particularly in relation to the risk compensation hypothesis [24,25]. This hypothesis suggests that individuals may adjust their behavior when they perceive themselves to be more protected. A systematic review [25] investigating the relationship between helmet use and risky cycling behavior evaluated more than 20 studies, including field observations, surveys, and controlled experiments. The review indicated that cyclists who wore helmets occasionally exhibited marginally riskier behavior. However, the overall evidence was insufficient and inconsistent, suggesting that the impact of helmets cannot be attributed solely to their physical protective function. The results, in contrast, suggest that the effects of helmets may interact with behavioral and contextual factors. A meta-analysis performed using over fifty studies revealed that helmet use significantly reduces the risk of severe head, traumatic brain, and facial injuries [24]. This protective effect was particularly evident in single-bicycle crashes. The study also found that external factors such as crash speed, vehicle type, and roadway characteristics influence the crash severity.

Research indicates that injury severity is different among VRUs (pedestrians and cyclists). Consequently, it would be inappropriate to include both categories into one category for analysis of safety. There are several studies [26–28] have demonstrated that police

records often underreport cyclist injuries, particularly those not involving motor vehicle crashes. In contrast, pedestrian crashes are more consistently recorded in crash databases. These results may provide a better understanding of pedestrian safety than underreporting cyclist crashes. They also observe that injury patterns for these two groups change with time. Studies indicated that cyclist crashes increased compared to pedestrian crashes during the COVID-19 pandemic. These changes are mostly explained by changes in exposure levels and travel behavior, for example, higher cycling activity during lockdown periods and less pedestrian traffic in urban areas. Additional analyses indicate that the characteristics and severity of injuries differ among VRUs. The results emphasized the importance of treating each group separately, particularly when assessing risk factors and developing safety treatments.

Other studies have examined pedestrians and cyclists within the same analytical framework to better understand how injury severity differs between these two groups [26,29,30]. Islam and Hossain [26] applied random parameter logit models to investigate injury severity in pedestrian–motor vehicle and bicycle–motor vehicle crashes in Alabama. In their analysis, they found some variables that had significant effects on the injury severity of pedestrians but not on that of cyclists, and vice versa. However, both groups found a few factors to be significant, like darkness and crashes occurring on low-speed roadways. Similarly, a study of national U.S. data on injury to child pedestrians and cyclists found clear differences in risk of injury and fatality when the two groups were examined in the same framework. The results showed that the influences of factors such as age, vehicle type, and crash circumstances on injury severity varied between pedestrians and cyclists. For example, while certain age groups resulted in pedestrians experiencing a higher rate of severe injuries and fatalities, cyclist injury outcomes were more strongly associated with crash location and vehicle speed [30]. A separate study [29] utilized other machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNNs) to assess the risk factors impacting VRUs. The findings indicate that Random Forest was the best classifier in all the cases and there were significant differences in factors related to injury severity among the VRU groups. The results also showed that motorcyclists are involved in more severe crashes than pedestrians and cyclists and, therefore, the different characteristics and risk mechanisms of each VRU category should be considered.

Road crashes have long been analyzed using traditional statistical approaches like the ordered logit model, ordered probit model, and multinomial logit model in traffic safety research [5,31–33]. However, real-world crash data often violates several restrictive statistical assumptions that these approaches rely on. For example, ordered models are based on the proportional odds assumption that explanatory factors have equal effects at all levels of injury severity. The assumption might not hold when comparing different VRU groups where the exposure conditions and risk mechanisms are different. This may result in potentially misleading conclusions about the effectiveness of safety interventions for each road user group. Furthermore, conventional statistical models are often limited in their capability to capture complex interactions of risk factors. In crash analysis, where fatal crashes occur far less frequently than minor injuries, they may also have trouble with highly unbalanced data [32]. Such limitations can lead to biased findings and misinterpretations of the crash data. This may thereby reduce the performance of these models in crash prediction, especially in the estimation of severe injuries for VRUs.

Previous safety studies have identified further limitations associated with conventional crash models, particularly their limited ability to account for unobserved heterogeneity [6,32,34]. This term refers to unaccounted for variables that affect the severity of injuries at each crash. Improperly dealing with this kind of variability can lead to biased or inconsistent estimates of parameters. The studies also highlight that conventional models

rely on strict statistical assumptions. One important assumption is that predictors have the same effect on all levels of injury severity. But this assumption may not fully consider the complexity of real crash conditions. In practice, the effect of risk factors is often different for different types of road users, roadway features, and surrounding conditions (e.g., weather conditions, time of the day, and traffic volume).

To overcome these limitations, recent safety studies have used a machine learning framework to better capture the complex relationships in crash data [29,35–37]. A study conducted by Al Hasanat et al. [35] employed cost-sensitive algorithms and special under-sampling methods to address the class imbalance in crash data. The authors applied Extreme Gradient Boosting (XGBoost) with a balanced loss function as well as Random Data Partitioning with Voting Rule (RDPVR). These techniques enhanced the predictive power of the crash models in terms of learning from the underrepresented severe and fatal crash cases while maintaining the integrity of the original data structure. In addition, mixed correlation analysis and statistical tests have revealed several significant factors influencing injury severity, including vehicle speed, road illumination, vehicle type, driver age, and road conditions. Overall, the implemented methods achieved significantly higher true positive rates for severe injury compared with conventional methods. Amiri et al. [38] studied the performance of different machine learning models for crash severity prediction, tackling the common problem of class imbalance in crash data. The machine learning models used by the authors were Random Forest, XGBoost and Support Vector Machine. The results showed that combining machine learning models with suitable data balancing techniques improved the prediction of severe crash outcomes compared with conventional modeling approaches.

However, Alobidan et al. [39] in a global systematic review of AI-based approaches for traffic crash severity analysis reported that ensemble and hybrid learning techniques generally outperform traditional statistical and ML methods. The review also identified important challenges regarding class imbalance, model validation, reproducibility, and explainability, thus emphasizing the importance of interpretable frameworks like SHAP to support trustworthy and transparent transportation safety applications. These observations further prompt the integration of explainable ML methods to the analysis of VRU injury severity. Another study [40] compared the performance of advanced ordinal ensemble models such as Ordered XGBoost, LightGBM and Random Forest with traditional probit and logit regression models. The analysis revealed that ensemble methods were significantly better than traditional methods in predicting the perceived safety levels.

Tree-based machine learning models, including XGBoost, Gradient Boosting, Random Forest and CatBoost, are used in different roadway environments to predict crash severity [37,38,41]. These models can automatically select important features and generate interpretable results. This capability makes them particularly valuable for guiding transportation safety policies through identifying high-risk locations and supporting targeted interventions to improve road safety. They can also handle missing crash data and identification of how each of the predictor variables contributes to the predicted outcomes by using measures such as feature importance. Tree-based machine learning models have also been used beyond crash severity analysis to more general VRU related safety issues. A recent study explored VRU perceptions of sharing the road with emerging automated driving technologies, employing XGBoost and Random Forest algorithms to identify factors influencing public acceptance. The results show that the most dominant determinants were the concerns on traffic injuries and fatalities, perceived safety in mixed-traffic environments, awareness of major crash incidents involving automated systems and operating speed limits [42].

In addition to the development of machine learning models, Explainable Artificial Intelligence (XAI) techniques have become increasingly important for interpreting com-

plex “black-box” models in transportation safety research. One of the most widely used methods is Shapley Additive Explanations (SHAP), which offers a consistent and theoretically grounded method for interpreting machine learning results [36,43]. Using the principles of cooperative game theory, SHAP calculates an importance value for each input variable and illustrates its increases or decreases in the predicted result. SHAP explains model predictions by attributing the additive contributions of each feature allowing researchers to study the overall model behavior and individual predictions. This ability improves the transparency of the model and allows for the identification of important variables and complex interactions affecting the crash outcomes. As a result, SHAP makes advanced predictive models more understandable and practically useful for transportation safety analysis [43].

In recent years, SHAP has been widely used to interpret models of injury severity developed using algorithms such as Random Forest, Decision Tree, XGBoost and CatBoost [37,44]. The researchers, through the use of SHAP-based analysis, have quantified the magnitude and direction of influencing factors, such as vehicle speed, roadway characteristics, lighting conditions, weather, crash location, and driver demographics. In addition, SHAP visualizations provide clear insights into how these variables interact and contribute to severe or fatal crash outcomes. The recent study of Budzyński and Cieśla [45] based on XAI has further demonstrated the applicability of XAI techniques for cyclist safety analysis. In the study on bicycle–vehicle crashes in Poland, the CatBoost model in combination with SHAP was used to identify that cyclist age, driver experience, lighting conditions, and roadway context were significant factors related to the injury severity. The results showed that explainable ML frameworks can accurately describe nonlinear relationships and provide clear explanations of the mechanisms that lead to severe cyclist injuries, thus enabling evidence-based road safety interventions and sustainable mobility policies. Despite the SHAP capability, the majority of literature commonly uses SHAP to plot marginal main effects, while there is a distinct lack of studies leveraging SHAP’s interaction values to compare how risk mechanisms shift between pedestrians and cyclists within a unified modeling space.

Overall, the existing literature indicates that injury severity among VRUs results from complex interactions among multiple risk factors. Conventional statistical models frequently fail to represent these relationships comprehensively, particularly when predicting severe and fatal crashes. In contrast, more recent approaches combine advanced modeling techniques with interpretable frameworks to better analyze crash severity. Methods such as the Partial Proportional Odds Model and tree-based machine learning algorithms like Random Forest, XGBoost, and CatBoost have shown improved predictive performance. In addition, explainability tools such as SHAP help stakeholders better understand the key factors influencing crash injury severity.

3. Descriptive Statistics

The crash severity of vulnerable users, specifically pedestrians and cyclists, was analyzed based on the individual level of five years of crash data that occurred in Michigan, USA, between 2016 and 2020 which obtained from the Michigan Department Of Transportation MDOT. From an initial sample of more than 19,000 VRU crashes, records were strictly filtered to exclude non-motor-vehicle incidents, private property crashes, and records with unclassifiable injury severity. Moreover, the categorical variables were retained as a “Missing” category to leverage CatBoost’s native handling, while continuous variables were imputed using the median from the training folds to prevent data leakage for missing data in explanatory variables. As a result, the final analytical sample that was conducted in the crash severity analysis for VRUs is consists of 14,898, crashes. The response variable of crash severity was classified into two primary severity categories for

each vulnerable user group, including severe crash, which includes fatal and serious injuries, and non-severe crash, which includes minor injury, possible injury, and no injury. An examination of the severity response variable indicated a significant class imbalance across the full dataset. The data showed that non-severe crashes accounted for 80.86% of the sample ($N = 12,047$), while severe crashes comprised only 19.14% ($N = 2851$). Regarding the subset groups used for the binary classification framework, this imbalance was distinctly present: pedestrian severe crashes accounted for 14.68% of all pedestrian cases ($N = 2187$), while cyclist severe crashes represented an even smaller proportion at 4.53% of all cyclist cases ($N = 675$). Table 1 illustrates severe and non-severe categories for each group of vulnerable users.

Table 1. Classification of the combined outcome variable.

Road User Type	Combined Category
Pedestrians	Pedestrians—Non-severe Injury
	Pedestrians—Severe Injury
Cyclists	Bicyclists—Non-severe Injury
	Bicyclists—Severe Injury

Modeling Strategy

To comprehensively analyze VRU crash severity, this study employed a two-tiered modeling strategy utilizing both multiclass and binary classification frameworks:

- **Binary Framework:** A subset of binary classification models (Severe vs. Non-Severe) were trained separately on the pedestrian-only subset and the cyclist-only subset to capture the individualized risk for each road user group without the potential impact caused by mixing populations. This approach provides focused SHAP summary plots that clearly delineate the distinct environmental and behavioral determinants of severity for pedestrians and cyclists independently.
- **Joint Multiclass Framework:** The analytical framework utilized a combined 4-class-dependent variable based on vulnerable road user type and the category of severity (Table 1). This framework provides direct cross-group comparisons by modeling pedestrians and cyclists simultaneously. Also, it captures SHAP interaction effects that identify how specific factors shift crash outcomes, not just from non-severe to severe but between user groups such as pedestrian severe vs. cyclist severe. This addresses the study's objective of capturing shared versus unique risk factors.

Table 2 summarizes the selected variables for vulnerable users' crash circumstances to analyze injury severity. The factors include spatial, temporal, environmental, and user-related characteristics attributed to each case. The table provides a clarification for variable distribution in the dataset and the ability to capture crash patterns that influence the likelihood of injury type occurrence for both pedestrians and cyclists in a combined modeling framework.

Table 2. Descriptive statistics of the explanatory variables used in the injury severity analysis of pedestrians and bicyclists.

Variable	Count	Mean	Std. Dev	Min	Max
Day of week					
Weekend	3233	0.217	0.412	0	1
Speed limit	14,898	33.643	11.7	5	75
Alcohol Involved					
Yes	1207	0.081	0.273	0	1
Drug Involved					
Yes	328	0.022	0.146	0	1

Truck Involved					
Yes	268	0.018	0.133	0	1
Emergency vehicle involved					
Yes	104	0.007	0.084	0	1
Distracted Driver Involved					
Yes	596	0.040	0.196	0	1
Number of motor vehicles involved					
Single	14,342	0.9630	0.190	0	1
Multiple	557	0.037	0.190	0	1
Number of non-motor vehicles involved					
Single	14,474	0.971	0.166	0	1
Multiple	425	0.029	0.166	0	1
Rural/Urban					
Rural	1177	0.079	0.270	0	1
Sub urban	3352	0.225	0.418	0	1
Urban	10,369	0.696	0.460	0	1
Intersection Involved					
Yes	8015	0.538	0.499	0	1
Hazardous Action					
At Fault	6868	0.461	0.498	0	1
Crash type					
Single	10,622	0.713	0.452	0	1
Head-on	3903	0.262	0.444	0	1
Angle	60	0.004	0.060	0	1
Rear	104	0.007	0.085	0	1
Sideswipe	89	0.006	0.075	0	1
Backing	15	0.001	0.038	0	1
Age					
Under 18	3427	0.230	0.421	0	1
18–24	2413	0.162	0.369	0	1
25–44	3963	0.266	0.442	0	1
45–64	3823	0.257	0.437	0	1
65 and more	1251	0.084	0.278	0	1
Gender					
Male	10,235	0.687	0.463	0	1
Vehicle type					
Passenger Car	10,816	0.726	0.446	0	1
Light Condition					
Daylight	9416	0.632	0.482	0	1
Dark	4752	0.319	0.466	0	1
Other	730	0.049	0.216	0	1
Weather Condition					
Clear	10,369	0.696	0.460	0	1
Traffic Control					
Signal	4648	0.312	0.463	0	1
Stop Sign	2667	0.179	0.383	0	1
Yield Sign	134	0.009	0.095	0	1
None	7449	0.500	0.500	0	1
Road Condition					
Dry	2875	0.193	0.395	0	1
Lane Number					
1	1117	0.075	0.264	0	1
2	6868	0.461	0.499	0	1
3 or More	6898	0.463	0.499	0	1
Prior Action					
Going straight ahead	3814	0.256	0.436	0	1
Crossing at the intersection	4246	0.285	0.451	0	1
Crossing at not allowed crossing area	2160	0.145	0.352	0	1
Other	4678	0.314	0.464	0	1

Data distribution indicates that most crashes occurred in urban areas (urban and sub-urban), which consist of 92.1%, while 7.9% occurred in rural areas. In particular, the data shows approximately 53.8% of crashes took place at intersections. Regarding driving under the influence, alcohol and drugs were responsible for 8.1% and 2.2% of crashes, respectively. Furthermore, distracted driving accounted for 4% of all crashes. In terms of environmental factors, the majority of crashes took place in daylight (63.2%) and clear weather (96.6%), which provides a preliminary understanding that visibility issues are less prevalent compared to other factors. The data also indicates that passenger cars dominate vulnerable users' crashes with a presence of 72.6% of total cases, while the average posted speed for all selected cases was approximately 33 mph. The demographic characteristics show a clear skew toward male vulnerable users with 68.7% of the sample. For age distribution, there is a notable balance in the middle age of vulnerable users, at approximately 26% even though a substantial percentage (23%) were under 18 years. However, pre-crash behavior also had a significant presence in crash circumstances. Crossing at the intersection area shows the most frequent prior crash action (28.5%), followed by going straight ahead (26%). Furthermore, it should be noticed that about 46.1% of incidents were assigned as the user's responsibility, which indicates the complexity of assignment in crash liability. In this dataset, the 'Hazardous Action' variable specifically indicates whether the vulnerable road user was reported by the responding police officer as committing a hazardous action such as failing to yield, disregarding a traffic signal, or crossing outside of a designated crosswalk or being formally determined to be at fault for the crash. This highlights the complexity of liability assignment in VRU crashes.

Overall, these statistics suggest that pedestrian and bicyclist crashes are concentrated in urban areas and at intersections, often occurring under favorable lighting and weather conditions. Behavioral factors such as alcohol involvement and hit-and-run incidents emerge as important elements potentially associated with a higher likelihood of severe injuries.

4. Model Description

To capture the influence of contributory features on vulnerable users' injury severity, this research utilized the CatBoost model. The boosting approaches improve prediction performance by combining several classifiers in one classifier model, such as CatBoost. Mainly, the boosting algorithms iteratively minimize a loss function by fitting new models to the gradients of previous errors [46–48]. The boosting process starts with a decision tree as a base learner and iteratively refines the model by fitting subsequent trees to minimize prediction errors. Specifically, CatBoost, as a boosting machine learning model developed by ref. [49], is defined to effectively handle both numerical and categorical features during the training period. Traffic data typically contains a mix of categorical and numerical features. Traditional ensemble models, such as XGBoost and LightGBM, require converting the categorical variables into numerical representations to capture the association with response variables, which may lead to increased dimensionality and potential information loss. In contrast, the CatBoost approach provides the chance to process the categorical features to be numerical based on the average target value for a specific category by conducting Ordered Target Statistics instead of simple encoding. Randomly, CatBoost implements ordered boosting by calculating the statistics for an occurrence depending on the preceding historical data rather than the full dataset. The following equation illustrates the mathematical expression for CatBoost prediction.

$$\hat{x}_k^i = \frac{\sum_{j:\sigma(j)<\sigma(k)} 1_{x_j^i=x_k^i} \cdot y_j + a \cdot p}{\sum_{j:\sigma(j)<\sigma(k)} 1_{x_j^i=x_k^i} + a}$$

where σ is a random permutation, x_k^i represents a specific feature value, y_j is the target value (crash severity) for previous record, a is a prior strength, and p is the prior that refers to the average crash severity across the entire dataset used to smooth the estimate for rare categories. Moreover, SHAP (SHapley Additive exPlanations) is an interpretable machine learning technique that illustrates how much each feature contributes to a model output [13]. The technique decomposes each model's prediction into the additive contribution of each feature relative to some baseline value, enabling both global and local interpretability [46]. This technique ensures consistency and locality in the explanations of model output, establishing SHAP as a robust framework for interpretation [13]. The additive SHAP technique was applied to quantify the contribution of each feature to the prediction of crash severity.

$$\phi_i(f, x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (|N| - |S| - 1)!}{|N|!} [f_x(S \cup \{i\}) - f_x(S)]$$

And the additive property:

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i$$

$f(x)$ is the predicted crash severity for a specific individual crash event x , ϕ_0 indicates the baseline for crash severity based on the average crash severity prediction for entire training dataset, ϕ_i is the impact of single feature on the response variable probability, $f_x(S)$ refers to the predicted crash severity if it only knew the factors in subset S , $[f_x(S \cup \{i\}) - f_x(S)]$ denotes an expression to capture the change (marginal effect) in the response crash severity when the unexpectedly specific feature i is changed. Due to its capability to natively process categorical variables and superiority over other boosting models by minimizing overfitting through ordered boosting. This makes CatBoost a reliable and efficient method for enhancing prediction accuracy and capturing complex nonlinear relationships in crash data [46,49].

Four different classification algorithms were utilized for comparison: Logistic Regression (LR), Random Forest (RF), LightGBM, and CatBoost. LR implemented as the traditional statistical baseline, while RF (300 trees) and LightGBM (learning rate = 0.05, 300 estimators) provided alternative ensemble and boosting architectures. CatBoost was configured with a depth of 5 and 500 iterations, leveraging its native ability to process categorical features without manual encoding. The Synthetic Minority Over-sampling Technique (SMOTE) was employed to generate synthetic instances of the minority (severe) class. To prevent data leakage and ensure realistic evaluation, SMOTE was applied exclusively to the training data within each fold of the 10-fold stratified cross-validation process. Although SMOTE is mostly developed for continuous numerical fields, the categorical variables were label encoded, so the process takes into account the whole feature matrix. Importantly, those encoded columns were then explicitly flagged as categorical features inside the CatBoost model. So, during the tree construction, CatBoost ended up using its own Ordered Target Statistics mechanism, which leans on category frequency counting rather than any numerical Euclidean distance. As a result, it essentially eliminates the interpolative that SMOTE could have introduced into the categorical data.

5. Results and Discussion

In addition to examining which variables are associated with crash severity for vulnerable road users (VRUs), this study utilized a machine learning approach known as ensemble machine learning utilizing CatBoost to compare the differences in crash severity at various levels. Namely, comparing crashes involving pedestrians versus cyclists as well

as comparing crashes that resulted in a high level of injury/property damage versus those that did not.

To ensure the reliability of the predictive performance and to evaluate the stability of the models under different data partitions, a 10-fold stratified cross-validation was utilized. Table 3 summarizes the average performance metrics across both binary classification frameworks (pedestrian and cyclist subsets). According to the illustrated results, the CatBoost model achieved the highest mean AUC-ROC (0.917) and the highest F1-Score (0.812) among all tested models, confirming its superior capability in distinguishing between severe and non-severe crashes. The POC curve in Figure 1 further shows that CatBoost has better discriminative performance than the other models assessed, due to a large area under the curve. Additionally, CatBoost demonstrated exceptional stability across the different data partitions and associated with low standard deviation of ± 0.006 for AUC-ROC. Based on F1-Score results, LightGBM indicated a marginally higher recall for severe crashes (0.747 vs. 0.734) while CatBoost maintained a better balance between precision and recall. The low variance across all metrics confirms that the CatBoost model's predictive capability is robust, highly reliable, and not an artifact of a specific random data partition. Therefore, due to its ability to rank prediction probabilities accurately (AUC-ROC), superiority in balance between precision and recall, and capability of handling categorical variable without manual coding, CatBoost forms the most reliable models for SHAP analysis of vulnerable road users crash severity. This result is consistent with previous studies that showed the CatBoost model outperformed both traditional statistical models and competing gradient boosting algorithms in the transportation safety field. The literature indicates an enhancement in prediction performance ranging from 2% to 15% when using CatBoost, depending on the dataset and comparison models [50–52]. For instance, one study examined pedestrians' behavior to improve crash severity prediction by comparing several models [52]. The study applied the hunter-prey algorithm to optimize the hyperparameters of the CatBoost model. Findings illustrate an accuracy exceeding 96% when using CatBoost, along with improved capability in capturing complex feature interactions. As a result, these findings further support the ability of CatBoost to effectively distinguish between different severity levels and to capture complex nonlinear relationships within the dataset. Therefore, CatBoost was chosen as the final model for examining crash severity among VRUs.

Table 3. Performance comparison of classification models for severity prediction.

Model	AUC-ROC	F1-Score (Severe)	Recall (Severe)	Accuracy
Logistic Regression	0.856 $\pm$ 0.010	0.736 $\pm$ 0.011	0.774 $\pm$ 0.020	0.768 $\pm$ 0.009
Random Forest	0.893 $\pm$ 0.007	0.781 $\pm$ 0.006	0.739 $\pm$ 0.011	0.827 $\pm$ 0.005
LightGBM	0.914 $\pm$ 0.006	0.802 $\pm$ 0.007	0.747 $\pm$ 0.014	0.846 $\pm$ 0.004
CatBoost	0.917 $\pm$ 0.006	0.812 $\pm$ 0.011	0.734 $\pm$ 0.019	0.846 $\pm$ 0.006

The predictive capability of the joint modeling approach was also evaluated by utilizing 10-fold stratified cross-validation which was applied to the 4-class multiclass framework. Table 4 summarizes these performance metrics, indicating that the performance metrics for the 4-class problem are lower than those observed in the binary frameworks. This reduction is a common outcome in crash severity modeling is expected during the transition to multiclass structures, driven by increased decision boundary complexity and severe class imbalance particularly for the “Bicyclists—Severe Injury” category, which constitutes only 4.53% of the total observations and heavily penalizes the Macro F1-Score. However, the LightGBM exhibited a marginal, statistically negligible advantage over CatBoost in this specific 4-class task (difference of <1). The decision to retain CatBoost for SHAP analysis was driven by its superior performance in the binary frameworks and abil-

ity to handle categorical variables which provides more reliable SHAP feature splits as well as the need for methodological consistency. Most importantly, the primary objective of the 4-class CatBoost model in this study is not isolated point-prediction, but rather to establish a unified feature space required to accurately calculate SHAP interaction values across different VRU groups.

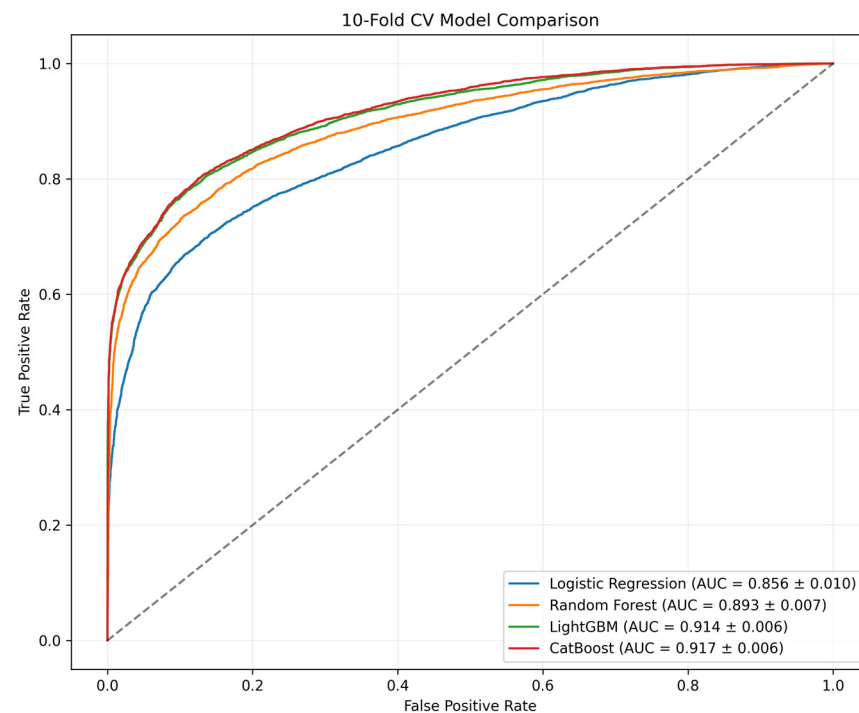


Figure 1. ROC curves for model performance comparison.

Table 4. Performance comparison of classification models for the 4-class joint multiclass.

Model	AUC-ROC	F1-Score (Severe)	Recall (Severe)	Accuracy
Logistic Regression	0.775 ± 0.013	0.446 ± 0.014	0.479 ± 0.018	0.527 ± 0.017
Random Forest	0.775 ± 0.009	0.472 ± 0.013	0.474 ± 0.013	0.614 ± 0.014
LightGBM	0.803 ± 0.011	0.503 ± 0.015	0.502 ± 0.014	0.647 ± 0.017
CatBoost	0.806 ± 0.012	0.502 ± 0.018	0.505 ± 0.016	0.640 ± 0.018

To better understand how different variables influence injury severity among VRUs, feature importance analysis was performed using interpretable machine learning techniques based on SHAP. Figures 2 and 3 present the ranking of variables according to their mean SHAP values, reflecting their relative contribution to the CatBoost model predictions.

The findings indicated that prior actions, such as crossing at intersections (whether authorized or not), are the most influential factor in explaining vulnerable users' severity, with a SHAP value of 0.30. This indicates that the actions taken by the parties involved play a critical role in determining crash severity. In addition, speed limit and hazardous action, which represent environmental conditions and user behavior (such as crossing behavior or responsibility in the crash), also exhibit high importance, with SHAP values of 0.25 and 0.24, respectively. These variables are associated with a higher predicted probability of severe crash outcomes among VRUs.

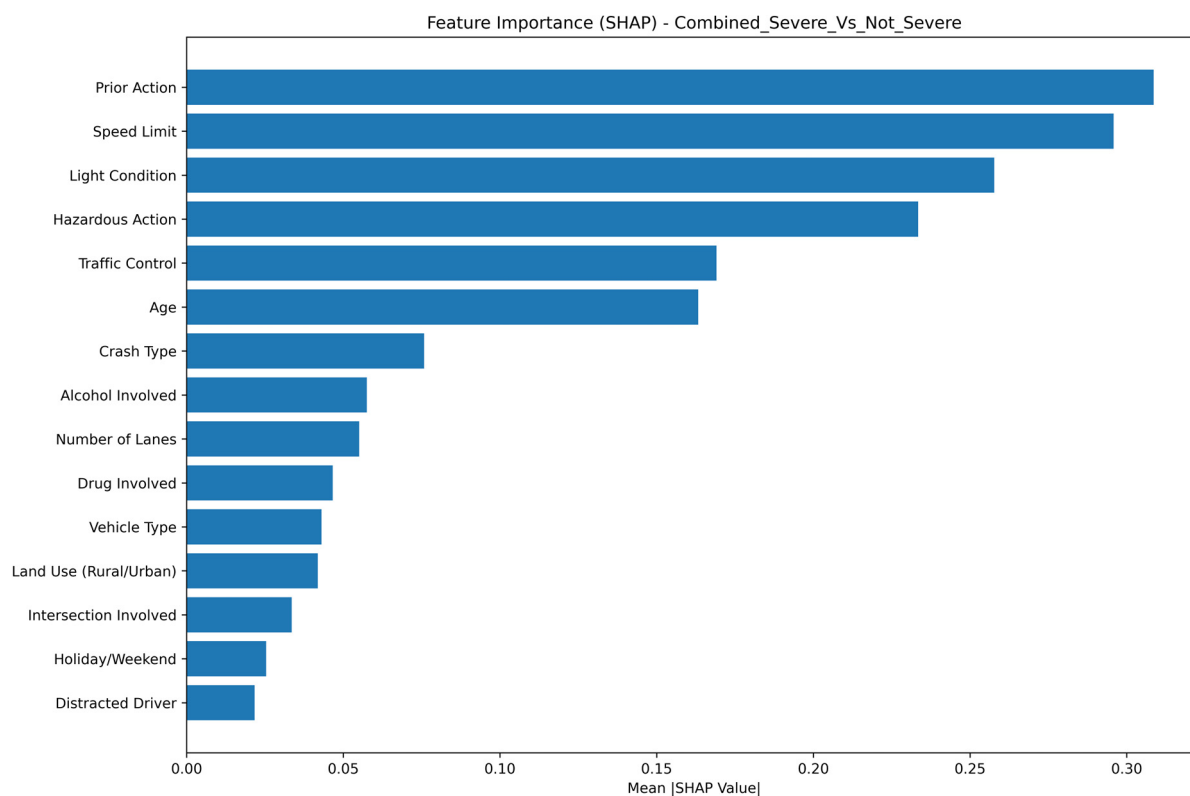


Figure 2. Feature importance based on SHAP values for pedestrian and bicycle injury severity.

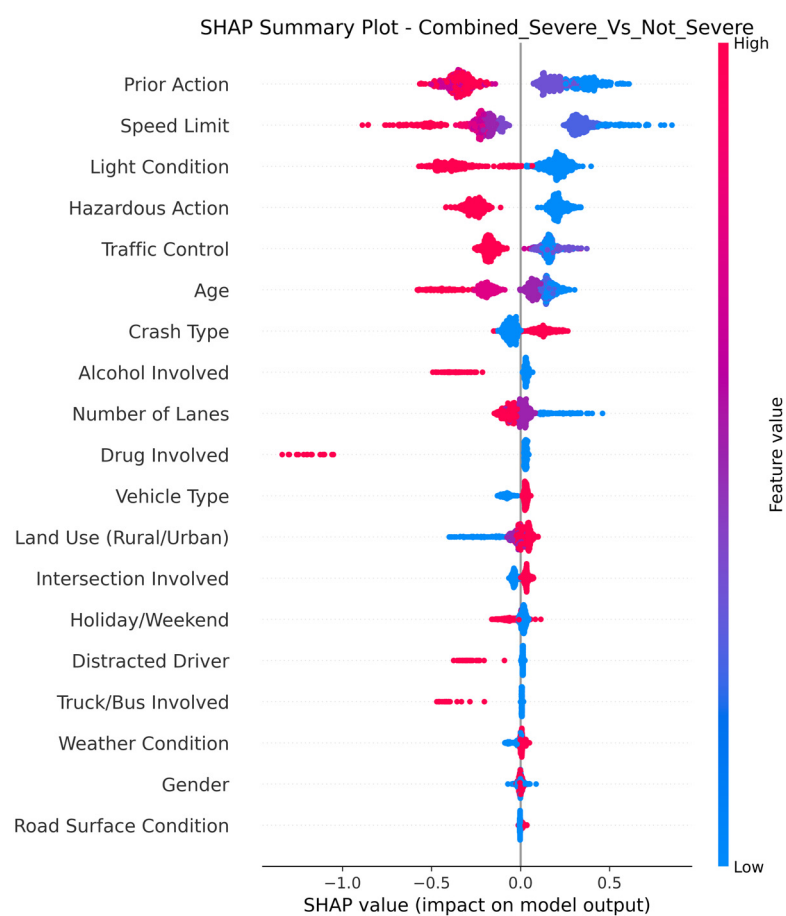


Figure 3. SHAP summary plot for pedestrian and bicycle injury severity prediction.

Moreover, the figure indicates the light and weather conditions as one of the top influential factors with a value of 0.17, which is related to the visibility of users. In terms of demographic characteristics, age was indicated as an important feature, ranking fifth with a value of 0.16. In addition, infrastructure and environmental factors, such as light condition and speed category, were found to play a role in explaining injury severity by 0.13 and 0.09, respectively. In contrast, factors like alcohol involvement, the number of lanes, and the type of vehicle had a smaller effect compared to the other main variables.

The SHAP analysis was also conducted at the within-subject level for each group of vulnerable users (pedestrians and cyclists). Results shown in Figure 4 illustrate the influence of factors on the prediction of pedestrian injury severity (severe vs. non-severe injury). The figure identifies lighting conditions, hazardous actions, and speed limits as the most important predictors of crash severity. The SHAP analysis indicates that higher values of reckless behavior are associated with higher predicted probabilities of severe pedestrian injury and those in high-speed zones due to the differences in kinetic energy and impact nature. This result can be correlated with pedestrian exposure conditions or infrastructure design, which can sometimes permit or even encourage unsafe crossing behaviors. These results are consistent with the previous literature that clearly states an increase in crash severity due to fault behavior from pedestrians when interacting with vehicles on the road, especially at intersecting areas [53,54]. For example, four years of crash data from North Carolina were analyzed to capture the direct and indirect effect of contributing factors on pedestrians crash severity [53]. The study utilized a path analysis by using an ordered logit model and a mixed logit model to identify factors impacting crash severity. The results indicated that pedestrians' behavior indirectly contributes to increased injury severity. Moreover, another study was conducted on signalized and unsignalized intersections to investigate and enhance pedestrian safety [54]. A machine learning algorithm, negative binomial, and order logit models were implemented on crowdsourced and mobile-based data to capture pedestrians' crash severity and frequency. The findings illustrated a consistent impact of pedestrians' behavior, such as interacting while impaired, as a key factor influencing both crash frequency and injury severity across different intersection types. Environmental factors were also identified as influential contributors, such as road condition, light condition, type of traffic control, and the number of lanes. The presence of a dry road surface is associated with lower injury severity for pedestrians compared with a slippery or icy road, because less surface friction can lead to loss of vehicle control, which is correlated with higher predicted crash severity [55]. This loss in control because of surface friction may contribute to higher impact speeds and the likelihood of fatal outcomes, especially for vital injuries such as head-striking injuries [56]. As expected, there is an increased probability of high-severity injuries under low-lighting conditions. Low visibility was identified as a major key factor contributing to crash severity, where the driver is unable to see vulnerable users in time; thus, the driver maintains their driving speed without caution towards pedestrians [57]. Therefore, the available recognition time will be reduced, which impacts the ability to practice appropriate braking or evasive maneuvers and may be associated with higher impact speeds and higher predicted injury severity [58]. On the other hand, the type of traffic signal control was associated with lower predicted severity with severe pedestrian injuries compared to other types, such as signs and uncontrolled intersections. This can be related to the power of traffic lights in regulating traffic and reducing the conflict and confusion between road users [59]. By providing clear right-of-way instructions, the traffic signal is linked to lower predicted severity, likely by regulating right-of-way and correlating with lower vehicle speeds at crossings compared with uncontrolled locations [60].

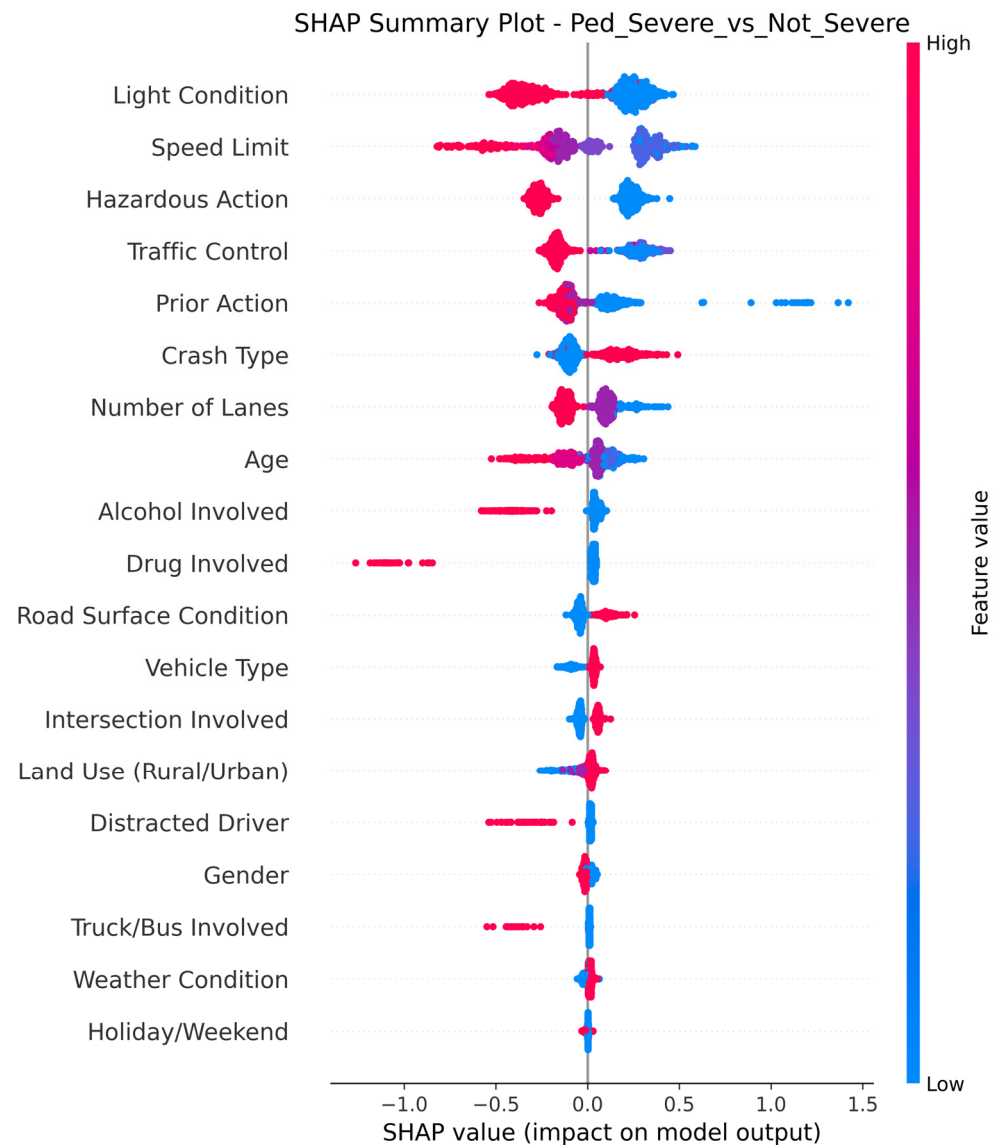


Figure 4. SHAP summary plot for pedestrian injury severity prediction.

Increasing the number of lanes showed a negative association with severe pedestrian injuries in the SHAP analysis. This unreasonable finding is highly associated with the confounding between lane numbers and the speed environment variables. One- and two-lane systems are mainly considered as rural roadways with high speed, whereas the three-lane system is characterized by urban roads with low speed, higher congestion, and frequent traffic signalization presence. Therefore, the finding is not claiming that adding lanes will reduce severity but rather that capturing the type of interaction environment where multi-lane roads exist is correlated with less severe pedestrian outcomes compared to high-speed rural roadways. Moreover, several studies indicated that increasing lane numbers may be associated with the openness of clear vision for drivers and provide them with a higher degree of freedom in roadway maneuvers [61]. This lateral space helps drivers to avoid severe crashes and reduce the impact force by providing space for avoidance, which causes less severe injury compared to narrow and restrictive roadways [62]. Moreover, regarding demographic variables, age showed a negative association with severe injury probability for elderly pedestrians. This association can be related to behavioral and environmental confounding effects in the crash dataset regardless of elderly pedestrians' physiological fragility, which typically increases injury risk. Elderly pedestrians are more likely to walk in dense, low-speed urban environments with shorter crossing distances

and lower vehicle speeds, which the model associates with non-severe outcomes. On the other hand, younger pedestrians may be overrepresented in higher-speed, suburban arterial environments where severe crashes occur. Thus, the model captures the protective effect of the environmental context typically associated with elderly pedestrian travel, rather than a direct mitigating effect of age itself [63,64]. As a conclusion, the SHAP analysis for pedestrian severity suggests that even though the behavioral factor impact is notable, the crash environment and speed-related variables provide the most consistent predictive power for identifying severe pedestrian crashes.

On the other hand, Figure 5 illustrates the SHAP summary plot for the cyclist crash severity (severe vs. not severe). The model results show that intersection involvement is the most important factor affecting cyclist injury, indicating a strong association within the model between cyclist crash severity and intersection involvement. High values on the scale (shown in red) are associated with higher predicted severity with severe outcomes. The SHAP analysis captured an inverse correlation with land use classification (rural, suburban, and urban). The figure suggests a noticeable shift toward the negative side of the SHAP scale when the crash occurred in urban settings. These results may correlate to the complexity and low speed of urban areas, where the kinetic energy can be mitigated compared to rural areas, thereby reducing injury severity [65]. Moreover, the existing literature indicates that using urban countermeasures, such as traffic calming, plays a key role in controlling vehicle speeds, which reduces the likelihood of the occurrence of severe and fatal injuries [66]. Moreover, variables such as crash type and the age of VRUs demonstrate a clear positive association with the probability of severe cyclist injuries. Side or angle crashes commonly result in higher impact energy to the cyclist's body, which causes critical injuries compared to other crash types, such as rear-end crashes. This effect is also influenced by age, where older cyclists are associated with physiological resilience and recovery capacity [67]. The number of lanes, conversely, shows a negative association, indicating that a roadway with a low number of lanes is correlated with severe crash injuries in the case of cyclist crashes. This result can be explained by the fact that narrow roadways force cyclists to operate closer to the vehicles, which limits the ability for drivers and cyclists to execute evasive maneuvers without risking a head-on crash with traffic [68].

However, while some factors, such as speed limit and vehicle type, have shown a moderate influence on severe cyclist injuries, other predictors, like gender, weather conditions, and lighting conditions, show relatively limited impact on cyclist model outcomes. This finding is unreasonable when compared to pedestrian vulnerability, but it is highly justified by the fundamental differences in user exposure and visual detection. Compared to pedestrians who rely on ambient infrastructure lighting due to their low profiles and static positions, cyclists indicate continuous lateral movements and typically travel at higher speeds [4]. This makes cyclists more visible to drivers regardless of the lighting conditions as well as their ability to adopt and utilize personal safety equipment such as reflectors and headlamps that mitigate the impact of poor lighting conditions. Therefore, the model correctly identifies that a cyclist's spatial dynamics and active lighting reduce their reliance on environmental weather and light conditions compared to pedestrians.

Figure 6 shows the top 10 feature interactions captured by SHAP analysis that determine the degree of injury to those involved in a crash. These feature interactions can be found in each of four types of crashes: Ped_Severe, Ped_Non_Severe, Cyc_Severe, and Cyc_Non_Severe. The feature interactions indicate the primary role speed limits have in determining crash severity, where speed emerged as one of the most influential features in the model in the most high-impact factor combinations [59]. For example, the interaction between speed and hazardous action, particularly when the VRU is at fault, illustrates that the predictive influence of speed is magnified when interacting with risky behavior

variables, suggesting that the combined presence of these factors is a strong indicator of severe crash risk [69]. Moreover, the results also observed that the model assigned higher predicted probabilities of severe injury to cyclists when they engaged in high-risk behaviors, such as crossing an intersection regardless of permission or not, and traveled at higher speeds than others [57,70].

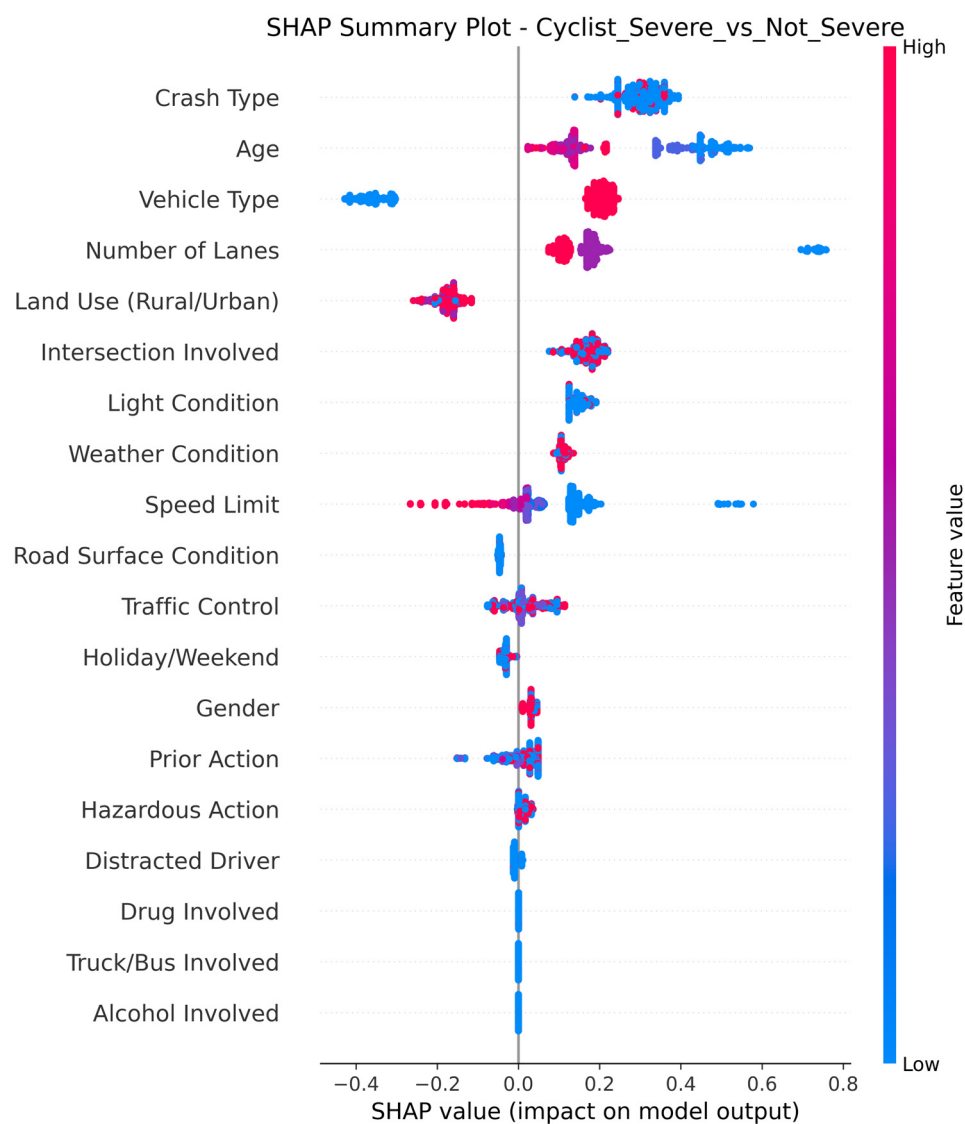


Figure 5. SHAP summary plot for predicting bicycle crash severity.

Infrastructure-related interactions also show notable effects. The results indicated that the predicted effect of speed on severe injury probability is notably higher in the presence of poor lighting and the location had a higher level of intersection control, such as traffic signals. This suggests that under the low-visibility intersection conditions, the benefits of traffic control can be compromised, especially in high-speed locations where the drivers have to control their time to perceive the vulnerable users [68]. In addition, several other important interactions were observed, including speed limit $\times$ gender, speed limit $\times$ number of lanes, and hazardous action $\times$ prior action. These findings indicate that the role of speed is not uniform; instead, it is highly correlated with geometric characteristics of the road and users' behavior [29]. These results concluded that while crash severity among VRUs may appear to be determined by a single factor, it is actually influenced by many different combinations of behavioral, environmental, and traffic-related feature interac-

tions. This confirmed the necessity for advanced models capable of capturing such complex dependencies and the multifaceted nature of crash causation.

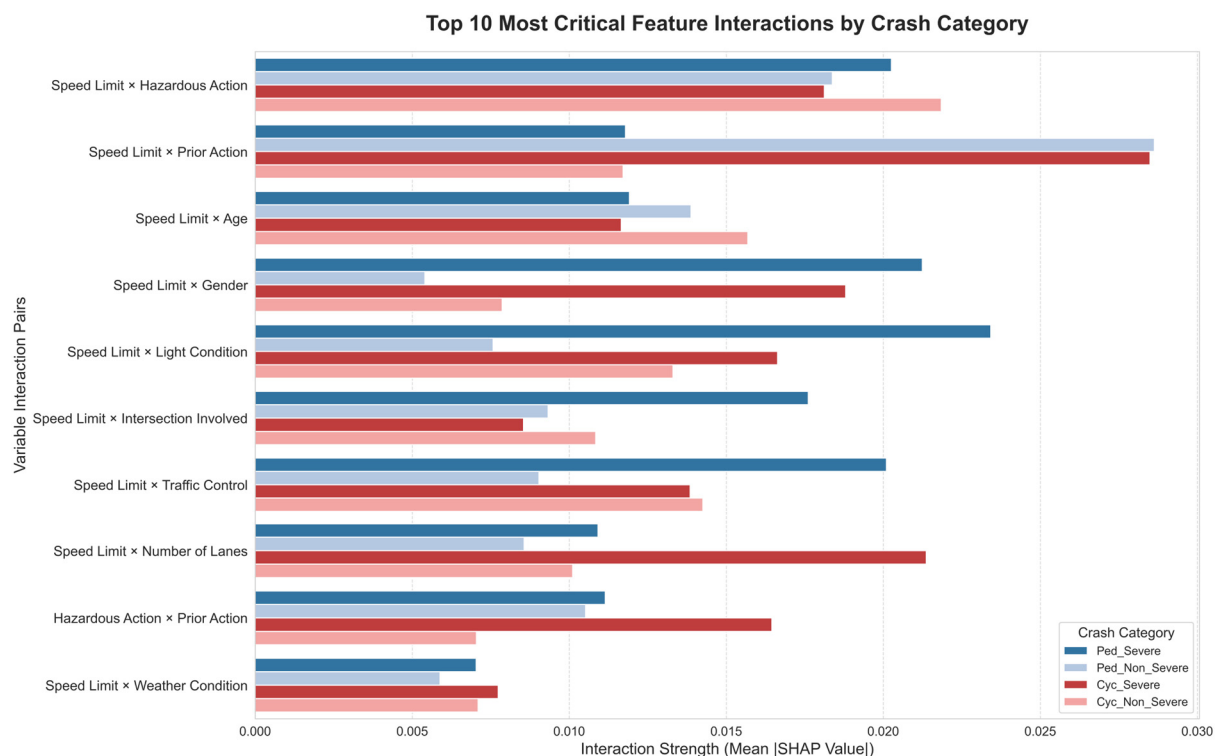


Figure 6. Top SHAP interactions affecting vulnerable users' crash severity.

In addition, a multiclass analysis was conducted using the CatBoost machine learning technique to enable cross-class comparisons of injury severity among VRUs. Overall, the results indicate that prior crash action is the most influential factor across all classes, indicating that the model suggests that injury severity predictions are strongly influenced by prior crash behavior and driven by specific risk-behavior rather than random interactions of features [69]. The correlation between prior action and injury severity varies among user groups. For pedestrians, the results suggest a negative association between low-value action categories, such as going straight ahead, and the probability of severe injury. This suggests that driver cognitive workload can be minimized by counting on predictability and path expectations, which reflect confusion and reduce the occurrence of severe injuries [71]. In contrast, cyclist crashes show a positive association between high-risk prior actions, such as crossing at intersections, and the probability of severe crashes. This result highlights the vulnerability of cyclists to lateral impact and driver expectations [72]. Furthermore, while non-severe crash causes are affected by several factors with small individual impact, severe injury crashes were more frequently associated with a distinct subset of dangerous user behaviors. This identifies an opportunity for focused interventions toward risky driver behavior to enhance overall safety results for VRUs.

Environmental risk factors play an important role in the differentiation between pedestrian and bicycle crash outcomes. Light condition factors were identified as a strong predictor for pedestrian crash injuries compared to cyclist crashes. This indicates that vulnerable pedestrians are more reliant on visibility when interacting with the traffic. Moreover, pedestrians with low profiles and slow speeds can be disproportionately affected by driver visual detection, unlike cyclists who can benefit from parallax motion [72]. Roadway characteristics also showed a significant influence on crash severity. However, the severe cyclist injury category was identified by only four features, which suggest that severe injuries are driven by specific high-risk circumstances, such as truck involvement.

Figure 7 also demonstrates that absolute SHAP values were significantly greater for severe crash classes. This indicates that severe outcomes are associated with more extreme factor values and stronger predictive relationships. Driver-related factors such as alcohol and drug involvement were more effective in predicting pedestrian crashes than cyclist crashes. Furthermore, age was found to have a stronger positive correlation with the severity of pedestrian injuries than with cyclist injuries; this finding has been previously supported through other studies regarding pedestrian vulnerabilities.

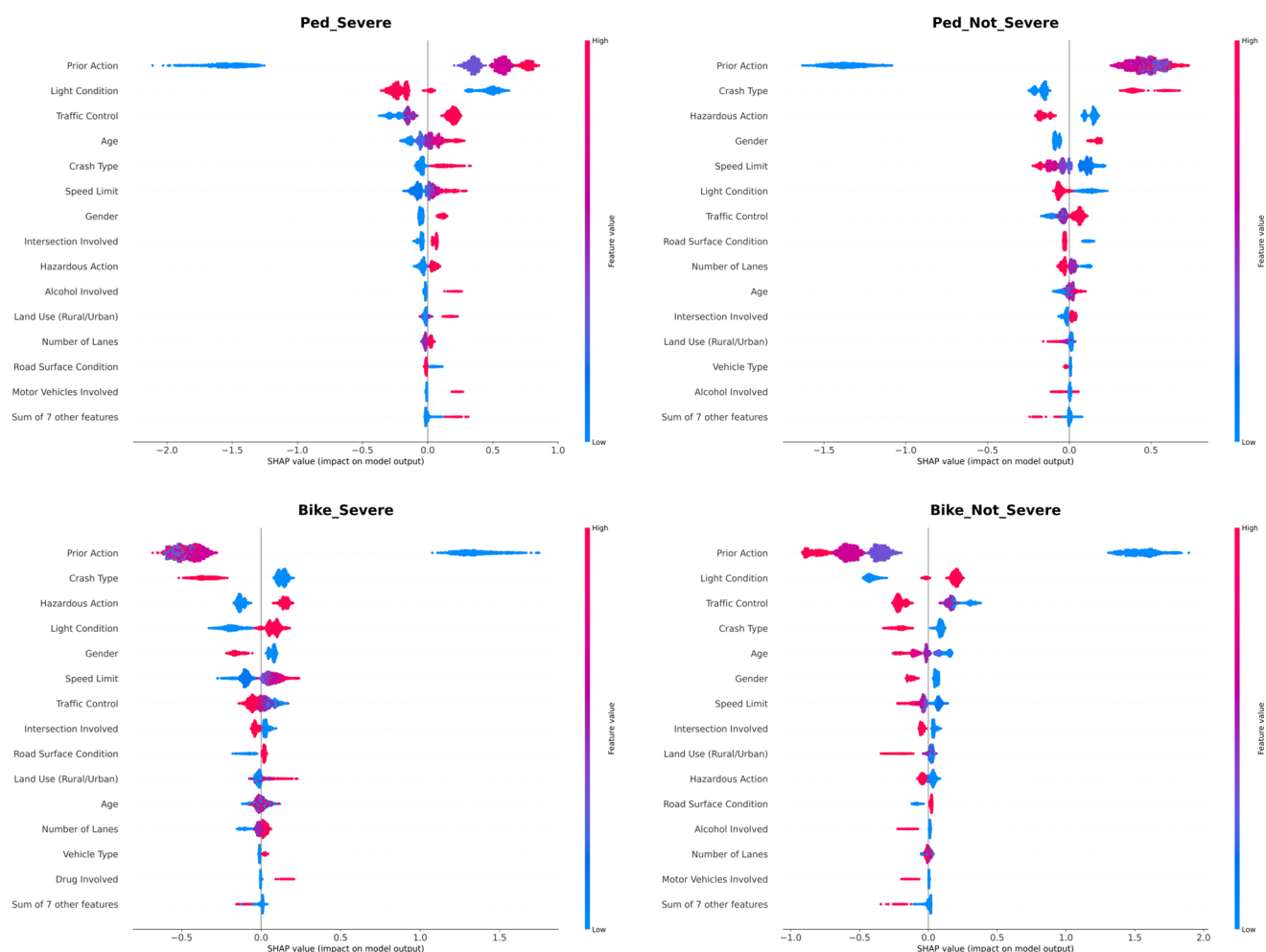


Figure 7. Feature contribution estimates across crash severity outcomes for pedestrians and cyclists.

The SHAP interaction results and summary plots reveal critical differences between pedestrians and cyclists that directly inform targeted countermeasures. A strong positive interaction between speed limits and poor lighting conditions on the severity for pedestrians. This suggests that passive visibility treatments such as street lighting enhancements are insufficient alone if speed limits remain high. Therefore, a combined engineering intervention for pedestrian corridors should be associated with sufficient illuminated pedestrian crossing systems and physical traffic calming measures such as curb extensions and raised crosswalks that force speed reduction at conflict points, rather than relying solely on visual cues.

Conversely, the finding indicated that cyclist severity outcome is significantly affected by infrastructure like intersection involvement and low lane numbers (narrow roadways), illustrating a spatial constraint issue. Compared to pedestrians, cyclists possess parallax motion and higher speeds, making them more vulnerable to lateral impact angles at intersections. Therefore, operational and engineering policies and countermeasures

ures should prioritize protected intersections by providing safety features such as corner refuge islands, which physically separate cyclists from turning vehicles and mitigate the severity of angle crashes identified in the SHAP analysis. Furthermore, the data suggesting severe cyclist injuries are highly concentrated in specific high-risk scenarios such as truck involvement, indicating that policy interventions could yield disproportionate safety benefits for cyclists.

Finally, the key importance of vulnerable road users actions across both user groups emphasizes a limitation of purely engineering-based solutions. According to study findings, user behavior like risky crossing increases the danger of environmental factors (speed/lighting). These findings support the implementation of users behavioral counter-measures such as pedestrian scramble phases at complex signalized intersections and public awareness campaigns that address the combined risk of behavioral violations in high-speed or low-light environments.

6. Conclusions

This study proposed an analytical framework to investigate the crash injury severity of vulnerable road users (VRUs) by utilizing a CatBoost ensemble boosting machine model integrated with Explainable Artificial Intelligence (SHAP). The model successfully captured the complex and nonlinear interaction between features dominating traditional statistical and other ensemble methods.

Overall, the findings emphasize that crash severity is not determined by single factors but is rather the result of sophisticated interactions that combine human behavior, environmental condition, and infrastructure. CatBoost outperformed all other models with respect to predicting crashes ($AUC = 0.921$) and can handle both large amounts of dimensional data and highly unbalanced crash data, especially as categorical variables. Additionally, the use of the SHAP additive explanation method provided evidence of how each variable contributed to the model's prediction, thus allowing for policy development and investments in transportation infrastructure to be made based on understandable evidence.

Behavioral and spatial factors played a significant role in injury severity prediction. Prior crash actions (especially those involving risky crossing behaviors) were identified as the most contributory factors in severity prediction for both vulnerable user groups. Moreover, the joint modeling framework highlighted important differences in vulnerability profiles for pedestrians and cyclists. For pedestrians, lighting conditions and speed limits were identified as critical features in severity prediction. The findings revealed that low visibility was associated with higher predicted probabilities of severe outcomes, while pedestrians' severe crash probability decreases as the number of lanes on the roadway increases. Conversely, there was a high association between intersection involvement and roadway geometry with higher predicted cyclist severity. For instance, cyclists are exposed to a higher probability of severe crashes on narrow roadways where the proximity to vehicles limits the ability to perform evasive maneuvers. Finally, urban environments were associated with a lower probability of severe injuries compared to rural environments due to lower speeds.

SHAP interaction analysis revealed a compounded effect of speed as a key risk factor. The results indicate that effect of speed increases significantly when combined with hazardous actions or poor lighting conditions. This finding shows that combining speed reduction strategies with behavioral interventions and visibility improvements will yield maximum effectiveness.

The results of this study can be utilized for practical implications by targeting specific user-safety interventions. The findings show a strong influence of prior crossing behaviors that emphasize the need to reduce the exposure time of pedestrians on the road by utilizing some practical interventions such as scramble phases at complex intersections and

raised crosswalks at unauthorized mid-block crossings. Additionally, the amplified risk caused by high speeds interacting with poor visibility indicates that standalone lighting improvements are inadequate where the illumination should be paired with physical traffic calming measures to force speed reduction at conflict points. Conversely, regarding the cyclist, a strong correlation between cyclist severity and intersection involvement and narrow lanes (characterized by the minimal impact of lighting) suggests shifting infrastructure resources away from visibility upgrades toward protected intersection designs such as corner refuge islands and dedicated physical separation. Finally, the analysis illustrates that lower-severity outcomes associated with dense urban areas suggest utilizing some interventions such as road diets and variable speed limits on high-speed and multilane suburban roads to replicate the low-speed protective conditions. Ultimately, pedestrian safety interventions should focus on the behavioral–lighting–speed nexus, whereas cyclist safety relies primarily on geometric separation.

Despite the valuable insights provided by this research, several limitations restricted and impacted the findings. Underreporting of specific minor crashes of VRUs, such as cyclists, that do not result in injury and rely on police reports can introduce a potential reporting bias. Moreover, exposure variables such as pedestrian and cyclist volumes or miles traveled are missing in the data, which can have a significant impact on the severity outcomes. Finally, as noted in the discussion, while SHAP values effectively quantify the predictive importance and directional associations of risk factors, they are derived from observational data and do not establish direct causal relationships between the variables and crash outcomes. Therefore, future research should integrate exposure data and causal inference frameworks to build upon these findings.

Author Contributions: Conceptualization, M.A., M.N.A.-M., R.A.-S., F.A. (Fadi Alhomaidat) and A.J.; methodology, M.A., M.N.A.-M., R.A.-S., F.A. (Fareh Abudawaba) and A.J.; software, M.A., M.N.A.-M., R.A.-S., F.A. (Fadi Alhomaidat) and A.J.; validation, M.A., M.N.A.-M., R.A.-S., F.A. (Fareh Abudawaba), F.A. (Fadi Alhomaidat) and A.J.; formal analysis, M.A., M.N.A.-M., R.A.-S., F.A. (Fareh Abudawaba) and A.J.; writing—original draft preparation, M.A., M.N.A.-M., R.A.-S., F.A. (Fadi Alhomaidat) and A.J.; writing—review and editing, M.A., M.N.A.-M., R.A.-S., F.A. (Fareh Abudawaba), F.A. (Fadi Alhomaidat) and A.J.; visualization, M.A., M.N.A.-M., R.A.-S., F.A. (Fareh Abudawaba), F.A. (Fadi Alhomaidat) and A.J.; supervision, M.A., M.N.A.-M., R.A.-S., F.A. (Fareh Abudawaba) and A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jaber, A.; Csonka, B. Towards a Sustainable and Safe Future: Mapping Bike Accidents in Urbanized Context. *Safety* **2023**, *9*, 60. <https://doi.org/10.3390/safety9030060>.
2. World Health Organization. *Global Status Report on Road Safety 2023*; World Health Organization: Geneva, Switzerland, 2023.
3. National Highway Traffic Safety Administration. *Overview of Motor Vehicle Traffic Crashes*; National Highway Traffic Safety Administration, U.S. Department of Transportation: Washington, DC, USA, 2023.
4. Eluru, N.; Bhat, C.R.; Hensher, D.A. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accid. Anal. Prev.* **2008**, *40*, 1033–1054. <https://doi.org/10.1016/j.aap.2007.11.010>.
5. Cappelli, G.; Nardoiani, S.; D’Apuzzo, M.; Nicolosi, V. Cyclist Safety in Complex Urban Environments: Infrastructure, Traffic Interactions, and Spatial Anomalies in Rome, Italy. *Urban Sci.* **2026**, *10*, 73. <https://doi.org/10.3390/urbansci10020073>.
6. Mannering, F.L.; Shankar, V.; Bhat, C.R. Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal. Methods Accid. Res.* **2016**, *11*, 1–16. <https://doi.org/10.1016/j.amar.2016.04.001>.

7. Alhadidi, T.I.; Al-Marafi, M.N.; Alazimi, A. Development of safety performance measures for different crashes severity at urban roundabouts. *Results Eng.* **2025**, *25*, 103680. <https://doi.org/10.1016/j.rineng.2024.103680>.
8. Abushattal, M.; Alhomaidat, F.; El-Yabroudi, M. Effects of intersection control types on driver yielding behavior to cyclists using mixed logit modeling. *Sci. Rep.* **2025**, *15*, 33928. <https://doi.org/10.1038/s41598-025-09801-6>.
9. Silva, P.B.; Andrade, M.; Ferreira, S. Machine learning applied to road safety modeling: A systematic literature review. *J. Traffic Transp. Eng. (Engl. Ed.)* **2020**, *7*, 775–790. <https://doi.org/10.1016/j.jtte.2020.07.004>.
10. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
11. Yumak, A.; Tercan, S.H.; Colak, U.C.; Ozcanan, S. A Machine Learning Approach to Identify High-Risk Road Segments and Accident Severity Patterns Based on Categorical Data. *Appl. Sci.* **2025**, *15*, 12824. <https://doi.org/10.3390/app152312824>.
12. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>.
13. Lundberg, S.M.; Allen, P.G.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. 2017. Available online: <https://github.com/slundberg/shap> (accessed on 1 February 2026).
14. Kim, J.-K.; Ulfarsson, G.F.; Shankar, V.N.; Mannering, F.L. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accid. Anal. Prev.* **2010**, *42*, 1751–1758. <https://doi.org/10.1016/j.aap.2010.04.016>.
15. Tulu, G.S.; Washington, S.; Haque, M.M.; King, M.J. Injury severity of pedestrians involved in road traffic crashes in Addis Ababa, Ethiopia. *J. Transp. Saf. Secur.* **2017**, *9*, 47–66. <https://doi.org/10.1080/19439962.2016.1199622>.
16. Zeng, Q.; Wang, Q.; Zhang, K.; Wong, S.C.; Xu, P. Analysis of the injury severity of motor vehicle–pedestrian crashes at urban intersections using spatiotemporal logistic regression models. *Accid. Anal. Prev.* **2023**, *189*, 107119. <https://doi.org/10.1016/j.aap.2023.107119>.
17. Junaid, M.; Jiang, C.; Alotaibi, S.; Wang, T.; Almarhab, Y. Investigating factors influencing injury severity in crashes involving vulnerable road users in Pakistan. *Sci. Rep.* **2025**, *15*, 32317. <https://doi.org/10.1038/s41598-025-16477-5>.
18. Zhai, X.; Huang, H.; Sze, N.N.; Song, Z.; Hon, K.K. Diagnostic analysis of the effects of weather condition on pedestrian crash severity. *Accid. Anal. Prev.* **2019**, *122*, 318–324. <https://doi.org/10.1016/j.aap.2018.10.017>.
19. Rodionova, M.; Skhvediani, A.; Kudryavtseva, T. Determinants of Pedestrian–Vehicle Crash Severity: Case of Saint Petersburg, Russia. *Int. J. Technol.* **2021**, *12*, 1427. <https://doi.org/10.14716/ijtech.v12i7.5403>.
20. Gutiérrez, M.; Ramos, R.; Soto, J.J.; Córdova, F. Factors influencing pedestrian injury severity in Chile: A hierarchical probit ordered model approach. *J. Saf. Res.* **2025**, *92*, 272–282. <https://doi.org/10.1016/j.jsr.2024.11.021>.
21. Jaber, A.; Csonka, B. Investigating the temporal differences among bike-sharing users through comparative analysis based on count, time series, and data mining models. *Alex. Eng. J.* **2023**, *77*, 1–13. <https://doi.org/10.1016/j.aej.2023.06.087>.
22. Ma, Z.; Lu, X.; Chien, S.I.-J.; Hu, D. Investigating factors influencing pedestrian injury severity at intersections. *Traffic Inj. Prev.* **2018**, *19*, 159–164. <https://doi.org/10.1080/15389588.2017.1354371>.
23. Khaki, A.M.; Mohammadnazar, A. A probabilistic model for pedestrian gap acceptance behavior at uncontrolled midblock crosswalks. *arXiv* **2021**, arXiv:2101.10903.
24. Høye, A. Bicycle helmets—To wear or not to wear? A meta-analyses of the effects of bicycle helmets on injuries. *Accid. Anal. Prev.* **2018**, *117*, 85–97. <https://doi.org/10.1016/j.aap.2018.03.026>.
25. Esmaeilikia, M.; Radun, I.; Grzebieta, R.; Olivier, J. Bicycle helmets and risky behaviour: A systematic review. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *60*, 299–310. <https://doi.org/10.1016/j.trf.2018.10.026>.
26. Islam, S.; Hossain, A.B. Comparative Analysis of Injury Severity Resulting from Pedestrian–Motor Vehicle and Bicycle–Motor Vehicle Crashes on Roadways in Alabama. *Transp. Res. Rec. J. Transp. Res. Board* **2015**, *2514*, 79–87. <https://doi.org/10.3141/2514-09>.
27. Wegman, F.; Zhang, F.; Dijkstra, A. How to make more cycling good for road safety? *Accid. Anal. Prev.* **2012**, *44*, 19–29. <https://doi.org/10.1016/j.aap.2010.11.010>.
28. Macpherson, A.K.; Zagorski, B.; Saskin, R.; Howard, A.W.; Harris, M.A.; Namin, S.; Rothman, L. Comparison of the number of pedestrian and cyclist injuries captured in police data compared with health service utilisation data in Toronto, Canada 2016–2021. *Inj. Prev.* **2024**, *30*, 161–166. <https://doi.org/10.1136/ip-2023-044974>.
29. Komol, M.M.R.; Hasan, M.M.; Elhenawy, M.; Yasmin, S.; Masoud, M.; Rakotonirainy, A. Crash severity analysis of vulnerable road users using machine learning. *PLoS ONE* **2021**, *16*, e0255828. <https://doi.org/10.1371/journal.pone.0255828>.
30. Swedler, D.I.; Ali, B.; Hoffman, R.; Leonardo, J.; Romano, E.; Miller, T.R. Injury and fatality risks for child pedestrians and cyclists on public roads. *Inj. Epidemiol.* **2024**, *11*, 15. <https://doi.org/10.1186/s40621-024-00497-2>.

31. Savolainen, P.T.; Mannering, F.L.; Lord, D.; Quddus, M.A. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accid. Anal. Prev.* **2011**, *43*, 1666–1676. <https://doi.org/10.1016/j.aap.2011.03.025>.
32. Ye, F.; Lord, D. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Anal. Methods Accid. Res.* **2014**, *1*, 72–85. <https://doi.org/10.1016/j.amar.2013.03.001>.
33. Alhomaidat, F.; Abushattal, M.; Kwayu, K.M.; Kwigizile, V. Investigating the interaction between age and liability for crashes at stop-sign-controlled intersections. *Transp. Res. Interdiscip. Perspect.* **2022**, *14*, 100612. <https://doi.org/10.1016/j.trip.2022.100612>.
34. Lord, D.; Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 291–305. <https://doi.org/10.1016/j.tra.2010.02.001>.
35. Al Hasanat, H.A.; Hassanat, A.B.; Alharasees, O.; Tarawneh, A.S.; Altarawneh, G.A.; Alhasanat, L.A.; Begović, M. Interpretable machine learning for imbalanced pedestrian injury severity prediction in urban Jordan. *J. King Saud Univ.–Eng. Sci.* **2025**, *37*, 39. <https://doi.org/10.1007/s44444-025-00053-3>.
36. Wang, C.; Serre, T. A Hybrid Approach to Investigating Factors Associated with Crash Injury Severity: Integrating Interpretable Machine Learning with Logit Model. *Appl. Sci.* **2025**, *15*, 10417. <https://doi.org/10.3390/app151910417>.
37. Champahom, T.; Se, C.; Watcharamaisakul, F.; Jomnonkwao, S.; Karoonsoontawong, A.; Ratanavaraha, V. Tree-based approaches to understanding factors influencing crash severity across roadway classes: A Thailand case study. *IATSS Res.* **2024**, *48*, 464–476. <https://doi.org/10.1016/j.iatssr.2024.09.001>.
38. Amiri, M.A.; Afshari, S.; Soltani, A. Machine learning approaches to traffic accident severity prediction: Addressing class imbalance. *Mach. Learn. Appl.* **2025**, *22*, 100792. <https://doi.org/10.1016/j.mlwa.2025.100792>.
39. Alobidan, Y.A.; Soh, B.; Li, A.; Almudayni, Z. Artificial intelligence for predicting road accident severity: A global systematic review. *Results Eng.* **2026**, *31*, 111289. <https://doi.org/10.1016/j.rineng.2026.111289>.
40. Maji, A.; Ghosh, I. What drives perceived safety concerns at roundabouts under disordered, heterogeneous traffic? An inquiry using explainable machine learning. *Can. J. Civ. Eng.* **2026**, *53*, 1–39. <https://doi.org/10.1139/cjce-2025-0538>.
41. Santos, K.; Dias, J.P.; Amado, C. A literature review of machine learning algorithms for crash injury severity prediction. *J. Saf. Res.* **2022**, *80*, 254–269. <https://doi.org/10.1016/j.jsr.2021.12.007>.
42. Xiao, Y.; Duan, Z. An explainable multi-task deep learning framework for crash severity prediction using multi-source data. *Sci. Rep.* **2025**, *15*, 21978. <https://doi.org/10.1038/s41598-025-09226-1>.
43. Kang, Y.; Khattak, A.J. Deep Learning Model for Crash Injury Severity Analysis Using Shapley Additive Explanation Values. *Transp. Res. Rec. J. Transp. Res. Board* **2022**, *2676*, 242–254. <https://doi.org/10.1177/03611981221095087>.
44. Budzyński, A.; Cieśla, M. Explainable Machine Learning for Cyclist Injury Severity in Bicycle–Vehicle Crashes in Poland: Association Patterns and Implications for Sustainable Road Safety. *Sustainability* **2026**, *18*, 5501. <https://doi.org/10.3390/su18115501>.
45. Ileri, K. Comparative analysis of CatBoost, LightGBM, XGBoost, RF, and DT methods optimised with PSO to estimate the number of k-barriers for intrusion detection in wireless sensor networks. *Int. J. Mach. Learn. Cybern.* **2025**, *16*, 6937–6956. <https://doi.org/10.1007/s13042-025-02654-5>.
46. Zahid, M.; Habib, M.F.; Ijaz, M.; Ameer, I.; Ullah, I.; Ahmed, T.; He, Z. Factors affecting injury severity in motorcycle crashes: Different age groups analysis using Catboost and SHAP techniques. *Traffic Inj. Prev.* **2024**, *25*, 472–481. <https://doi.org/10.1080/15389588.2023.2297168>.
47. Li, J.; Wang, X.; Yang, X.; Zhang, Q.; Pan, H. Analyzing Freeway Safety Influencing Factors Using the CatBoost Model and Interpretable Machine-Learning Framework, SHAP. *Transp. Res. Rec. J. Transp. Res. Board* **2024**, *2678*, 563–574. <https://doi.org/10.1177/03611981231208903>.
48. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical feature. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6639–6649.
49. Hassija, V.; Chamola, V.; Mahapatra, A.; Singal, A.; Goel, D.; Huang, K.; Hussain, A. Interpreting black-box models: A review on explainable artificial intelligence. *Cogn. Comput.* **2024**, *16*, 45–74. <https://doi.org/10.1007/s12559-023-10179-8>.
50. Li, G.; Wu, Y.; Bai, Y.; Zhang, W. ReMAHA–CatBoost: Addressing Imbalanced Data in Traffic Accident Prediction Tasks. *Appl. Sci.* **2023**, *13*, 13123. <https://doi.org/10.3390/app132413123>.
51. Aziz, K.; Chen, F.; Khattak, A. Improvements aiming at a safer living environment by analyzing crash severity through the use of boosting-based ensemble learning techniques. *Int. J. Crashworthiness* **2026**, *31*, 37–52. <https://doi.org/10.1080/13588265.2025.2492984>.
52. Okafor, S.; Liu, J.; Adanu, E.K.; Jones, S. Behavioral pathway analysis of pedestrian injury severity in pedestrian-motor vehicle crashes. *Transp. Res. Interdiscip. Perspect.* **2023**, *18*, 100777. <https://doi.org/10.1016/j.trip.2023.100777>.

53. Alqahtani, F. Enhancing Vulnerable Road User Safety Using Machine Learning and Crowdsourced Data: A Study of Pedestrian Crashes at Signalized and Non-Signalized Intersections. Ph.D. Thesis, Purdue University Graduate School, West Lafayette, IN, USA, 2025.
54. Rosén, E.; Sander, U. Pedestrian fatality risk as a function of car impact speed. *Accid. Anal. Prev.* **2009**, *41*, 536–542. <https://doi.org/10.1016/j.aap.2009.02.002>.
55. Kim, J.-K.; Ulfarsson, G.F.; Kim, S.; Shankar, V.N. Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accid. Anal. Prev.* **2013**, *50*, 1073–1081. <https://doi.org/10.1016/j.aap.2012.08.011>.
56. Sullivan, J.M.; Flannagan, M.J. The role of ambient light level in fatal crashes: Inferences from daylight saving time transitions. *Accid. Anal. Prev.* **2002**, *34*, 487–498. [https://doi.org/10.1016/S0001-4575\(01\)00046-X](https://doi.org/10.1016/S0001-4575(01)00046-X).
57. Wood, J.M.; Tyrrell, R.A.; Carberry, T.P. Pedestrian visibility at night: Effects of pedestrian clothing, driver age, and headlamp beam setting. In Proceeding of the 2003 Annual Meeting of the Transportation Research Board, Washington, DC, USA, 12–16 January 2003.
58. Aziz, H.M.A.; Ukkusuri, S.V.; Hasan, S. Exploring the determinants of pedestrian–vehicle crash severity in New York City. *Accid. Anal. Prev.* **2013**, *50*, 1298–1309. <https://doi.org/10.1016/j.aap.2012.09.034>.
59. Zajac, S.S.; Ivan, J.N. Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. *Accid. Anal. Prev.* **2003**, *35*, 369–379. [https://doi.org/10.1016/S0001-4575\(02\)00013-1](https://doi.org/10.1016/S0001-4575(02)00013-1).
60. Campos-Requena, N.; Vásquez-Lavin, F.; Fernández, F.; Barrientos, M.; Gelcich, S.; Oliva, R.D.P. A comparison of mixed logit and latent class models to estimate market segments for seafood faced with ocean acidification. *Aquac. Econ. Manag.* **2023**, *27*, 282–314. <https://doi.org/10.1080/13657305.2022.2100005>.
61. Yasmin, S.; Eluru, N.; Ukkusuri, S.V. Alternative Ordered Response Frameworks for Examining Pedestrian Injury Severity in New York City. *J. Transp. Saf. Secur.* **2014**, *6*, 275–300. <https://doi.org/10.1080/19439962.2013.839590>.
62. Roudsari, B.S.; Mock, C.N.; Kaufman, R.; Grossman, D.; Henary, B.Y.; Crandall, J. Pedestrian crashes: Higher injury severity and mortality rate for light truck vehicles compared with passenger vehicles. *Inj. Prev.* **2004**, *10*, 154–158. <https://doi.org/10.1136/ip.2003.003814>.
63. Wilmut, K.; Purcell, C. Why Are Older Adults More at Risk as Pedestrians? A Systematic Review. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2022**, *64*, 1269–1291. <https://doi.org/10.1177/0018720821989511>.
64. Omer, I.; Gitelman, V.; Rofé, Y.; Lerman, Y.; Kaplan, N.; Doveh, E. Evaluating Crash Risk in Urban Areas Based on Vehicle and Pedestrian Modeling. *Geogr. Anal.* **2017**, *49*, 387–408. <https://doi.org/10.1111/gean.12128>.
65. Chen, P.; Zhou, J. Effects of the built environment on automobile-involved pedestrian crash frequency and risk. *J. Transp. Health* **2016**, *3*, 448–456. <https://doi.org/10.1016/j.jth.2016.06.008>.
66. Esmaili, A.; Aghabayk, K.; Shiwakoti, N. Latent Class Cluster Analysis and Mixed Logit Model to Investigate Pedestrian Crash Injury Severity. *Sustainability* **2022**, *15*, 185. <https://doi.org/10.3390/su15010185>.
67. Behnood, A.; Mannering, F. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. *Anal. Methods Accid. Res.* **2017**, *16*, 35–47. <https://doi.org/10.1016/j.amar.2017.08.001>.
68. Kim, J.-K.; Kim, S.; Ulfarsson, G.F.; Porrello, L.A. Bicyclist injury severities in bicycle–motor vehicle accidents. *Accid. Anal. Prev.* **2007**, *39*, 238–251. <https://doi.org/10.1016/j.aap.2006.07.002>.
69. Kaplan, S.; Vavatsoulas, K.; Prato, C.G. Aggravating and mitigating factors associated with cyclist injury severity in Denmark. *J. Saf. Res.* **2014**, *50*, 75–82. <https://doi.org/10.1016/j.jsr.2014.03.012>.
70. Räsänen, M.; Summala, H. Attention and expectation problems in bicycle–car collisions: An in-depth study. *Accid. Anal. Prev.* **1998**, *30*, 657–666. [https://doi.org/10.1016/S0001-4575\(98\)00007-4](https://doi.org/10.1016/S0001-4575(98)00007-4).
71. Asgarzadeh, M.; Verma, S.; Mekary, R.A.; Courtney, T.K.; Christiani, D.C. The role of intersection and street design on severity of bicycle-motor vehicle crashes. *Inj. Prev.* **2017**, *23*, 179–185. <https://doi.org/10.1136/injuryprev-2016-042045>.
72. Balk, S.A.; Tyrrell, R.A.; Brooks, J.O.; Carpenter, T.L. Highlighting Human Form and Motion Information Enhances the Conspicuity of Pedestrians at Night. *Perception* **2008**, *37*, 1276–1284. <https://doi.org/10.1068/p6017>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.