

## Article

# Traffic Safety Assessment and Injury Severity Analysis for Undivided Two-Way Highway–Rail Grade Crossings

Qiaoqiao Ren <sup>1</sup>, Min Xu <sup>1,\*</sup> , Bojian Zhou <sup>2</sup> and Sai-Ho Chung <sup>1</sup>

<sup>1</sup> Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China; qiaoqiao.ren@connect.polyu.hk (Q.R.); nick.sh.chung@polyu.edu.hk (S.-H.C.)

<sup>2</sup> School of Transportation, Southeast University, Nanjing 210096, China; zbj387@gmail.com

\* Correspondence: min.m.xu@polyu.edu.hk

**Abstract:** The safety and reliability of undivided two-way highway–rail grade crossings (HRGCs) are of paramount importance in transportation systems. Utilizing crash data from the Federal Railroad Administration between 2020 and 2021, this study aims to predict crash injury severity outcomes and investigate various factors influencing injury severities. The  $\chi^2$  test was first used to select variables that were significantly associated with injury outcomes. By employing the eXtreme Gradient Boosting (XGBoost) model and interpretable SHapley Additive exPlanations (SHAP), a cross-category safety assessment that offers an evidence-based hierarchy and statistical inference of risk factors associated with crashes, crossings, vehicles, drivers, and environment was provided for killed, injured, and uninjured outcomes. Some significant predictors overlapped between the killed and injured models, such as old driver, driver was in vehicle, main track, went around the gate, adverse crossing surface, and truck, while the other different significant factors revealed that the model could distinguish between different severity levels. Additionally, the results suggested that the model has varying performances in predicting different injury severities, with the killed model having the highest accuracy of 93.36%. The SHAP dependency plots for the top three features also ensure reliable predictions and inform potential interventions aimed at strengthening traffic safety and risk management practices, such as enhanced warning systems and targeted educational campaigns for older drivers.



**Citation:** Ren, Q.; Xu, M.; Zhou, B.; Chung, S.-H. Traffic Safety Assessment and Injury Severity Analysis for Undivided Two-Way Highway–Rail Grade Crossings. *Mathematics* **2024**, *12*, 0. <https://doi.org/10.3390/math12040519>

Academic Editor: Kamel Rekab

Received: 5 January 2024

Revised: 18 January 2024

Accepted: 3 February 2024

Published: 7 February 2024



**Copyright:** © 2024 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:** transportation system; traffic safety; reliability estimation and mathematical statistics; risk management; injury severity; HRGC crashes; XGBoost; SHAP

**MSC:** 62G86; 62H10; 62H12; 68T09

## 1. Introduction

Highway–rail grade crossings (HRGCs) represent pivotal junctures within the transportation network where the controlled flow of vehicular traffic intersects with rail infrastructure. Although integral to multimodal transportation systems, HRGCs continue to pose substantial safety challenges worldwide. For example, in the United States, the Federal Railroad Administration reports there were 2146 HRGC crashes in 2021, resulting in considerable casualties and property damage [1]. Within this context, it warrants special attention that the interplay between undivided two-way traffic highways and railways carries an increased risk. The lack of separation between opposing traffic flows could introduce a variety of hazard factors that can lead to catastrophic accidents. Vehicles may inadvertently drift into the opposing lane, leading to head-on collisions, or encroach upon railway tracks, placing them directly in the path of oncoming trains. In this regard, Ref. [2] pointed out that varying median separations on rural two-lane roads have a considerable impact on how drivers behave, specifically in terms of their lateral positioning and driving trajectories, along with the related safety risks. Furthermore, it has been observed that

collisions tend to happen more frequently on undivided two-way roads, which is attributed to the fluctuating traffic circumstances [3]. The statistics also reveal that nearly 95% of HRGC incidents take place in two-way traffic scenarios without divisions [1], highlighting the critical necessity to identify associated determinants and improve crossing safety in this scenario.

Studies on the safety of HRGCs have traditionally focused on the statistical analysis of accident data to model injury severity outcomes [4,5]. For instance, Ref. [6] employed an ordered probit approach to investigate how age and gender influence the severity of injuries sustained by motor vehicle drivers at HRGCs, uncovering significant differences in the behavioral and physiological attributes of male and female drivers engaged in collisions. Based on the structural equation model, Ref. [7] analyzed the self-reported data from Nebraska households to identify socioeconomic, personality, and attitudinal factors linked to inattentive driving at HRGCs. The results revealed that factors such as female drivers, younger drivers, higher incomes, and limited exposure to safety information regarding HRGCs caused a higher risk of inattentive driving at HRGCs. In the study of Ref. [8], the zero inflated negative binomial model alongside the empirical Bayes technique was used to forecast incidents at HRGCs for various warning mechanisms. Their outcomes showed improved congruence with actual field data when compared to the latest model from the Federal Railroad Administration. Subsequently, discrete choice models have also been utilized to identify the determinants considering the discrete nature of HRGC crash variables. For example, Ref. [9] investigated the HRGC crashes in the United States by employing a random parameters logit model with heterogeneity in means to account for varied factors influencing crashes. It highlighted specific behaviors and driver attributes, such as non-compliance with stop signals, old driver, and female driver, would raise the possibility of severe injuries.

Compared to statistical or econometric models, machine learning approaches have the advantage of being capable of processing big data and capturing intricate relationships between input variables and outcome predictions [10,11]. Although machine learning models have been extensively studied for real-time risk prediction [12–14], the application of these models to analyze injury severities at undivided two-way traffic HRGCs has not been investigated to the best of our knowledge. As a typical technique in the domain of machine learning, the XGBoost model has gained recognition as a powerful algorithm for classification and regression tasks. In traffic safety research, XGBoost has been utilized for its efficiency and accuracy in predicting road accident severity and identifying critical incident-related features [15,16]. Additionally, it is worth noting that some previous studies have utilized the XGBoost model to evaluate highway crash injury severities, although these studies may not specifically focus on non-divided two-way highway and railway crashes. For example, Ref. [17] proposed a data-driven consolidation model for highway–rail grade crossings in the United States utilizing the XGBoost algorithm and incorporating various engineering variables. The developed model achieved an overall accuracy of 0.991 and provided insights into the relative importance of the input variables. Through a 10-fold empirical analysis of various performance metrics, Ref. [18] found that the XGBoost technique had superior collective predictive performance and individual class accuracies for injury severity compared to other machine learning models. The XGBoost feature importance analysis identified collision type, weather status, road surface conditions, on-site damage type, lighting conditions, and vehicle type as sensitive variables for predicting crash injury severity outcomes. Numerous studies have compared the performance of machine learning techniques with statistical models, finding that machine learning approaches have comparable predictive accuracy metrics [10,19,20]. Given that the XGBoost model could provide useful insights into crash severity analysis, this study specifically focuses on the application of this algorithm to investigate the influence of various determinants on injury severity outcomes at undivided two-way traffic HRGCs.

Despite these remarkable achievements, it should be noted that data-driven approaches may lack the interpretability needed for direct application in safety enhancement

strategies (i.e., deeply comprehend the causal impact of certain characteristics on HRGC accident likelihoods and the corresponding probability of injury outcome). The interpretability of predictive models is critical, especially in domains where decision making can have life-altering consequences. The advancement of machine learning methodologies has made it possible for SHAP values to provide a consistent and cohesive method of interpreting models [21]. By attributing prediction output to individual features, SHAP values provide transparency to otherwise ‘black-box’ models in the analysis of HRGC crashes. While SHAP’s use in various predictive modeling scenarios has been documented, its application in conjunction with the XGBoost model for undivided two-way traffic HRGC safety analysis remains minimal.

Therefore, to predict crash injury severity outcomes and investigate various factors influencing injury severities at undivided two-way traffic HRGCs, the current study employed an advanced machine learning algorithm, XGBoost, alongside a comprehensive interpretative approach using SHAP values. This could assist policymakers and transportation authorities in reducing the frequency and severity of accidents at undivided two-way traffic HRGCs. The paper is organized as follows: Section 2 delivers a statistical analysis of independent variables. Section 3 delineates the analytical methods employed in the research. Detailed model outputs, interpretation, and policy implications are discussed in Section 4. The paper concludes with Section 5, which summarizes the key findings and pinpoints avenues for future research.

## 2. Data Description

In this study, the crash data were gathered from the Safety Analysis System of the Federal Railroad Administration [1]. HRGC crash records in the United States between 2020 and 2021 were collected from two distinct sources: (1) HRGC accident data and (2) current crossing inventory data. Comprehensive information on variables, including speed, vehicle, time, weather, visibility, and driver demographics, is provided by the accident database. The inventory database offers detailed information on each crossing, such as the location, illumination, traffic volumes, and signals.

A three-point ordinal scale is used in this study to code the injury severity outcomes: 1—uninjured, 2—injured, or 3—killed. This discrete variable is referred to as the dependent variable in the following analysis. Using distinct identification numbers assigned to each crossing, the information contained in the current crossing inventory is combined with the extracted crash database.

This combination facilitates a holistic identification of factors contributing to the severity of crash injuries. Subsequently, the merged data were subjected to thorough verification and screening to remove samples with missing information in variables. The variables were subsequently categorized into several classifications, including crash, crossing, vehicle, driver, and environmental characteristics. These exploratory variables were then converted into binary indicator variables, i.e., Yes (1) or No (0). A detailed partition and descriptive analysis of these independent variables can be seen in Table 1. After eliminating the records with incomplete variable information, the final dataset used for the following analysis comprised 1503 accident records, including 1009 records of no injuries (67.13%), 384 records of injuries (25.55%), and 110 records of fatalities (7.32%). Figure 1 delineates the geospatial distribution of these HRGC crashes.

**Table 1.** The descriptive statistics of independent variables.

Category	Variables	Frequency	Percentage (%)
Crash characteristics	Train speed $\leq$ 45 MPH	1256	83.57
	Train speed $>$ 45 MPH	247	16.43
	Unobstructed view	1444	96.10
	Obstructed view *	59	3.90
	Estimated vehicle speed $\leq$ 25 MPH	1374	91.40
	Estimated vehicle speed $>$ 25 MPH	129	8.60
	Did not stop	574	38.19
	Stopped and then proceeded	333	22.16
	Stopped on the crossing	214	14.24
	Went around the gate	290	19.29
	Other actions	92	6.12
	Private crossing type	32	2.13
	Public crossing type	1471	97.87
	Highway speed limit $\leq$ 25 MPH	405	26.95
Highway speed limit $>$ 25 MPH	1098	73.05	
Crossing characteristics	Annual average daily traffic (AADT) $\leq$ 5000	1047	69.66
	Annual average daily traffic (AADT) $>$ 5000	456	30.34
	Estimated percent of trucks $\leq$ 10%	1093	72.72
	Estimated percent of trucks $>$ 10%	410	27.28
	Crossing without signs or signals	32	2.10
	Crossing with signs or signals	1471	97.90
	Unpaved Highway	148	9.85
	Paved Highway	1355	90.15
	Land without commercial power	144	9.58
	Land with commercial power	1359	90.42
	Both-side crossing warning	1423	94.68
	Single-side crossing warning	80	5.32
	Crossing without illumination	928	61.74
	Crossing with illumination	575	38.26
Vehicle characteristics	Dry crossing surface	1290	85.83
	Adverse crossing surface *	213	14.17
	Industry track	82	5.46
	Main track	1328	88.36
	Siding track	9	0.60
	Yard track	84	5.59
	Auto	790	52.56
	Bus	2	0.13
	Motorcycle	9	0.60
	Truck	32	2.13
	Van	1374	91.40
	Other vehicles	527	35.06
	Middle driver	643	42.78
	Old driver	483	32.14
Driver characteristics	Young driver	377	25.08
	Female driver	392	26.08
	Male driver	1111	73.92
	Driver was not in vehicle	236	15.70
Environmental characteristics	Driver was in vehicle	1267	84.30
	Day	847	56.35
	Dusk	167	11.11
	Dark	360	23.95
	Dawn	129	8.58
	Clear	1060	70.53
	Cloudy	313	20.83
	Fog	13	0.86
	Rain	90	5.99
	Sleet	3	0.20
Snow	24	1.60	

\* Adverse crossing surface encompasses conditions such as moisture, ice, snow, slush, sand, mud, or oil on roadways. Obstructed view arises from factors including vehicles, trains, railroad apparatus, infrastructure, terrain, flora, or similar obstructions.

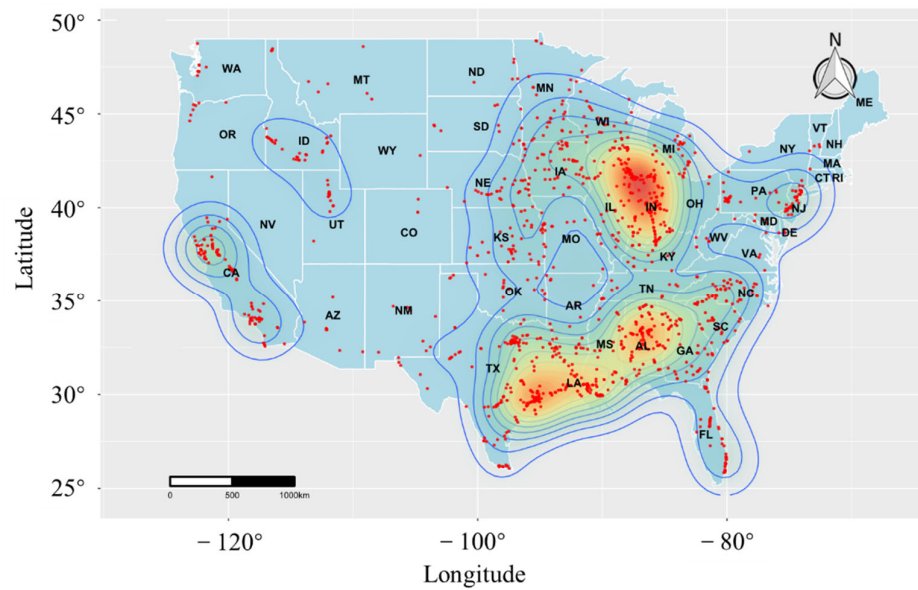


Figure 1. The geospatial distribution of HRGC crashes.

3. Method

The XGBoost model, an ensemble method that aggregates the forecasts from numerous decision trees, was applied in assessing the severity of injuries from HRGC accidents in this study [10,20,22]. The objective function of XGBoost is defined as follows [23]:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{t=1}^K \Omega(h_t) \tag{1}$$

where  $l(y_i, \hat{y}_i)$  is the loss function that measures the difference between the predicted value  $\hat{y}_i$  and the actual injury severity label  $y_i$  and  $K$  is the number of weak learners in the ensemble. The regularization term that penalizes the models' complexity is

$$\Omega(h_t) = \gamma J + \frac{1}{2} \lambda \sum_{j=1}^J w_{tj}^2 \tag{2}$$

where  $J$  represents the total leaf count;  $\gamma$  is the parameter that controls the leaf quantity; the coefficient for the L2 regularization term, represented by  $\lambda$ , increases the model's conservatism with each increment; and  $w_{tj}$  signifies the  $j$ -th leaf node's optimum value in the  $t$ -th decision tree. By minimizing the sum of the regularization term and the loss function, the objective function is optimized. The objective function changes to the following form when the second-order Taylor expansion is applied to it:

$$Obj = \sum_{j=1}^J \left[ \frac{1}{2} (H_{tj} + \lambda) w_{tj}^2 + G_{tj} w_{tj} \right] + \gamma J \tag{3}$$

where  $H_{tj}$  represents the sum of the second-order derivatives for each leaf node, and  $G_{tj}$  denotes the total of the first-order derivatives for each individual leaf node. The optimal value for  $w_{tj}$  is equal to  $-\frac{G_{tj}}{H_{tj} + \lambda}$  at the point when the derivative of the objective function is 0. Subsequently, the optimal  $w_{tj}$  is used to split the tree. This process of branching continues until the depth of the node matches the predetermined maximum depth.

In this study, three dichotomous crash severity levels are further split and determined as follows:

$$y_i = \begin{cases} 1, & \text{if the sample belongs to } i, i = \text{Uninjured, Injured, or Killed} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

To evaluate the model's performance, the five-fold cross-validation method is used [22]. The accuracy is also calculated to measure the performance of the model:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where  $TP$  represents a correct prediction of the specific injury severity;  $FP$  represents a false classification of an injury severity case being predicted as a non-injury-severity case;  $TN$  denotes a correct prediction of a non-injury-severity case; and  $FN$  indicates a non-injury-severity case incorrectly predicted as an injury case.

Additionally, the performance of the model can be assessed by the Area Under the Receiver Operating Characteristic Curve (AUC), which evaluates the model's capacity to discern between instances that are positive and negative. A higher AUC indicates better performance [20].

SHAP is a unified measure of feature importance that leverages cooperative game theory to allocate the contribution of each feature to the final prediction in a fair and computationally efficient manner [21]. Each feature's SHAP value can be interpreted as the average marginal contribution of that feature across all possible feature subsets [21]. It provides a powerful tool to interpret machine learning models such as XGBoost and deliver transparent and trustworthy predictions. Therefore, SHAP was utilized in this study to better interpret the results of the XGBoost models:

$$\phi_i = \sum_{S \subseteq N(i)} \frac{|S|!(k-|S|-1)!}{k!} (v(S \cup \{i\}) - v(S)) \quad (6)$$

where  $k$  represents the total number of features;  $\phi_i$  denotes the influence of feature  $i$ ;  $S$  signifies a subset of features; and  $|S|$  is the total feature number.

The Shapley value is an additive attribution-assignment approach, with the prediction from the model equating to the aggregate of these attributed values for every feature. This results in the following representative linear model:

$$\zeta(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (7)$$

where  $\phi_0$  indicates the average prediction across all training samples;  $M$  is the total feature count; and  $z'_i$  is a binary indicator that equals 1 if the feature is present, and 0 if not.

To ascertain the relative significance of each feature, the absolute values of the feature attributions across all samples are summed. Given  $n$  in the total sample count, the cumulative influence of feature  $i$  is encapsulated by the following equation:

$$I_i = \sum_{j=1}^n |\phi_{i,j}| \quad (8)$$

where  $\phi_{i,j}$  denotes the attribution value of the  $i$ -th attribute for the  $j$ -th sample.

#### 4. Results and Discussions

Table 2 presents the results of the  $\chi^2$  test to examine the association between each explanatory variable and the injury severity. A small  $p$ -value suggests that the observed categorical data are unlikely to occur by chance. The variables with a significant association with injury severity are listed in descending order based on the magnitude of the



$\chi^2$  value and the corresponding  $p$ -value. Variables such as driver was in vehicle, train speed > 45 MPH, and went around the gate show a strong correlation with injury severity outcomes, as evidenced by their high  $\chi^2$  values and negligible  $p$ -values. Only these indicators less than the chosen significance level of 0.05 were selected to predict the severity of the injury outcome in the XGBoost model.

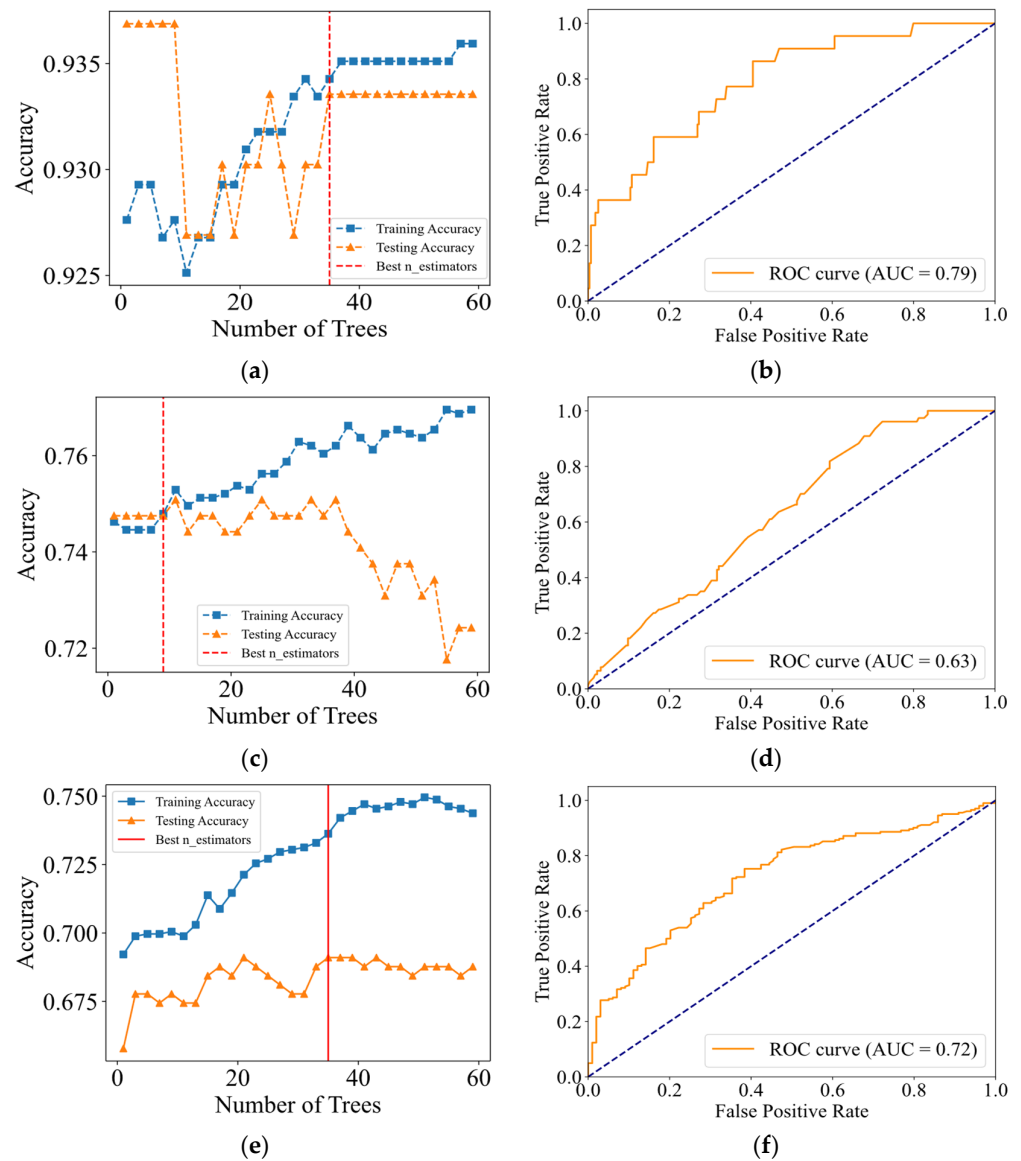
**Table 2.** Test results of explanatory variables and the injury severity.

Variables	$\chi^2$	$p$ -Value
Driver was in vehicle	112.657	<0.01
Train speed > 45 MPH	73.143	<0.01
Went around the gate	45.152	<0.01
Stopped on the crossing	28.614	<0.01
Did not stop	24.230	<0.01
Estimated vehicle speed > 25 MPH	16.451	<0.01
Truck	16.424	<0.01
Main track	15.913	<0.01
Motorcycle	12.358	<0.01
Land with commercial power	10.242	<0.01
Old driver	9.721	<0.01
Yard track	9.608	<0.01
Dark	9.288	<0.01
Unpaved Highway	8.216	0.016
Went through the gate	7.754	0.021
Middle driver	7.422	0.024
Day	7.288	0.026
Female driver	7.215	0.027
Adverse crossing surface	6.785	0.034
Crossing without illumination	5.996	0.050

Modeling injury severities in HRGC crashes was approached as a binary classification problem in this study, encompassing a range of categorical variables and dichotomous outcomes (1 for Yes, 0 for No) corresponding to three distinct injury severity levels. Selection and refinement of hyperparameters within the XGBoost framework, specifically `max_depth` set to [3,5,7] and `n_estimators` within the range of 1 to 61 in steps of 2, were conducted to enhance model accuracy. These selected parameters and their corresponding searching ranges were subjected to preliminary tests to ascertain their influence on model sensitivity. Additionally, to balance computational efficiency and predictive accuracy, the grid search technique with selected parameters was adopted. The strategy includes a 5-fold cross-validation method to ensure the robustness of the optimized model. Each dataset was split into training and testing sets using an 80:20 ratio. This process involves training and comparing the model with different parameter combinations to find the optimal configuration. The validation set was used to measure the accuracy of the model for each parameter combination. The parameter combination that yielded the highest accuracy on the validation set was selected as the best set of parameters.

Figure 2 presents the model performance of the hyper-tuning process for the killed, injured, and uninjured datasets. Each row contains two subplots corresponding to the hyperparameter tuning curve and the ROC curve for each dataset. Subplots Figure 2a,c,e show the variation in the chosen hyperparameters' performance with respect to the respective dataset. The best `max_depths` (3, 3, and 3 for the killed, injured, and uninjured categories, respectively) and `n_estimators` (35, 9, and 35 for the killed, injured, and uninjured categories, respectively) were utilized for the subsequent analysis. The trade-off between the true positive rate and the false positive rate at various classification thresholds is displayed by the ROC curves in subplots Figure 2b,d,f. It can be seen that the XGBoost model achieved the best performance in predicting the killed category, with an accuracy of 93.36% and an AUC of 0.79. However, the performance of the models in predicting the injured and uninjured datasets was less impressive, with accuracy values of 74.75%

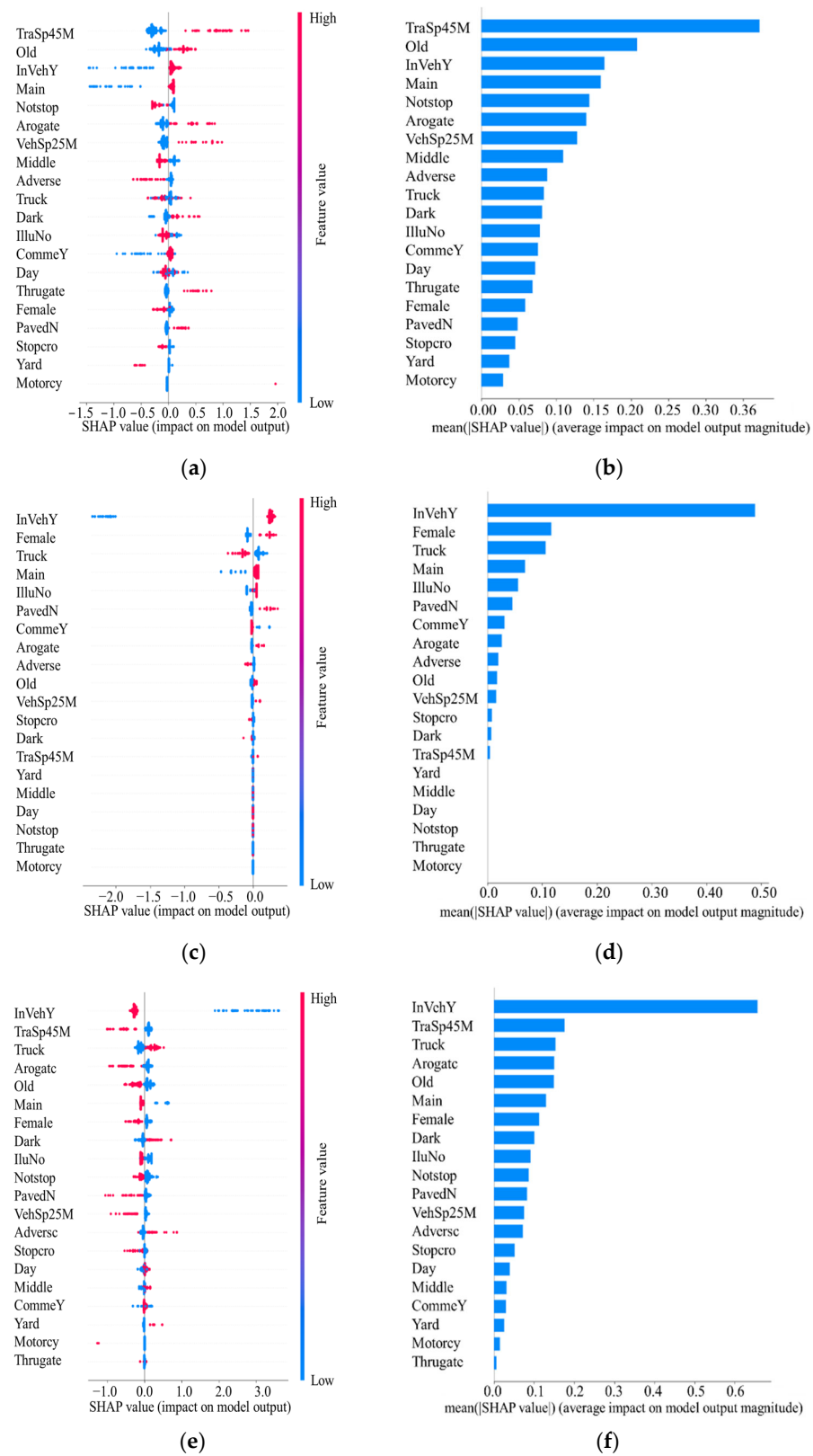
and 69.10%, respectively. Due to the imbalanced dataset, the XGBoost model shows varying levels of accuracy across different injury severities, which is consistent with previous studies [20,22].



**Figure 2.** Model performance for killed, injured, and uninjured datasets: (a) Hyperparameter tuning in XGBoost: killed; (b) ROC curve: killed; (c) Hyperparameter tuning in XGBoost: injured; (d) ROC curve: injured; (e) Hyperparameter tuning in XGBoost: uninjured; (f) ROC curve: uninjured.

The impact of each feature variable on the predictions made by the XGBoost model is shown in Figure 3. The beeswarm plot was employed to show each feature’s distribution of SHAP values, with individual dots representing an undivided two-way traffic HRGC crash. The respective features are delineated along the vertical axis of subplots Figure 3a,c,e, and the SHAP values pertaining to each feature for the three datasets are plotted along the horizontal axis. The sign of the SHAP value—be it positive or negative—denotes a positive or negative impact on the predicted injury severity outcome. The feature value’s magnitude is illustrated using different colors, where blue denotes lower values (interpreted as 0 for each variable) and red signifies higher values (interpreted as 1 for each variable). For example, it can be seen that the train speed > 45 MPH indicator was the most crucial impact factor and was positively associated with killed crashes.





**Figure 3.** SHAP summary plot of input features: (a) Impact of the feature on killed crashes; (b) Feature importance in killed crashes; (c) Impact of the feature on injured crashes; (d) Feature importance in injured crashes; (e) Impact of the feature on uninjured crashes; (f) Feature importance in uninjured crashes.

The subplots in Figure 3b,d,f list the features by their importance for each injury severity dataset. According to the SHAP values, the top ten important feature indicators in the killed model are as follows: train speed > 45 MPH, old driver, driver was in vehicle, main track, did not stop, went around the gate, estimated vehicle speed > 25 MPH, middle driver, adverse crossing surface, and truck. The top ten important feature indicators in the injured model include driver was in vehicle, female driver, truck, main track, crossing without illumination, unpaved highway, land with commercial power, went around the gate, adverse crossing surface, and old driver. The corresponding top ten important feature indicators in the uninjured model are driver was in vehicle, truck, train speed > 45 MPH, went around the gate, old driver, main track, female driver, dark, crossing without illumination, and did not stop.

Comparing the important features between the killed and injured models, it can be seen that some important features overlapped, such as old driver, driver was in vehicle, main track, went around the gate, adverse crossing surface, and truck. However, there are also some differences, such as train speed > 45 MPH, did not stop, middle driver, and estimated vehicle speed > 25 MPH, which were identified as important features in the killed model but not in the injured model, while female driver, crossing without illumination, unpaved highway, and land with commercial power were significant features in the injured model but not in the killed model. The results also indicated that the XGBoost model could distinguish between different severity levels.

The feature train speed > 45 MPH stood out in the killed model, which aligns with the findings of previous studies emphasizing high train speed as a critical factor in the severity of accidents at HRGCs [24]. This relationship is intuitive, as higher speeds reduce the time available for a driver to react to an approaching train and come to a stop once a hazard is noticed. For the injured model, the presence of female driver is notable, which suggests gender-specific differences in driving behavior or risk exposure at crossings. This shows that there are inherent differences in how different genders perceive and react to railway crossing risks. The existing literature frequently presents opposite findings regarding the role of gender as a risk factor in vehicular accidents, with some studies suggesting that although males tend to be involved in more accidents, females may be at a higher risk due to their vulnerability [25].

The driver was in vehicle indicator appeared as a significant feature in the XGBoost models across all severity levels of HRGC crashes. This variable's consistent significance suggests that the presence of the driver during the accident plays a crucial role in the injury severity outcome. One possible explanation is that any impact can directly affect the driver's safety due to their vulnerability to injury or fatality during the crash. Furthermore, the implication of driver was in vehicle as a significant feature can also reflect the scenario where a driver might not have the situational awareness or the reaction time necessary to respond effectively to an approaching train [26]. This could be due to various factors such as distractions, impairment, or other cognitive limitations. The literature supports the notion that driver behavior is critical in road safety outcomes [7,25]. A study by Ref. [27] of drivers' response to railroad crossings indicated that driver performance, including their ability to detect an oncoming train and respond appropriately, was a decisive factor in accident occurrences and injury severity outcomes. Therefore, measures such as increased visibility and warning signals at HRGCs can help improve the driver's situational awareness and response. Additionally, automated systems that can take control to prevent a vehicle from stalling or staying on the tracks when a train is approaching could also be effective.

The old driver indicator was also significant across all models, which may be indicative of the declined perception and reaction ability that accompany older drivers, as documented by Ref. [25]. These declines can impair older drivers' ability to perceive and respond in a timely manner to unexpected or complex driving situations, such as those encountered at HRGCs. This suggests that special interventions at crossings are needed to account for the age-related decline in driving performance. Potential measures may include extending train warning times, launching educational campaigns targeted at older drivers to raise

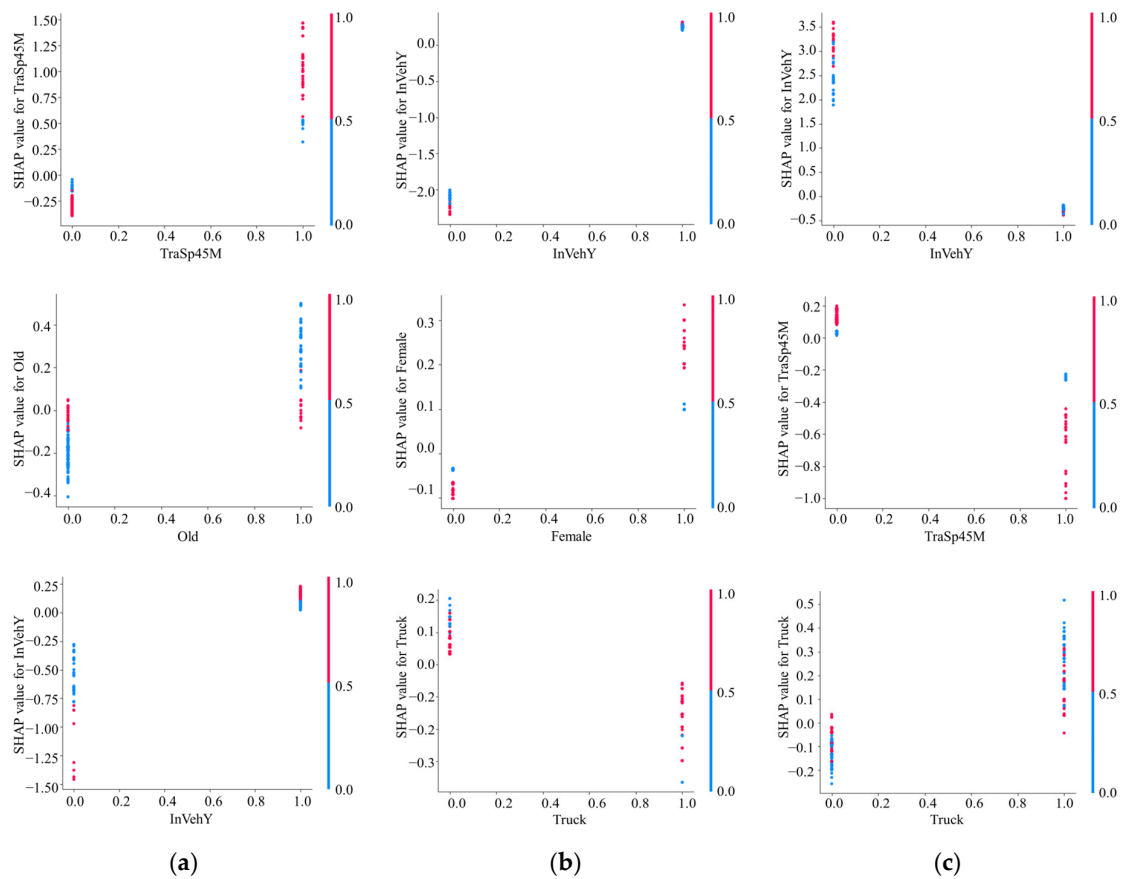
awareness about the specific risks at HRGCs, and implementing engineering controls like grade separation to obviate direct crossings.

Intriguingly, unpaved highway and crossing without illumination emerged as significant factors in the injured model but not in the killed model. This could imply that while these factors contribute to the likelihood of an accident, they may not necessarily increase its fatality rate, possibly due to a compensation mechanism and lower speeds typically associated with such conditions. The literature corroborates that environmental conditions markedly influence driver behavior, which in turn affects the accident severity [28].

Features such as went around the gate and did not stop were indicative of driver non-compliance with traffic control devices, a well-documented issue in traffic safety research [29]. These findings suggest that educational and enforcement measures targeting driver behavior could be instrumental in reducing both the frequency and severity of highway–rail crossing accidents. The presence of adverse crossing surfaces for both killed and injured outcomes pointed to infrastructure issues that complicate navigation over crossings and potentially contribute to accidents. This is supported by studies that have identified poor crossing surface conditions as a hazard for road users [30]. Lastly, truck being a common feature underscores the important role of vehicle type in HRGC accidents. It might be that the proficiency of truck drivers, who generally follow safety regulations and are careful at rail crossings, leads to the reduced severity of incidents observed [9]. Furthermore, trucks and other larger cars may offer their occupants superior protection, reducing the risk of injuries or fatalities at crossings.

More detailed information regarding the relationships between determinants and injury severity outcomes can be extracted by investigating the dependency plots of the SHAP values. It can not only be used to understand the overall importance of a feature, but also to explore the detailed distribution and the context-dependent nature of feature influences on model predictions, thereby ensuring more reliable predictions and informing potential interventions based on the predictive factors. Figure 4 displays SHAP dependency plots for the three most influential features in killed, injured, and uninjured crashes, respectively. If the SHAP values are highly concentrated within a certain range, this may indicate that the model has a high sensitivity to the feature values within that range. In contrast, a sparse distribution of SHAP values across a wide range of variables may suggest that there is a significant variation in the model's predictive influence for this feature. For example, the driver was in vehicle indicator exhibited a very consistent and concentrated distribution of SHAP values across three datasets representing varying levels of injury severity. This observation implies that the influence of this particular feature on the model's prediction is rather stable and reliable across different injury severity outcomes.

For killed crashes, the partial levitation observed in Figure 4a suggests that there were specific values of train speed > 45 MPH that led to higher SHAP values, indicating a stronger impact on the prediction outcome. In addition, the SHAP values for certain values of the old driver fluctuate throughout the subplot; this could indicate the presence of interactions with other features, such as time of day (which might affect visibility) or the type of area (urban vs. rural, which impacts driving speed and emergency medical service response time). For injured crashes, a sparse distribution of SHAP values can also be observed in Figure 4b for female driver and truck, reflecting the variability in the impact of these factors. Notably, the partial levitation can only be seen for the truck indicator in injured crashes, suggesting that this feature may have a heterogeneous impact on predicting uninjured outcomes. Meanwhile, the train speed > 45 MPH indicator was negatively associated with no injuries in Figure 4c, implying that higher train speeds were less likely to be related to no-injury scenarios.



**Figure 4.** SHAP dependency plots for top three features: (a) killed crashes; (b) injured crashes; (c) uninjured crashes.

**5. Conclusions**

Leveraging the crash data from the Federal Railroad Administration for the years 2020–2021, this study provides a predictive and statistical analysis of undivided two-way traffic HRGC crashes by utilizing a robust machine learning approach, the XGBoost algorithm, along with the interpretative power of SHAP values to investigate the dominant factors affecting injury severity outcomes. Multiple variables, including crash characteristics, crossing characteristics, vehicle characteristics, driver characteristics, and environmental characteristics, were analyzed to ascertain their impacts on three injury severity outcomes: killed, injured, and uninjured.

The findings underscore the complex interplay between various influencing factors and their respective contributions to the severity of injuries sustained in HRGC accidents. The relationship between each explanatory factor and the injury severity was first examined using the  $\chi^2$  test, and only the factors with a significant association with injury severities were chosen to predict the outcomes. The XGBoost model, optimized through hyperparameter tuning and validated by five-fold cross-validation, demonstrated proficiency in distinguishing different injury severities, particularly in predicting fatal outcomes, with a highest degree of accuracy of 93.36% and an AUC of 0.79. The model’s performance in predicting the injured and uninjured datasets was less impressive, with accuracy values of 74.75% and 69.10%, respectively. Furthermore, the top ten most important feature indicators for the killed, injured, and uninjured datasets were identified. It was found that some of them overlapped, such as old driver, driver was in vehicle, main track, went around the gate, adverse crossing surface, and truck, while there were also some differences. The diverse features highlight the nuanced nature of risk factors across different severity levels. The application of SHAP values further enhanced the interpretability of XGBoost models, offering a granular perspective on the influence of individual features. The detailed

distributions of the respective top three significant variables that emerged from the analysis were presented, including train speed > 45 MPH, driver was in vehicle, truck, old driver, and female driver.

The generated insights could provide a compelling foundation for policymakers and practitioners to formulate strategic interventions that can significantly enhance safety at HRGCs. For example, the driver was in vehicle indicator indicates a potential direction for intervention, suggesting that the presence and attentiveness of the driver are vital. The implementation of more effective warning systems, including advanced detection and communication technologies, can provide drivers with timely and clear warnings, thereby reducing the likelihood of severe crashes. For older drivers, who emerged as a vulnerable group, extended warning times and targeted awareness campaigns could be beneficial. Additionally, driver education programs tailored to improve awareness of HRGC risks, combined with strategic enforcement campaigns, can mitigate risk behaviors such as bypassing active warnings at crossings.

This study illustrates the potential of machine learning techniques in enhancing our understanding of traffic safety and strengthening traffic risk management practices. Despite these efforts, the study is not without limitations. One limitation is that the scope of the data is confined to two years of HRGC incidents. In the context of crash injury severity analysis, the impact of certain factors can vary at different times of the day, on certain days of the week, or throughout the seasons [25]. Meanwhile, examining the fluctuations over time can underscore the significance of driver behavior patterns. Elements like the driver's age, levels of distraction, and physical impairment can have varying impacts on the severity of injuries. As accident dynamics and transportation practices evolve, continuous data collection is paramount, and the model would benefit from using an expanded temporal dataset to devise more targeted and effective safety strategies [25]. Additionally, while the XGBoost model illustrates solid predictive performance, the nature of machine learning algorithms can obscure the causal relationship between variables and outcomes. Therefore, the machine learning method could be coupled with statistical models to guide the implementation of safety measures in the future. This integration could enhance predictive accuracy and uncover complex, nonlinear relationships between variables that affect crash injury severities [20]. By employing sensors, IoT devices, and data analytics platforms, future studies could also integrate advanced analytics with real-time road infrastructure monitoring to predict and prevent possible conflict points between vehicular traffic and trains by adjusting signaling systems and improving the design of crossings.

**Author Contributions:** Conceptualization, Q.R. and M.X.; methodology, Q.R.; software, Q.R.; validation, Q.R., M.X., B.Z. and S.-H.C.; writing—original draft preparation, Q.R.; writing—review and editing, Q.R., M.X., B.Z. and S.-H.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the National Natural Science Foundation of China (Grant No. 72071041), the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 15222822), and the Hong Kong Polytechnic University (G-UARN).

**Data Availability Statement:** The authors do not have permission to share data.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Federal Railroad Administration. Highway/Rail Grade Crossing Incidents. Available online: <https://railroads.dot.gov/accident-and-incident-reporting/highwayrail-grade-crossing-incidents/highwayrail-grade-crossing> (accessed on 1 March 2023).
2. Calvi, A.; Cafiso, S.D.; D'Agostino, C.; Kieć, M.; Petrucci, G. A Driving Simulator Study to Evaluate the Effects of Different Types of Median Separation on Driving Behavior on 2 + 1 Roads. *Accid. Anal. Prev.* **2023**, *180*, 106922. [CrossRef]



3. Mahmud, S.M.S.; Ferreira, L.; Hoque, M.S.; Tavassoli, A. Overtaking Risk Modeling in Two-Lane Two-Way Highway with Heterogeneous Traffic Environment of a Low-Income Country Using Naturalistic Driving Dataset. *J. Saf. Res.* **2022**, *80*, 380–390. [[CrossRef](#)]
4. Ghomi, H.; Bagheri, M.; Fu, L.; Miranda-Moreno, L.F. Analyzing Injury Severity Factors at Highway Railway Grade Crossing Accidents Involving Vulnerable Road Users: A Comparative Study. *Traffic Inj. Prev.* **2016**, *17*, 833–841. [[CrossRef](#)] [[PubMed](#)]
5. Haleem, K. Investigating Risk Factors of Traffic Casualties at Private Highway-Railroad Grade Crossings in the United States. *Accid. Anal. Prev.* **2016**, *95*, 274–283. [[CrossRef](#)] [[PubMed](#)]
6. Hao, W.; Kamga, C.; Daniel, J. The Effect of Age and Gender on Motor Vehicle Driver Injury Severity at Highway-Rail Grade Crossings in the United States. *J. Saf. Res.* **2015**, *55*, 105–113. [[CrossRef](#)] [[PubMed](#)]
7. Zhao, S.; Khattak, A.J. Factors Associated with Self-Reported Inattentive Driving at Highway-Rail Grade Crossings. *Accid. Anal. Prev.* **2017**, *109*, 113–122. [[CrossRef](#)] [[PubMed](#)]
8. Mathew, J.; Benekohal, R.F. Highway-Rail Grade Crossings Accident Prediction Using Zero Inflated Negative Binomial and Empirical Bayes Method. *J. Saf. Res.* **2021**, *79*, 211–236. [[CrossRef](#)] [[PubMed](#)]
9. Ren, Q.; Xu, M. Injury Severity Analysis of Highway-Rail Grade Crossing Crashes in Non-Divided Two-Way Traffic Scenarios: A Random Parameters Logit Model. *Multimodal Transp.* **2024**, *3*, 100109. [[CrossRef](#)]
10. Goswamy, A.; Abdel-Aty, M.; Islam, Z. Factors Affecting Injury Severity at Pedestrian Crossing Locations with Rectangular RAPID Flashing Beacons (RRFB) Using XGBoost and Random Parameters Discrete Outcome Models. *Accid. Anal. Prev.* **2023**, *181*, 106937. [[CrossRef](#)]
11. Mannering, F.; Bhat, C.R.; Shankar, V.; Abdel-Aty, M. Big Data, Traditional Data and the Tradeoffs between Prediction and Causality in Highway-Safety Analysis. *Anal. Methods Accid. Res.* **2020**, *25*, 100113. [[CrossRef](#)]
12. Arvin, R.; Khattak, A.J.; Qi, H. Safety Critical Event Prediction through Unified Analysis of Driver and Vehicle Volatilities: Application of Deep Learning Methods. *Accid. Anal. Prev.* **2021**, *151*, 105949. [[CrossRef](#)]
13. Khan, M.N.; Ahmed, M.M. Trajectory-Level Fog Detection Based on in-Vehicle Video Camera with TensorFlow Deep Learning Utilizing SHRP2 Naturalistic Driving Data. *Accid. Anal. Prev.* **2020**, *142*, 105521. [[CrossRef](#)]
14. Sun, P.; Aljeri, N.; Boukerche, A. Machine Learning-Based Models for Real-Time Traffic Flow Prediction in Vehicular Networks. *IEEE Netw.* **2020**, *34*, 178–185. [[CrossRef](#)]
15. He, L.; Yu, B.; Chen, Y.; Bao, S.; Gao, K.; Kong, Y. An Interpretable Prediction Model of Illegal Running into the Opposite Lane on Curve Sections of Two-Lane Rural Roads from Drivers' Visual Perceptions. *Accid. Anal. Prev.* **2023**, *186*, 107066. [[CrossRef](#)] [[PubMed](#)]
16. Ma, Y.; Zhang, J.; Lu, J.; Chen, S.; Xing, G.; Feng, R. Prediction and Analysis of Likelihood of Freeway Crash Occurrence Considering Risky Driving Behavior. *Accid. Anal. Prev.* **2023**, *192*, 107244. [[CrossRef](#)] [[PubMed](#)]
17. Soleimani, S.; Mousa, S.R.; Codjoe, J.; Leitner, M. A Comprehensive Railroad-Highway Grade Crossing Consolidation Model: A Machine Learning Approach. *Accid. Anal. Prev.* **2019**, *128*, 65–77. [[CrossRef](#)]
18. Jamal, A.; Zahid Khattak, M.; Rahman, M.T.; Hasan, A.; Almoshaogeh, M.; Farooq, D.; Ahmad, M. Injury Severity Prediction of Traffic Crashes with Ensemble Machine Learning Techniques: A Comparative Study. *Int. J. Inj. Control Saf. Promot.* **2021**, *28*, 408–427. [[CrossRef](#)]
19. Islam, Z.; Abdel-Aty, M.; Mahmoud, N. Using CNN-LSTM to Predict Signal Phasing and Timing Aided by High-Resolution Detector Data. *Transp. Res. Part C Emerg. Technol.* **2022**, *141*, 103742. [[CrossRef](#)]
20. Yuan, C.; Li, Y.; Huang, H.; Wang, S.; Sun, Z.; Li, Y. Using Traffic Flow Characteristics to Predict Real-Time Conflict Risk: A Novel Method for Trajectory Data Analysis. *Anal. Methods Accid. Res.* **2022**, *35*, 100217. [[CrossRef](#)]
21. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
22. Yan, X.; He, J.; Zhang, C.; Liu, Z.; Qiao, B.; Zhang, H. Single-Vehicle Crash Severity Outcome Prediction and Determinant Extraction Using Tree-Based and Other Non-Parametric Models. *Accid. Anal. Prev.* **2021**, *153*, 106034. [[CrossRef](#)] [[PubMed](#)]
23. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
24. Hao, W.; Kamga, C.; Yang, X.; Ma, J.; Thorson, E.; Zhong, M.; Wu, C. Driver Injury Severity Study for Truck Involved Accidents at Highway-Rail Grade Crossings in the United States. *Transp. Res. Part F Traffic Psychol. Behav.* **2016**, *43*, 379–386. [[CrossRef](#)]
25. Ren, Q.; Xu, M. Exploring Variations and Temporal Instability of Factors Affecting Driver Injury Severities between Different Vehicle Impact Locations under Adverse Road Surface Conditions. *Anal. Methods Accid. Res.* **2023**, *40*, 100305. [[CrossRef](#)]
26. Hickey, A.R.; Collins, M.D. Disinhibition and Train Driver Performance. *Saf. Sci.* **2017**, *95*, 104–115. [[CrossRef](#)]
27. Larue, G.S.; Filtner, A.J.; Wood, J.M.; Demmel, S.; Watling, C.N.; Naweed, A.; Rakotonirainy, A. Is It Safe to Cross? Identification of Trains and Their Approach Speed at Level Crossings. *Saf. Sci.* **2018**, *103*, 33–42. [[CrossRef](#)]
28. Yan, X.; He, J.; Zhang, C.; Wang, C.; Ye, Y.; Qin, P. Temporal Instability and Age Differences of Determinants Affecting Injury Severities in Nighttime Crashes. *Anal. Methods Accid. Res.* **2023**, *38*, 100268. [[CrossRef](#)]



29. Larue, G.S.; Naweed, A. Evaluating the Effects of Automated Monitoring on Driver Non-Compliance at Active Railway Level Crossings. *Accid. Anal. Prev.* **2021**, *163*, 106432. [[CrossRef](#)]
30. Amin, S. Backpropagation—Artificial Neural Network (BP-ANN): Understanding Gender Characteristics of Older Driver Accidents in West Midlands of United Kingdom. *Saf. Sci.* **2020**, *122*, 104539. [[CrossRef](#)]

**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.