

Article

Injury Risk Assessment and Interpretation for Roadway Crashes Based on Pre-Crash Indicators and Machine Learning Methods

Chenwei Gu *, Jinliang Xu * , Shuqi Li, Chao Gao  and Yongji Ma

School of Highway, Chang'an University, Xi'an 710061, China

* Correspondence: 2020021051@chd.edu.cn (C.G.); xujinliang@chd.edu.cn (J.X.)

Abstract: Pre-crash injury risk (IR) assessment is essential for guiding efforts toward active vehicle safety. This work aims to conduct crash severity assessment using pre-crash information and establish the intrinsic mechanism of IR with proper interpretation methods. The impulse–momentum theory is used to propose novel a priori formulations of several severity indicators, including velocity change (ΔV), energy equivalent speed (EES), crash momentum index (CMI), and crash severity index (CSI). Six IR models based on different machine learning methods were applied to a fusion dataset containing 24,082 vehicle-level samples. Prediction results indicate that the pre-crash indicators (PCIs) are more influential than the commonly used basic crash information because the average accuracy of six models can be improved by 14.35% after utilizing PCIs. Furthermore, the features' importance and their marginal effects are interpreted based on parameter estimation, Shapley additive explanation value, and partial dependence. The ΔV , EES, and CMI are identified as the determinant indicators of the potential IR, and their partial distributions are significantly influenced by the crash type and impact position. Based on partial dependence probabilities, the study establishes decision thresholds for PCIs for each severity category for different impact positions, which can serve as a useful reference for developing targeted safety strategies. These results suggest that the proposed method can effectively improve pre-crash IR assessment, which can be readily transferred to safety-related modeling in an active traffic management system.



Citation: Gu, C.; Xu, J.; Li, S.; Gao, C.; Ma, Y. Injury Risk Assessment and Interpretation for Roadway Crashes Based on Pre-Crash Indicators and Machine Learning Methods. *Appl. Sci.* **2023**, *13*, 6983. <https://doi.org/10.3390/app13126983>

Academic Editor: Marek Krawczuk

Received: 15 May 2023

Revised: 7 June 2023

Accepted: 8 June 2023

Published: 9 June 2023



Copyright: © 2023 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/>).

Keywords: injury risk assessment; a priori crash analysis; crash severity; machine learning; model interpretation

1. Introduction

Roadway crashes have been a primary cause of fatalities worldwide, directly resulting in approximately 1 million deaths and 50 million serious injuries annually [1]. Reducing the injury risk (IR) from crashes, especially the potential severe outcome, has been a major focus of extensive academic research and engineering projects related to roadway safety [2]. Many studies have attempted to model roadway safety by analyzing massive amounts of crash information and parsing injury outcomes with crash reconstruction analysis [3,4]. The current well-established estimates of crash severity are based on IR functions that are modeled on post-crash variables, such as the change in velocity (ΔV) sustained by a vehicle before and after a crash. Although posterior analysis can infer crash mechanisms, real-time assessment of safety maneuvers in the pre-crash phase remains challenging.

Recently, innovative risk assessment technologies that use real-time information have become an emerging trend in active roadway safety, driven by the continuous advancements in vehicle safety technology. Accordingly, many researchers have suggested that active safety measures can greatly benefit from credible applications that use identifiable pre-crash information as indicators for potential crashes [5–7]. This approach could complement the traditional post-crash analysis and enable real-time assessment of safety maneuvers, which is currently difficult to achieve.

The pre-crash phase of a vehicle refers to the period from the occurrence of a hazardous situation to the moment when the collision becomes unavoidable. From the perspective of preventive assessment, two criteria must be satisfied to evaluate the latent dynamic risk for this phase: the proximity to a crash and the potential severity [8]. A priori variables, such as pre-impact velocity and closing angle, must be considered because they are crucial in retrieving IR based on pre-crash phase information [9]. Traffic conflict is commonly used to interpret risky threats with real-time operation information. However, the lack of a clear connection between conflict parameters and the prevalent severity scale has resulted in limited research on measuring injury severity within the pre-crash framework [10].

Arun established the explicit relationships between the conflict index and the actual crash severity measured by the maximum abbreviated injury scale (MAIS) to assess the potential risk [11]. This finding demonstrates that the gap between the pre-events and injury severity can be filled by applying in-depth crash datasets that contain adequate variables and critical operation data prior to a crash. Synthetic crash data provide the key elements in tuning models that can describe the complexity of real impacts and represent a valuable tool for the a priori assessment of IR [12]. Several studies have investigated the causal factors of crash severity and their influence on crash characteristics. Based on the statistical techniques, evident findings indicate that factors involving vehicles, operating conditions, occupant attributes, and dynamic collision mechanics are considered to influence the severity of injury [13–15]. However, the effect of the same variable in different studies is skewed. Atkinson noted that, unlike several existing studies, seat position had a significant influence on injury outcomes in frontal crashes [16]. Vadeby indicated that the passengers' gender and age have no particular relationship with injury severity [17], and Newgard emphasized the elevated effect of these two factors in assessing exposure to severe injuries [18]. According to potential interpretation, these factors are undoubtedly relevant to the IR, but they do not directly dictate the outcomes in comparison with the impact characteristics [19].

Crash-related studies have demonstrated that the primary factor determining the passenger injury level is the subjected force during the crash, which can be estimated based on multiple severity parameters, such as the velocity change (ΔV) and energy equivalent speed (EES) [20]. When combined with the crash characteristics, only a few indicators are sufficient to obtain accurate results in predicting the severity outcomes [21,22]. However, the main drawback of these parameters is that their initial formulations are derived from the post-crash consequence, which limits the applications to the reduction in potential risk before crash [23].

To fulfill this requirement, Ji adopted a new a priori factor to measure the energy dissipation of a vehicle during a crash and evaluated its effect in injury assessment [24]. Laureshyn presented a theoretical Extended Delta-V as a measure of conflict severity in site-based observations and validated its performance in identifying severe traffic events [8]. The other variables, such as closing velocity (V_r), crash momentum index (CMI), and crash severity index (CSI), are proposed in a priori formulations, which measure the impact speed and eccentricity of the potential crashes [25]. A single crash-related indicator cannot be expected to capture all risk-relevant information considering the variability of crash dynamics. In cases of multiple alternative crash indicators, the severity-variable relationships must be further explored. Hence, the aim of this study was to explore several pre-crash indicators (PCIs) based on in-depth crash datasets that fully consider the impact of crash mechanism with adequate data samples.

From the perspective of assessment methods, machine learning (ML) has been widely employed in the roadway safety field, especially for complex crash modeling, such as causation analysis and risk assessment [26–28]. ML models have displayed advantages in dealing with the limitations in data generalization and prediction performance compared with traditional statistical methods. The hidden associations between risk factors and model results can be effectively explained through novel post hoc analysis, thus guiding roadway safety [29]. ML has been successfully applied in risk modeling. Wang employed a classification and regression technique to identify the important factors affecting driving risk in terms of driver, vehicle, and road environment [13]. Wen compared the performance of light gradient boosting and extreme gradient boosting models in risk prediction and quantified the importance, total effect, and main and interaction effects of risk factors by Shapley additive explanation (SHAP) [30]. The novel visualization methods can effectively measure the evident distribution between crash characteristics and severity probabilities, providing detailed interpretation of the results [31]. To this end, this study attempts to explore the nonlinear associations between crash indicators and severity levels using ML and interpretation methods. Therefore, the key factors may be extracted from the model results, and efficient suggestions are correspondingly presented for road safety.

This study proposes a framework combining the PCIs and ML models to explore relevant factors of IR with full consideration of the impact mechanics and the a priori information. The formulations of several PCIs are proposed on the basis of the impulse–momentum theory, and six ML models were applied with a fusion dataset containing adequate crash cases. The model results are further investigated to identify the intrinsic crash patterns for different impact types with proper interpretation methods. This framework provides a major overview of the benefits and drawbacks of several PCIs, showcases their applicability in understanding the complex crash patterns, and explores favorable predictive methods for IR assessment in pre-crash scenarios.

2. Materials and Methods

This study proposes the PCI-ML method for the IR assessment of crash severity based on the pre-crash analysis and ML interpretation. The basic research framework is shown in Figure 1, which consists of four parts: (1) PCI derivation based on pre-crash analysis; (2) data fusion and preprocessing; (3) IR assessment modeling; and (4) feature analysis. First, a simplified formula for vehicle deformation energy was derived in the collision plane, and the a priori formulation for PCIs, including ΔV , CMI, ESS, CSI, and V_r , was further provided, which can be calculated based on the vehicle motion and configuration prior to a crash. Second, this study collected a total of 24,082 vehicle-level crash data from two independent crash databases. The collected data were fused and cleaned for appropriate evaluation variables. Third, this study combined the PCIs with various ML models to propose the framework for injury severity assessment. The predictive performance of different models was compared after Bayesian optimization and cross-validation. Finally, we analyzed the association between the influential variables and injury severity based on parameter estimation and post hoc analysis. The key features can be determined by the rank of feature importance, and their detailed impact on the likelihood of IR was further quantified based on the partial dependency plots.

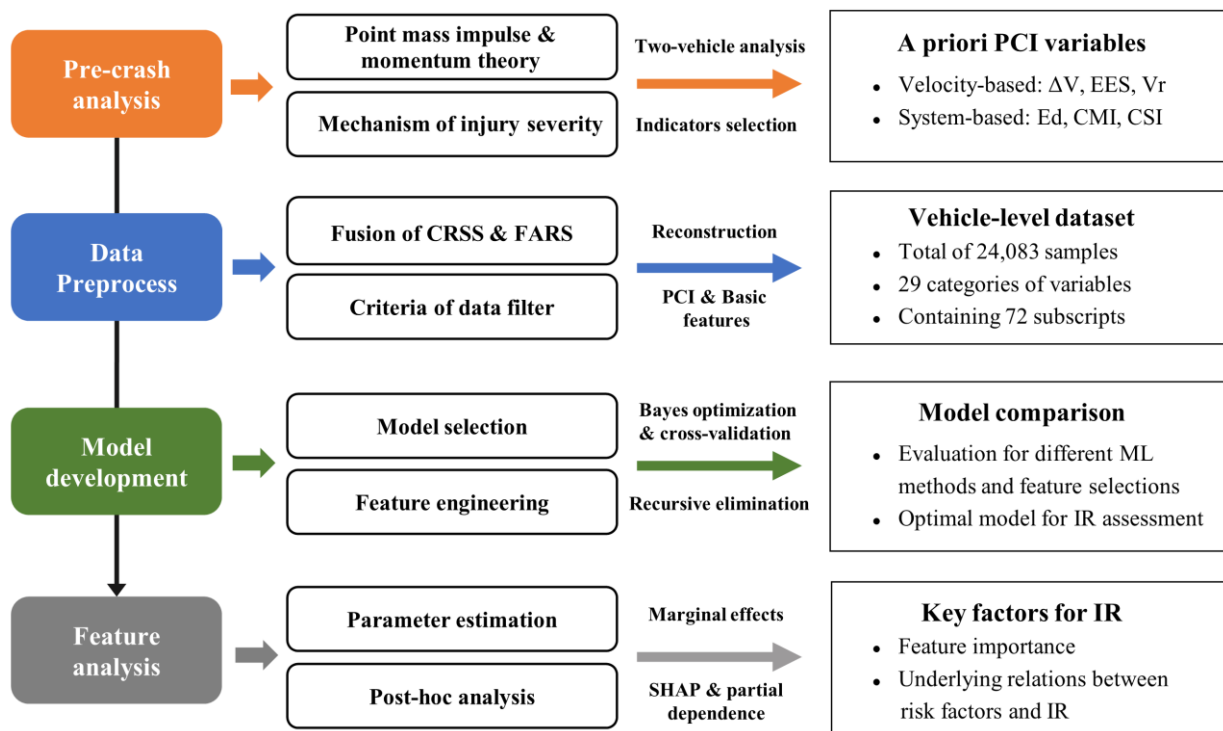


Figure 1. Framework of the PCI-ML method.

2.1. Theory

In most traffic conflict analysis, the predetermined pre-crash conditions are typically used as prerequisites for parameter calibration [32,33]. On the basis of this assumption, this section will describe the injury mechanism from the perspective of pre-crash analysis and derive several indicators in a prior formulation to measure the potential risk.

2.1.1. Mechanism of Crash Injury

Most kinetic energy carried by vehicles is converted into deformation of the vehicles during the compression phase until the relative velocity of the contact areas in the normal direction is lowered to zero [34]. During this process, the uncrashed structure is subjected to a violent acceleration in an instant, which will also be imposed on the car-occupant compartment. This equivalent acceleration, denoted as $\bar{a}_c$, directly reflects the force on the occupant and threatens the safety of the interior occupants [34]. Most initial studies typically used the posteriori $\bar{a}_c$ as a direct measure of the injury severity, which can be derived by Newtonian mechanics as follows:

$$\bar{a}_c = \frac{E_d}{m_i \cdot l} \quad (1)$$

where E_d is the deformation energy absorbed by the vehicle body, m_i is the mass of the vehicle, and l is the motion distance after the crash. In Equation (1), the deformation energy has a larger scale of magnitude distribution (typically 10^3 – 10^7 J) compared with m_i and l . Accordingly, the equivalent acceleration to the occupants is assumed to depend mainly on the energy conversion during the crash. This notion also suggests that the deformation energy of the vehicle body is proportional to the injury severity.

According to the crash reconstruction studies, most severity parameters must be calculated based on the deformation energy combined with vehicle relative motion analysis. The post-crash variables, such as ΔV , have been shown to be effective in measuring the IR of a crash. However, the estimation of these parameters is mostly derived from posteriori

variables, such as post-crash velocity, deflection angle, and degree of deformation, which are difficult to apply in the a priori analysis under near-crash conditions.

2.1.2. Crash Analysis under Pre-Crash Conditions

In this study, the generalization of a non-perfect elastic collision analysis is simplified on the basis of the point mass impulse–momentum theory. The a priori parameters, such as pre-crash speed, approaching direction, and vehicle configuration, were used to further deduce the near-crash condition. The pre-crash analysis was performed in the critical state where vehicles are in contact with each other to facilitate the explanation of the formula, as shown in Figure 2.

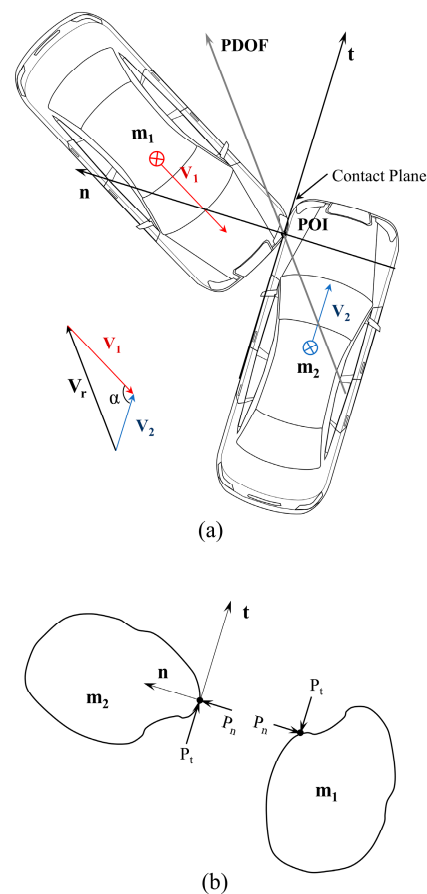


Figure 2. Planar visualization of an impact between vehicles. (a) Point of impact (POI) and (b) conservation of impulse in the point mass model.

Given that the vehicle does not rotate at the moment of collision, the velocity of the barycenter during the impact coincides with the velocity of the collision center. In this case, the vehicle can be regarded as a mass point moving in a 2D plane. Figure 2b shows the impact system for vehicles of masses m_1 and m_2 with velocities of v_1 and v_2 .

An n – t reference coordinate system is introduced in Figure 2a according to the moving direction of vehicles. The contact plane reflects the relative motion of vehicles, which can be determined based on geometric rules [35]. The coordinate origin is located at the POI, which represents the position of momentum exchange between vehicles. The t axis is tangential to the contact surface, and the n axis is its normal direction. In addition, V_r is the closing velocity between two vehicles, which is assumed to be aligned with the principal direction of force (PDOF). Figure 2b divides the impulse distribution of each

particle along the n - t direction. The following expression can be obtained by applying the impulsion–momentum conservation to m_1 and m_2 along the n and t directions:

$$\begin{aligned}(m_1 V_{1n} - m_1 v_{1n}) + (m_2 V_{2n} - m_2 v_{2n}) &= 0 \\ (m_1 V_{1t} - m_1 v_{1t}) + (m_2 V_{2t} - m_2 v_{2t}) &= 0\end{aligned}\quad (2)$$

where v_{1n} , v_{1t} , v_{2n} , and v_{2t} are the velocity components of m_1 and m_2 along the n and t axis before the impact and V_{1n} , V_{2n} , V_{1t} , and V_{2t} are the velocity components of m_1 and m_2 along the n and t axis after the impact. In Equation (2), the velocity of this two-vehicle system is equal to zero along each direction, and the momentum is conserved at the instant of crash.

The restitution coefficient ε , which characterizes the elastic recovery of the vehicle, is applied for analysis to further determine the motion condition before and after the impact. The velocity of the vehicle before and after the impact complies with the following relationship:

$$V_{2n} - V_{1n} = -\varepsilon(v_{2n} - v_{1n}) \quad (3)$$

where the restitution coefficient ε can be estimated by using V_r at POI based on the empirical formula by Antonetti [35]. The post-crash V_{2n} and V_{2t} can be calculated by applying the impulse ratio $\mu = P_t/P_n$. Assuming that the other types of dissipated energy can be neglected, the deformation energy (Equation (5)) can be initially deduced according to conservation of energy (Equation (4)):

$$E_d = \frac{1}{2}m_1(v_{1n}^2 + v_{1t}^2) + \frac{1}{2}m_2(v_{2n}^2 + v_{2t}^2) - \frac{1}{2}m_1(V_{1n}^2 + V_{1t}^2) - \frac{1}{2}m_2(V_{2n}^2 + V_{2t}^2) \quad (4)$$

$$E_d = \frac{1}{2}m_c(v_{2n} - v_{1n})^2(1 + \varepsilon) \left[(1 - \varepsilon) + 2\mu r - (1 + \varepsilon)\mu^2 \right] \quad (5)$$

where m_c is the system mass of the two-vehicle crash, which can be calculated by m_1 and m_2 :

$$m_c = \frac{m_1 m_2}{m_1 + m_2}$$

Variable r is the ratio of velocity component in the n and t directions, denoted as follows:

$$r = \frac{v_{2t} - v_{1t}}{v_{2n} - v_{1n}}$$

The impulse ratio μ in Equation (5) depends on the sliding state of the vehicle and ranges among $0 \leq \mu \leq \mu_0$. The impulse ratio, μ_0 , called the critical impulse ratio, characterizes the maximum kinetic energy loss within the system and is defined as $\mu_0 = r/(1 + \varepsilon)$ [25]. In the case of side impact with unknown post-impact parameters, the potential deformation energy can be obtained by substituting μ_0 into Equation (5) as follows:

$$E_d = \frac{1}{2}m_c(v_{2n} - v_{1n})^2(1 - \varepsilon^2 + r^2) \quad (6)$$

In the case of a frontal or rear-end crash, no tangential force exists in the vehicle contact plane, and the tangential impulse $P_t = 0$. The deformation energy can be simplified as follows:

$$E_d = \frac{1}{2}m_c(v_{2n} - v_{1n})^2(1 - \varepsilon)^2 \quad (7)$$

In addition, the energy distribution within the system depends on the mass ratio of the vehicle. Vehicles with smaller mass will absorb more energy and suffer more potential severe damage. The kinetic energy absorbed by each vehicle during the compression phase can be expressed as follows:

$$\frac{E_{d1}}{E_{d2}} = \left(\frac{m_1}{m_2} \right)^\alpha \quad (8)$$

The value of the negative constant α mainly depends on the impact position. A previous study has concluded that α is -5.0 for the side impact and -0.8 for the frontal or rear impacts, indicating that lighter vehicles subjected to side impacts can result in more severe outcomes [24].

Based on Equations (2)–(8), the simplified deformation energy under pre-crash conditions can be achieved using the pre-impact indicators, including approaching speed, moving direction, and vehicle configuration.

2.1.3. Pre-Crash Variable Selection

The simplified deformation energy formulation for pre-crash conditions provides the a priori form for most post-crash parameters. Crash parameters used for IR assessment can be broadly classified into two categories: velocity-based and system-based indicators. Velocity-based indicators define the kinematic features by characterizing the vehicle motion at the impact and have always been considered as a primary index for risk evaluation. Accordingly, the system-based indicators highlight the differences in impact configurations of crash systems with a dimensionless form, which focus on the effects of crash types, vehicle parameters, and impact eccentricity on occupant injuries. Both types of indicators have been proven to have a substantial influence on the injury outcome because they effectively characterize the main factors affecting the crash consequences: crash morphology and energy transformation [19,36,37]. Therefore, the combination of these indicators can explain the crash mechanism from different dimensions.

The following crash parameters were selected as the PCIs based on previous findings:

- (1) Closing velocity at impact (V_r) represents the effective velocity of the crash system along the direction of PDOF, reflecting the potential maximum damage of the collision [23]. V_r , as an a priori variable, can be calculated by vector subtraction of the vehicle speed (Figure 2a).
- (2) Velocity change (ΔV) refers to the change in a velocity vector experienced by a road user during a crash. Larger ΔV implies stronger external force on the car-occupant compartment and has been widely used in in-depth investigation crash databases. Unlike the traditional posteriori method, this study summarizes the a priori form of ΔV , which can be derived from the deformation energy and the system mass as follows:

$$\Delta V_i = \frac{1}{m_i} \sqrt{2E_d m_c \frac{(1+\varepsilon)}{(1-\varepsilon)}} \quad (9)$$

- (3) EES refers to the equivalent velocity for a given deformation energy under the rigid crash conditions [38]. Under the assumption that kinetic energy is completely converted into deformation energy, EES measures the velocity level of the vehicle at the time of impact. This variable can be derived from the deformation energy and the vehicle mass:

$$\text{EES} = \sqrt{2E_d / m_i} \quad (10)$$

- (4) CMI and CSI are system-based variables used to measure vehicle configuration and crash characteristics of two-vehicle crash systems. CMI characterizes the momentum change of the subject vehicle and depends on the mass ratio and elastic modulus [39]; CSI measures the relationship between the dissipated deformation energy and the approaching velocity and depends on the mass ratio and stiffness ratio between vehicles [40]. In the same energy dissipation, impact eccentricity significantly affects the impact outcomes, which can be measured by the dimensionless parameters. These indicators are widely used in risk analysis and can be calculated by the following equations:

$$\text{CMI} = \frac{1+\varepsilon}{1+R_m} = \frac{\Delta V}{V_r} \quad (11)$$

$$CSI = \sqrt{\frac{1 - \epsilon^2}{(1 + R_m)(1 + R_K)}} = \frac{EES}{V_r} \quad (12)$$

where $R_m = m_1/m_2$ and $R_K = K_1/K_2$ are the mass and stiffness ratios between vehicles, respectively, denoted by the vehicle configuration of the crash system.

The majority of the proposed PCIs require the key variable of deformation energy, which is derived in combination with vehicle configurations and contact plane. The PCIs will specifically refer to E_d , V_r , ΔV , EES, CMI, CSI, and M_c in the following for illustration.

2.2. Data Preparation

2.2.1. Data Source

This study integrates the crash data of the Crash Reporting Sampling System (CRSS, 2019 and 2020) and the Fatal Accident Analysis Reporting System (FARS, 2016–2020) as part of the NHTSA program of the U.S. Department of Transportation to better validate and analyze the risk factors in a data-driven method [41]. A total of 275,380 separate crash records are covered. The CRSS and FARS databases provide adequate information related to crashes, which is essential for in-depth investigations. Given that the detailed information in both databases is extracted from police records, the authenticity of the dataset is guaranteed.

Specifically, CRSS collects approximately 55,000 representative crash records from a total sample of 6 million crashes each year, ranging from property loss to fatal injuries. Meanwhile, FARS is a census of all fatal crashes in the US. A crash will be coded in FARS if at least one person involved dies within 30 consecutive days of the crash. Considering the difference for the property of severity cases, the fusion of these two databases could result in a balanced dataset. In both databases, crash information related to the driver and passenger, crash characteristics, vehicle, environmental factors, and injury severity is coded in three aspects, namely levels of accident, vehicle, and person, with approximately 120 explanatory variables. The key features, such as relative motion relationships and collision deformation classification (CDC), prior to the crash can be obtained.

Considering that the PCIs rely on detailed motion information, the evaluation was performed based on the vehicle level in this study. The following criteria were used to filter the crash data from the FARS and CRSS databases:

- (1) Vehicles that cannot provide external structural protection, such as motorcycles and non-motorized vehicles, were excluded. In addition, only vehicles within 15 years of registration were extracted. This is because the structures of the vehicle body frames determine the passive protection.
- (2) Cases with vacant data were excluded because vehicle motion, passenger information, road, and environmental conditions may significantly affect the outcome of crashes.
- (3) Crashes involving rollovers, roadside run-off, and secondary accidents were excluded. These patterns are associated with completely different deformation patterns, making it difficult to measure the degree of severity with general parameters.
- (4) Only the events in which the use of belt and airbags by occupants is reported were considered. The influence of these factors should be ruled out because seat belts and airbags have been shown to significantly enhance the passive safety performance.
- (5) Three impact configurations were extracted, namely head-on, rear-end, and side-impact crashes.

In addition, this study encoded the levels of injury severity as a categorical variable using three levels, namely minor level (SEV1, property damage only), moderate level (SEV2, non-disabling injuries), and major level (SEV3, disabling and fatal injuries), to measure the risk of injury under different conditions.

The imbalanced data for the different severity levels represent a prevailing issue for studies involving FARS or CRSS datasets, causing the prediction results to be lopsided toward the majority category [42]. To tackle this issue, a combination of the two datasets

was utilized, leveraging the larger volume of data in the CRSS database and the inclusion of fatal accidents in the FARS database. The timeframe of 2019–2020 was chosen for CRSS data, while the timeframe of 2016–2020 was chosen for FARS data. This approach aimed to achieve a balanced distribution of severity levels, resulting in controlled proportions of 43.1% for SEV1, 27.9% for SEV2, and 29.0% for SEV3. By adopting this approach, the issue of over-sampling or under-sampling involved could be effectively addressed.

2.2.2. Data Preprocessing

After filtering, the final data set contained a total of 33,529 occupants and 24,082 vehicles with more than 120 variables (Table S1). The crash data must be pre-processed to obtain the appropriate predictor features to evaluate the effects of crash information on IR. All explanatory variables can be divided into two categories, namely the basic variables involving original crash patterns and the proposed PCIs. The PCIs can be derived with vehicle speed, vehicle configuration, impact type, and CDC. Additional insights regarding several PCIs are provided in Figure 3 with density plot. The density distribution provides valuable indications that preliminarily reflect the correlation between continuous parameters and IR despite its simplicity. The visualization results indicate that severe injuries have a tendency to develop at elevated values of PCIs, which supports the basic hypothesis of this study.

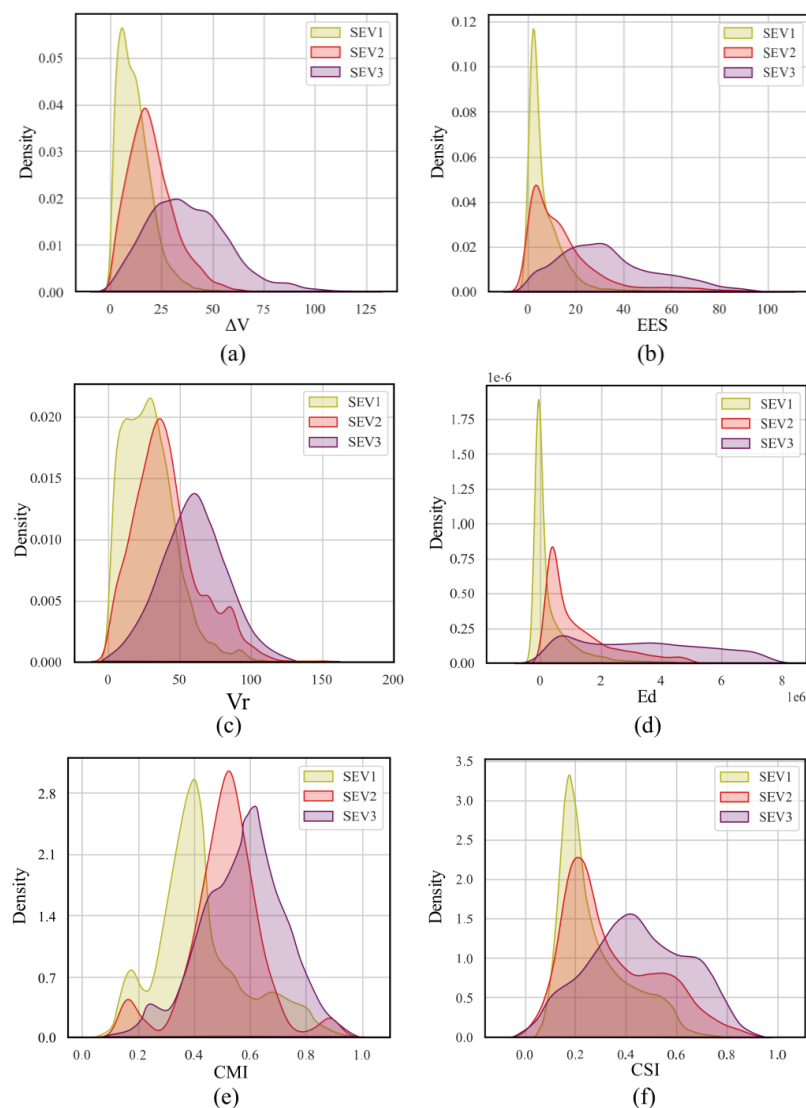


Figure 3. Density plots of PCIs.

According to previous studies, the basic crash variables were selected from the integration of person and accident levels in the database containing the following information: (1) occupant: gender, age, and seat position; (2) driver: driving behavior, drinking, speeding, and illegal behavior; (3) vehicle: vehicle type and approaching speed; (4) roadway: road alignment, section, road type, and speed limit; and (5) environmental conditions: weather, time, and lighting conditions. The preprocessed dataset contains 29 explanatory variables with 72 subscripts. The general overview of the main variables is provided in Table 1.

Table 1. Descriptive statistics of crash dataset.

I. Descriptive Statistics for Continuous Variables								
Variable	Mean	S.D.	Variable	Mean	S.D.			
ΔV (km/h)	26.25	21.01	Mc (kg)	995.80	444.37			
EES (km/h)	20.93	23.05	Mass (kg)	2831.23	3299.98			
Vr (km/h)	52.71	33.24	Speed (km/h)	55.19	32.96			
CMI	0.49	0.16	Speed limit (km/h)	72.96	22.38			
CSI	0.40	0.23	Age	43.23	19.02			
Ed (Joules)	4.10×10^6	1.57×10^7	Lane number	2.85	/			
II. Descriptive Statistics for Discrete Variables								
Variable	Subscript	Count	%	Variable	Subscript	Count	%	
Vehicle type	Light-duty	13,697	56.87	Time	0:00~6:00	2508	10.41	
	Pick up/SUV	6431	26.70		6:00~12:00	6435	26.72	
	Truck	3856	16.01		12:00~18:00	10,342	42.94	
	Bus	98	0.41		18:00~24:00	4797	19.92	
Behavior before crash	Turning left/right	4562	18.94	Crash type	Rear-end	4672	19.40	
	Proceeding straight	15,022	62.38		Head-on	5596	23.24	
	Decelerating/stopping	2700	11.21		Side impact	13,814	57.36	
	Overtaking/merging	559	2.32	Impact position	Front side	15,096	62.68	
	Negotiating a curve	1947	8.08		Back side	3042	12.63	
	Intersecting	4694	19.49		Near side	3367	13.98	
	Braking and steering	1525	6.33		Far side	2577	10.70	
Violation	Skidding	2557	10.62					
	Traveling on the wrong side	7010	29.11	Alignment	Straight	20,648	85.74	
	Alcohol	1231	5.11					
	Speeding	3159	13.12		Curve	3434	14.26	
Sex				Profile	Level	18,554	77.04	
	Male	14,158	58.79		Uphill	708	2.94	
					Downhill	1935	8.03	
	Female	9924	41.21		Hillcrest/sag	1060	4.40	
					Unknown	1825	7.58	
Road type	One-way	384	1.59	Median barrier type	No barrier	15,405	63.97	
	Divided two-way	8436	35.03		Unprotected	4365	18.12	
	Undivided two-way	13,427	55.75		Positive protected	4312	17.90	
	Ramp	241	1.00					
Season	Spring	5141	21.35	Seat position	First row	17,486	72.61	
	Summer	6008	24.95		Second row	5249	21.80	
	Autumn	6729	27.94					
	Winter	6204	25.76		Others	1347	5.59	
Weather	Clear	17,275	71.73	Roadway surface	Dry	19,658	81.63	
	Not clear	6870	28.53		Wet	3288	13.65	
					Slippery	1136	4.72	
Severity	SEV1	10,309	42.81	Traffic control	No control	15,641	64.95	
	SEV2	6654	27.63		Traffic sign	5445	22.61	
	SEV3	7119	29.56		Traffic device	2996	12.44	

2.3. Model Development

Considering the large number of crash cases and explanatory variables in the dataset, evaluation methods that combine statistical models and ML have better adaptability. The

accuracy of injury assessment depends on several aspects, mainly including model selection, feature selection, and hyperparameter optimization. This section introduces detailed establishment of the evaluation models.

2.3.1. Model Selection

The model performance between existing models must be compared to validate the effectiveness of the proposed framework on crash severity assessment. Unsuitable assessment methods may result in underfitting or overfitting of the prediction results. Six ML models were used for severity classification in this study. Each model is briefly described as follows:

- (1) The Bayesian ordered logit model (BOL) is an elegant parametric model that has the basic formulation of logit models with parameter estimation conducted by Bayesian inference [15]. The Bayes method overcomes the uncertainty of the maximum likelihood and reduces overfitting. The parametric modeling could analytically measure the influence of factors based on parameter estimation compared with other ML methods, which helps in the interpretation of the marginal effects on the probabilities of injury severity.
- (2) A multilayer perceptron (MLP) is a fully connected feed-forward artificial neural network, which uses back propagation for multilayer training with a relatively simple structure and is widely used in roadway safety evaluation. In this study, this approach can be regarded as the baseline model for comparison.
- (3) A support vector machine (SVM) is a supervised ML method that uses kernel functions to map data features to a high-dimensional feature space, which in turn enables classification prediction [43]. The SVM model has excellent generalization and efficiently solves nonlinear problems with medium-sized training sets. In addition, this model can obtain satisfactory results when dealing with datasets with a large number of features.
- (4) Random forest (RF) is a well-known tree-based method that uses multiple decision trees for classification and prediction [13]. The model randomly selects the same amount of data for training and several features for tree construction. With these two stochastic selections, this model does not require complex hyperparameter tuning and efficiently performs with large sample sizes, which is consistent with the datasets in this study.
- (5) A deep neural network (DNN) is a deep learning algorithm with more than one hidden layer in feedforward networks [44]. This method interacts with features through several fully connected layers that are interlinked by weighted connections. Each connected node is regarded as a nonlinear calculation module that converts the input information with an optimization algorithm to form a critical decision threshold. In this study, dropout regularization was applied to leave the biases unregularized and avoid overfitting in training.
- (6) The local cascade ensemble model (LCE) is a hybrid ensemble method for handling the bias–variance tradeoff. This method combines the bagging method to generate the aggregate predictors and the boosting method to learn weak classifiers. Furthermore, a divide-and-conquer approach is utilized in LCE to learn different parts of the training data, which can be locally applied for cascade generalization. The latest studies have proven that LCE outperforms the state-of-the-art classifiers on the UCI and UEA datasets [45]. Although, to our knowledge, limited studies have applied this model for roadway safety modeling, its utilization in IR assessment is promising.

Former applications have proven that model predictions can be enhanced by 5–20% through hyperparameter tuning and cross-validation (CV) regardless of the model used [46]. To identify the best model, the Bayes optimization method was adopted in this framework to generate information from previous optimization iterations and tune the model with multiple hyperparameters. In addition, a ten-fold CV method was applied to mitigate

the overfitting. The final prediction ability of each model was verified by aggregating the prediction results of all CV groups.

2.3.2. Performance Evaluation

Four metrics were used in this study for injury severity classification, namely accuracy, true positive rate (TPR), false positive rate (FPR), and area under the curve (AUC)–ROC. The overall performance accuracy of the model was measured for a given test dataset by calculating the proportion of the correct predictive results of all samples. TPR characterizes the number of identified positive samples over all positive samples. FPR refers to the proportion of identified negative cases to all negative cases. AUC is a widely used indicator that quantifies the classification performance by calculating the area under the ROC curve, where ROC can be obtained by plotting TPR to FPR. The formulation of accuracy, TPR, and FPR is shown in Equations (13)–(15).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$TPR = TP / (TP + FN) \quad (14)$$

$$FPR = FP / (TN + FP) \quad (15)$$

where TP , TN , FP , and FN are the number of true positive, true negative, false positive, and false negative samples, respectively.

2.3.3. Feature Selection

The collected dataset is high-dimensional, with more than 60 sub-variables, which may contain several redundant features. A considerable number of redundant features will increase the complexity of the model and even affect the prediction results. Accordingly, this study applied recursive feature elimination (RFE) based on the wrapper method for the best feature combination [47]. Several combinations are evaluated with expected accuracy by treating the selection as a search optimization issue.

All features (AF) are divided into two categories, namely PCI features and basic features (BFs). The optimal selections in these three categories were further determined through a global comparison. Table 2 summarizes the selected features and compares the accuracy based on the RF models.

Table 2. Feature selection results.

Type	Selected Features	Accuracy (%)	
		Before	After
PCI	EES, ΔV , V_r , CSI, CMI, M_c , Ed.	77.64	77.64
BF	Speed, Mass, Vehicle type, Violation, Age, Drinking, Crash type, Lanes, Skidding, Impact position, Speed limit.	62.37	64.04
AF	EES, ΔV , V_r , CSI, CMI, MC, Ed, Impact position, Seat position, Crash type, Age, Skidding, Speed, Mass, Vehicle type	75.59	79.74

Despite the minor differences among the models, the feature combinations in Table 2 have good performance in all six types of models. Blindly adding redundant features may degrade the prediction performance. All PCIs are proven to have a positive contribution. After the RFE selection, the RF model improves the prediction with the BF by 1.67% and the overall accuracy by approximately 4.15%.

In the abovementioned assessment method, the RFE selection contributes to the removal of redundant variables, while the iteration concept and supervised processes in different ML models further alleviate the noisy influence in the features. Thus, such a

compound approach helps in investigating more potential factors that may be neglected in the common method.

2.4. Feature Analysis

Model interpretability is a major challenge to the applications of ML methods due to the lack of transparency and interpretability. This section will introduce the three-step evaluation method based on the marginal effects, SHAP value, and partial dependency plot.

2.4.1. Marginal Effect Based on BOL

The parametric methods have advantages in quantifying the marginal probability of certain explanatory factors on injury severity. However, the contributing factors cannot be directly identified based on the estimated parameters because the probability is generally nonlinear with most features. Accordingly, the marginal effects are calculated based on BOL in this study, which represents the influence of a single unit change in a feature on the injury probabilities. Specifically, the marginal effects of the categorical features are calculated as the difference in the estimated probabilities with the variable changing (with all other variables equal to their means). The effects of the continuous variables can be calculated by taking the first-order derivative with respect to a certain factor [48]. The marginal effects are identified for each sample based on the BOL results, and the average marginal effects of all the observations are then reported.

2.4.2. SHAP

SHAP is an innovative ML explanatory method that combines the optimal likelihood distribution with local performance based on the Shapley value from game theory [49]. This method indicates that the sum of contributions equals the specific output of the model, which can be understood from the following mathematical formula:

$$F(x_i) = G(z_i) = \Phi_0 + \sum_{j=1}^m \Phi_{ij} z_{ij} \quad (16)$$

where $F(x_i)$ is the model output and $G(z_i)$ is an explanatory model for the interpretation of $F(x_i)$. $z_{ij} \in \{0, 1\}$, and when the feature j is observed, $z_{ij} = 1$; otherwise, $z_{ij} = 0$. Φ_0 is the initial output without features, and Φ_i is the Shapley value of feature i . In the end, Φ_{ij} can be defined as the difference between the Shapley value of factor i with and without feature j , which is weighted by the sum of the marginal contribution of feature j and reflects the feature importance. Therefore, the feature importance can be sorted with the calculated SHAP value. Thus, the model results can be consistently interpreted.

2.4.3. Partial Dependence Plot

The partial dependence plot is a visualization method for explaining the relationship between crash factors based on the prediction of the marginal benefit in an ML model [30]. This method explains the influence of values of the dependent features on the classification of severity levels. Similar to the marginal effects of BOL, this method presents the probability of the predicted outcome with continuous variable distribution, which reveals the detailed information about the risk factors. The mathematical equation is presented as follows:

$$f_s(X_s) = E_{X_c} f(X_s, X_c) \quad (17)$$

where S is the variable from which the partial dependence is derived and C denotes the dependent variable to S in the model, and E_{X_c} is the expected value of $f_s(X_s, X_c)$. According to the formulation, the partial dependence f_s identifies the marginal effects of variable S by marginalizing the ML model results over the distribution of factor C . Therefore, this method can be used in most ML models to evaluate the IR of the specific crash patterns.

3. Results

This study compared the performance of several ML models and identified the key factors related to IR based on feature analysis and feature importance in four parts: (1) comparing the predictive performance of several models with different feature selection (PCI, BF, and AF); (2) quantifying the marginal effects of crash factors based on the parameter estimation; (3) analyzing the feature importance that clarifies the effects of factors on injury severity; and (4) identifying thresholds of IR based on the partial dependence distribution.

3.1. Model Comparison

A requirement of this work is to compare the practical application of each selected model to verify the effectiveness of the proposed PCIs in crash severity evaluation. The comparison results for different severity levels after the 10-fold cross-validation and parameter optimization are shown in Table 3. In each ML method, the AF combination outperforms other combinations in severity classification, with an average of accuracy of 78.05% and an average AUC of 91.59%. By contrast, the average accuracy using initial crash information (BF feature set) is only 63.70%. This result means that the average accuracy of the six models could increase by 14.35% after the PCIs are applied, with the LCE–PCI model improving by 25.32%. The accuracy of the AF model could only be improved by 1.65% after the continued addition of nine BF variables compared with PCI features. This result demonstrates that the enhancement of the PCIs in classification and the proposed PCI-ML method is sufficient to obtain a good prediction accuracy with a concise form.

Table 3. Model performance of the different ML methods.

Feature Selection	Model	SEV1 (%)			SEV2 (%)			SEV3 (%)			Total (%)		
		TPR	FPR	ACC ¹	TPR	FPR	ACC	TPR	FPR	ACC	ACC	AUC	TPR
BF	BOL	81.67	35.22	71.94	23.59	11.24	71.17	67.77	14.07	80.37	63.74	80.80	20.18
	MLP	80.72	33.28	72.66	23.54	10.82 ²	71.46	69.62	15.91	79.67	61.89	79.93	20.00
	SVM	85.09	35.16	73.44	24.87	10.92	71.75	67.54	11.98	81.76	62.74	82.17	19.35
	RF	81.01	30.47	74.41	28.71	13.33	71.03	71.66	12.52	82.64	64.04	82.00	18.77
	DNN	81.63	27.39	76.44	30.97	12.88	71.97	74.96	12.84	83.43	65.92	84.02	17.70
	LCE	75.29	25.80	75.23	35.87	17.77	70.43	71.80	13.36	83.10	63.89	81.12	18.98
	Mean	80.90	31.22	74.02	27.93	12.83	71.30	70.56	13.45	81.83	63.70	81.67	19.17
PCI	BOL	85.68	35.78	74.83	20.51	11.07	71.96	70.39	11.38	84.54	71.41	85.51	19.41
	MLP	87.70	27.36	80.53	32.36	11.20	75.06	78.09	8.99	88.56	69.31	85.84	15.85
	SVM	89.27	27.94	80.86	34.51	10.45	76.11	79.81	6.88	90.55	73.55	89.24	15.09
	RF	87.11	19.50	84.80	48.25	11.78	78.93	83.88	6.54	92.03	77.64	91.30	12.61
	DNN	90.90	17.27	87.71	56.61	9.31	82.99	84.88	4.99	93.41	81.82	93.54	10.52
	LCE	92.88	9.43	93.05	73.84	7.32	89.09	90.43	3.23	96.32	88.98	97.03	6.66
	Mean	88.92	22.88	83.63	44.35	10.19	79.02	81.25	7.00	90.90	76.45	90.08	13.36
AF	BOL	85.23	33.16	76.14	26.10	10.39	73.97	74.78	10.45	86.54	73.58	87.14	18.00
	MLP	83.79	29.17	77.83	30.26	12.40	73.62	77.14	9.87	87.66	71.32	86.91	17.15
	SVM	94.68	27.94	83.16	37.38	7.15	79.38	83.97	3.97	93.84	77.44	93.27	13.02
	RF	88.25	16.62	86.95	54.87	10.92	81.34	84.97	6.09	92.68	79.74	92.86	11.21
	DNN	90.73	16.07	88.32	57.79	9.20	83.39	86.37	5.09	93.80	82.01	94.11	10.12
	LCE	93.37	9.21	93.38	74.51	7.27	89.32	90.49	2.88	96.60	89.21	97.23	6.45
	Mean	89.34	22.03	84.30	46.82	9.56	80.17	82.95	6.39	91.85	78.05	91.59	12.66
Total (mean)		86.39	25.38	80.65	39.70	10.86	76.83	78.25	8.95	88.19	72.74	87.78	15.06

¹ ACC stands for accuracy. ² The best performance for different categories is shown in bold.

The classification performance widely varied among the different models. The LCE-AF model can be considered the best evaluation method with the best accuracy of 89.21%, best AUC of 97.23%, and lowest TPR of only 6.45%. The prediction results also show that (1) the MLP models cannot effectively solve the complex nonlinear classification, and the accuracy is still lower than 72% after the addition of PCI features. (2) The RF and SVM methods perform better than the traditional methods and outperform the LCE model for the BF features. (3) The deep learning method (DNN) has been the most advanced technique due to its complex structure in the field of traffic prediction [44], but the misspecification issue

(high FPR) renders it an unsuitable algorithm for this dataset. (4) The LCE-based model can effectively identify the high-dimensional relationships between pre-crash information and injury level. The model accuracy is improved by 25.32% with proper feature engineering, which can be effectively used for IR assessment in pre-crash conditions.

The model performance for certain severity levels considerably varies when each level is regarded as a binary variable. SEV3 exhibited more evident crash patterns and higher accuracy, while the accuracy and sensitivity are lower at SEV2 due to the ambiguous mechanism for moderate severity. In addition, the SEV1 achieves the highest average TPR (86.39%) and the highest average FPR (25.38%), and the misclassification rate can be moderated with LCE. This finding suggests that the LCE method effectively reduces misclassification by mitigating the noise effect among factors.

3.2. Interpretation of the Marginal Effects

In contrast with the better prediction performance of the data mining models, the importance of parametric method (e.g., BOL) is in quantifying the effect of the underlying variables on the injury probabilities. Table 4 summarizes the parameter estimations and average marginal effects after excluding multicollinearity among factors. All the listed variables have statistically significant (95% confidence level) effects on crash severity.

Table 4. Results of the parameter estimation and average marginal effects.

Factor ¹	Parameter Estimation		Average Marginal Effects		
	Estimate	Std	SEV1 (%)	SEV2 (%)	SEV3 (%)
ΔV	0.0638	0.0002	−0.68	0.18	0.49
CSI	1.4531	0.0152	−22.6	6.2	16.39
CMI	1.3271	0.0210	−20.4	7.3	13.1
EES	0.0584	0.0004	−0.54	0.10	0.44
Light-duty	0.0357	0.2453	−0.57	0.16	0.4
Truck	−0.2947	0.2421	2.12	−0.62	−1.49
Proceeding	−0.2558	0.0893	4.02	−1.1	−2.91
Decelerating	−0.3289	0.0805	5.19	−1.42	−3.76
Overtake/merge	−0.6687	0.1356	10.4	−2.85	−7.54
Wrong side	0.4047	0.0494	−6.21	1.71	4.5
Intersecting	0.1716	0.0486	−2.79	0.76	2.02
Age	0.2157	0.0094	−0.34	0.09	0.24
Female	0.3739	0.0298	−5.79	1.59	4.2
Drink/drugs	0.4838	0.0756	−7.50	2.06	5.44
Speeding	0.1035	0.0347	−3.20	0.88	2.32
Speed limit	0.0062	0.0012	−0.09	0.03	0.06
Head-on	0.1445	0.0695	−5.57	1.53	4.04
Rear-end	−0.3594	0.0541	1.16	−0.34	−0.82
Lane number	−0.1358	0.0151	2.10	−0.57	−1.53
Curve	−0.1257	0.1615	1.95	−0.53	−1.41
Down slope	0.4302	0.1082	−6.67	1.83	4.83
Two-way, undivided	0.3709	0.0803	−5.75	−0.18	5.93
Ramp	0.3362	0.0799	−5.21	−0.33	5.54
0:00~6:00	0.2234	0.0561	−3.46	0.95	2.51
12:00~17:00	−0.0996	0.0384	1.55	−0.42	−1.12
Far side	−0.5598	0.0501	8.68	−2.38	−6.29
Rear side	−0.4606	0.0747	7.14	−1.96	−5.18
Near side	0.5286	0.0556	−8.19	2.25	5.94
Cut1	4.3961	0.3376			
Cut2	6.4116	0.3395			
Prob > χ^2 = 0.0001					
Pseudo R^2 = 0.2971					

¹ Considering the multicollinearity among factors, Vr, Ed, and other features are not included in this table.

Although most BF variables are not applied for ML models, various factors directly or indirectly influence the probability of severity. PCI, occupant attributes, driving behavior, crash type, environment, and roadway factors may influence the probability of injury severity. Specifically, higher speed-based (ΔV and EES) and system-based (CMI and CSI) crash indicators will result in more severe injuries. Taking ΔV as an example, the probability of moderate and major severity will increase by about 0.18% and 0.49%, respectively, for every 1 km/h increase in ΔV . Meanwhile, the probability of minor severity will decrease by 0.57%, with all other factors being equal. This notion explains the positive correlation between PCIs and IR.

Furthermore, the driving maneuver before crash events directly influenced the severity thresholds. Deceleration before crashes and driving on wrong side may result in an increase and decrease in the probability of minor injury by about 5.19% and 6.21%, respectively. Although some studies regard improper overtaking as risky behavior, lane changing or merging will decrease the probability of SEV3 by 7.54%. A plausible reason is that overtaking may result in the proximity to a crash, but it is not highly related to potential overcomes.

The impact position has a non-negligible effect on the threshold of IR. The results indicate that a vehicle subjected to the far and back-sides will increase the probability of minor injury by 8.68% and 7.14%, while that subjected to the near-side will increase the probability of moderate and major outcomes by 2.25% and 5.94%, respectively. In addition, head-on crashes are more prone to causing serious injury than rear-end ones. Considering that the crash characteristics (crash type and impact position) significantly influence the damage mechanism, the relationships between impact type, IR, and PCIs were further analyzed in detail as follows.

3.3. Feature Importance

The contribution of factors to the IR is explained by global and local SHAP values based on the outperforming LCE-CF model. Figure 4 depicts the global effect of each explanatory variable on severity levels. The PCIs have greater (or comparable) importance value than BF variables. Variable ΔV has the highest mean SHAP value, followed by CMI, EES, and V_r , indicating their superior utility in the IR assessment. In addition, most features are not sensitive to the SEV2 levels (low percentage of green bars), which indirectly explains the poor prediction performance of SEV2.

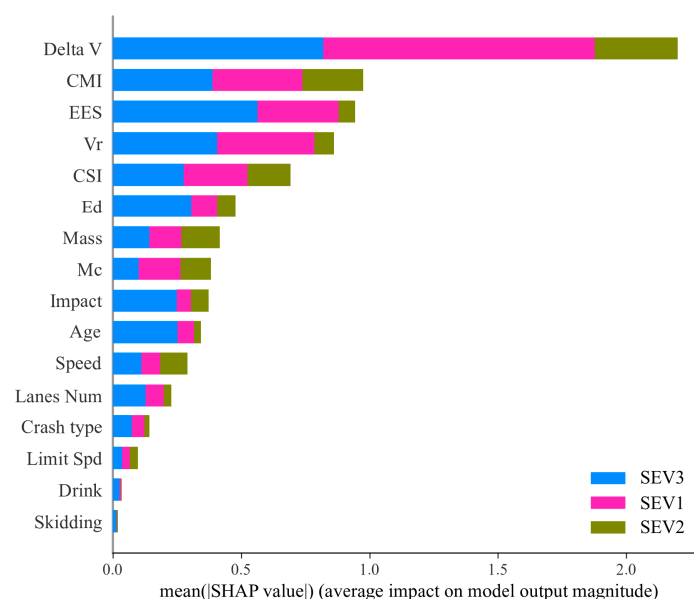


Figure 4. Global feature importance based on SHAP values.

Figure 5 provides a detailed interpretation of the different variables for the injury severity assessment. Each subplot provides three dimensions of information: (1) the features are ranked in a descending order according to their importance (average SHAP value); (2) the color of each point represents the corresponding feature value; and (3) the distribution of SHAP values in each row provides a local interpretation of features on a given severity level, with a positive SHAP value implying the enhancement of the corresponding variable value on the severity category, and vice versa.

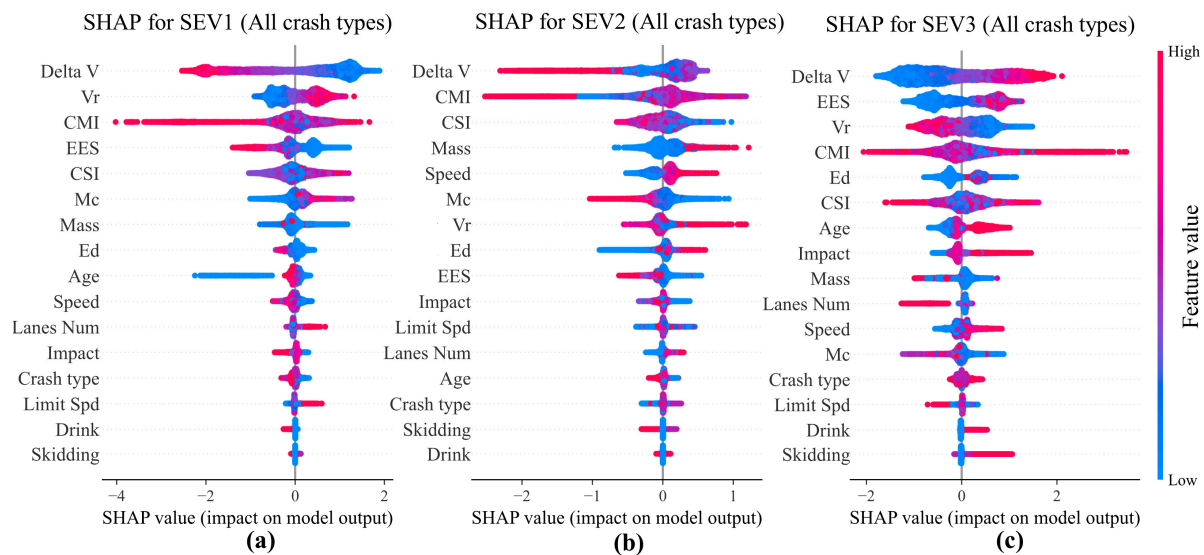


Figure 5. Distribution of local feature importance.

ΔV provides the best explanation for all severity levels: the potential outcomes become more severe with the ΔV increase, which is in agreement with relevant posterior crash studies and proves the validity of the proposed PCI method [50]. Although EES and CSI have similar effects to ΔV , they cannot be fully applied to SEV1 and SEV2. A possible reason is that the definition of EES neglects the influence of restitution coefficient ϵ (Equation (10)), which cannot be neglected during a low-speed impact.

In terms of basic features, mass and M_c (referring to system mass) indirectly respond to the stiffness and structural properties of the vehicle body, which are proven to be effectively used for the identification of moderate outcomes. Moreover, occupant age and crash characteristics significantly affect the prediction of severe crashes. In particular, near-side impacts (corresponding to a high value of impact) and older occupants are prone to major injury, which is consistent with the marginal effect of BOL. Fewer lanes, higher approaching speed, skidding, and drinking also exacerbated the injury severity, but such information is relatively limited in the IR assessment.

To better investigate the difference in crash types, Figure 6 shows the SHAP value for head-on, rear-end, and side impact crashes. The potential influence of most features on the different severity levels is similar to that in Figure 6. However, the rank of features greatly varies for different types. The results indicate that judging the potential risk of side-impact crashes only through velocity-based indicators is insufficient because the CMI and CSI provide a better interpretation of the degree of injury. The potential reason is that the eccentricity of impact significantly influences the energy conversion for the complex angle crashes, which can be directly reflected with the distribution of the system-based indicators. The impact position also shows high association with the injury severity for side-impact and rear-end crashes.

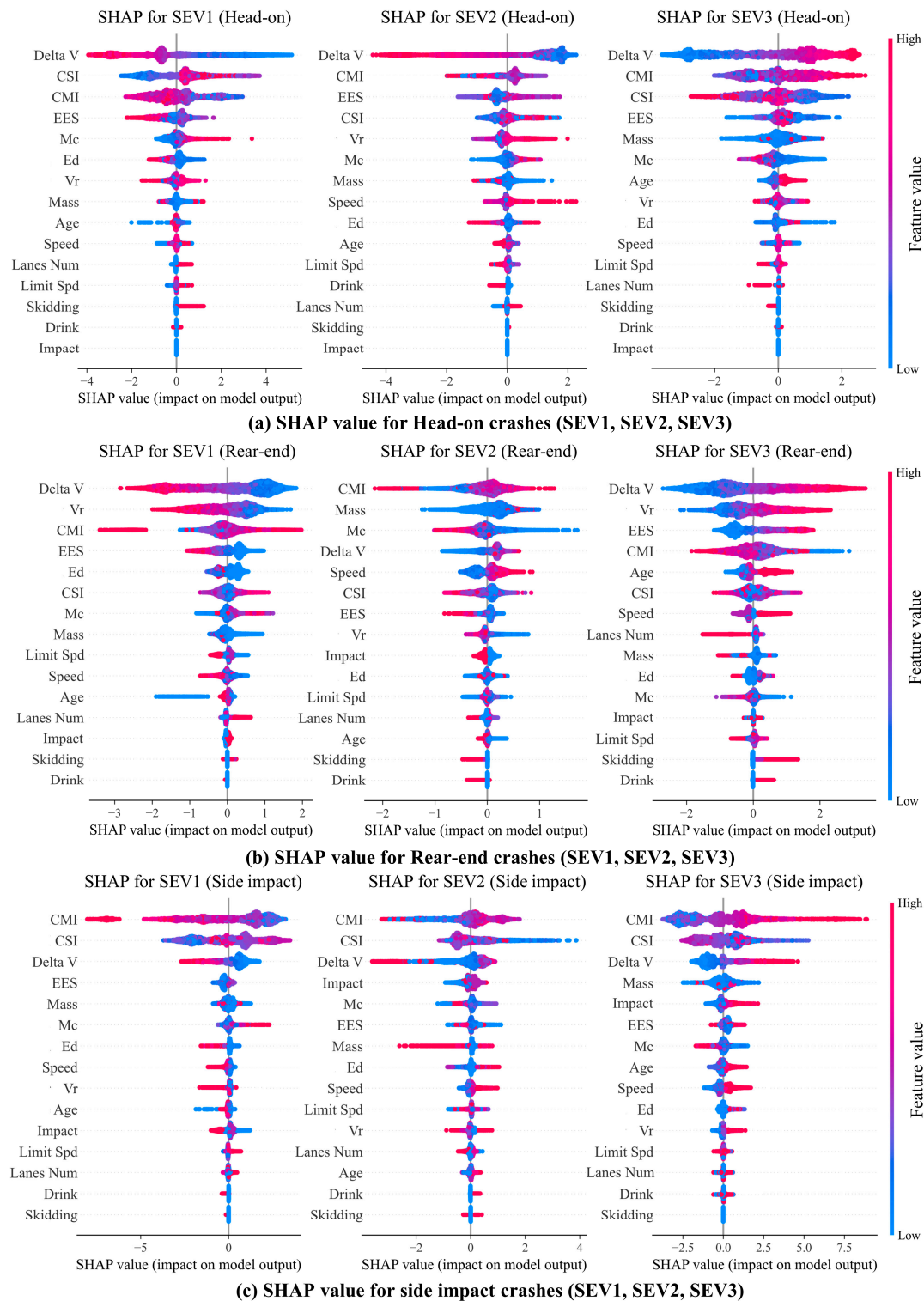


Figure 6. Ranking of SHAP values of the different crash types.

3.4. Analysis of Partial Dependency

Figures 4–6 qualitatively analyze the association between injury severity and risk factors. To further quantify the influence of the PCIs on crash risk, Figure 7 shows the partial dependence of ΔV , CMI, and EES derived from the LCE-CF model, reflecting the injury severity probability for the specific feature value. The increase in these three indicators results in more severe outcomes, and the distribution will stabilize after a certain value. We

labeled the 40th and 70th percentiles of each variable with a dashed line in Figure 7, which is close to the proportion of each severity category in the overall dataset (43.1% vs. 27.9% vs. 29.0% for SEV1–SEV3). The partial dependence of each injury level almost peaks within the expected interval, which indicates that the delineated critical intervals correspond to high probability area. In addition, these critical values also converge on the inflection points of distribution curves.

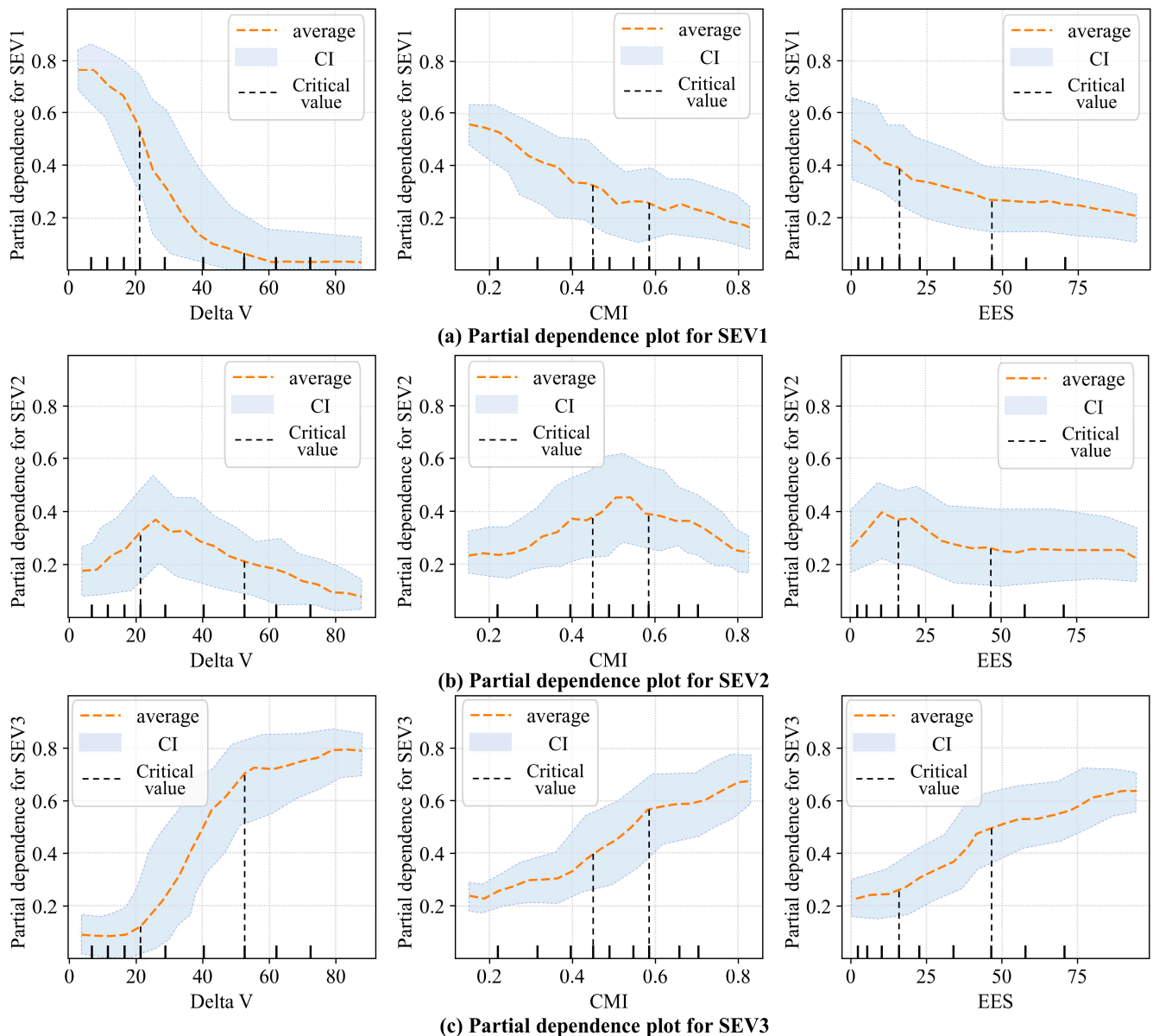


Figure 7. Partial dependence plot of ΔV , CMI, and EES. CI is the confidence interval of individual expectation distribution.

ΔV , as the key factor identified, directly affects the probability of injury severity: when ΔV is greater than 17.6 km/h, the dependence curves of SEV2 and SEV3 sharply increase; when ΔV equals 25.8 km/h, the probability of SEV2 reaches its maximum, close to 37.6%; and when ΔV is greater than 53.1 km/h, the probability of minor injury tends to 0%–4.5%, and the probability of a severe injury is greater than 70%, indicating a rather serious IR. This actual probability distribution is similar to the fitted S-curve in a previous study [50] and further provides more details about the moderate level (SEV2). Larger EES and CMI also elevated the likelihood of serious injury, but their sensitivity is lower than ΔV .

To verify the effect of impact position on casualty risk, Figure 8 shows the partial dependence of ΔV for the different impact positions. The trend of the IR for these four impacts can be ranked in a descending order as near, front, far, and back sides. When a vehicle is exposed to a backside impact, the potential IR is lower than otherwise because more than 80% of backside impacts result in only minor or moderate severity. By contrast, near-side impact tends to the highest casualty risk probability. When ΔV is greater than 50 km/h, the probability of a near-side crash resulting in a serious accident is higher than 80%, while the corresponding probability of a back side impact is only 12.4%. In addition, the partial dependence distribution of the front side and near side impacts is similar to the overall probability in Figure 8.

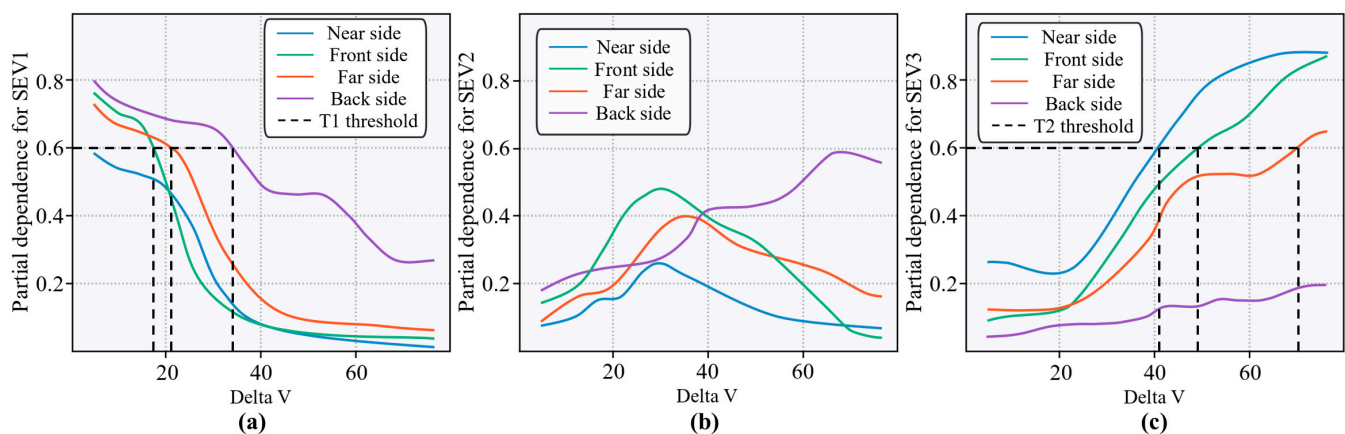


Figure 8. Partial dependence plot of ΔV for the different impact positions.

The partial dependence at each severity level has been integrated to further quantify the impact of the proposed PCIs. According to the threshold setting guidelines from the United States CDC expert panel [4], the critical thresholds T1 and T2 of ΔV , CMI, and EES for three severity levels are summarized in Table 5. The criteria for threshold delineation are represented by the dashed lines labeled as T1 and T2 in Figure 8. These lines serve as a reference for the 60% acceptance rate of dependence distribution for the different severity probabilities. Values below T1 or above T2 imply that the probability of SEV1 or SEV3 will surpass the other cases.

Table 5. Thresholds of injury risk for ΔV , CMI, and EES.

	ΔV (km/h)		CMI (%)		EES (km/h)	
	T1	T2	T1	T2	T1	T2
Near side	/	40.16	33.58	53.29	5.31	33.61
Front side	17.76	48.71	36.64	56.82	10.13	40.95
Far side	21.34	70.04	39.46	57.96	13.97	57.81
Back side	35.83	/	41.01	/	23.20	/

Higher thresholds for T1 and T2 correspond to increased safety margins for the corresponding impact type. This means that in the case of a far-side impact (T2 with $\Delta V = 70.04$ km/h) or a back-side impact (T1 with $\Delta V = 35.83$ km/h), the risk of severe injury to occupants is significantly reduced. For these two types, the vehicle body provides additional buffer space, allowing the structure and safety systems to better absorb and distribute the impact energy. As a result, the force exerted on the occupants is mitigated, reducing the potential for severe injuries. These findings align with previous studies [12] and provide further support to the robustness of the conclusions regarding crash mechanisms.

4. Discussion

In the above sections, this work demonstrates the optimization effect of the proposed severity parameters based on the pre-crash information in the IR assessment of two-vehicle crashes. According to the results in Table 3, the PCIs significantly improve the prediction performance compared with the underlying crash information, regardless of the ML model used. This finding not only verifies the validity of a priori severity evaluation for latent traffic conflicts but also corroborates the positive effect of the proposed PCIs on crash risk assessment. In terms of model selection, the advanced bagging–boosting model (LCE) exhibits a higher applicability, higher accuracy, and lower misclassification rate than other models. More accurate models usually better capture the underlying relationship between injury severity and risk factors. Accordingly, the LCE model combining PCIs obtains satisfactory prediction performance and can acquire complex patterns under the pre-crash conditions. This model offers a robust alternative for the further analysis of roadway safety.

ML methods are frequently criticized as “black boxes” because of their lack of clarity and explicability [51]. These drawbacks have affected the widespread use of non-parametric models, even though they are flexible in data adaptation and have advantages in outstanding predictive accuracy. This study calculates the marginal effects of different features based on BOL and ranks the feature importance with SHAP value to better interpret the model results. Partial dependence plots were applied to visualize the injury probability for the different impact positions for the key PCIs. The proposed framework provides a new perspective on the interpretability of predictive models, which effectively quantify the corresponding effects between crash indicator values and IR (Figures 5–8).

The SHAP value suggests that all six PCIs have the largest contribution in the injury severity assessment. Such a priori PCIs are derived from the well-established crash reconstruction theory, which has been proven to reflect the crash risk to a certain extent [52]. Based on the results, ΔV provides the best interpretation for all severity levels. We considered a sideswipe collision between two low-speed vehicles, in which the likelihood of injury is relatively minor, and the ΔV of both vehicles during the crash is equally low. When two vehicles with large differences in mass and speed have potential conflicts in the approach trajectory, the higher ΔV increases the IR in this case. With regard to system-based dimensionless parameters, the predictive performance of the CMI is higher than that of the CSI. Both indicators are related to vehicle geometry and structural stiffness. However, CSI mainly measures the energy absorbed during the crash, while different forces and deformations may correspond to the same energy loss [53]. By contrast, CMI is directly related to the impact eccentricity, and the increase in the impact eccentricity results in a decrease in the rate of kinetic energy converted into translation. Thus, CMI outperforms most of the other variables in side-impact crashes (Figure 7c).

The partial dependence plots depict a nonlinear relationship between PCIs and IR with intuitive visualization compared with the marginal effects of logit-based estimation. The effect of PCIs varies for the different crash types due to the variation in crash mechanisms, especially for head-on and side-impact crashes. In the side-impact cases, the passengers are impacted in a complex way with the safety protection systems and the internal cabin, and the high-dimensional parameters are instrumental in optimizing crash analysis. This finding also demonstrates that different analytical approaches should be applied considering crash types and impact position. Therefore, the thresholds of IR with ΔV , CMI, and EES are presented for different impact positions based on the prevailing threshold setting guidelines. The threshold division addresses a critical issue in a previous study in that it provides a certain border for a serious or non-serious outcome conflict [8]. Further application can be conducted with an evident threshold criterion considering various acceptance levels and under-triage control in an actual scene.

In all impact types, near-side impact is more likely to result in severe outcomes. The marginal effect may be due to the relatively limited side protection capacity of light-duty cars. Automobile manufacturers should strengthen the side protection design of vehicles, including a side airbag, forced door, and B-pillar structure [54,55]. Volvo’s report showed

that enforced side protection can reduce the overall risk of injury by more than 70% [56]. According to the proposed threshold value, the ideal vehicle design should effectively protect the safety of the near side passengers with the ΔV of 20 km/h. Correspondingly, the advanced driver-assistance (ADAS) system or AACN system should also evaluate the risk of traffic conflict induced by the impact side.

The findings of this work are mainly based on the KABCO severity criterion, while similar conclusions have been validated in the post-crash study within other evaluation frameworks (e.g., MAIS, AIS, and ISS) [19,57]. The a priori assessment approach can be applied to any active safety system based on crash analysis because all the required parameters (speed, angle, passengers, and vehicles attributes) can be retrieved from the operation information and on-board vehicle sensors. Given the emerging trend of the traffic conflict identification toward automated analyses [58], the combination of trajectory prediction and a priori risk assessment holds the potential to effectively improve the active safety performance.

Furthermore, the implications of this study extend beyond the realm of pre-crash injury risk assessment, offering opportunities to improve overall roadway safety. The accurate assessment of potential conflicts or crashes and the identification of key indicators of injury risk enable the development of proactive safety measures and targeted interventions. Transportation agencies and policymakers can leverage the insights gained from this study to prioritize infrastructure improvements, identify hazardous driving behaviors and black-point roadway locations, and allocate resources effectively for traffic enforcement [59,60]. Additionally, automotive manufacturers can utilize the identified determinants of injury risk to inform the design of vehicle structures and incorporate advanced safety technologies that address specific crash scenarios and impact positions [61,62]. By translating the research findings into practical applications, this study contributes to the ongoing efforts aimed at enhancing road safety and mitigating the human and economic costs associated with crash injuries.

5. Conclusions

Examination of the pre-crash factors can offer insights into the underlying relationships between crash mechanisms and injury severity. This work highlights the importance of pre-crash analysis and ML interpretation in a priori IR evaluation. The calculation of PCIs under pre-crash conditions is introduced on the basis of the impulse–momentum theory. Considering many existing criteria for injury severity, an ML-based framework is proposed to compare the explanatory variables related to the crash severity with considerable samples. Six models were used to assess the injury severity within a fusion FARS/CRSS dataset and investigate the potential influence of the different crash characteristics. The differences between the basic features and the PCIs in margin effects were also compared based on parameter estimation of BOL. The key crash factors for different crash types were then identified from the outperforming LCE-AF model by interpreting the SHAP value and partial dependence plot. The thresholds of PCIs for several impact positions are also proposed to evaluate the IR, which can be used as a reference for targeted safety strategies. The following conclusions are drawn:

- (1) The PCIs derived from the pre-crash conditions display an excellent optimization effect in injury severity assessment. Among the selected models, the LCE method with PCI features is effective in handling complex pre-crash features and a priori information and can obtain the best performance with an AUC of 97.23% and an accuracy of 89.21%.
- (2) Based on the SHAP value and partial dependence distribution, ΔV is identified as the most valuable factor for synthetical IR evaluation, followed by CMI, EES, V_r , and CSI. We also determined the nonlinear relationship between the built PCI features and the likelihood of IRs, which is often disregarded in traditional statistical methods.
- (3) With regard to the impact side, occupants subjected to near-side impacts more severely suffered from crashes, and the protective effect reflects that the back side of vehicles

has better-designed structures to reduce the impulse from the impacted vehicle. Based on the partial dependence analysis, the probability decision thresholds of ΔV , CMI, and EES were developed considering different impact sides.

The key finding from this work is that the indicators derived from the initial pre-crash information are found to be stronger than the more commonly used basic features in IR assessment. This feature may be attributed to the severity indices obtained from crash momentum analysis that can reflect energy transformations and mechanisms of vehicle-to-vehicle collision. The developed PCI formulation provides useful insights, which can serve as a reference for the transfer from post-crash passive disposal to pre-crash active prevention.

Nowadays, a large amount of plausible data can be provided through the V2V communications and ADAS system due to the development of the smart transportation and information technologies. As a potential issue for further research, we aim to better model IRs and discover unobserved heterogeneity for complex crash characteristics on high-dimensional datasets based on the proposed framework. Therefore, the combination of explainable ML models and safety field knowledge is considered a promising research direction. In addition, this study only compared the expected variables of injury severity due to the lack of vehicle operational data. Further study can explore a novel framework for measuring the conflict risk considering the proximity to a crash and the severity level of the potential outcomes.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app13126983/s1>, Table S1: Crash data.

Author Contributions: Conceptualization, C.G. (Chenwei Gu) and J.X.; Data curation, S.L., C.G. (Chao Gao) and Y.M.; Formal analysis, C.G. (Chenwei Gu); Funding acquisition, J.X.; Investigation, C.G. (Chenwei Gu) and Y.M.; Methodology, C.G. (Chenwei Gu) and S.L.; Project administration, J.X.; Resources, J.X. and C.G. (Chao Gao); Software, C.G. (Chenwei Gu) and S.L.; Supervision, C.G. (Chao Gao); Validation, C.G. (Chenwei Gu); Visualization, C.G. (Chenwei Gu); Writing—original draft, C.G. (Chenwei Gu); Writing—review and editing, C.G. (Chenwei Gu) and J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Fundamental Research Funds for the Central Universities, Chang'an University, grant number: 300102212107.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://cdan.nhtsa.gov/> (accessed on 6 June 2023).

Acknowledgments: The authors would like to thank the reviewers and the editors for their valuable comments, which improved the quality of this paper.

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

References

1. World Health Organization. *Global Plan Decade of Action for Road Safety 2021–2030*; WHO: Geneva, Switzerland, 2021.
2. Suarez-del Fuego, R.; Junge, M.; Lopez-Valdes, F.; Gabler, H.C.; Woerner, L.; Hiermaier, S. Cluster analysis of seriously injured occupants in motor vehicle crashes. *Accid. Anal. Prev.* **2021**, *151*, 105787. [[CrossRef](#)] [[PubMed](#)]
3. Oorni, R.; Goulart, A. In-Vehicle Emergency Call Services: eCall and Beyond. *IEEE Commun. Mag.* **2017**, *55*, 159–165. [[CrossRef](#)]
4. CDC. *Advanced Automatic Collision Notification and Triage of the Injured Patient*; CDC: Atlanta, GA, USA, 2008.
5. Shi, X.; Wong, Y.D.; Li, M.Z.F.; Chai, C. Key risk indicators for accident assessment conditioned on pre-crash vehicle trajectory. *Accid. Anal. Prev.* **2018**, *117*, 346–356. [[CrossRef](#)] [[PubMed](#)]
6. Yue, L.; Abdel-Aty, M.; Wu, Y.; Ugan, J.; Yuan, C. Effects of forward collision warning technology in different pre-crash scenarios. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *76*, 336–352. [[CrossRef](#)]
7. Bálint, A.; Fagerlind, H.; Kullgren, A. A test-based method for the assessment of pre-crash warning and braking systems. *Accid. Anal. Prev.* **2013**, *59*, 192–199. [[CrossRef](#)]

8. Laureshyn, A.; De Ceunynck, T.; Karlsson, C.; Svensson, Å.; Daniels, S. In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator. *Accid. Anal. Prev.* **2017**, *98*, 46–56. [\[CrossRef\]](#)
9. Gulino, M.S.; Fiorentino, A.; Vangi, D. Prospective and retrospective performance assessment of Advanced Driver Assistance Systems in imminent collision scenarios: The CMI-Vr approach. *Eur. Transp. Res. Rev.* **2022**, *14*, 3. [\[CrossRef\]](#)
10. Arun, A.; Haque, M.M.; Washington, S.; Sayed, T.; Mannering, F. How many are enough?: Investigating the effectiveness of multiple conflict indicators for crash frequency-by-severity estimation by automated traffic conflict analysis. *Transp. Res. Part C Emerg. Technol.* **2022**, *138*, 103653. [\[CrossRef\]](#)
11. Arun, A.; Haque, M.M.; Bhaskar, A.; Washington, S.; Sayed, T. A bivariate extreme value model for estimating crash frequency by severity using traffic conflicts. *Anal. Methods Accid. Res.* **2021**, *32*, 100180. [\[CrossRef\]](#)
12. Yasmin, S.; Eluru, N.; Pinjari, A.R. Pooling data from fatality analysis reporting system (FARS) and generalized estimates system (GES) to explore the continuum of injury severity spectrum. *Accid. Anal. Prev.* **2015**, *84*, 112–127. [\[CrossRef\]](#)
13. Wang, J.; Zheng, Y.; Li, X.; Yu, C.; Kodaka, K.; Li, K. Driving risk assessment using near-crash database through data mining of tree-based model. *Accid. Anal. Prev.* **2015**, *84*, 54–64. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Jeong, H.; Jang, Y.; Bowman, P.J.; Masoud, N. Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. *Accid. Anal. Prev.* **2018**, *120*, 250–261. [\[CrossRef\]](#)
15. Zeng, Q.; Wang, Q.; Wang, X. An empirical analysis of factors contributing to roadway infrastructure damage from expressway accidents: A Bayesian random parameters Tobit approach. *Accid. Anal. Prev.* **2022**, *173*, 106717. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Atkinson, T.; Gawarecki, L.; Tavakoli, M. Paired vehicle occupant analysis indicates age and crash severity moderate likelihood of higher severity injury in second row seated adults in frontal crashes. *Accid. Anal. Prev.* **2016**, *89*, 88–94. [\[CrossRef\]](#)
17. Vadeby, A.M. Modeling of relative collision safety including driver characteristics. *Accid. Anal. Prev.* **2004**, *36*, 909–917. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Newgard, C.D. Defining the “older” crash victim: The relationship between age and serious injury in motor vehicle crashes. *Accid. Anal. Prev.* **2008**, *40*, 1498–1505. [\[CrossRef\]](#)
19. Gulino, M.-S.; Gangi, L.D.; Sortino, A.; Vangi, D. Injury risk assessment based on pre-crash variables: The role of closing velocity and impact eccentricity. *Accid. Anal. Prev.* **2021**, *150*, 105864. [\[CrossRef\]](#)
20. Riviere, C.; Lauret, P.; Ramsamy, J.F.M.; Page, Y. A Bayesian Neural Network approach to estimating the Energy Equivalent Speed. *Accid. Anal. Prev.* **2006**, *38*, 248–259. [\[CrossRef\]](#)
21. Kusano, K.; Gabler, H.C. Comparison and Validation of Injury Risk Classifiers for Advanced Automated Crash Notification Systems. *Traffic Inj. Prev.* **2014**, *15*, S126–S133. [\[CrossRef\]](#)
22. Vangi, D. Impact severity assessment in vehicle accidents. *Int. J. Crashworthiness* **2014**, *19*, 576–587. [\[CrossRef\]](#)
23. Vangi, D.; Gulino, M.-S.; Fiorentino, A.; Virga, A. Crash momentum index and closing velocity as crash severity index. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2019**, *233*, 3318–3326. [\[CrossRef\]](#)
24. Ji, A.; Levinson, D. An energy loss-based vehicular injury severity model. *Accid. Anal. Prev.* **2020**, *146*, 105730. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Brach, M.; Mason, J.; Brach, R.M. *Vehicle Accident Analysis and Reconstruction Methods*; Sae International: Pittsburgh, PA, USA, 2011.
26. Montella, A.; Aria, M.; D’Ambrosio, A.; Mauriello, F. Data-Mining Techniques for Exploratory Analysis of Pedestrian Crashes. *Transp. Res. Rec. J. Transp. Res. Board* **2011**, *2237*, 107–116. [\[CrossRef\]](#)
27. Montella, A.; Mauriello, F.; Perneti, M.; Riccardi, M.R. Rule discovery to identify patterns contributing to overrepresentation and severity of run-off-the-road crashes. *Accid. Anal. Prev.* **2021**, *155*, 106119. [\[CrossRef\]](#)
28. Jiang, F.; Ma, J. A comprehensive study of macro factors related to traffic fatality rates by XGBoost-based model and GIS techniques. *Accid. Anal. Prev.* **2021**, *163*, 106431. [\[CrossRef\]](#)
29. Murdoch, W.J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22071–22080. [\[CrossRef\]](#)
30. Wen, X.; Xie, Y.; Jiang, L.; Li, Y.; Ge, T. On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development. *Accid. Anal. Prev.* **2022**, *168*, 106617. [\[CrossRef\]](#)
31. Guo, M.; Zhao, X.; Yao, Y.; Yan, P.; Su, Y.; Bi, C.; Wu, D. A study of freeway crash risk prediction and interpretation based on risky driving behavior and traffic flow data. *Accid. Anal. Prev.* **2021**, *160*, 106328. [\[CrossRef\]](#)
32. Chen, A.Y.; Chiu, Y.-L.; Hsieh, M.-H.; Lin, P.-W.; Angah, O. Conflict analytics through the vehicle safety space in mixed traffic flows using UAV image sequences. *Transp. Res. Part C Emerg. Technol.* **2020**, *119*, 102744. [\[CrossRef\]](#)
33. Zheng, L.; Sayed, T. From univariate to bivariate extreme value models: Approaches to integrate traffic conflict indicators for crash estimation. *Transp. Res. Part C Emerg. Technol.* **2019**, *103*, 211–225. [\[CrossRef\]](#)
34. Wood, D.P. Safety and the car size effect: A fundamental explanation. *Accid. Anal. Prev.* **1997**, *29*, 139–151. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Kolk, H.; Tomasch, E.; Sinz, W.; Bakker, J.; Dobberstein, J. Evaluation of a momentum based impact model and application in an effectivity study considering junction accidents. In Proceedings of the ESAR—7th International Conference: “Expert Symposium on Accident Research”, Hanover, Germany, 7–8 September 2017.
36. Rapp van Roden, E.; Zolock, J. Using the Instantaneous Center of Rotation to Examine the Influence of Yaw Rate on Occupant Kinematics in Eccentric Planar Collisions. *SAE Int. J. Adv. Curr. Pract. Mobil.* **2022**, *5*, 266–283.
37. Fatzinger, E.; Landerville, J. *Using Vehicle EDR Data to Calculate Motorcycle Delta-V in Motorcycle-Vehicle Lateral Front End Impacts*; SAE Technical Paper: Warrendale, PA, USA, 2020.

38. Miltner, E.; Salwender, H.J. Influencing factors on the injury severity of restrained front seat occupants in car-to-car head-on collisions. *Accid. Anal. Prev.* **1995**, *27*, 143–150. [[CrossRef](#)] [[PubMed](#)]
39. Lu, Y.; Liu, Y.; Shu, Y.; Ma, L. Crash recognition algorithm of automatic crash notification system with adaptive discrimination threshold. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2022**, *237*, 410–425. [[CrossRef](#)]
40. Paul, M.; Ghosh, I. Development of conflict severity index for safety evaluation of severe crash types at unsignalized intersections under mixed traffic. *Saf. Sci.* **2021**, *144*, 105432. [[CrossRef](#)]
41. Harris, W.; Trueblood, A.B.; Brooks, R.D.; Brown, S. *Fatal and Nonfatal Transportation Injuries in the Construction Industry, 2011–2020*; CDC: Atlanta, GA, USA, 2022.
42. Topuz, K.; Delen, D. A probabilistic Bayesian inference model to investigate injury severity in automobile crashes. *Decis. Support Syst.* **2021**, *150*, 113557. [[CrossRef](#)]
43. Goh, Y.M.; Ubeynarayana, C.U. Construction accident narrative classification: An evaluation of text mining techniques. *Accid. Anal. Prev.* **2017**, *108*, 122–130. [[CrossRef](#)]
44. Zhang, Z.; He, Q.; Gao, J.; Ni, M. A deep learning approach for detecting traffic accidents from social media data. *Transp. Res. Part C Emerg. Technol.* **2018**, *86*, 580–596. [[CrossRef](#)]
45. Fauvel, K.; Fromont, É.; Masson, V.; Faverdin, P.; Termier, A. XEM: An explainable-by-design ensemble method for multivariate time series classification. *Data Min. Knowl. Discov.* **2022**, *36*, 917–957. [[CrossRef](#)]
46. Gumustekin, S.; Senel, T.; Cengiz, M.A. A Comparative Study on Bayesian Optimization Algorithm for Nutrition Problem. *J. Food Nutr. Res.* **2014**, *2*, 952–958. [[CrossRef](#)]
47. Chen, X.W.; Jeong, J.C. Enhanced recursive feature elimination. In Proceedings of the Sixth International Conference on Machine Learning and Applications (ICMLA 2007), Washington, DC, USA, 13–15 December 2007; pp. 429–435.
48. Jalayer, M.; Shabanpour, R.; Pour-Rouholamin, M.; Golshani, N.; Zhou, H. Wrong-way driving crashes: A random-parameters ordered probit analysis of injury severity. *Accid. Anal. Prev.* **2018**, *117*, 128–135. [[CrossRef](#)] [[PubMed](#)]
49. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
50. Nishimoto, T.; Mukaigawa, K.; Tominaga, S.; Lubbe, N.; Kiuchi, T.; Motomura, T.; Matsumoto, H. Serious injury prediction algorithm based on large-scale data and under-triage control. *Accid. Anal. Prev.* **2017**, *98*, 266–276. [[CrossRef](#)] [[PubMed](#)]
51. Warner, B.; Misra, M. Understanding Neural Networks as Statistical Tools. *Am. Stat.* **1996**, *50*, 284–293. [[CrossRef](#)]
52. Hauschild, H.; Halloway, D.; Pintar, F. Delta-v slope as an indicator of injury. *Traffic Inj. Prev.* **2021**, *22*, S165–S169. [[CrossRef](#)] [[PubMed](#)]
53. Husted, D.C.; Biss, D.J.; Heverly, D.E. *The Appropriate Use of “Delta-V” in Describing Accident Severity*; SAE Technical Paper; SAE International: Warrendale, PA, USA, 1999.
54. Gaylor, L.; Junge, M.; Abanteriba, S. Efficacy of seat-mounted thoracic side airbags in the German vehicle fleet. *Traffic Inj. Prev.* **2017**, *18*, 852–858. [[CrossRef](#)]
55. Pal, C.; Okabe, T.; Sakurai, M.; Masashi, M.; Vimalathithan, K. Development of High Efficiency Load Path Structure to Enhance Side Impact Safety Performance. In Proceedings of the 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration, Gothenburg, Sweden, 8–11 June 2015.
56. Jakobsson, L.; Isaksson-hellman, I.; Lindman, M. WHIPS (Volvo Cars’ Whiplash Protection System)—The Development and Real-World Performance. *Traffic Inj. Prev.* **2008**, *9*, 600–605. [[CrossRef](#)]
57. Sobhani, A.; Young, W.; Logan, D.; Bahrololoom, S. A kinetic energy model of two-vehicle crash injury severity. *Accid. Anal. Prev.* **2011**, *43*, 741–754. [[CrossRef](#)]
58. Saunier, N.; Sayed, T.; Ismail, K. Large-Scale Automated Analysis of Vehicle Interactions and Collisions. *Transp. Res. Rec.* **2010**, *2147*, 42–50. [[CrossRef](#)]
59. Gu, C.; Xu, J.; Gao, C.; Mu, M.E.G.; Ma, Y. Multivariate analysis of roadway multi-fatality crashes using association rules mining and rules graph structures: A case study in China. *PLoS ONE* **2022**, *17*, e0276817. [[CrossRef](#)]
60. Wu, J.; Chen, X.; Bie, Y.; Zhou, W. A co-evolutionary lane-changing trajectory planning method for automated vehicles based on the instantaneous risk identification. *Accid. Anal. Prev.* **2023**, *180*, 106907. [[CrossRef](#)]
61. Sahraei, E.; Digges, K.; Marzougui, D.; Roddis, K. High strength steels, stiffness of vehicle front-end structure, and risk of injury to rear seat occupants. *Accid. Anal. Prev.* **2014**, *66*, 43–54. [[CrossRef](#)] [[PubMed](#)]
62. Song, Y.; Chitturi, M.V.; Noyce, D.A. Intersection two-vehicle crash scenario specification for automated vehicle safety evaluation using sequence analysis and Bayesian networks. *Accid. Anal. Prev.* **2022**, *176*, 106814. [[CrossRef](#)] [[PubMed](#)]

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.