

Article

Enhancing Intelligent Transportation Safety with Explainable AI: A Framework for Uncovering Crash Severity Factors at Highway–Rail Grade Crossings

Dongming Wang¹, Qin He², Jinwen Peng^{3,*} and Gen Li^{2,*} ¹ College of Automotive Engineering, Jiangsu Vocational College of Electronics and Information, Huai'an 223003, China; 023102@jsei.edu.cn² College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China; heqin@njfu.edu.cn³ School of Urban Railway Transportation, Shanghai University of Engineering Science, Shanghai 201620, China

* Correspondence: M405123301@sues.edu.cn (J.P.); ligen@njfu.edu.cn (G.L.)

Abstract

Improving road safety is a fundamental goal of Intelligent Transportation Systems (ITS). However, the complex interplay of factors in accident-prone scenarios, such as highway–rail grade crossings, poses significant challenges for conventional analysis. This paper addresses this gap by proposing and validating a novel explainable artificial intelligence (XAI) framework, which integrates Extreme Gradient Boosting (XGBoost) with Shapley Additive Explanations (SHAP), to enhance safety analysis within ITS. Applying this framework to a comprehensive dataset of highway–rail grade crossing collisions, our research moves beyond simple correlation to uncover the nonlinear relationships and interaction effects governing injury severity. The model identifies speed-related factors, driver age, and traffic exposure as primary predictors. More critically, the SHAP analysis quantitatively reveals significant synergistic risks, demonstrating that the combination of non-dry road surfaces and poor lighting conditions drastically amplifies injury severity. These findings offer granular insights for the “smart management” and development of “resilient infrastructures,” enabling targeted interventions like adaptive lighting systems and dynamic risk warnings. This study not only provides critical safety solutions for grade crossings but also showcases the power of XAI as a robust tool for “advanced analysis” across various complex transportation safety problems, ultimately contributing to the creation of safer and more reliable ITS.

Keywords: highway–rail grade crossing accidents; interaction effects; injury severity; XGBoost; SHAP



Academic Editors: Zonghai Chen,
Yujie Wang, Zhiling Wang,
Kailong Liu and Jikai Wang

Received: 6 October 2025

Revised: 15 November 2025

Accepted: 17 November 2025

Published: 20 November 2025

Citation: Wang, D.; He, Q.; Peng, J.; Li, G. Enhancing Intelligent Transportation Safety with Explainable AI: A Framework for Uncovering Crash Severity Factors at Highway–Rail Grade Crossings. *World Electr. Veh. J.* **2025**, *16*, 637. <https://doi.org/10.3390/wevj16110637>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the World Electric Vehicle Association. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid growth in vehicular traffic is placing unprecedented strain on global transportation networks, leading to escalating concerns over road safety, traffic congestion, and environmental impact [1,2]. In response to these pressing issues, Intelligent Transportation Systems (ITS) have emerged as a transformative paradigm, promising to revolutionize road service and operational efficiency [3]. By integrating cutting-edge technologies such as the Internet of Things (IoT), big data, and Artificial Intelligence (AI), ITS aims to create a new generation of smart, resilient, and sustainable civil infrastructure. A primary objective of

this vision is to proactively enhance road safety, moving from reactive incident response to predictive risk mitigation.

However, the realization of a fully functional ITS is fraught with challenges, particularly in managing the high complexity of real-world traffic environments. Highway–rail Grade Crossings (HRGCs) represent a quintessential example of such a complex transportation scenario. Such crossings are those locations where a roadway intersects one or more railroad tracks, primarily involving conventional trains rather than high-speed rail systems. According to data from the Federal Railroad Administration (FRA, 2025) [4], over 2000 collisions occur annually at highway–rail grade crossings in the United States, resulting in more than 200 fatalities and 700 injuries each year. Thus, the HRGCs are always considered as high-risk zones, making them a significant global safety concern affecting regions such as the United States and the European Union. As critical junctures where road and rail traffic intersect, HRGCs are characterized by a unique confluence of high-speed trains, variable road user behaviors, and dynamic environmental conditions, making them persistent hotspots for severe accidents. Effectively modeling and mitigating risks at HRGCs is therefore not just a niche safety problem but a litmus test for the analytical power and practical applicability of advanced ITS methodologies. Understanding the intricate, often nonlinear, factors that contribute to crash severity in these environments is fundamental to developing effective smart management and control strategies.

Previous research has significantly contributed to understanding crash severity at HRGCs, employing various statistical and machine learning models [5–13]. Traditional statistical methods, such as multinomial logit models, have been instrumental in identifying key risk factors but often struggle to capture the complex, nonlinear relationships and interaction effects inherent in crash data [5–9]. While more advanced machine learning models, like neural networks, offer superior predictive accuracy, their “black-box” nature often hinders practical application. The lack of transparency makes it difficult for traffic engineers and policymakers to understand why a certain risk level is predicted, thereby limiting the ability to formulate precise, evidence-based safety interventions [10–13]. This gap highlights a critical need within ITS for analytical tools that are not only powerful but also inherently interpretable.

To bridge the identified research gaps, this paper proposes and validates a novel explainable AI (XAI) framework for advanced safety analysis in complex transportation systems, using highway–rail grade crossings (HRGCs) as a case study. Traditional statistical and machine learning models have achieved high predictive performance but often lack interpretability, making it difficult to uncover the mechanisms underlying crash severity. In contrast, explainable methods can disentangle these complex, nonlinear relationships by quantifying and visualizing how each variable—and their interactions—affect model outcomes.

Our proposed framework integrates the predictive capability of the Extreme Gradient Boosting (XGBoost) algorithm with the interpretive strength of Shapley Additive Explanations (SHAP). This synergy not only enables accurate prediction of injury severity but also opens the “black box,” revealing how environmental, operational, and human factors jointly contribute to crash outcomes.

Unlike previous studies that applied XGBoost–SHAP primarily for feature ranking or general crash prediction, this research advances their use to quantitatively interpret interaction effects in the safety-critical context of HRGCs—a domain that has received limited attention in explainable AI research. The XAI framework provides a systematic means to uncover how combinations of conditions amplify or mitigate crash severity, thus extending the analytical capability beyond single-variable interpretation. In doing so, this study contributes a new, transparent approach for transforming predictive modeling into

interpretable, actionable insights that can inform the design and management of safer intelligent transportation systems (ITS).

The remainder of this paper is organized as follows. Section 2 gives the literature review. Section 3 introduces the data sources, variable reclassification methods, and the categorization of road surface and lighting conditions. Section 4 presents the XGBoost-SHAP analytical framework, detailing model construction and evaluation metrics. Section 5 reports the model results, analyzing feature importance and comparing key influencing factors under different conditions. Section 6 discusses the mechanisms of accident severity based on model findings and proposes targeted policy implications such as speed management and lighting improvements. Finally, Section 7 summarizes the research outcomes, highlights the limitations in data and methodology, and outlines directions for future work.

2. Literature Review

Due to the severity of accidents at highway–rail grade crossings, an increasing number of studies in recent years have focused on identifying and extracting the key factors associated with such accidents. Previous studies based on accident data analysis have identified key influencing factors such as vehicle type, train/vehicle speed, roadway conditions and geometric features, lighting conditions, weather conditions, as well as driver age and gender [5,6]. Wei Hao et al. [7] applied an ordered probit model using FRA crash data (2002–2011) to investigate the influence of time of day on driver injury severity at HRGCs. They found that crashes occurring during early morning, a.m. peak, and p.m. peak periods were associated with significantly higher injury severity compared to other times. Vehicle speeds and train speeds exceeding 50 mph significantly increased the probability of severe injuries. Wei Hao and Camille Kamga [8] examined the effects of lighting on driver injury severity at highway–rail grade crossings (HRGCs) in the United States using a mixed logit model. The study found that the fatality rate at unlit crossings was significantly higher than at lit crossings. Factors such as high-speed driving (vehicle or train speeds exceeding 50 mph), adverse weather conditions (such as rain, snow and fog), open space areas, and unpaved road surfaces were shown to significantly increase injury severity, with these effects being more pronounced under unlit conditions.

Yang et al. [9] employed a LightGBM-SHAP framework using FRA crash data (2012–2021) to identify the importance of road surface conditions in accident severity at HRGCs. They found that wet, icy, and gravel surfaces significantly exacerbated injury severity compared with dry conditions. Wei Hao and Camille Kamga [10] used an ordered probit model to analyze driver injury severity at highway–rail grade crossings, focusing on rural and urban differences. They found that vehicle speed, train speed, lighting conditions, and road surface significantly affect injury severity, with rural areas showing higher severity levels. Their study highlights the need to consider environmental factors and area type in accident severity analysis. Hao and Daniel [11] investigated driver injury severity at highway–rail grade crossings under different weather conditions using mixed logit models. The results revealed that inclement weather (rain, snow, fog) significantly exacerbates crash severity, particularly when combined with factors like high train/vehicle speeds (increasing fatality risk by 65–132%), unpaved roads, and poor lighting. Older drivers (65+) were found to be 16–51% more likely to suffer fatal injuries compared to younger drivers. The study also identified open areas and low-traffic crossings as high-risk environments.

Zhao and Khattak [12] applied a random parameters binary logit model using Nebraska crash data (2002–2013) to examine the role of driver inattention in injury severity at HRGCs. They identified inattentive driving as a critical factor, finding it significantly increased the probability of severe driver injuries, comparable to DUI and aggressive driving. Gabree et al. [13] based on field video observations identified the importance of dynamic

envelope pavement markings and signage. They found that these countermeasures significantly reduced the proportion of vehicles stopping on tracks and increased safe stopping behind stop lines, indicating improved driver safety behavior. Wei Hao et al. [14] utilized an ordered probit model to explore how driver age and gender influence injury severity at highway–rail grade crossings (HRGCs) in the United States. Based on 25,945 accident records from 2002 to 2011, they found significant variations across different age and gender groups. Older drivers were more prone to severe or fatal injuries due to slower reaction times and physical limitations, while younger male drivers were more likely to sustain serious injuries during rush hours when driving at high speeds.

The above description addresses the influence of single factors on crashes; however, in many cases, crashes may result from the combined effects of multiple factors [15]. Given the complexity of traffic accidents, potential interactions among different influencing factors should be considered [16]. Morgan and Mannering [17] employed a mixed logit model using Indiana single-vehicle crash data (2007–2008) to examine the interactive effects of road-surface conditions, driver age, and gender. They found that wet and snow/ice surfaces significantly increased severe injury risks for females and older males, whereas younger males showed reduced severity under the same conditions. Khales et al. [18] employed mixed logit models to analyze the joint effects of visibility and warning devices on driver injury severity at HRGCs. They found that poor visibility combined with passive warnings substantially increased fatality risk, underscoring the compounded hazards of adverse conditions and inadequate control measures. Hu and Lin [19] analyzed the interaction effects of three advanced devices—obstacle detector, LED train approaching indicator, and law enforcement camera—on HRGC safety. They found that combined deployment significantly enhanced crash and gate-breaking reduction, with the indicator–camera pairing and all three devices together producing the strongest synergistic safety benefits. Zhao and Khattak [20] employed factor analysis and structural equation modeling to investigate inattentive driving at HRGCs. They identified gender, age, income, driving frequency, residence years, and safety information exposure as significant determinants, and revealed interactions between these factors and drivers' patience and safety attitudes, showing that females, younger, and high-income drivers with lower patience and weaker safety attitudes were more prone to inattentive driving. The above several studies indicate that significant interactions exist among different influencing factors; however, current research remains insufficient. In particular, the interaction between road surface conditions and lighting has not been examined. Previous studies have shown that both factors exert significant effects individually, yet under interaction their mechanisms may change substantially, which could affect policy implementation.

Common methods used in the study of investigating the influencing factors and severity of highway–rail grade crossing accidents can be broadly divided into two categories: parametric models and machine learning methods [21,22]. Previous HRGC safety studies have primarily focused on analyzing crash severity, typically employing parametric models. Fan et al. [23] applied a multinomial logit model to examine the injury severity of 7414 vehicle–rail crashes at U.S. highway–rail grade crossings. Their results indicated that higher train and vehicle speeds, pickup trucks, and older or male drivers increased crash severity, while higher traffic volumes and developed-area crossings were associated with reduced severity. Kang and Khattak [24] employed a variational Bayesian latent class clustering approach combined with ordered logit models to analyze HRGC crash injury severity, and found that the cluster-based method better captured unobserved heterogeneity and identified severity factors specific to certain crash subgroups. Kutela et al. [25] applied a mixed multinomial logit model with a Bayesian approach to 20 years of FRA crash data to investigate pre-crash driver behaviors at highway–rail grade crossings. Their

results indicated that demographic, vehicle, environmental, and crossing factors significantly influenced risky behaviors, with strong regional heterogeneity requiring random effects. Ren and Xu [26] employed a random parameters logit model with heterogeneity in means to investigate factors influencing injury severity in HRGC crashes under non-divided two-way traffic conditions. Their analysis revealed that risky driving behaviors, higher vehicle and train speeds, and vulnerable road users such as older, female, and motorcyclist drivers significantly increased the likelihood of severe outcomes. Wu et al. [15] adopted a random parameters logit model with heterogeneity in means and variances to identify the significant factors contributing to the crash injury severity involving truck and automobile drivers, finding that this model better captured unobserved heterogeneity and revealed significant differences in the factors influencing injury severity between automobile and truck crashes. Parametric models have the advantages of simple formulations and clear interpretability; however, they are limited in identifying variable interactions, requiring extensive trial and error, and the interactions ultimately discovered may be highly constrained by the data.

Recent years, machine learning methods have increasingly attracted the attention of researchers. Keramati et al. [27] used data from North Dakota (1990–2018) to analyze HRGC crash severity with a Random Survival Forest (RSF) model. The study identified train speed, traffic volume, control type, and roadway classification as key factors. Compared with traditional statistical models, RSF captured nonlinear relationships and interactions more effectively, offering higher predictive accuracy. Rana et al. [28] employed ensemble machine learning models, including Random Forest (RF), AdaBoost, and XGBoost, to analyze Canadian highway–rail grade crossing (HRGC) crashes. Using the ExtraTrees classifier for feature selection, they identified key predictors and applied SMOTE to balance imbalanced datasets. The models were evaluated through cross-validation with metrics such as precision, recall, and F1-score, showing that XGBoost achieved the highest accuracy. Additionally, ArcGIS hotspot analysis was used to detect spatial clusters of HRGC crashes.

Lu et al. [29] employed a Gradient Boosting (GB) model with a functional gradient descent algorithm to predict crashes at highway–rail grade crossings. They conducted a comprehensive evaluation using six prediction metrics, compared the GB model with a decision tree benchmark, and demonstrated its superior accuracy. In addition to generating variable importance rankings, they applied partial dependence plots to interpret nonlinear interactions, offering clearer insights into how key factors influence crash likelihood. Zheng et al. [21] applied a neural network (NN) model to predict crash risk at highway–rail grade crossings using 19 years of historical data. They developed variable importance rankings based on connection weights and mean-square error, and conducted marginal effect analysis to interpret nonlinear relationships between contributory factors and crash likelihood. Compared with decision tree models, the NN approach achieved higher predictive accuracy and provided deeper insights into the dynamic, nonlinear interactions among explanatory variables. Keramati et al. [27] developed a hybrid Analytic Hierarchy Process–Hazard Index (AHP-HI) model to evaluate crash severity at highway–rail grade crossings. The method integrates a Competing Risk Model (CRM) to estimate cumulative probabilities of fatal, injury, and property-damage-only crashes, which are then synthesized using AHP to generate a comprehensive hazard index. This approach enables a more refined ranking of crossings based on severity likelihood, offering decision-makers a data-driven tool for prioritizing high-risk locations. Machine learning methods possesses strong feature extraction capabilities, demonstrates robust generalization, and is well-suited for handling high-dimensional data. However, it remains relatively opaque as a “black box” [30], and especially the interactions between variables are not easy to express.

In summary, despite these advances, the interaction between road surface and lighting conditions at HRGCs remains unexplored and warrants further investigation. Moreover, there is a lack of research leveraging interpretable models to differentiate risk factors under varying environmental scenarios. This gap is particularly relevant given that adverse surface conditions and poor illumination often coincide, creating compounded safety risks that require tailored countermeasures. Therefore, this study proposes the use of the XGBoost-SHAP framework to analyze the severity of collision injuries at HRGCs, with a particular focus on four combined scenarios and on examining the interaction between road surface and lighting conditions.

3. Date Preparation

3.1. Data Source

The data used in the paper primarily originate from two databases maintained by the Federal Railroad Administration (FRA): the Highway–Rail Crossing Database Files and the Highway–Rail Accident/Incident Reports. The Highway–Rail Crossing Database, updated through April 2025, contains records dating back to 1975 and covers all 50 U.S. states and territories. It provides essential information about highway–rail crossings, including control devices, control systems, geometric features, land use types, and annual average daily traffic (AADT). Meanwhile, the Highway–Rail Accident/Incident Reports focus on detailed information recorded at the time of accidents, such as weather conditions, time of occurrence, train and vehicle speeds, driver gender, and driver behavior. By using a unique identifier assigned to each crossing, data from the two sources were merged to form a more complete and consistent dataset of HRGC accident records.

This study focuses on data from 2012 to 2021, extracting 21,163 accidents. To ensure the quality of the dataset, data cleaning was performed. Data cleaning involves processing the raw data to correct identifiable errors and inaccuracies within the dataset. The main procedures include filtering invalid entries and handling missing values. Before handling missing values, this study removed abnormal records of relevant variables based on the Federal Railroad Administration (FRA) standards and Highway–Rail Crossing Inventory Instructions. In addition, since some information may be lost or difficult to obtain during data collection, certain variables contained missing values. When the proportion of missing data for a given variable was less than 2%, the missing entries were directly deleted in this study. Following the optimization of the dataset, a total of 14,197 accident records were retained, with the yearly distribution presented in Figure 1 below.

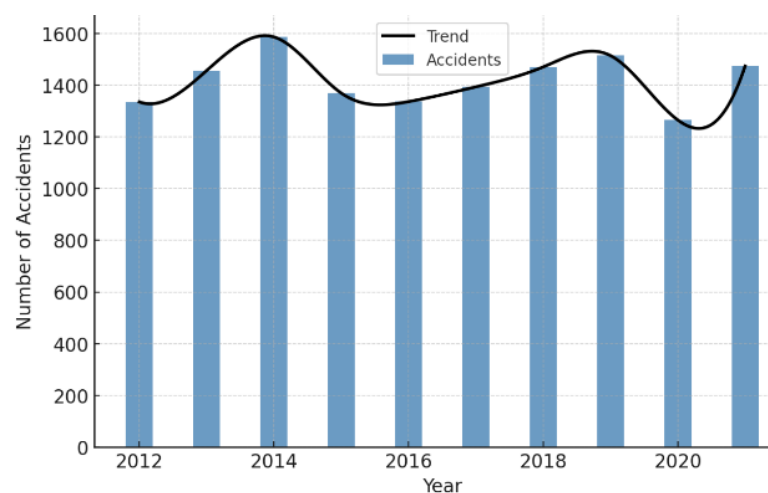


Figure 1. Annual number of accidents.

3.2. Data Description

To analyze the impact of various features on the severity of highway–rail grade crossing accidents under different road surface conditions and their interaction with nighttime lighting conditions, the dataset was partitioned, and the process is shown in Figure 2. Instead of simply including surface and lighting as categorical variables, the dataset was partitioned into four combined conditions to explicitly analyze their interaction effects. The original six road surface conditions (A = dry, B = wet, C = snow/mud, D = ice, E = sand/mud/dust/oil/gravel, F = water (standing/flowing)) were reclassified into two categories: (1) dry (A = dry); (2) Non-dry (B = wet, C = snow/mud, D = ice, E = sand/mud/dust/oil/gravel, F = water (standing/flowing)). Lighting conditions were divided into two categories: with or without lighting. Under these two classification schemes, variable cross-tabulation was performed, and the processed dataset was divided into four subsets: (1) Condition 1: Dry road surface with lighting; (2) Condition 2: Dry road surface without lighting; (3) Condition 3: Non-dry road surface with lighting; (4) Condition 4: Non-dry road surface without lighting. The sample classification under these four different conditions is shown in Figure 3.

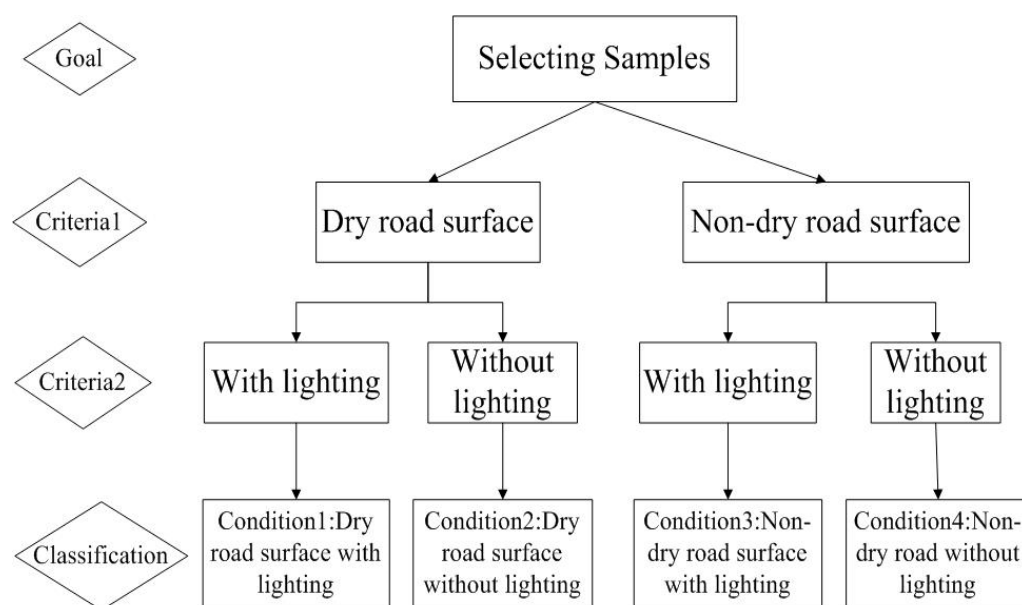


Figure 2. Dataset partitioning process.

■ Condition1 ■ Condition2 ■ Condition3 ■ Condition4

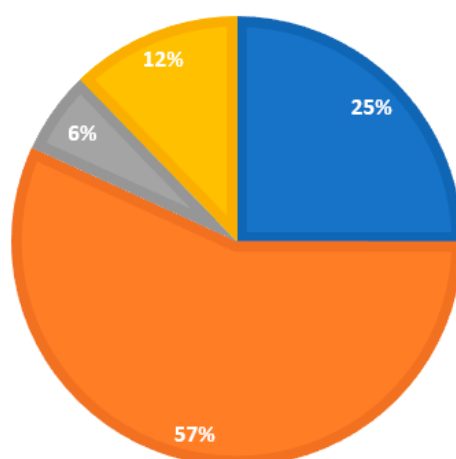


Figure 3. Sample distribution by category.

Because the FRA database contains a substantial amount of missing data, data pre-processing is particularly important. In this study, the unique identifier assigned to each crossing was used to link the accident records with crossing characteristic information. Crossings without valid identification numbers were removed from the dataset. The study utilized raw data from the Highway–Rail Grade Crossing Accident database and the Crossing Inventory maintained by the FRA, comprising a total of 282 variables. Among these, the accident dataset contains 102 variables, with “DRIVER” used to define the classification of crash severity. Based on previous studies on highway–rail grade crossing safety and the FRA accident reporting standards, the independent variables were selected and categorized into four major groups: driver characteristics, vehicle characteristics, crossing characteristics, and environmental conditions. On this basis, twelve representative variables were selected for this study, and a description of the variables used in this analysis is shown in Table A1.

In this paper, the injury severity as the dependent variable is ranked as 0—property damage only (PDO), 1—injury, and 2—killed. For the dataset of dry road surface with lighting conditions, the percentage distribution of collisions across three injury severity levels is as follows: 73.70% PDO, 21.88% injured, and 4.42% killed. For the dataset of dry road surface without lighting conditions, the percentage distribution of collisions across three injury severity levels is as follows: 68.23% PDO, 25.27% injured, and 6.50% killed. For the dataset of non-dry road surface with lighting conditions, the percentage distribution of collisions across three injury severity levels is as follows: 79.55% PDO, 16.55% injured, and 3.90% killed. For the dataset of non-dry road surface without lighting conditions, the percentage distribution of collisions across three injury severity levels is as follows: 74.34% PDO, 21.11% injured, and 4.55% killed. The frequency and percentage distribution of these variables are shown in Table A2.

4. Methodology

To improve the accuracy of interpreting and analyzing the interactions among influencing factors, this study selected four machine learning models for analysis: Random Forest (RF), Gradient Boosting Decision Tree (GBDT), XGBoost, and LightGBM. All four models share comparable ensemble-learning architectures: XGBoost, LightGBM, and GBDT are based on gradient boosting mechanisms, whereas Random Forest represents a bagging-based approach. This structural alignment allows for a balanced comparison across different ensemble paradigms. Furthermore, these ensemble learning models are highly effective in handling high-dimensional data, capturing nonlinear relationships, and modeling complex interactions among variables. The study integrates the XGBoost algorithm with SHAP (Shapley Additive Explanations) principles to interpret and analyze influential factors in the model, and the framework diagram is shown in Figure 4.

Although the combination of XGBoost and SHAP has been widely applied in safety analysis, most existing studies have focused primarily on global feature ranking or marginal effect visualization. In contrast, this study does not rely on a single global model but instead partitions the dataset into four combined road-surface–lighting scenarios, allowing for context-specific mechanism analysis. This approach enables the model to capture environmental heterogeneity that conventional global frameworks tend to overlook.

Moreover, rather than limiting the analysis to global feature importance, this study employs SHAP interaction values to quantitatively reveal how pairs of variables jointly influence crash severity. This approach uncovers synergistic and interaction effects that have not been identified in previous XGBoost–SHAP studies.

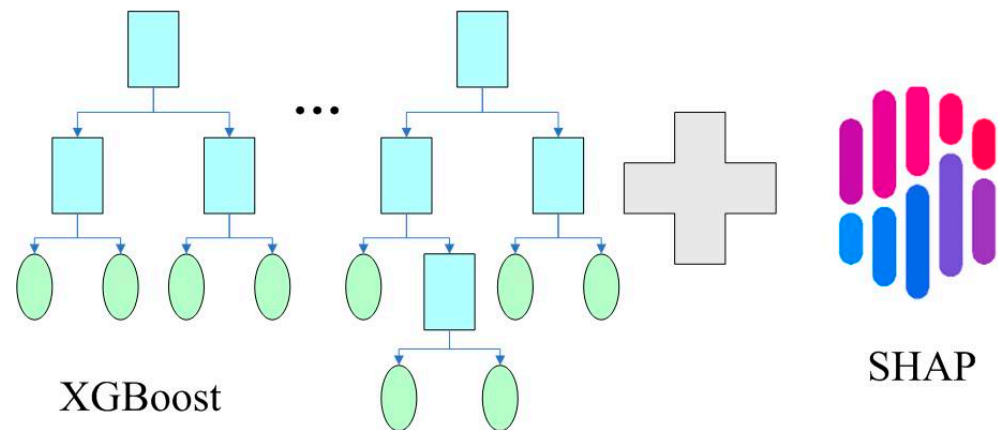


Figure 4. Diagram of the XGBoost-SHAP framework.

Overall, the proposed framework is specifically tailored to the safety-critical context of highway–rail grade crossings (HRGCs), where environmental, operational, and human factors interact in complex nonlinear ways. By integrating SHAP-based interpretations with established transportation safety principles, the study extends prior XGBoost–SHAP applications and develops a context-sensitive, interaction-aware interpretability framework for complex transportation systems.

4.1. XGBoost

The XGBoost algorithm incorporates regularization terms into its objective function to smooth the learned weights and prevent overfitting, while employing second-order Taylor expansion of the loss function to enhance predictive accuracy. The formula of predicted value is as follows [31]:

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (1)$$

where $\hat{y}_i^{(t)}$ is the predicted value for sample i after the t -th iteration in the XGBoost algorithm, $\hat{y}_i^{(t-1)}$ is the predicted output from the $(t - 1)$ -th tree, $f_t(x_i)$ is the model prediction from the t -th tree.

The model's objective function consists of a loss function and a regularization term.

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{j=1}^t \Omega(f_j) \quad (2)$$

where $Obj^{(t)}$ denotes the model's objective function; $\sum_{i=1}^n l(y_i, \hat{y}_i^{(t)})$ represents a loss function that measures the discrepancy between predicted values $\hat{y}_i$ and target values y_i ; $\sum_{j=1}^t \Omega(f_j)$ is the regularization term, also known as the penalty function.

The dataset was randomly divided into a training set (80%) and a testing set (20%), stratified by crash severity to maintain the original class proportions. To enhance the stability of the estimates, a parameter grid was established for hyperparameter tuning, and a five-fold cross-validation scheme was employed to evaluate model performance and select the optimal parameter combination.

This study optimized several key hyperparameters of the XGBoost model to maximize its performance. The resulting hyperparameters are summarized in Table 1. Each tree was trained using 70% of the samples and 70% of the features, effectively reducing overfitting while maintaining high predictive performance and improving generalization capability. The learning rate was set to 0.1, ensuring a stable and efficient convergence process. The maximum tree depth was set to 4, allowing the model to capture complex feature interac-

tions without excessive complexity. A total of 100 trees were used, which, in combination with the relatively low learning rate, provided an accurate and stable fit to the data.

Table 1. Optimized Hyperparameter Values for XGBoost Model.

Parameter	Parameter Meaning	Optimum Value
learning_rate	Controls model training time	0.1
n_estimators	Number of trees	100
max_depth	Maximum tree depth	4
subsample	Subsample ratio	0.7
colsample_bytree	Feature subsampling ratio	0.7

4.2. Shapley Additive exPlanations

SHAP (Shapley Additive exPlanations) is a game-theory-based approach for interpreting machine learning model predictions, proposed by Lundberg et al. [32]. It quantifies feature contributions to model outputs by assigning each feature its marginal impact, thereby enhancing model interpretability. The SHAP value calculates the average marginal contribution of each feature i to the model's prediction, as expressed by [32]:

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (3)$$

where φ_i denotes the SHAP value (contribution) of the i -th feature; N represents the complete feature set; S indicates a feature subset (excluding feature i); $f(S)$ is the model's predicted output using feature subset S ; $f(S \cup \{i\})$ quantifies the marginal contribution of feature i .

The SHAP (SHapley Additive exPlanations) model reformulates the interpretability problem as an additive feature attribution task, establishing a linear combination-based explanatory framework. Its core principle decomposes complex predictions into the sum of individual feature contributions, enabling users to quantify each feature's impact. The formulation is as follows:

$$f(x) = \varphi_0 + \sum_{i=1}^M \varphi_i \quad (4)$$

where $f(x)$ is the model's predicted value for input sample x ; M is the number of features; φ_0 is the mean prediction across all samples; φ_i is the SHAP value for feature i , also known as the contribution value of feature i to the model's prediction.

4.3. Evaluation Indicators of the Model

In XGBoost, a confusion matrix is a tool used to evaluate the performance of a classification model. It is a two-dimensional matrix that compares the model's predicted results with the actual class labels. The confusion matrix is particularly useful for understanding the model's performance across different classes.

The confusion matrix consists of four primary components:

- (1) True Positives (TP): The quantity of samples correctly predicted as positive by the model, where actually positive instances are accurately classified as positive.
- (2) True Negatives (TN): The number of samples correctly predicted as negative by the model, where truly negative instances are accurately classified as negative.
- (3) False Positives (FP): The number of samples incorrectly predicted as positive by the model, where truly negative instances are misclassified as positive.
- (4) False Negatives (FN): The number of samples incorrectly predicted as negative by the model, where truly positive instances are misclassified as negative.

The confusion matrix enables the computation of various performance metrics (such as accuracy, precision, recall, and F1-score), thereby facilitating a comprehensive evaluation of XGBoost's classification performance.

Accuracy measures the ratio of correct predictions (both true positives and true negatives) to the total number of samples evaluated, with higher values denoting superior classification performance. The formulation is as follows:

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

Recall: The proportion of actual positive samples that are correctly predicted as positive. Higher recall indicates better performance. The formulation is as follows:

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Precision (also known as positive predictive value) measures the model's ability to avoid false positives, calculated as the ratio of true positives to the sum of true positives and false positives. Maximizing precision is essential when the cost of misclassifying negative instances is high. The formulation is as follows:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

The F1-score, as the harmonic mean of precision and recall, provides a single metric to evaluate classifier performance when class distribution is imbalanced. Unlike the arithmetic mean, it penalizes extreme disparities between precision and recall, making it particularly valuable in scenarios where both false positives and false negatives carry significant costs. The formulation is as follows:

$$F1 - Score = \frac{2TP}{2TP + (FP + FN)} \quad (8)$$

5. Results

5.1. Model Comparison

The study evaluates the fitting performance of the constructed XGBoost model by introducing LightGBM, Random Forest, and GBDT as comparative models. The performance of different machine learning models is comprehensively compared using metrics such as Accuracy, Precision, Recall, and F1-Score, with the results presented in Table 2.

Table 2. Model performance comparison.

Model	Accuracy (Train)	Accuracy (Test)	Recall	Precision	F1-Score	AUC
LightGBM	0.932	0.731	0.731	0.693	0.703	0.804
Random Forest	0.710	0.723	0.723	0.553	0.608	0.788
GBDT	0.742	0.743	0.743	0.703	0.704	0.808
XGBoost	0.769	0.741	0.741	0.699	0.705	0.812

In the complete dataset, the XGBoost model achieved an accuracy (training) of 76.9%, accuracy (test) of 74.1%, recall of 74.1%, precision of 69.9%, F1-score of 70.5%, and AUC of 0.812. XGBoost exhibits exceptional generalization capability, as manifested by consistent performance across both training and testing datasets (Δ accuracy = 2.8%) without evidence of overfitting. The model achieves the highest ROC AUC score (0.91) among comparative methods, demonstrating superior class separability. With an F1-score of 0.705, marginally

exceeding GBDT (0.704), XGBoost maintains the most favorable precision-recall balance, indicative of optimal model stability for real-world deployment.

Figure 5 presents the ROC (a) and PR (b) curves of all classifiers. All four models exhibited high and comparable predictive performance, with AUC and PR-AUC values around 0.90–0.91. Among them, XGBoost and LightGBM slightly outperformed the others, showing smoother and more stable curves across different thresholds, while Random Forest achieved marginally lower scores (AUC = 0.90). These results demonstrate that XGBoost provides superior stability and generalization compared with traditional ensemble methods.

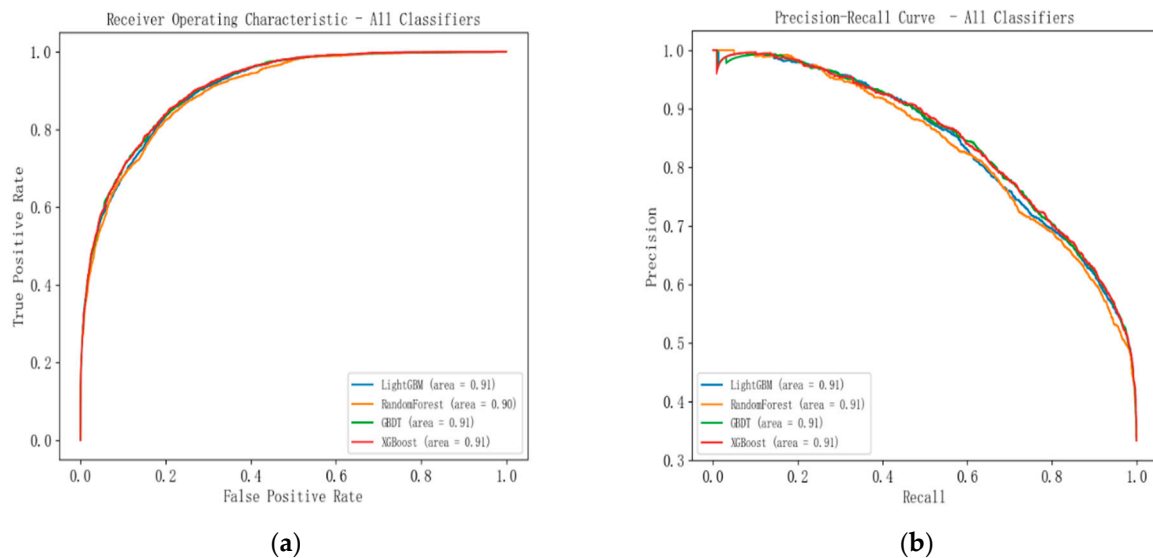


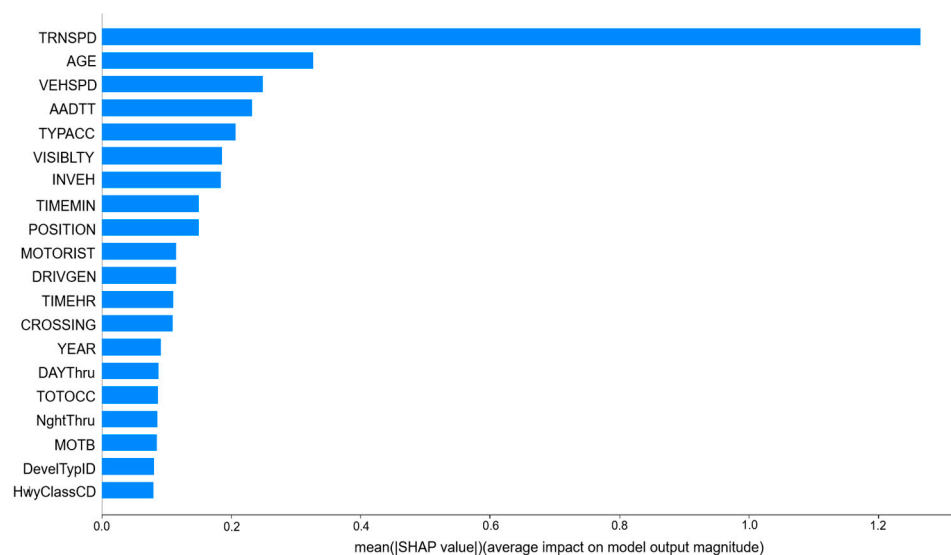
Figure 5. Performance evaluation: (a) ROC curves; (b) PR curves.

Among the evaluated machine learning models, XGBoost demonstrated superior overall performance. Therefore, the study selects the XGBoost algorithm for analyzing determinant factors influencing accident severity at highway–rail grade crossings.

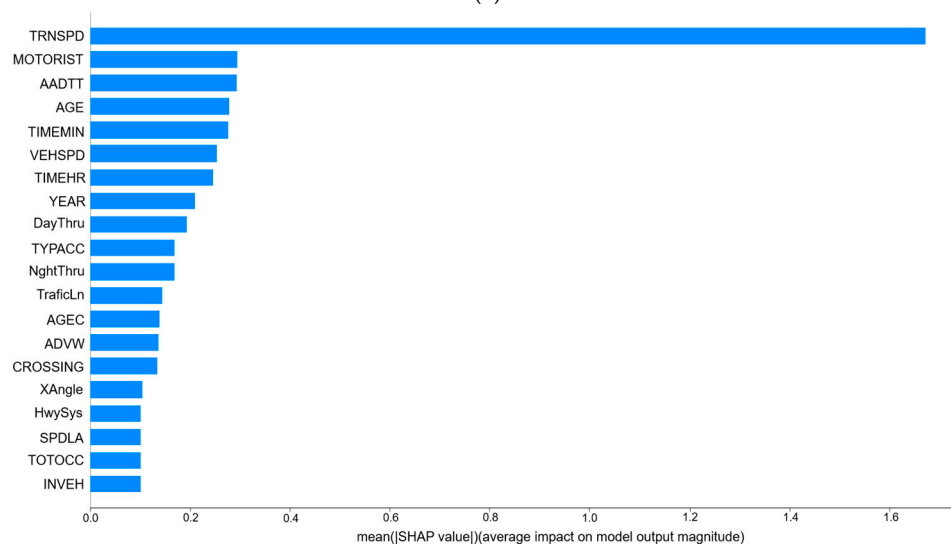
5.2. Importance of Risk Factors

This study employs the SHAP method to conduct global and local interpretability analysis of the XGBoost model, investigating the influence of different feature inputs on the model's prediction outcomes.

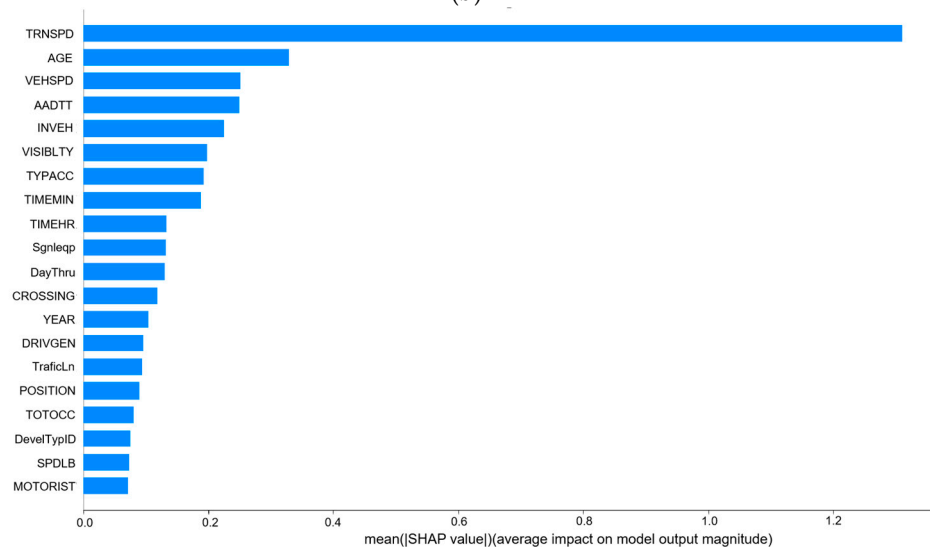
The relative importance of various features affecting accident severity at highway–rail grade crossings was quantitatively assessed through SHAP (SHapley Additive exPlanations) attribution analysis, as illustrated in Figure 6. The abscissa denotes the mean absolute SHAP values, representing the magnitude of feature contributions, while the ordinate displays the predictive factors ranked in descending order based on their relative importance to the model's interpretability. The analysis of risk factors revealed several critical variables that consistently influence accident severity across all conditions. Figure 6 illustrates the global feature importance derived from the XGBoost model. The results demonstrate that the five most influential features are: "TRNSPD" (train speed), "AGE" (driver age), "VEHSPD" (vehicle speed), "AADTT" (average annual daily traffic), and "TYPACC" (type of accident).



(a)



(b)



(c)

Figure 6. Cont.

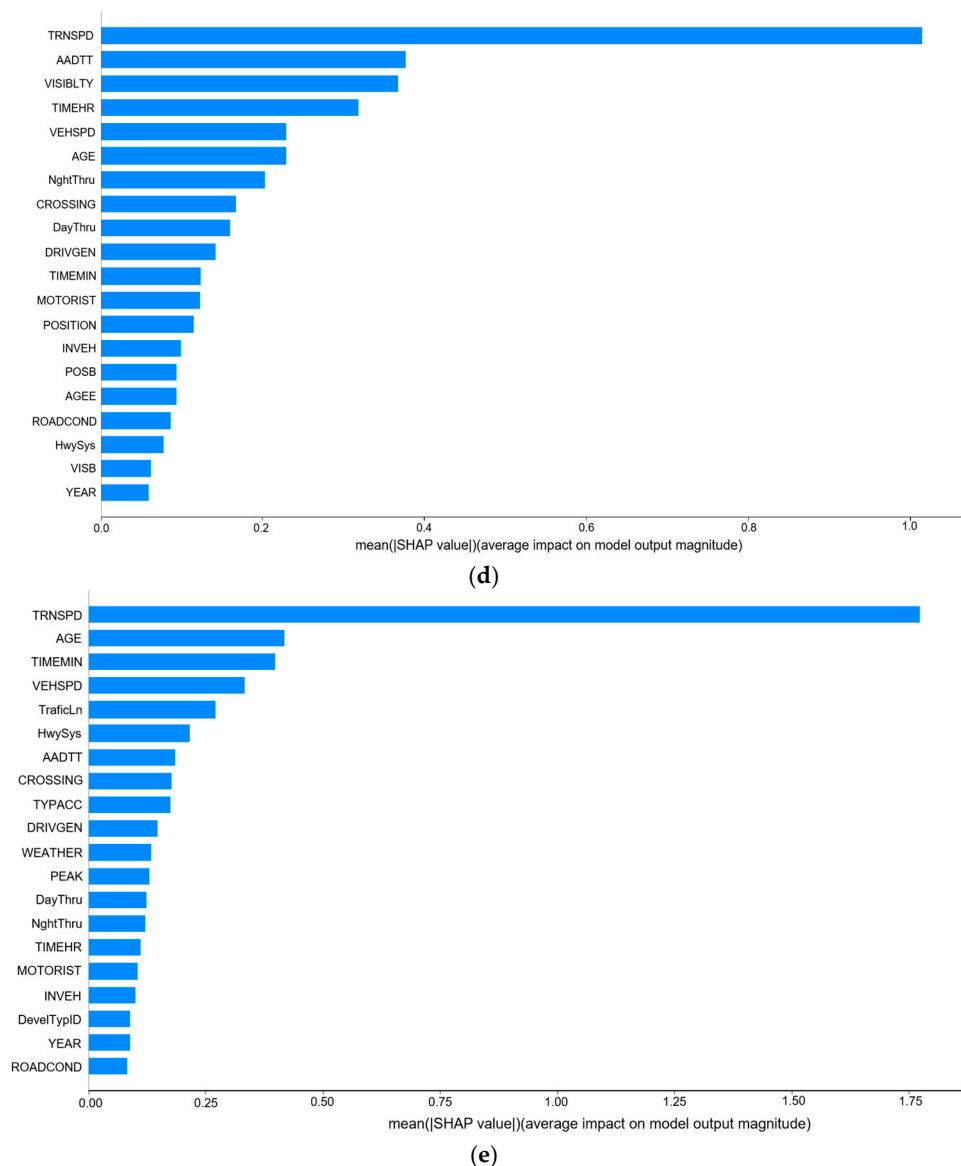


Figure 6. Feature variable importance based on the SHAP model: (a) Not Grouped; (b) Condition 1; (c) Condition 2; (d) Condition 3; (e) Condition 4.

TRNSPD (Train Speed)

The “train speed” variable represents the operational speed reported at the moment of collision, typically ranging from 10 to 50 mph. The crashes analyzed in this study involve conventional trains operating at speeds below 80 mph, rather than high-speed rail systems. In the ungrouped analysis (Figure 6a), train speed was identified as the single most influential factor determining accident severity. Its SHAP values consistently showed strong positive contributions, reflecting the exponential increase in collision force and the limited reaction time at higher velocities. When conditions were disaggregated, train speed remained the top determinant across all four pavement–lighting combinations, but its relative importance varied. Under dry-lit conditions (Figure 6b), its influence was high yet somewhat moderated by visibility and advanced safety infrastructure. In dry-unlit and non-dry-lit settings, the contribution of train speed intensified, while in the non-dry-unlit scenario, it reached its maximum, clearly indicating a compounding effect of poor visibility and reduced friction. This pattern highlights that while train speed is globally dominant, its risks are disproportionately magnified in adverse environmental contexts.

VEHSPD (Vehicle Speed)

Globally, vehicle speed ranked within the top five variables in Figure 6, showing a strong positive association with severe outcomes. When analyzed across conditions, vehicle speed displayed clear context dependence. On dry-unlit pavements, vehicle speed ranked among the top three predictors, reflecting how the absence of lighting makes high vehicular velocity especially hazardous. In contrast, under dry-lit conditions, its importance was lower, as visibility partially mitigated risks. In non-dry-unlit scenarios, vehicle speed again became one of the top contributors, reinforcing the interaction effect whereby slippery surfaces and poor visibility exacerbate the dangers of excessive speed. This pattern confirms that vehicular speed is not uniformly risky but is strongly modulated by pavement and lighting conditions.

AGE (Driver Age)

Driver age demonstrated a consistent effect across all scenarios, with older drivers (65+) more prone to severe outcomes. This effect reflects age-related declines in visual acuity, cognitive processing, and motor response. Notably, age gained greater importance in unlit conditions, where diminished visibility compounded the limitations of older drivers, further delaying hazard recognition and reaction. In non-dry and unlit environments, age effects interacted with speed-related factors, producing a marked escalation of severity risk. These findings underscore the need for targeted interventions for elderly drivers, particularly under poor lighting and adverse surface conditions.

MOTORIST (Motorist Behavior)

Although not among the top five variables in the ungrouped analysis (Figure 6a), motorist behavior emerged as a critical factor when disaggregated by condition. Under dry-lit pavements, it ranked second, demonstrating that risky behaviors such as gate violations or failure to stop contribute substantially to severity even when environmental conditions are favorable. However, under dry-unlit conditions, its relative influence declined, as speed and age factors became more dominant. This shift implies that in poor lighting, the direct risks of environmental and physiological constraints overshadow risky behaviors, though violations remain critical. Thus, motorist behavior is particularly important in contexts where drivers perceive visibility as sufficient, and violations become the decisive human factor in accident outcomes.

AADTT (Average Annual Daily Traffic)

AADTT (average annual daily traffic) played a substantial role in shaping severity outcomes but with notable variation across conditions. On dry-lit and non-dry-lit pavements, AADTT ranked among the top three variables, suggesting that crossings with higher traffic volumes face elevated risks due to queuing and congestion. However, in non-dry-unlit conditions, its importance declined, as environmental and infrastructural factors became more decisive. This finding reflects a context-dependent effect: while traffic exposure can increase risks under favorable conditions, its influence is partly overshadowed when environmental hazards dominate.

Taken together, Table 3 highlights that while speed-related factors (train and vehicle velocity) remain dominant under all conditions, their contributions are disproportionately magnified under non-dry and unlit scenarios. Driver age and motorist behavior exhibit context-specific importance, with age effects amplified in unlit environments and risky driving behavior most critical under adequate lighting. Traffic exposure also shows conditional influence, further illustrating that the interaction between road surface and lighting plays a central role in shaping the relative importance of risk factors.

Table 3. Importance ranking of key risk factors under different road surface and lighting conditions.

Variable	Condition 1	Condition 2	Condition 3	Condition 4	Overall (Not Grouped)
TRNSPD	1	1	1	1	1
VEHSPD	6	3	5	4	3
AGE	4	2	6	2	2
MOTORIST	2	20	12	16	10
AADTT	3	4	2	7	4

5.3. SHAP Analysis

The SHAP analysis is conducted to interpret the marginal contributions of each variable to crash severity predictions. Figure 7 presents the SHAP summary plot for different conditions, illustrating the mechanistic influences of various features on accident severity at highway–rail grade crossings. Each data point in the plot represents an individual sample observation. The x-axis displays SHAP values, calculated as the mean absolute contribution across all samples, which quantifies each feature’s relative impact on model predictions. Higher absolute SHAP values indicate greater influence on the predictive outcome, where positive values denote risk-increasing effects and negative values represent protective effects. The study concentrates on analyzing the top five features influencing accident severity:

Train speed has a strong positive influence on accident severity, meaning that as train speed increases, the probability of severe or fatal outcomes rises. This is because higher speed not only shortens the available reaction time but also increases collision energy exponentially. Under non-dry and unlit conditions, this positive effect is even stronger, as wet or icy surfaces reduce braking efficiency and the absence of lighting further limits visibility, together amplifying the hazard.

Vehicle speed also shows a consistent positive effect. Faster vehicle movement increases the likelihood of severe crashes, primarily by reducing stopping distance and elevating impact force. On dry and well-lit roads, this effect is less pronounced because visibility and traction allow drivers to compensate. However, in unlit or non-dry conditions, the positive contribution of vehicle speed becomes more evident, reflecting how slippery surfaces and poor lighting combine with high velocity to raise risk.

Driver age demonstrates a positive relationship with severity, particularly for older drivers. Higher age is linked to greater crash severity due to slower reactions and reduced hazard perception. In lit conditions, the positive effect of age is weaker, as better visibility partly offsets age-related disadvantages. Yet in unlit and non-dry scenarios, the positive effect grows stronger, showing that poor visibility magnifies the vulnerabilities of elderly drivers.

Motorist behavior has a positive effect when violations occur, such as driving around gates or failing to stop. Under dry and lit conditions, this variable contributes strongly, as reckless actions become the dominant trigger of severity even when the environment is favorable. In unlit conditions, however, its relative contribution is reduced, as the effects of speed and age overshadow behavior. Nevertheless, when combined with high speeds, unsafe driving actions still significantly push outcomes toward more severe levels.

AADTT exerts a mixed influence. At low-traffic crossings, its SHAP values are more positive, indicating greater severity, likely because drivers underestimate train arrival frequency and are more prone to risky maneuvers. At high-traffic crossings, SHAP values tend to be negative, suggesting that congestion and slower approach speeds lower the risk of severe outcomes. In adverse scenarios, such as non-dry and unlit conditions, the influence of AADTT weakens, as environmental hazards dominate.

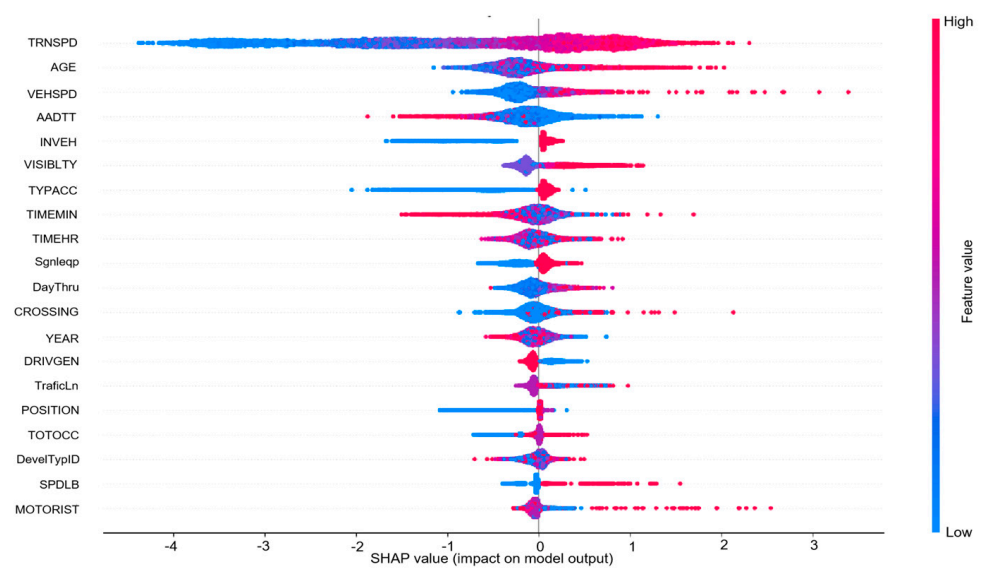
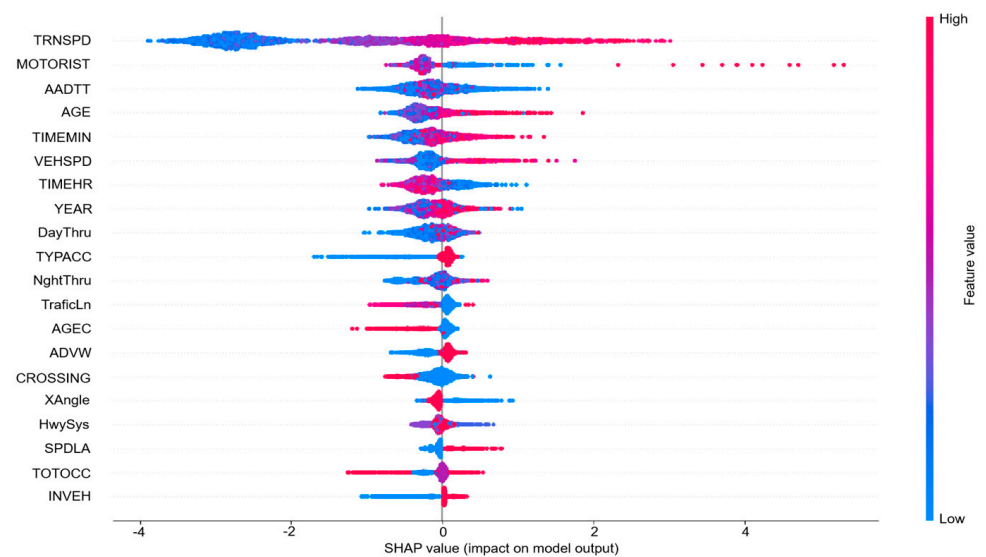
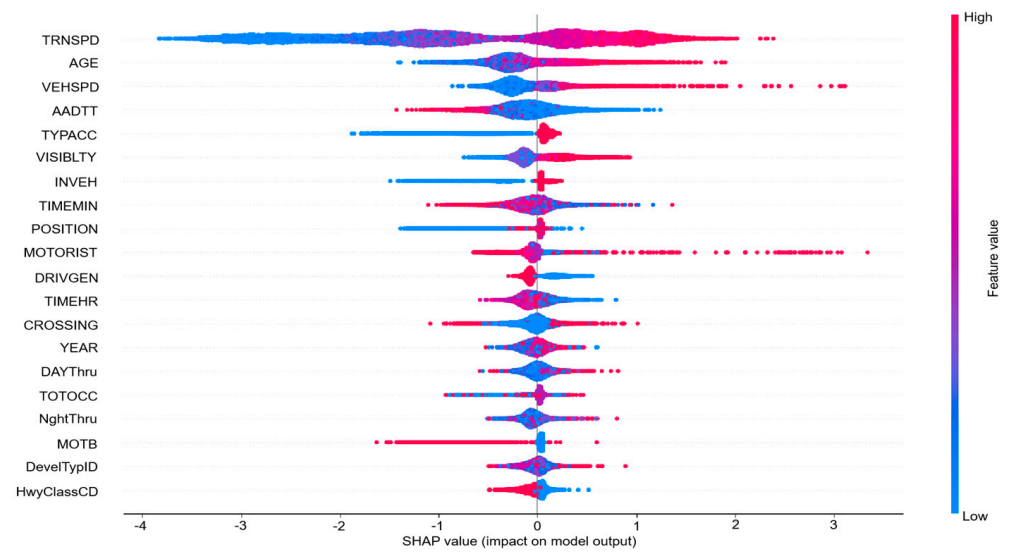


Figure 7. Cont.

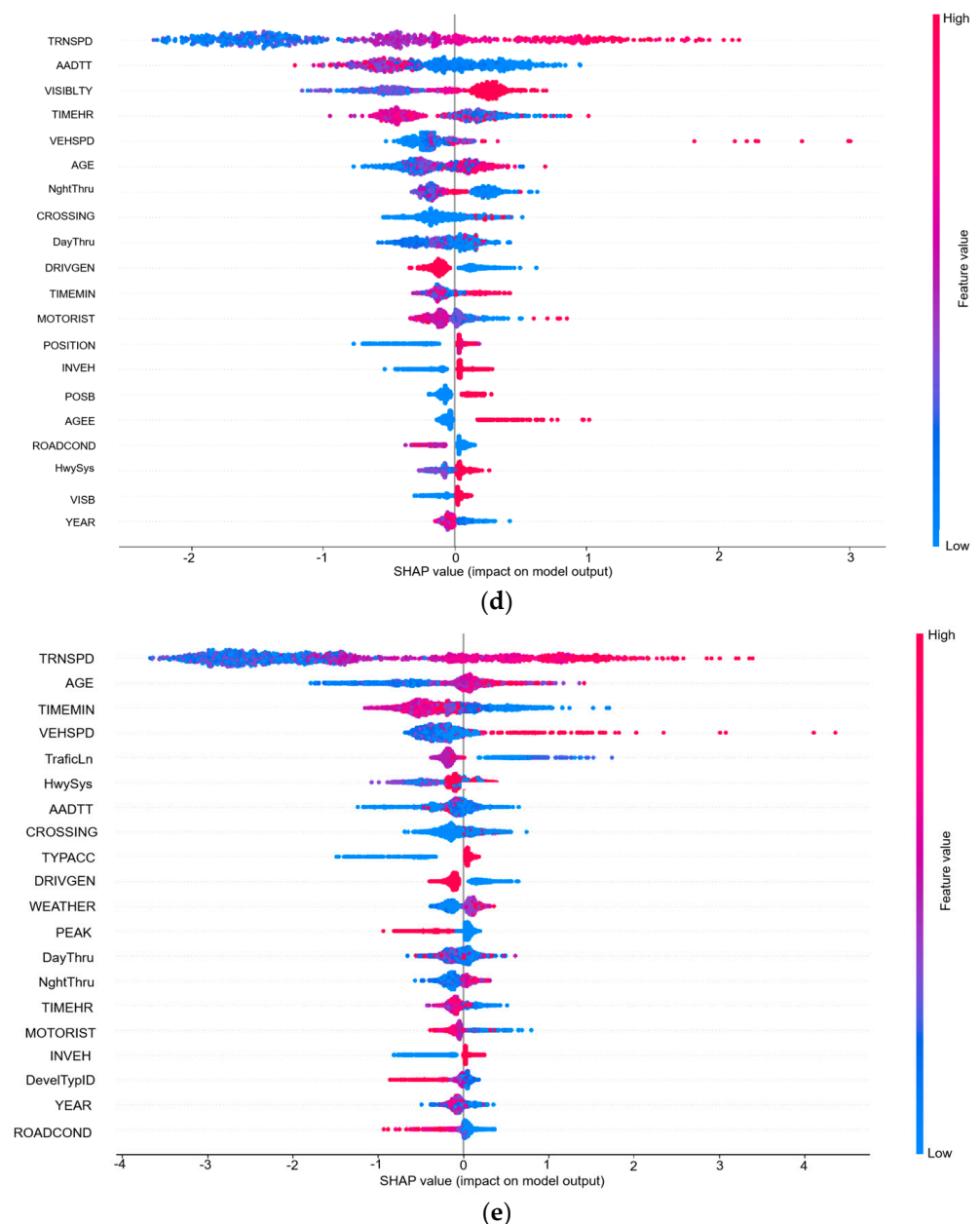


Figure 7. SHAP summary plots: (a) Not Grouped; (b) Condition 1; (c) Condition 2; (d) Condition 3; (e) Condition 4.

Weather and visibility also show distinct effects in grouped analyses. Poor visibility and adverse weather conditions have positive contributions, meaning they increase the likelihood of severe accidents, especially when paired with high speed. In contrast, clear weather and good visibility produce negative contributions, serving as protective factors. The most critical condition is the combination of non-dry surfaces and absent lighting, where positive contributions from speed, age, and environmental factors converge, producing the highest severity.

To study the contribution of various factors to the severity of driver injuries in accidents at highway–railroad grade crossings, SHAP force plots were generated for the accident dataset in each scenario, as shown in Figure 8. The following is a detailed analysis:

Condition 1: AGE = 41 (driver age) and MOTORIST = 1 (Motor vehicle driver's action at highway–rail grade crossing) increase the severity of driver injuries. In contrast, TRNSPD = 79 (train speed), AADTT = 8503 (average annual daily traffic) and

NightThru = 15 (Number of through trains passing the grade crossing at night) reduce the predicted severity. A potential explanation is that elevated train speeds are predominantly observed at crossings with advanced safety infrastructure, optimal visibility conditions, and substantial traffic flow. These factors may synergistically mitigate collision severity.

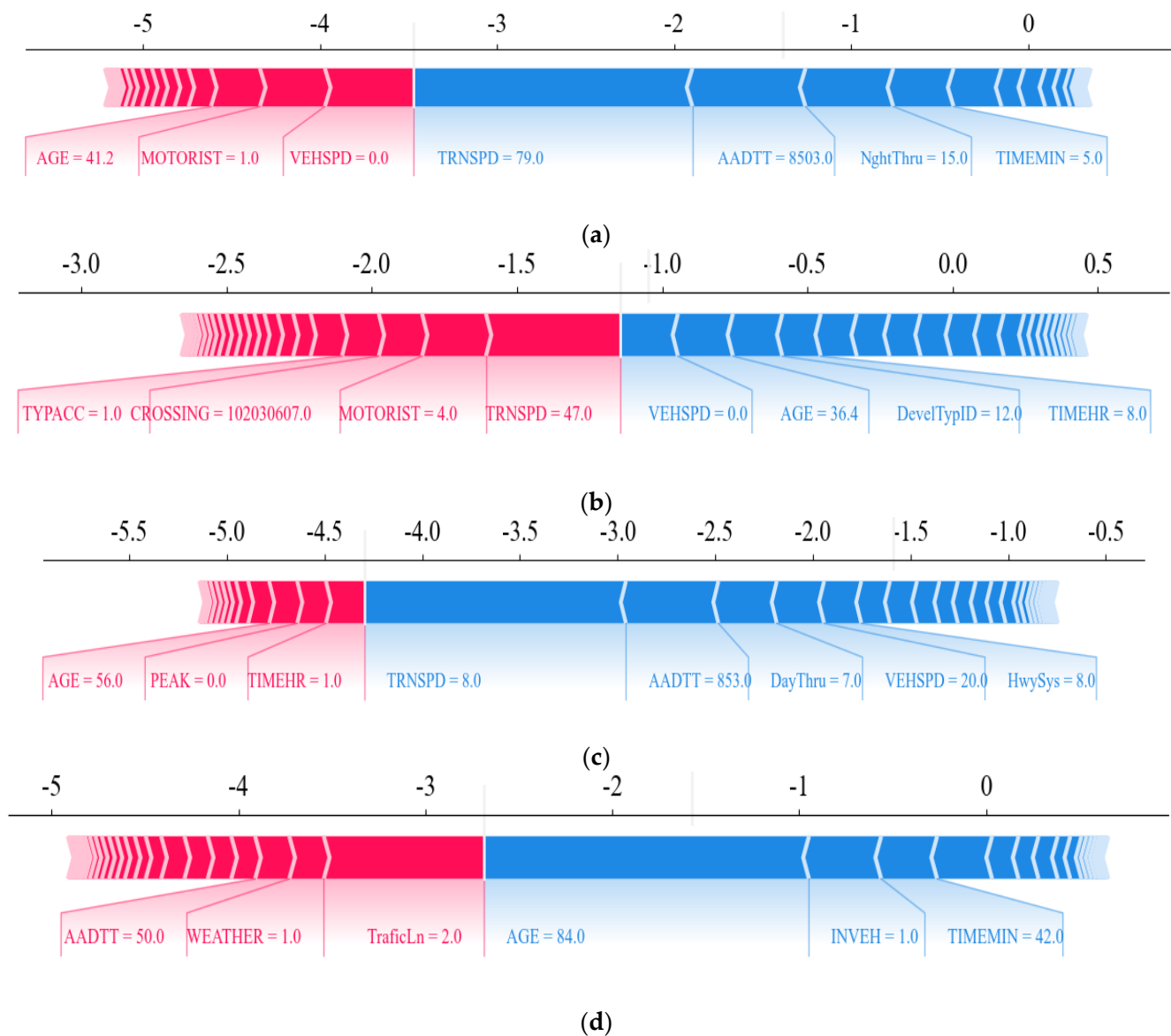


Figure 8. SHAP force plots: (a) Condition 1; (b) Condition 2; (c) Condition 3; (d) Condition 4.

Condition 2: Factors such as TYPACC = 1 (train colliding with a motor vehicle), MOTORIST = 4 (Motor vehicle driver's action at highway–rail grade crossing) and TRNSPD = 47 (train speed) elevate severity. Conversely, VEHSPD = 0 (vehicle speed) and AGE = 36.4 (driver age) reduce the severity of injuries. This suggests that on unlit dry road surfaces, accident severity is predominantly determined by collision direction, driving behavior, and train speed, while lower vehicle speeds and younger drivers are associated with reduced risk.

Condition 3: Under this condition, AGE = 56 (driver age) increases the severity of driver injuries. However, TRNSPD = 8 (train speed) and VEHSPD = 20 (vehicle speed) show a protective effect against severe collisions.

Condition 4: Factors such as AADTT = 50 (average annual daily traffic) and WEATHER = 1 (adverse weather conditions) are significantly associated with increased

injury severity. On the contrary, factors such as INVEH = 1 (driver inside the vehicle), AGE = 56 (driver age), and TIMEMIN = 42 (time in minutes) attenuate the severity of injuries. This suggests that non-dry pavement without lights represents the most hazardous condition, where environmental factors and road infrastructure predominantly determine injury severity levels.

In summary, SHAP analysis indicates that speed-related factors, driver age, and risky behaviors all have strong positive effects on crash severity, but their marginal influence is shaped by pavement and lighting conditions. Non-dry and unlit scenarios create a compounding effect in which environmental hazards and human vulnerabilities interact, leading to the greatest escalation of crash risk. The findings demonstrate that absent illumination consistently amplifies the adverse effects of speed and driving behaviors, while non-dry pavement conditions shift criticality to environmental and infrastructural variables. Additionally, The SHAP force plots reveal significant local interaction effects, facilitating the development of context-specific safety strategies.

6. Discussion and Policy Implication

This study advances the understanding of crash severity at highway–rail grade crossings by demonstrating that pavement condition and lighting do not simply exert additive influences but interact in ways that fundamentally reshape severity outcomes. Prior research has consistently identified the independent effects of lighting [8] and road-surface conditions [9] on injury severity, yet these factors have been modeled separately and often treated as parallel contributors. The present findings provide the first empirical evidence, using SHAP interaction analysis, that the joint effect of non-dry surfaces and absent illumination produces a super-additive risk pattern—a mechanism that earlier studies largely implied but never quantified.

Compared with Hao and Kamga's work [8], which attributed higher severity levels at unlit crossings to reduced visibility, our results show that this effect is substantially intensified when darkness co-occurs with wet, icy, or contaminated surfaces. Similarly, Yang et al. [9] highlighted the role of surface slipperiness but did not examine how its influence changes under different illumination conditions. The present analysis shows that under poor lighting, the marginal contribution of non-dry pavement nearly doubles, indicating that illumination is not merely a background factor but a condition that alters the operational meaning of surface friction. This aligns with Morgan and Mannering's broader observation [17] that environmental factors interact with demographic characteristics but goes further by quantifying the structural form of these interactions.

According to the model results, accident severity at highway–rail grade crossings arises from the intertwined effects of environmental conditions, operational characteristics, and driver behavior. Rather than acting independently, these elements interact in ways that reshape the risks associated with each factor. The most striking finding is the amplified influence of train and vehicle speeds under non-dry and unlit conditions, revealing that the combination of reduced surface friction and poor visibility produces a level of danger that exceeds the simple sum of their individual effects. This super-additive pattern reflects the underlying physical dynamics of crash development: moisture or debris on the pavement lengthens braking distance, while darkness delays hazard recognition. When these two conditions occur together, the remaining reaction window becomes sharply compressed, and even small increases in speed can lead to disproportionately severe outcomes. The interaction uncovered by the SHAP analysis makes clear that environmental factors cannot be treated as isolated contributors but instead form a coupled system that significantly alters the consequences of driver and train actions.

The behavior of older drivers further illustrates how human limitations interact with adverse environmental conditions. Age-related declines in visual processing and reaction time do not simply add to the challenges created by poor lighting; the two constraints reinforce each other. In low-light settings, drivers already struggle to detect the presence or speed of an approaching train, and this perceptual gap is widened when visibility is further reduced by rain, snow, or road contamination. As a result, older drivers face a compounded disadvantage: they recognize hazards later and have less effective braking capacity available once they do. The model's output captures this combined effect, with driver age contributing more strongly to severe outcomes in the scenarios where illumination and surface conditions are jointly unfavorable.

The influence of risky driver actions also varies with environmental context, providing insight into how behavior adapts or fails to adapt to changing conditions. In well-lit and dry situations, violations such as driving around gates or failing to stop play a decisive role in determining severity. Under such conditions, drivers may perceive the crossing environment as predictable and manageable, leading them to engage more confidently in unsafe actions. However, once visibility deteriorates or the pavement becomes slippery, the relative impact of these behaviors shifts. The environment itself becomes a dominant source of danger, but when violations still occur under these conditions, their consequences escalate sharply. The model captures this shift by showing that the marginal contribution of driver behavior to severity varies across the four scenarios, reinforcing the need to interpret behavioral data within its environmental context rather than in isolation.

As a result, based on the identified high-risk factors, the following targeted improvement measures are proposed. Firstly, speed management measures for both trains and vehicles should be implemented. Under conditions of poor lighting and adverse weather, dynamic speed limits should be applied, with real-time adjustments based on weather, visibility, and traffic conditions to reduce the likelihood of accidents. Secondly, authorities should strengthen driver safety education and enforcement. For high-risk behaviors such as illegal crossing or driving around barriers, enhanced video surveillance and automated enforcement should be adopted, along with targeted safety training for elderly drivers. Additionally, low-traffic but high-risk crossings should be prioritized for intelligent warning systems and stricter safety management measures [33]. Last but not least, high-intensity LED lighting and anti-skid facilities should be installed at high-risk crossings to improve visibility and road safety. Under wet or slippery road conditions, additional lighting can be activated, while under dry conditions, lighting can be reduced. This adaptive lighting strategy not only ensures adequate visibility to enhance traffic safety but also contributes to energy conservation and reduced carbon emissions, aligning with both safety and sustainability objectives.

7. Conclusions and Future Works

This study systematically investigates the key determinants of accident severity at highway–rail grade crossings (HRGCs) using an integrated XGBoost–SHAP framework, with a particular focus on the interaction effects of road surface and lighting conditions. By analyzing a decade of FRA accident records, the results demonstrated that speed-related factors (train and vehicle velocities), AGE, MOTORIST, and AADTT consistently emerged as the most influential determinants of crash outcomes. SHAP analysis further revealed that these factors exert context-dependent contributions, with their marginal effects substantially altered by pavement and illumination conditions.

A key finding is the critical role of interaction effects. Although train speed remains globally the dominant predictor, its positive contribution to crash severity was disproportionately magnified under non-dry and unlit scenarios. Vehicle speed and driver age

exhibited similar amplification, with older drivers facing particularly high risks when visibility was poor. Motorist violations, while always hazardous, were most decisive under dry and lit environments where human error becomes the primary trigger. Traffic exposure showed a nonlinear influence: low-volume crossings paradoxically displayed higher severity risks, particularly when adverse environmental conditions were present. Together, these findings highlight a “super-additive effect” in which non-dry pavement and absent illumination jointly elevate severity beyond the sum of their independent risks.

The methodological contribution of this research lies in combining XGBoost with SHAP interpretability. The XGBoost model outperformed traditional ensemble approaches, offering superior predictive accuracy and stability, while SHAP analysis enabled clear attribution of risk contributions and uncovered the mechanisms underlying variable interactions. This dual approach not only enhanced predictive capacity but also provided interpretable evidence for policy and engineering interventions.

There are some limitations to this study. The dataset used in this study lacks quantitative information on driver behavioral states (such as distraction and psychological stress), which may limit the comprehensiveness of interpreting human factors in accident severity. Future research could incorporate real-time environmental data to enhance predictive accuracy and explore the interactions between driver behavioral characteristics and environmental variables. Additionally, although lighting conditions partially capture visibility effects, the dataset does not explicitly include the “time of day” variable. This may limit the differentiation between natural daylight and artificial illumination. Future work could incorporate a temporal visibility index or categorize crashes by time segments (e.g., dawn, day, dusk, night) to better distinguish between natural and artificial lighting effects. Moreover, given the temporal effects of accident data, conducting temporal stability analyses would provide deeper insights and strengthen the robustness of the findings.

Author Contributions: Conceptualization, D.W. and G.L.; methodology, J.P.; software, Q.H.; validation, D.W., J.P. and G.L.; formal analysis, D.W.; investigation, J.P.; resources, G.L.; data curation, D.W.; writing—original draft preparation, Q.H.; writing—review and editing, G.L.; visualization, Q.H.; supervision, J.P.; project administration, D.W.; funding acquisition, D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Jiangsu Province Industry-Academia-Research Cooperation Project, grant number BY20240965.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on demand from the Corresponding author at M405123301@sues.edu.cn; ligen@njfu.edu.cn or 023102@jsei.edu.cn.

Acknowledgments: The authors thank the Jiangsu Province Industry-Academia-Research Cooperation Project (grant number BY20240965). We also thank the anonymous reviewers for providing critical comments and suggestions that improved the manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

SHAP	Shapley Additive Explanations
XGBoost	Extreme Gradient Boosting

Appendix A

Table A1. Variable description.

Variable	Classification	Classification Description
Driver	0	PDO
	1	Injured
	2	Killed
TRNSPD	0	Less than 50
	1	More than 50
AGE	0	Age 65 and over
	1	Otherwise
VEHSPD	0	Less than 50
	1	More than 50
AADT	0	Less than 10,000
	1	More than 10,000
TYPACC	0	Vehicle collision with train
	1	Train collision with vehicle
VISIBLTY	0	Day
	1	Night
INVEH	0	Driver not in vehicle
	1	Driver in vehicle
POSITION	0	Stalled or stuck on crossing
	1	Stopped on crossing
	2	Moving over crossing
	3	Trapped on crossing by traffic
	4	Blocked on crossing by gates
	0	Went around the gates
	1	Stopped and then proceeded
	2	Did not stop
MOTORIST	3	Stopped on crossing
	4	Other
	5	Went around/through temporary barricade
	6	Went through the gate
	7	Suicide/attempted suicide
DRIVGEN	0	Female
	1	Male
DayThru	0	Daytime train traffic volume less than 50
	1	Daytime train traffic volume more than 50
TOTOCC	0	Vehicle occupancy less than or equal to 2
	1	Vehicle occupancy greater than or equal to 3

Table A2. Description of highway–rail collision characteristics.

		Condition 1 (Total Sample Size = 3552)	Condition 2 (Total Sample Size = 8065)	Condition 3 (Total Sample Size = 846)	Condition 4 (Total Sample Size = 1734)
Description Dependent Variable		Percentage	Percentage	Percentage	Percentage
Driver	0	73.70%	68.23%	79.55%	74.34%
	1	21.88%	25.27%	16.55%	21.11%
	2	4.42%	6.50%	3.90%	4.55%
Independent variables					
TRNSPD	0	88.23%	86.21%	90.07%	88.24%
	1	11.77%	13.79%	9.93%	11.76%
AGE	0	10.78%	10.79%	10.17%	10.32%
	1	89.22%	89.21%	89.83%	89.68%

Table A2. Cont.

		Condition 1 (Total Sample Size = 3552)	Condition 2 (Total Sample Size = 8065)	Condition 3 (Total Sample Size = 846)	Condition 4 (Total Sample Size = 1734)
VEHSPD	0	98.82%	98.54%	98.94%	98.50%
	1	1.18%	1.46%	1.06%	1.50%
AADTT	0	72.35%	90.82%	78.72%	92.04%
	1	27.65%	9.18%	21.28%	7.96%
TYPACC	0	20.75%	17.78%	21.28%	19.61%
	1	79.25%	82.22%	78.72%	80.39%
VISIBLTY	0	43.19%	100%	30.38%	100%
	1	56.81%	0	69.62%	0
INVEH	0	23.34%	19.54%	34.87%	26.41%
	1	76.66%	80.46%	65.13%	73.59%
POSITION	0	14.84%	13.11%	26.48%	19.84%
	1	25.14%	22.06%	22.93%	19.15%
	2	56.56%	62.79%	49.29%	60.09%
	3	2.45%	1.29%	0.83%	0.69%
	4	1.01%	0.75%	0.47%	0.23%
	0	18.24%	9.54%	12.07%	6.52%
MOTORIST	1	5.49%	7.00%	4.73%	4.79%
	2	24.01%	41.44%	24.82%	43.89%
	3	27.06%	28.23%	33.92%	31.95%
	4	18.22%	10.02%	18.20%	9.69%
	5	0.22%	0.18%	0.24%	0.11%
	6	6.25%	3.04%	5.20%	2.59%
	7	0.51%	0.55%	0.82%	0.46%
DRIVGEN	0	29.87%	25.08%	30.85%	28.60%
	1	70.13%	74.92%	69.15%	71.40%
DayThru	0	96.23%	99.26%	96.10%	99.25%
	1	3.77%	0.74%	3.90%	0.75%
TOTOCC	0	95.66%	96.26%	96.10%	97.06%
	1	4.34%	3.74%	3.90%	2.94%

References

- Fang, S.; Lu, W.; Ma, J.; Shen, L. Analysis of Moving Work Vehicles on Traffic Flow in City Tunnel. *World Electr. Veh. J.* **2025**, *16*, 491. [\[CrossRef\]](#)
- Zhu, W.; Dai, J.; Zhou, X.; Gao, X.; Cheng, R.; Yang, B.; Li, E.; Lü, Q.; Wang, W.; Tan, Q. A Multi-Fusion Early Warning Method for Vehicle–Pedestrian Collision Risk at Unsignalized Intersections. *World Electr. Veh. J.* **2025**, *16*, 407. [\[CrossRef\]](#)
- Dimitrakopoulos, G.; Demestichas, P. Intelligent Transportation Systems. *IEEE Veh. Technol. Mag.* **2010**, *5*, 77–84. [\[CrossRef\]](#)
- FRA. All Highway-Rail Incidents at Public and Private Crossings, 1981–2025. 2025. Available online: <https://oli.org/about-us/news/collisions-casulties> (accessed on 3 May 2025).
- Haleem, K.; Gan, A. Contributing Factors of Crash Injury Severity at Public Highway-Railroad Grade Crossings in the US. *J. Saf. Res.* **2015**, *53*, 23–29. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ghomi, H.; Bagheri, M.; Fu, L.; Miranda-Moreno, L. Analyzing Injury Severity Factors at Highway Railway Grade Crossing Accidents Involving Vulnerable Road Users: A Comparative Study. *Traffic Inj. Prev.* **2016**, *17*, 833–841. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hao, W.; Kamga, C.; Wan, D. The Effect of Time of Day on Driver’s Injury Severity at Highway-Rail Grade Crossings in the United States. *J. Traffic Transp. Eng.* **2016**, *3*, 37–50. [\[CrossRef\]](#)
- Hao, W.; Kamga, C. The Effects of Lighting on Driver’s Injury Severity at Highway-Rail Grade Crossings. *J. Adv. Transp.* **2016**, *50*, 446–458. [\[CrossRef\]](#)
- Yang, Z.; Zhang, C.; Li, G.; Wang, Y. Analysis of the Impact of Different Road Conditions on Accident Severity at Highway-Rail Grade Crossings Based on Explainable Machine Learning. *Symmetry* **2025**, *17*, 147. [\[CrossRef\]](#)
- Hao, W.; Kamga, C. Difference in Rural and Urban Driver-Injury Severities in Highway–Rail Grade Crossing Accidents. *Int. J. Inj. Control Saf. Promot.* **2017**, *24*, 174–182. [\[CrossRef\]](#)
- Hao, W.; Daniel, J. Driver Injury Severity Related to Inclement Weather at Highway–Rail Grade Crossings in the United States. *Traffic Inj. Prev.* **2016**, *17*, 31–38. [\[CrossRef\]](#)
- Zhao, S.; Khattak, A.J. Injury Severity in Crashes Reported in Proximity of Rail Crossings: The Role of Driver Inattention. *J. Transp. Saf. Secur.* **2018**, *10*, 507–524. [\[CrossRef\]](#)
- Gabree, S.; Chase, S.; da Silva, M. Effect of Dynamic Envelope Pavement Markings on Vehicle Driver Behavior at a Highway-Rail Grade Crossing. In Proceedings of the ASME/IEEE Joint Rail Conference, Colorado Springs, CO, USA, 15–17 April 2014; American Society of Mechanical Engineers: New York, NY, USA, 2014; p. V001T06A005.

14. 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. [\[CrossRef\]](#)
15. Wu, L.; Zhou, J.; Min, H.; Zhang, Y. Injury-Severity Analysis of Highway-Railroad Grade Crossing Crashes Based on a Random Parameters Logit Model with Heterogeneity in Means and Variances. *Adv. Civ. Eng.* **2025**, *2025*, 6682353. [\[CrossRef\]](#)
16. Li, G.; Ji, S.; Cheng, W.; Wang, W.; Yan, X.; Yang, Z.; Yang, Y.; Zhang, J. A New Framework for Modelling Traffic Conflict Interval Time Based on Correlated Random Parameter Duration Model with Heterogeneity in Means and Variances. *Accid. Anal. Prev.* **2026**, *224*, 108282. [\[CrossRef\]](#)
17. Morgan, A.; Mannering, F.L. The Effects of Road-Surface Conditions, Age, and Gender on Driver-Injury Severities. *Accid. Anal. Prev.* **2011**, *43*, 1852–1863. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Khales, S.D.; Chien, S.; Lee, J.; Han, L.D. Analysis of the Effects of Visibility and Warning Devices on Driver Injury Severity at Highway-Rail Grade Crossings Considering Temporal Transferability of Data. *Int. J. Inj. Control Saf. Promot.* **2020**, *27*, 243–252. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Hu, S.R.; Lin, J.P. Effects of Three Advanced Devices on Preventing Crashes and Gate-Breaking Incidents at Highway–Rail Grade Crossings. *Transp. Res. Rec.* **2013**, *2384*, 109–117. [\[CrossRef\]](#)
20. Zhao, S.; Khattak, A.J. Factors Associated with Self-Reported Inattentive Driving at Highway-Rail Grade Crossings. *Accid. Anal. Prev.* **2017**, *109*, 113–122. [\[CrossRef\]](#)
21. Zheng, Z.; Lu, P.; Pan, D. Predicting Highway–Rail Grade Crossing Collision Risk by Neural Network Systems. *J. Transp. Eng. Part A Syst.* **2019**, *145*, 04019033. [\[CrossRef\]](#)
22. Keramati, A.; Lu, P.; Iranitalab, A.; Khattak, A.J. A Crash Severity Analysis at Highway-Rail Grade Crossings: The Random Survival Forest Method. *Accid. Anal. Prev.* **2020**, *144*, 105683. [\[CrossRef\]](#)
23. Fan, W.; Kane, M.R.; Haile, E. Analyzing Severity of Vehicle Crashes at Highway-Rail Grade Crossings: Multinomial Logit Modeling. *J. Transp. Res. Forum* **2015**, *54*, 39–56.
24. Kang, Y.; Khattak, A. Cluster-Based Approach to Analyzing Crash Injury Severity at Highway–Rail Grade Crossings. *Transp. Res. Rec.* **2017**, *2608*, 58–69. [\[CrossRef\]](#)
25. Kutela, B.; Kidando, E.; Kitali, A.E.; Mwende, S.; Langa, N.; Novat, N. Exploring Pre-Crash Gate Violations Behaviors of Drivers at Highway-Rail Grade Crossings Using a Mixed Multinomial Logit Model. *Int. J. Inj. Control Saf. Promot.* **2022**, *29*, 226–238. [\[CrossRef\]](#)
26. 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. [\[CrossRef\]](#)
27. Keramati, A.; Lu, P.; Moatari-Kazerouni, A. Evaluating Crash Severity at Highway-Rail Grade Crossings Using an Analytic Hierarchy Process-Based Hazard Index Model. *Accid. Anal. Prev.* **2025**, *211*, 107918. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Rana, P.; Sattari, F.; Lefsrud, L.; Zhang, L. Machine Learning Approach to Enhance Highway Railroad Grade Crossing Safety by Analyzing Crash Data and Identifying Hotspot Crash Locations. *Transp. Res. Rec.* **2024**, *2678*, 1055–1071. [\[CrossRef\]](#)
29. Lu, P.; Zheng, Z.; Ren, Y.; Pan, D. A Gradient Boosting Crash Prediction Approach for Highway-Rail Grade Crossing Crash Analysis. *J. Adv. Transp.* **2020**, *2020*, 6751728. [\[CrossRef\]](#)
30. Liao, C.; Chen, X.M. A Meta-Learning Approach to Improving Transferability for Freeway Traffic Crash Risk Prediction. *Digit. Transp. Saf.* **2024**, *4*, 21–30. [\[CrossRef\]](#)
31. 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.
32. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process Syst.* **2017**, *30*, 4765–4774.
33. Gazdar, U.; Ahmed, A.; Abdulhusain, B.H.; Mohamed, A.H.; Ratrou, N. Estimating the Severity Levels of Road Traffic Crashes in Bahrain with Crash Costs Estimated with Different Approaches. *Digit. Transp. Saf.* **2023**, *2*, 278–283. [\[CrossRef\]](#)

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.