

Article

Behavioral and Traffic-Related Determinants of Road Crash Severity: Supporting the Safety Dimension of Sustainable Transportation Systems

Paraskevas Nikolaou *  and George Markou 

Department of Civil Engineering and Geomatics, Cyprus University of Technology, Limassol 3036, Cyprus

* Correspondence: paraskevas.nikolaou@cut.ac.cy

Abstract

Road crashes undermine the sustainability of transportation systems by generating substantial human, social, and economic costs. Identifying the factors associated with severe crash outcomes is therefore essential for supporting safer and more sustainable mobility, in accordance with Vision Zero and Sustainable Development Goals 3.6 and 11.2. This study investigates the influence of behavioral, environmental, and infrastructural factors on road crash severity through a comparative evaluation of three modeling approaches: Ordinal Logistic Regression (OLR), eXtreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost). The analysis was based on 5194 police-reported crash records from Cyprus covering the period 2015–2024. The dataset included crash severity outcomes together with traffic violations and roadway and environmental conditions. The results identified speeding, substance use, and improper turning maneuvers as influential predictors of crash severity. Among the evaluated models, CatBoost achieved the strongest overall predictive performance, obtaining the highest accuracy (0.52), weighted area under the curve (AUC) (0.72), weighted precision (0.50), and weighted recall (0.52), while CatBoost and XGBoost tied for the highest weighted F1-score (0.48). CatBoost's overall performance reflects its ability to efficiently process categorical variables and capture nonlinear relationships without extensive preprocessing. The findings highlight the dominant role of driver behavior in determining crash severity and demonstrate how the comparative modeling framework can serve as an analytical decision-support tool for identifying and monitoring crash-severity risk factors, prioritizing targeted interventions, and supporting progress towards safer and more sustainable transportation systems.

Keywords: road crash severity; CatBoost; XGBoost; ordinal logistic regression; sustainable development goals



Academic Editors: Athanasios (Akis)
Theofilatos, Apostolos Anagnostopoulos
and Athanasios Galanis

Received: 28 July 2026

Revised: 22 August 2026

Accepted: 29 August 2026

Published: 2 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

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

1. Introduction

Road traffic accidents remain one of the most significant public health and transportation challenges worldwide, resulting in substantial human, social, and economic consequences. According to the World Health Organization, road crashes are among the leading causes of death globally, particularly among younger age groups, while severe injuries continue to place considerable pressure on healthcare systems and national economies [1]. Consequently, improving road safety and reducing traffic-related fatalities are central objectives of international initiatives, including the European Union's Vision Zero strategy and the United Nations Sustainable Development Goals (UN-SDG).

Road safety constitutes an essential social and economic dimension of transport sustainability. Fatal and serious crashes not only result in loss of life and long-term health consequences but also impose substantial costs on healthcare systems, households, public authorities, and national economies. Consequently, identifying, quantifying, and monitoring the factors associated with crash severity can support progress towards SDG 3.6, which concerns the reduction of road traffic deaths and injuries, as well as SDG 11.2, which promotes access to safe, affordable, accessible, and sustainable transport systems.

Road crash severity is influenced by behavioral, roadway, environmental, and traffic-related factors. Understanding their interactions is essential for developing effective road safety policies. Ordinal Logistic Regression (OLR) has been widely applied in crash severity analysis because it provides interpretable estimates for ordered injury outcomes. However, crash severity data often exhibit nonlinear relationships, complex interactions, and imbalanced class distributions, limiting the ability of traditional statistical models to fully capture crash mechanisms.

Machine learning (ML) techniques have increasingly been adopted to address these limitations by modeling complex nonlinear relationships and improving predictive performance. Among them, ensemble tree-based methods have demonstrated strong performance. eXtreme Gradient Boosting (XGBoost) offers high predictive accuracy and computational efficiency, whereas CatBoost is particularly suitable for crash severity analysis because it directly handles categorical variables, such as traffic violations, roadway characteristics, weather, lighting, and driver behavior, without extensive preprocessing.

Despite these advances, opportunities remain to compare interpretable statistical models and modern machine-learning approaches within a consistent analytical framework. The performance and interpretation of these methods may vary according to the geographical context, class distribution, and characteristics of the available variables. Accordingly, this study compares OLR, XGBoost, and CatBoost using the same police-reported crash dataset, with particular attention paid to behavioral violations and crash-severity patterns in Cyprus.

To address these gaps, this study develops and compares three modeling approaches, OLR, XGBoost, and CatBoost, using police-reported crash data from Cyprus for the period 2015–2024. The objectives are to identify the factors associated with crash severity, evaluate whether ML models outperform a conventional statistical approach, and assess the suitability of CatBoost for crash severity prediction. By applying the three methods to the same dataset and evaluation framework, the study provides comparative evidence on their explanatory and predictive performance, including their ability to identify fatal crashes under class imbalance and the complementary information provided by OLR coefficients and machine-learning feature importance.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents the dataset, methodology, modeling framework, and evaluation metrics. Section 4 reports the modeling results and compares the three approaches. Section 5 discusses the findings and their policy implications, while Section 6 summarizes the main conclusions, limitations, and directions for future research.

2. Literature Review

To be able to achieve the road safety goals set by the European Union such as the elimination of road fatalities by 2050 (“Vision Zero”) [2] and by the United Nations, such as Goal 3 (“Good Health and Well-being”) and Goal 11 (“Sustainable Cities and Communities”) [3], it is important to identify realistic factors that affect road crashes and increase crash severity.

Although contemporary research generally recognizes that road accidents cannot be attributed to a single dominant cause, earlier studies indicated that a substantial proportion of road crashes, reaching up to 90%, were associated with factors related to user behavior [4]. Therefore, this study incorporates traffic violations of commuters, along with environmental and infrastructural factors for analyzing how these factors affect crash severity.

A classic method for analyzing crash severity is Ordinal Logistic Regression, which can be used for measuring the effect of different factors on the different levels of crash severity. For example [5], it identified darkness as one of the most significant factors associated with crash severity level in the city of Addis Ababa in Ethiopia.

Notwithstanding the statistical significance of logistic regression models when analyzing different data characteristics, decision tree models offer superior accuracy for complex datasets [6]. Since ML methodologies operate based on different underlying mechanisms, road safety analysis can benefit from the application of multiple modeling approaches. In particular, decision tTree (DT) methods are considered to be especially useful for crash severity classification problems, mainly when the objective is to examine and interpret the interactions among contributing factors [7].

eXtreme Gradient Boosting (XGBoost) is a relatively new method that was initially proposed by Chen and Guestrin in 2016 [8]; this method typically achieves high predictive accuracy and fast execution while remaining computationally efficient and relatively simple in structure. Several studies have developed the XGBoost method for predicting delays at signalized intersections (e.g., Ref. [9]), structural mechanics (e.g., Ref. [10]), or network traffic anomaly detection (e.g., Ref. [11]), but some have also adopted it in road safety (e.g., Ref. [12]).

To reinforce confidence in models' ability to analyze crash severity, it is important to develop an evaluation strategy and estimate some critical metrics such as accuracy, precision, recall, area under the curve, and F1-score [13].

CatBoost (Categorical Boosting) is a gradient boosting algorithm developed by Yandex that is specifically designed for the efficient analysis of datasets containing categorical variables [14]. Unlike many conventional ML approaches that require extensive preprocessing or encoding of categorical features prior to model development, CatBoost can directly process categorical data through internally implemented encoding procedures during training, making it particularly suitable for datasets with mixed variable types and complex relationships among explanatory factors. Similar to other gradient-boosting techniques, CatBoost constructs an ensemble of decision trees sequentially, where each newly generated tree attempts to reduce the prediction errors produced by the previously developed trees. Through this iterative process, the model progressively improves its predictive capability by minimizing the selected loss function and refining the estimation of residual errors during training.

The CatBoost method has been applied in several fields, such as finance [15] the environment [16] road safety [17], and others. The CatBoost method has been used for analyzing, for example, the severity of cyclists' crashes using vehicle characteristics, crash-related factors, infrastructural and environmental factors [18], and the analysis of motorcycle crash severity using environmental, geometric, infrastructural, and demographic factors [19]. Therefore, several studies have applied CatBoost to different road-safety contexts, including cyclist- and motorcycle-crash severity analysis.

Alongside statistical and machine-learning approaches, traffic simulation provides a complementary framework for examining road-safety conditions under controlled scenarios. Akopov and Beklaryan [20] developed a simulation approach to examine accident rates involving unmanned ground vehicles within a smart-city transportation system, whereas Ref. [21] used SUMO to evaluate the effects of road incidents on traffic congestion.

More broadly, Ref. [22] reviewed developments in road-safety simulation modeling, while Ref. [23] demonstrated how microscopic simulation outputs can be used to derive surrogate safety measures, including time-to-collision, post-encroachment time, and deceleration rate. Such approaches facilitate the evaluation of alternative driver-behavior, traffic, and network scenarios, and therefore complement analyses based on historical crash records. However, simulation studies require detailed and calibrated information on traffic demand, road geometry, signal control, vehicle trajectories, and driver behavior, while their outputs commonly represent traffic conflicts or other surrogate safety indicators rather than the severity of observed police-reported crashes. Accordingly, the simulation represents a valuable complementary direction for future research rather than a direct substitute for the OLR, XGBoost, and CatBoost models applied in the present study.

Crash-severity research has progressively expanded from traditional statistical models to more flexible machine-learning approaches. Models such as OLR, Multinomial Logit, and Mixed Logit remain valuable because they provide interpretable coefficients, confidence intervals, and formal hypothesis tests. OLR accounts for the ordered nature of severity outcomes, while Mixed Logit can represent unobserved heterogeneity. Nevertheless, these models depend on assumptions concerning functional form, proportional odds, independence, or parameter distributions, and their estimated associations should not automatically be interpreted as causal effects [24]. The increasing availability of large and heterogeneous crash datasets has motivated the application of Random Forest, XGBoost, LightGBM, and CatBoost, which can capture nonlinear relationships and complex interactions and may improve predictive performance [25]. However, machine-learning models generally provide less direct interpretation, may be affected by overfitting, and can remain biased toward frequent severity categories when class imbalance is not adequately considered. Although resampling, class weighting, and cost-sensitive learning can improve minority-class identification, feature-importance measures indicate predictive contribution rather than effect direction or causality. Consequently, the two paradigms are complementary: statistical models support interpretable association analysis and hypothesis testing, whereas machine-learning models emphasize flexible prediction and pattern recognition.

Building on these studies, this present research compares CatBoost with XGBoost and OLR and focuses on police-recorded behavioral violations, environmental conditions, and roadway characteristics in Cyprus.

3. Methodological Framework

This methodological framework presents a straightforward approach for the analysis of road crash severity by providing a comparative assessment among a traditional statistical modeling approach, namely Ordinal Logistic Regression (OLR), and two decision-tree-based ML approaches, specifically eXtreme Gradient Boosting (XGBoost) and Categorical Boosting (CatBoost). The analysis was conducted using R free software with the 4.5.2 version.

3.1. Data

For the purposes of this study, a road crash database obtained from the Cyprus national Traffic Police Department (TFD) was used, covering the period from 2015 to 2024. Table 1 presents the variables considered in the analysis, along with their corresponding description, measurement units, and type used throughout the study.

Figure 1 presents the box plots of the numeric variables included in the dataset, categorized by crash severity. Box plots provide a concise visualization of the distribution of the data by illustrating the median value (horizontal line within each box), the first quartile (lower boundary of the box), and the third quartile (upper boundary of the box), which together form the interquartile range (IQR). In addition, the whiskers represent the

spread of the data toward the minimum and maximum non-outlier values. Observations located outside the so-called “inner limits” are considered outliers and are illustrated as individual points in the figure [26,27].

Table 1. Accident data from Cyprus’ traffic police department.

Variable	Description	Unit	Type
Crash severity	Severity level of the crash	Categorical	Property damage only; slight injuries; serious injuries; fatal
No. of vehicles	Number of vehicles involved in the crash	Count	Numerical
Conjunction type	Type of the conjunction at the crash location at the time of the crash	Categorical	“T” Junction; Not at junction; Intersection of ≥ 2 roads; Staggered “T”; Roundabout; “Y” Junction; Motorway accel/decel lane; Other
Lighting	Lighting conditions at the crash location at the time of the crash	Categorical	Night—with road lighting; Night—without road lighting; Day; Dawn; Dusk; Unknown
Pavement	Condition of pavement at the crash location at the time of the crash	Categorical	Dry; Wet; Gravel/sand/dirt; Snow/ice; Muddy; Other
Weather	Weather conditions at the crash location at the time of the crash	Categorical	Clear; Rain/hail; Fog; Snow; Other
Year	Year of the crash	Years	Continuous
Accident month	Month of the crash	Month	Continuous
Accident day	Day of the crash	Day	Continuous
Factor	Primary, secondary, tertiary, and quaternary contributing factors identified by the police as causes of the crash	Categorical	Careless/distracted driving; Right turn without care; Failure to keep safe distance; Failure to give way to vehicles; Traffic sign violation; Failure to keep left; Speed; Inexperienced driver; Traffic lights violation; Illegal overtaking on right; Reversing; Failure to keep proper lane; Run-off road; U-turn; Alcohol; Drugs; Drowsiness; Physical disability; Loss of consciousness; Sudden illness; Medicines (prescribed); Pedestrian used crossing carelessly; Pedestrian crossed carelessly; Other pedestrian error; Careless crossing at uncontrolled junction
Speed limit	Speed limit of the road that the crash occurred	km/h	Continuous
No. Injured	Number of injured people involved in the crash	Count	Numerical

In detail, the distributions indicate that the majority of crashes, regardless of severity level, occurred on roads with a posted speed limit of 50 km/h. The initial dataset comprised 6229 road crashes, of which 5321 (85.42%) occurred on roads with a posted speed limit of 50 km/h. This observation was considered in the present study, and therefore the analysis focused specifically on roads with a speed limit of 50 km/h, as they represent the dominant roadway environment within the dataset. Because of their low frequencies, records classified as Dawn, Dusk, or Unknown for lighting; Other, Fog, or Snow for weather; and Other, Gravel/Sand/Dirt, Snow/Ice, or Muddy for pavement condition were excluded. This procedure removed 127 records, resulting in a final analytical dataset of 5194 crashes. The final sample comprised 1647 property-damage-only crashes (31.71%), 1241 slight-injury

crashes (23.89%), 2014 serious-injury crashes (38.78%), and 292 fatal crashes (5.62%). This distribution demonstrates the class imbalance of the outcome, particularly the relatively small proportion of fatal crashes. Additionally, the variable referring to the number of injuries was excluded from the modeling procedures since it is inherently and directly associated with the dependent variable, namely crash severity. Including this variable could introduce redundancy and potential bias into the analysis. Regarding temporal characteristics, the distributions of accident day and month appear relatively consistent across the different crash severity categories. Crashes were observed throughout all days of the week and across all months of the year, suggesting the absence of strong temporal concentration patterns associated with a specific day or season. Nevertheless, slight variations in dispersion indicate that certain periods may experience marginally higher crash frequencies, although no clear differentiation among severity levels is evident from the figure.

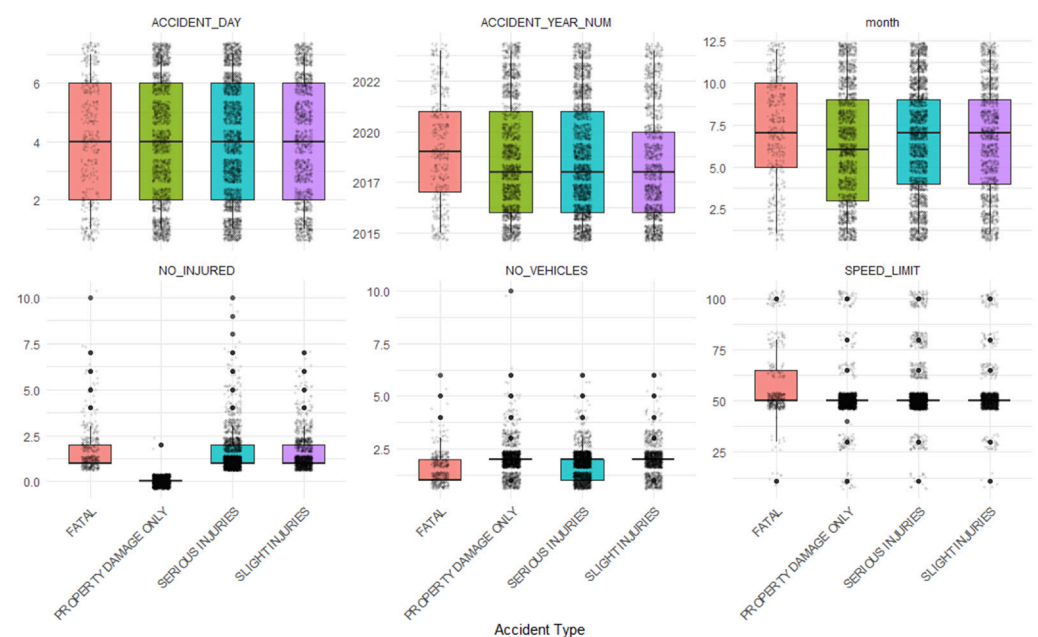


Figure 1. Box plots of the numeric variables by crash severity. The y-axis units are variable-specific: accident day (days 1–7), accident year (calendar year 2015–2024), month (months 1–12), number of injured persons (count), number of vehicles (count), and speed limit (km/h).

Figure 2 presents the environmental and infrastructural conditions under which crashes occurred, including junction type, lighting, pavement status, and weather conditions. Most crashes were recorded under favorable conditions, namely on dry pavements, during clear weather, and under daylight or roads with lighting infrastructure. In contrast, relatively few crashes occurred under adverse conditions such as wet pavements, rain/hail, or roads without lighting. These findings suggest that crash occurrence is more strongly associated with human-related factors, such as risky driving behavior and traffic violations, than with unfavorable environmental or infrastructural conditions. Police records also identified up to four contributing factors for each crash.

Prior to the analysis, a data preprocessing procedure was conducted to improve model robustness by addressing sparse categories. For the Lighting variable, the categories Dawn, Dusk, and Unknown were omitted. For Weather, the categories Other, Fog, and Snow were excluded, while for Pavement Status, the categories Other, Gravel/Sand/Dirt, Snow/Ice, and Muddy were removed because of their low frequencies. For Junction Type, the categories Staggered “T”, Roundabout, “Y” Junction, Motorway Acceleration/Deceleration Lane, and Other were merged into a single Other category to improve model stability.

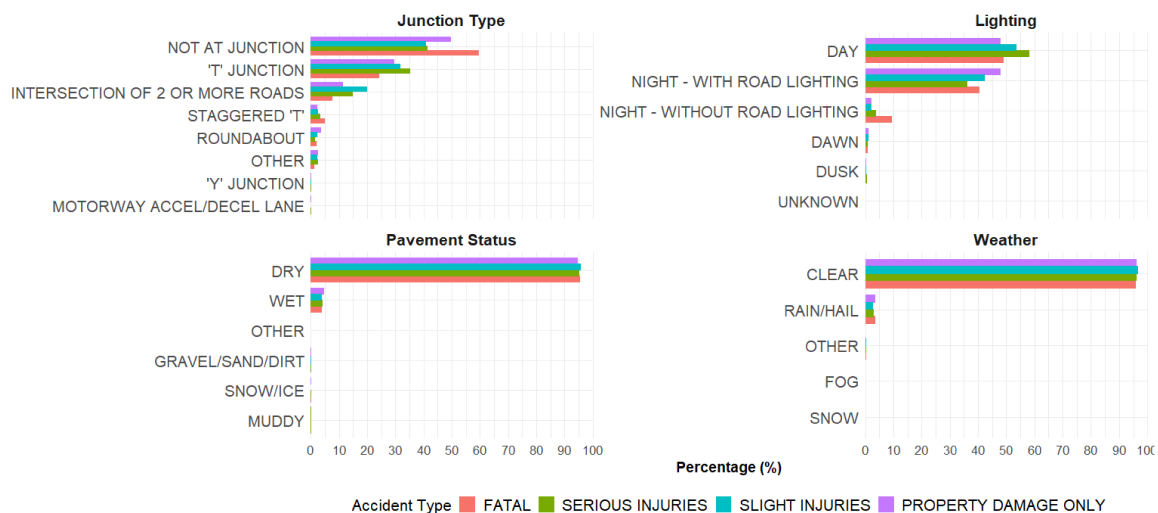


Figure 2. Natural and infrastructural conditions per accident type.

The same procedure was applied to the Factor variable. The categories Drowsiness, Physical Disability, Loss of Consciousness, Sudden Illness, and Medicines (Prescribed) were combined into a single Driver Impairment category, while Pedestrian Used Crossing Carelessly, Pedestrian Crossed Carelessly, Other Pedestrian Error, and Careless Crossing at Uncontrolled Junction were grouped as Pedestrian Error. In addition, because multiple contributing factors could be recorded for each crash, each violation type was transformed into a separate binary variable to estimate its independent effect on crash severity. The final processed dataset comprised 5194 crashes across the Cyprus road network.

3.2. Ordinal Logistic Regression Model

For modeling the outcomes of a categorical dependent variable, such as overall liking scores measured on a hedonic scale (e.g., crash severity), logistic regression is considered a better alternative to classical linear models [28]. The Ordinal Logistic Regression (OLR) model for an outcome Y with levels $l = 1, 2, \dots, L$ is:

$$\ln\left(\frac{P(Y \leq l)}{P(Y > l)}\right) = \zeta_l - \eta_1 X_1 - \eta_2 X_2 - \dots - \eta_K X_K \text{ For each level, } l = 1, 2, \dots, L - 1 \quad (1)$$

where

ζ_l is an intercept and represents the log-odds of $Y \leq l$ when all the predictors are at 0 or their reference level. Thus, $P(Y \leq l)$ is the inverse-logit of ζ_l . Unlike other forms of regression, an Ordinal Logistic Regression model has multiple intercepts, one for each level of Y from 1 to $L - 1$.

$-\eta_k$ for $k = 1, 2, \dots, K$ is the log of the odds ratio comparing the odds of $Y \leq l$ between individuals who differ by one unit in X_k (or comparing individuals at a level of X_k and the reference level).

$e^{-\eta_k}$ is the odds ratio comparing the odds of $Y \leq l$ between those differing by one unit in X_k .

e^{η_k} is the odds ratio comparing the odds of $Y > l$ between those differing by one unit in X_k .

Furthermore, the left-hand side of Equation (1) is the cumulative probability of an outcome being at any level up to and including a specified level [29]. The proportional-odds assumption, which requires each explanatory variable to have a constant coefficient across the cumulative crash-severity thresholds, was evaluated using the Brant test. The null hypothesis of the test is that the parallel-regression, or proportional-odds, assumption

holds. A significance level of 0.05 was used. Multicollinearity among the explanatory variables was assessed using variance inflation factors (VIFs) and adjusted generalized variance inflation factors (GVIFs) for predictors with multiple degrees of freedom.

3.3. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) combines numerous small decision tree (DT) models to make predictions, which differentiate from each other. Boosting is an iterative process where, in each step, the individual data points are weighted. The iterative process is called gradient boosting, where a gradient is a mathematical term that refers to the vector field of partial derivatives that point in the direction of the steepest slope. Gradient boosting algorithms are among the best-performing ML tools on the market. XGBoost is one of the most used ML algorithms and is considered a quick way to achieve good performance [30]. For a dataset with n examples and m features $D = \{(x_i, y_i)\}$ ($|D| = n, x_i \in R$), a tree ensemble model uses K additive functions to predict the output [8].

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (2)$$

where

$F = \{f(x) = w_{q(x)}\}$ ($q : R^m \rightarrow T, w \in \mathcal{R}^T$) is the space of regression trees.

q represents the structure of each tree that maps an example to the corresponding leaf index.

T is the number of leaves in the tree.

f_k corresponds to an independent tree structure q and leaf weight w .

To learn the set of functions used in the model, we minimize the following regularized objective [8].

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (3)$$

where

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

l is differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i .

The second term Ω penalizes the complexity of the model. The additional regularization term helps to smooth the final learned weight to avoid over-fitting. Intuitively, the regularized objective will tend to select a model employing simple and predictive functions [8].

The tree ensemble in Equation (3) includes functions as parameters and cannot be optimized using traditional optimization methods in a Euclidean space. Let $\hat{y}_i^{(t)}$ be the prediction of the i -th instance at the n -th iteration, then f_t is added for improving the model, according to Equation (3).

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (4)$$

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (5)$$

Equation (5) may serve as a scoring metric for evaluating the quality of a given tree structure q . In practice, exhaustively exploring all possible tree configurations is computationally infeasible. Therefore, a greedy approach is typically adopted, where the process begins with a single leaf node and progressively expands the tree by introducing splits.

Let I_L and I_R denote the sets of instances assigned to the left and right child nodes after a split, respectively. If we define the parent node's instance set as $I = I_L \cup I_R$, then the improvement in the objective (i.e., the reduction in loss) resulting from this split can be expressed as [8]:

$$\mathcal{L}_{split} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] \quad (6)$$

3.4. Categorical Boosting (CatBoost)

Categorical Boosting (CatBoost) is a gradient boosting framework developed by Yandex [14], which addresses several limitations of traditional boosting methods, such as XGBoost and LightGBM. Although the core principles of boosted decision trees have already been discussed, CatBoost incorporates additional mechanisms that enhance its capability in handling categorical variables and producing stable predictive performance with relatively limited hyperparameter tuning requirements.

The name of the framework itself originates from its strong emphasis on categorical feature processing. In contrast to many other gradient boosting approaches, which typically require preprocessing techniques such as one-hot encoding or label encoding, CatBoost can directly utilize categorical variables and internally determine suitable encoding representations during the training process. This functionality, together with its regularization mechanisms and symmetric tree architecture, makes the method particularly appropriate for datasets containing mixed variable types and for applications where robust predictive performance is desired without extensive manual preprocessing. Furthermore, CatBoost mitigates overfitting through symmetric ("oblivious") decision trees and additional regularization strategies designed to improve model stability and generalization capability. Overall, the framework is designed to provide reliable and competitive predictive performance while reducing the need for complex hyperparameter optimization procedures [31].

CatBoost builds upon the fundamental gradient boosting framework, which is presented in Equation (7):

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (7)$$

where

$F_m(x)$ is the ensemble model at iteration m (cumulative prediction after adding m trees).

$F_{m-1}(x)$ is the ensemble model from the previous iteration (cumulative prediction after $m - 1$ trees).

γ_m is the learning rate at iteration m (controls the contribution of each new tree).

$h_m(x)$ the new tree added at iteration m (trained on residuals or gradients).

x is the input feature vector for a given observation.

One of the most distinctive characteristics of CatBoost is its ability to effectively process categorical variables. Rather than requiring prior preprocessing or manual encoding of categorical features, the framework internally estimates target-based statistical representations during the model training procedure. Therefore, for the categorical feature x^k with categories $c_1, c_2, \dots, c_k$, CatBoost computes an encoded value for each observation as follows:

$$\hat{x}_i^k = \frac{\sum_{j=1}^n \left[x_j^k = x_i^k \right] \cdot y_j + \alpha \cdot p}{\sum_{j=1}^n \left[x_j^k = x_i^k \right] \cdot y_j + \alpha} \quad (8)$$

where

$\hat{x}_i^k$ is the encoded numerical value for the i -th observation's category in feature x^k .

x_i^k the categorical value of feature k for observation i .
 $\mathbb{I}[x_j^k = x_i^k]$ is the indicator function that equals 1 if observation j has the same category as observation i , and 0 otherwise.
 y_j is the target value for observation j .
 n is the total number of observations in the dataset.
 α is the smoothing parameter that controls regularization (typically set to 1).
 p is the prior probability of the target (usually the mean of the target variable across the entire dataset).

CatBoost uses symmetric trees, also known as oblivious trees, where all nodes at the same depth (level) split on the same feature and threshold. This is in contrast to traditional decision trees, where each node can split on any feature independently.

3.5. Model Training and Evaluation

To ensure a consistent comparison, the processed dataset was divided into training (70%) and testing (30%) sets using stratified random sampling based on crash-severity class. The same partition was used for all three models. For XGBoost and CatBoost, the training set was further divided into 80% for model training and 20% for validation and hyperparameter selection. K-fold cross-validation was not conducted; instead, a holdout validation approach was applied. Random search evaluated 60 hyperparameter combinations for XGBoost and 50 for CatBoost. Early stopping was applied during tuning, with patience values of 50 iterations for XGBoost and 120 iterations for CatBoost. After hyperparameter selection, the final models were trained using the complete training set and evaluated on the independent testing set.

For assessing the performance of the three models (OLR, XGBoost, and CatBoost), metrics such as accuracy, area under curve, precision, recall, and F1-score were used. In road safety studies, crash severity datasets often exhibit substantial class imbalance, as severe and fatal crashes generally represent a limited share of the total observations. In these cases, relying solely on overall accuracy can lead to an inaccurate interpretation of a model's predictive capability [32].

Let C be the confusion matrix, where C_{jk} denotes instances of true class j predicted as class k , and $N_k = \sum_{j=1}^K C_{jk}$ is the support of class k ; therefore, accuracy is defined as the overall correctness across all classes as presented in Equation (9):

$$Accuracy = \frac{\sum_{k=1}^K C_{kk}}{N} \quad (9)$$

The area under the curve (AUC) is computed for each class k against all others, then averaged by class support as expressed in Equation (10).

$$AUC_{weighted} = \sum_{k=1}^K AUC_k \frac{N_k}{N} \quad (10)$$

where

AUC_k is the binary AUC for class k versus the remaining classes.

N_k is the support (true instances) of class k .

The weighted precision is given in Equation (11):

$$Precision_{weighted} = \sum_{k=1}^K \left(\frac{TP_k}{TP_k + FP_k} \right) \frac{N_k}{N} \quad (11)$$

where

$TP_k = C_{kk}$ (true positives for class k).

$FP_k = \sum_{j \neq k} C_{jk}$ (false positives for class k).

The recall metric is given in Equation (12):

$$Recall_{weighted} = \sum_{k=1}^K \left(\frac{TP_k}{TP_k + FN_k} \right) \frac{N_k}{N} \quad (12)$$

where

$FN_k = \sum_{j \neq k} C_{jk}$ (false negatives for class k).

Finally, the F1-score (weighted) metric is expressed in Equation (13):

$$F1_{weighted} = \sum_{k=1}^K \left(\frac{2 \cdot Precision_k \cdot Recall_k}{Precision_k + Recall_k} \right) \frac{N_k}{N} \quad (13)$$

The evaluation metrics presented above provide a comprehensive framework for assessing the predictive capability and reliability of crash severity classification models. In general, higher values of accuracy, weighted AUC, weighted precision, weighted recall, and weighted F1-score indicate improved model performance, as they reflect stronger classification capability, improved discrimination between severity categories, and a more balanced prediction of both majority and minority classes. In particular, the weighted AUC metric is especially important in imbalanced crash severity datasets because it evaluates the ability of the model to correctly distinguish between the different crash severity classes independently of class frequency. Similarly, the weighted F1-score provides a balanced assessment between precision and recall and is therefore considered particularly important when severe crash categories, such as fatal crashes, constitute only a limited proportion of the observations.

4. Application and Results

4.1. Ordinal Logistic Regression Model

The results of the Ordinal Logistic Regression model are presented in Table 2. The model achieved a McFadden's pseudo- R^2 of 0.09, indicating limited overall model fit and only a modest improvement over the intercept-only model [33]. The adjusted GVIF values ranged from 1.011 to 1.091, with the highest value observed for failure to maintain a safe distance. Because all values were close to 1, there was no evidence of problematic multicollinearity among the explanatory variables. The Brant test rejected the proportional-odds assumption ($\chi^2 = 463.78$, $p < 0.001$), demonstrating that some explanatory variables did not have constant effects across the crash-severity thresholds. Variable-specific tests identified significant violations for the number of vehicles, substance use, inexperienced driving, traffic-light violations, failure to maintain a safe distance, reversing, and speeding ($p < 0.05$). Therefore, the single odds ratios reported for these variables represent coefficients constrained to be equal across the severity thresholds and should be interpreted with caution.

Within this constrained OLR specification, speeding had the largest positive coefficient (OR = 4.18), indicating an overall positive association with higher crash severity [34]. Pedestrian error (OR = 3.40), right turn without care (OR = 3.27), and nighttime conditions without road lighting (OR = 2.05) were also positively associated with more severe outcomes. Positive associations were additionally found for failure to keep left, illegal overtaking on the right, failure to give way, and traffic-light violations. However, because the proportional-odds assumption was violated for speeding and traffic-light violations, their effects should not be assumed to have the same magnitude across all severity thresholds.

Table 2. Ordinal Logistic Regression estimates and odds ratios for crash severity.

Variable	Estimate	Odds Ratio	Confidence Interval	
			Lower	Upper
No_vehicles	−0.54 *** (0.05)	0.59	0.53	0.64
Lighting: Night—With Road Lighting	−0.24 *** (0.06)	0.79	0.71	0.88
Lighting: Night—Without Road Lighting	0.72 *** (0.16)	2.05	1.51	2.78
Substance_use	−0.85 *** (0.08)	0.43	0.37	0.50
Inexperienced_driver	−0.61 *** (0.12)	0.54	0.43	0.69
Failure_to_give_way_to_vehicles	0.32 *** (0.09)	1.38	1.15	1.66
Traffic_lights_violation	0.31 ** (0.12)	1.37	1.08	1.73
Failure_to_keep_safe_distance	−0.91 *** (0.10)	0.40	0.33	0.49
Reversing	−0.96 *** (0.17)	0.38	0.28	0.54
Failure_to_keep_left	0.55 *** (0.11)	1.74	1.40	2.16
Speed	1.43 *** (0.14)	4.18	3.20	5.46
Other	0.16 * (0.07)	1.17	1.02	1.35
Right_turn_without_care	1.19 *** (0.09)	3.27	2.76	3.88
Pedestrian_error	1.22 *** (0.13)	3.40	2.66	4.35
Illegal_overtaking_on_right	0.53 *** (0.14)	1.70	1.29	2.25
Intercept				
Property Damage Only Slight Injuries	−1.88 *** (0.10)	0.15	0.12	0.19
Slight Injuries Serious Injuries	−0.69 *** (0.10)	0.50	0.41	0.61
Serious Injuries Fatal	2.24 *** (0.11)	9.37	7.54	11.65
Log-Likelihood: −5836.655				
AIC: 11,709.31				
McFadden's R ² : 0.09				

Note: Parentheses denote standard errors. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Failure to maintain a safe distance, reversing, substance use, number of involved vehicles, and inexperienced driving produced negative coefficients in the constrained model. Because all these variables exhibited significant violations of the proportional-odds assumption, their single odds ratios should not be interpreted as consistent protective effects across the severity categories. The direction and magnitude of their associations may differ depending on the severity threshold being examined. Contextual factors such as lower operating speeds during reversing or congestion-related crashes and limitations in police reporting may also contribute to these apparently counterintuitive results.

4.2. Extreme Gradient Boosting (XGBoost)

In addition to the Ordinal Logistic Regression model, an XGBoost model was developed to evaluate the predictive capability of machine learning for crash severity classification. Figure 3 presents the observed versus predicted crash severity for the training and testing datasets. The close concentration of observations around the ideal correlation line indicates good predictive performance, although some misclassification occurs between adjacent severity levels, particularly for higher-severity crashes. Similar prediction patterns in the training and testing datasets suggest satisfactory generalization without evidence of substantial overfitting. The model performed better for the more frequent intermediate severity categories than for fatal crashes, which is expected given the class imbalance of the dataset.



Figure 3. XGBoost observed vs. predicted crash severity—test and train sets.

Figure 4 presents the feature importance rankings obtained from the XGBoost model. The number of involved vehicles (*no_vehicles*) was identified as the most influential predictor of crash severity, followed by *right_turn_without_care*, *substance_use*, and *speed*, highlighting the dominant role of driver behavior and traffic exposure in crash severity prediction. Lighting conditions and variables associated with unsafe driving behavior, including *failure_to_keep_safe_distance*, *failure_to_keep_left*, *pedestrian_error*, and *reversing*, also contributed to the model's predictive performance. These findings indicate that XGBoost primarily relied on behavioral factors, while roadway and environmental characteristics provided a secondary contribution to crash severity classification.

4.3. Categorical Boosting (CatBoost)

Figure 5 presents the observed versus predicted crash severity for the CatBoost model using the training and testing datasets. The concentration of observations around the ideal correlation line indicates good predictive performance, particularly for the PDO and serious injury categories. Similar prediction patterns across the training and testing datasets suggest satisfactory generalization without evidence of substantial overfitting. Some misclassification is observed for the slight injury and fatal categories, primarily into adjacent severity levels, reflecting the ordinal nature of crash severity and the class imbalance associated with the relatively small number of fatal crashes. Nevertheless, the absence of substantial misclassification between extreme categories (e.g., PDO and fatal) indicates that the model effectively preserves the ordinal structure of the severity outcomes.

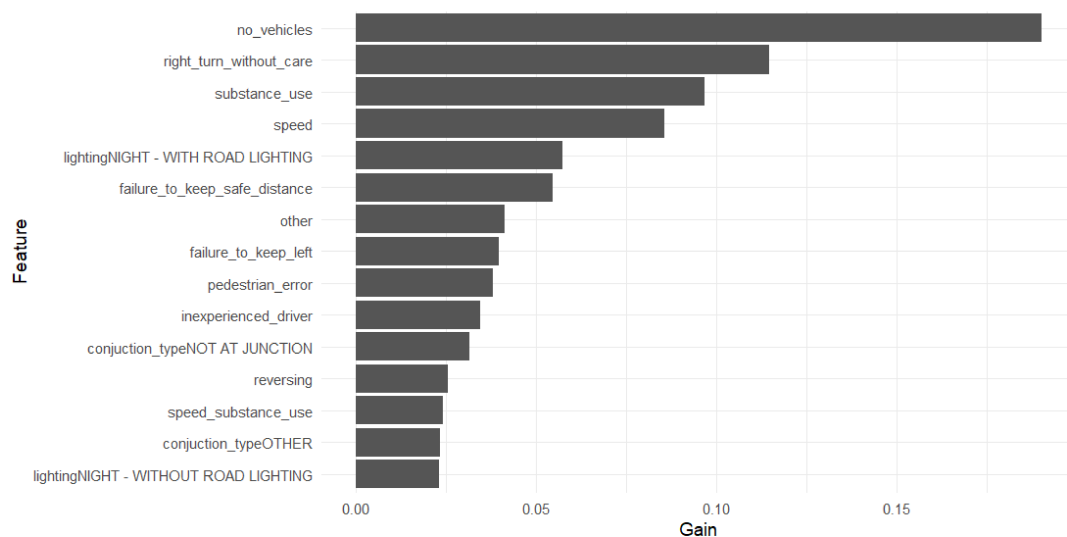


Figure 4. Relative importance of explanatory variables in XGBoost model.

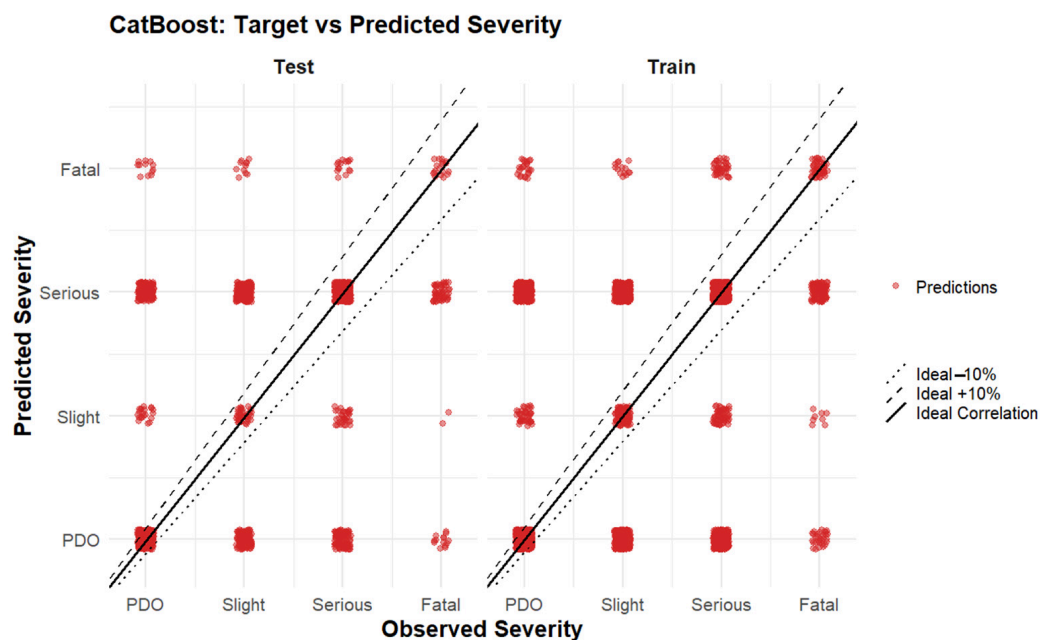


Figure 5. CatBoost observed vs. predicted crash severity—test and train sets.

Figure 6 presents the feature importance rankings generated by the CatBoost model. The variable `no_vehicles` was identified as the most influential predictor of crash severity, followed by `substance_use` and `speed`, highlighting the importance of traffic exposure, impaired driving, and excessive speed in determining crash outcomes. Behavioral variables, including `right_turn_without_care`, `failure_to_keep_safe_distance`, `failure_to_keep_left`, and `pedestrian_error`, also demonstrated high importance, emphasizing the dominant role of unsafe driving behavior and human error. Roadway and environmental variables, such as `lighting` and `junction_type`, showed moderate contributions, whereas `illegal_overtaking_on_right` and `failure_to_give_way_to_vehicles` were among the least influential predictors. Overall, the results indicate that the CatBoost model primarily relied on behavioral and traffic-related factors to predict crash severity, consistent with previous transportation safety research.

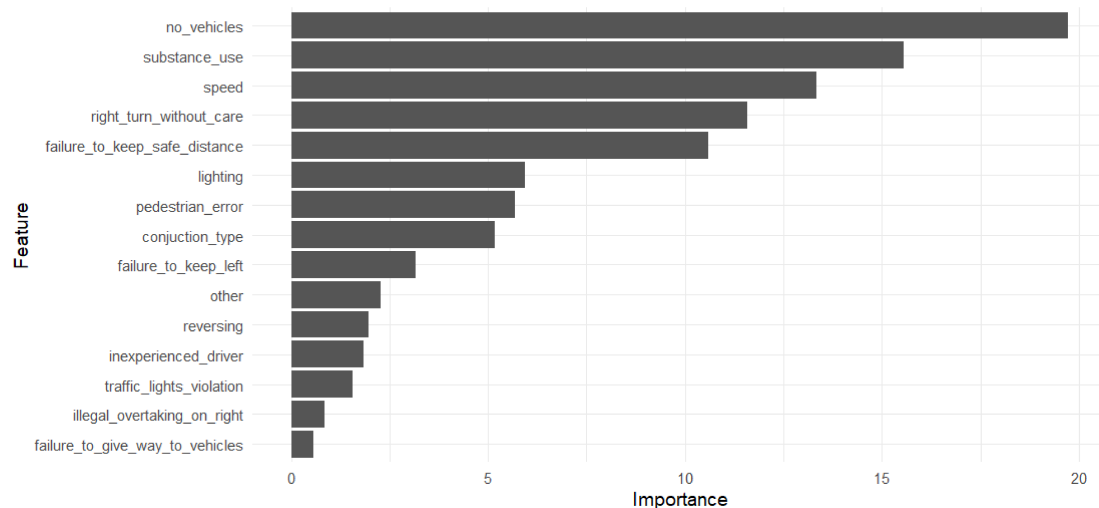


Figure 6. Relative importance of explanatory variables in CatBoost model.

4.4. Model Selection

The comparative performance of the three models is presented in Figure 7, using accuracy, weighted AUC, weighted F1-score, weighted precision, and weighted recall. Overall, the results indicate that CatBoost achieved the best predictive performance among the examined approaches. Specifically, it obtained the highest accuracy (0.52), weighted AUC (0.72), weighted precision (0.50), and weighted recall (0.52), while CatBoost and XGBoost achieved identical weighted F1-scores (0.48), both outperforming the OLR model.

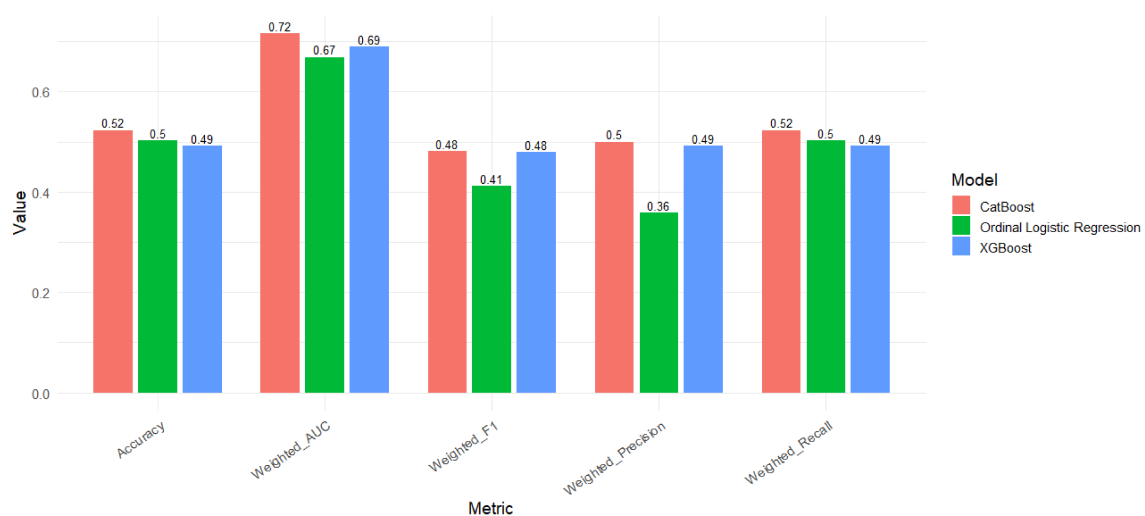


Figure 7. Comparative analysis of weighted classification metrics of CatBoost, OLR, and XGBoost models.

The higher weighted AUC achieved by CatBoost demonstrates its stronger ability to distinguish between the different crash severity categories, whereas its higher precision and recall indicate a better balance between minimizing false classifications and correctly identifying the true severity classes. The comparable F1-scores of CatBoost and XGBoost further suggest that both ML approaches achieved a more balanced predictive performance than the traditional OLR model, particularly in the presence of class imbalance.

Because weighted evaluation metrics may be influenced by the more frequent severity categories, class-specific performance was also examined, with particular attention given to fatal crashes. Among the 87 fatal crashes in the testing set, OLR did not correctly identify any fatal cases and did not produce any fatal-crash predictions; consequently, its recall

was 0, while precision and F1-score were undefined. XGBoost correctly identified six fatal crashes, achieving a precision of 0.50, recall of 0.07, and F1-score of 0.12. CatBoost identified the largest number of fatal crashes, correctly classifying 18 cases, with a precision of 0.32, recall of 0.21, and F1-score of 0.25. Nevertheless, the relatively low recall values indicate that fatal-crash identification remains challenging because of the limited number of fatal observations and the substantial class imbalance.

The stronger overall predictive performance achieved by CatBoost is likely attributable to its ability to capture nonlinear relationships and interaction effects while directly processing categorical variables without extensive preprocessing. Similar findings have also been reported by Ref. [35], who demonstrated the competitive predictive performance of machine-learning approaches relative to traditional statistical models for crash severity classification. Furthermore, the similarity between the training and testing results indicates satisfactory generalization capability, suggesting that CatBoost maintained stable predictive performance without evidence of substantial overfitting.

Although CatBoost generally achieved higher values for most evaluation metrics, it tied with XGBoost for the weighted F1-score (0.48). The remaining differences were relatively modest, indicating that the OLR model remains competitive despite its simpler statistical formulation. Importantly, OLR provides interpretable parameter estimates and odds ratios, allowing for the direct quantification of associations between the explanatory variables and crash severity. However, because the proportional-odds assumption was rejected, the coefficients of variables exhibiting assumption violations represent constrained common-slope estimates and should be interpreted cautiously rather than as constant effects across all severity thresholds. Consequently, while CatBoost was selected as the preferred model based on its overall predictive performance, OLR remains a valuable interpretable benchmark for examining associations in transportation safety analyses.

5. Discussion

The present study compared three methodological approaches for analyzing and predicting road crash severity: Ordinal Logistic Regression (OLR), XGBoost, and CatBoost. Across the three approaches, driver-behavior and traffic-violation variables repeatedly emerged as important explanatory or predictive factors. Nevertheless, consistency across the models should be understood in terms of variable relevance rather than identical effect directions or magnitudes. This distinction is important because machine-learning feature importance does not indicate the direction of an association, while the Brant test demonstrated that several OLR coefficients varied across the crash-severity thresholds. Speeding, substance use, pedestrian-related errors, improper turning maneuvers, unsafe driving behavior, and lighting conditions were among the principal variables identified by the models. Similar findings have been reported using logistic regression [36,37] and machine-learning approaches [38], highlighting the recurring importance of behavioral factors in crash-severity analysis.

To spatially contextualize these findings, Figure 8 presents the distribution of crashes associated with three influential driver-behavior violations in Limassol (the second-largest city in Cyprus): speeding, right turns without care, and substance use. The mapped crashes were distributed across the urban road network, including major urban and coastal corridors. These patterns provide a descriptive spatial illustration of the recorded violations and should not be interpreted as statistically significant hotspots or location-specific risk estimates, as they were not adjusted for traffic exposure. Moreover, because the database contains only crash observations, the maps cannot provide a comparison with careful or fully compliant driving.



Figure 8. Spatial distribution of police-reported crashes associated with selected driver-behavior violations in Limassol during 2015–2024: (a) speeding; (b) right turn without care; and (c) substance use. Each point represents a recorded crash location.

Importantly, the findings should be interpreted as conditional on the occurrence of a police-reported crash. The examined models estimate associations with and predictions of crash severity; they do not estimate whether a factor increases the overall number or probability of crashes. In the OLR model, odds ratios above one indicate an association with higher crash-severity levels, whereas odds ratios below one indicate an association with lower severity. Conversely, the feature-importance measures produced by XGBoost and CatBoost indicate the relative predictive contribution of each variable but do not, by themselves, establish the direction of its effect or a causal relationship. Consequently, the present results cannot quantify the protective effects of careful driving, as the database does not include non-crash trips or driving-exposure observations. Examining crash occurrence, traffic conflicts, and counterfactual changes in driver behavior would require additional exposure data or a calibrated microscopic simulation supported by detailed traffic-demand, signal-control, trajectory, and conflict data. Additionally, because the analysis was restricted to crashes occurring on roads with a posted speed limit of 50 km/h, the findings primarily characterize this roadway environment and should not be generalized directly to roads with different speed limits.

The OLR model provided a transparent statistical framework through its parameter estimates and odds ratios. However, the Brant test rejected the proportional-odds assumption ($\chi^2 = 463.78, p < 0.001$), with significant violations identified for the number of vehicles, substance use, inexperienced driving, traffic-light violations, failure to maintain a safe distance, reversing, and speeding. Consequently, the single coefficients estimated for these variables should be viewed as constrained common-slope estimates rather than effects that remain constant across all crash-severity thresholds. This limitation may partly explain some of the counterintuitive OLR findings and its comparatively lower predictive performance. The relatively low McFadden's pseudo- R^2 of 0.09 further indicates limited overall model fit. However, the adjusted GVIF values, which were close to 1, demonstrate that the

limited fit was not attributable to multicollinearity among the explanatory variables. The common-slope and linear specifications of OLR also limit its ability to represent threshold-specific effects, nonlinear relationships, and complex interactions. Nevertheless, OLR remains useful as an interpretable traditional benchmark, while partial proportional-odds or generalized ordered-logit models could provide a valuable extension in future research.

Both XGBoost and CatBoost demonstrated improved predictive performance over OLR in terms of weighted AUC, weighted precision, and weighted F1-score. CatBoost also achieved higher accuracy and weighted recall, whereas XGBoost recorded slightly lower values for these two metrics than OLR. The similarity of the training and testing prediction patterns did not indicate clear evidence of substantial overfitting. Furthermore, both ML approaches identified an overlapping set of influential variables, including the number of involved vehicles, speeding, substance use, right-turn violations, failure to maintain a safe distance, and pedestrian errors. Nevertheless, fatal-crash identification remained challenging. OLR did not identify any of the 87 fatal crashes in the testing set, while XGBoost and CatBoost achieved fatal-crash recall values of 0.069 and 0.207, respectively. These results demonstrate that improvements in overall weighted performance do not necessarily translate into strong identification of the rarest severity category.

Among the ML approaches, CatBoost achieved the strongest overall predictive performance, obtaining the highest accuracy, weighted AUC, weighted precision, and weighted recall. However, CatBoost and XGBoost achieved the same weighted F1-score of 0.48; therefore, CatBoost did not outperform XGBoost across every evaluation metric. Previous research has similarly demonstrated the competitive performance of CatBoost relative to other ML approaches, including XGBoost [39]. CatBoost's overall performance may be related to its ability to process categorical variables directly while capturing nonlinear relationships and interaction effects without extensive preprocessing. Given that the dataset consisted predominantly of categorical behavioral, environmental, and infrastructural variables, CatBoost appears to be particularly suitable for predictive crash-severity modeling. The stable performance observed between the training and testing datasets further supports its predictive applicability, although its relatively low fatal-crash recall should be considered.

An important finding of this study is that CatBoost assigned substantial predictive importance primarily to behavioral variables. Substance use, speeding, right-turn violations, pedestrian errors, failure to keep left, and failure to maintain a safe distance ranked among the model's most influential predictors. These findings indicate that behavioral variables contributed substantially to crash-severity prediction within the examined dataset. However, feature importance does not indicate the direction of an association or establish that these variables causally increase crash severity. Therefore, the results support prioritizing these factors for further investigation but do not demonstrate that behavioral interventions would necessarily produce greater safety benefits than infrastructural or environmental measures.

From a policy perspective, the findings indicate that road safety improvements should primarily target behavioral risk factors. Since speeding emerged as one of the strongest predictors across all three modeling approaches, strengthening speed management should constitute a priority, particularly in urban areas with high pedestrian activity. Measures such as automated speed enforcement and the implementation of 30 km/h speed limits have been shown to reduce fatalities by more than 40% while simultaneously providing environmental and public health benefits through lower fuel consumption and increased walking and cycling [40]. Likewise, the predictive importance assigned to substance use by the machine-learning models indicates that this variable warrants continued monitoring and further threshold-specific investigation. However, this finding does not establish that

substance use had a positive effect on crash severity, and the negative OLR coefficient should not be interpreted as a consistent protective effect because the proportional-odds assumption was violated for this variable. Alcohol- and drug-enforcement measures remain relevant to broader road-safety policy, but their recommendation should not be interpreted as resulting from the direction-neutral feature-importance values obtained in this study.

The findings also emphasize the importance of improving pedestrian safety. Given the prominence of pedestrian-related errors and improper driver maneuvers across the models, transportation authorities should prioritize safer pedestrian infrastructure through improved crossing facilities, traffic calming measures, protected intersections, enhanced signage, and improved nighttime visibility. Furthermore, the positive association between inadequate roadway lighting and severe crash outcomes highlights the need to improve roadway illumination, particularly in urban areas and at intersections with elevated pedestrian activity. The findings demonstrate that combining interpretable statistical models with advanced machine learning approaches provides a comprehensive framework for road crash severity analysis. While ML models improve predictive performance, statistical models remain valuable for estimating interpretable associations between individual risk factors and crash severity. Together, these approaches can support transportation planners and policymakers in developing more targeted and evidence-based road safety interventions.

6. Conclusions

This study comparatively evaluated the performance of Ordinal Logistic Regression (OLR), XGBoost, and CatBoost for analyzing and predicting road crash severity using police-reported crash data from Cyprus covering the period 2015–2024. Behavioral, environmental, infrastructural, and traffic-related factors were examined to identify the principal determinants of severe crash outcomes. The results demonstrated that driver behavior and traffic violations are the dominant contributors to crash severity. Speeding, substance use, unsafe turning maneuvers, and failure to maintain safe driving behavior consistently emerged as the most influential factors across all models, highlighting the importance of behavioral interventions and effective traffic law enforcement.

The comparative analysis showed that the ML approaches outperformed the traditional OLR model, with CatBoost achieving the best overall predictive performance. Its ability to directly process categorical variables and capture nonlinear relationships makes it particularly suitable for crash severity modeling using transportation safety data. From a policy perspective, the findings support strengthening speed management, alcohol and drug enforcement, pedestrian protection measures, safer intersection design, and improved roadway lighting. Measures such as 30 km/h urban speed limits and traffic calming interventions could contribute substantially to reducing severe crashes.

This study is subject to several limitations, including its reliance on police-reported crash data and the absence of traffic volume, driver demographic, and spatial exposure information. Moreover, the Brant test indicated that the proportional-odds assumption of the OLR model was not satisfied. Therefore, the reported OLR coefficients, particularly those for the variables exhibiting significant assumption violations, should not be interpreted as effects that remain constant across all severity thresholds. Future research should examine partial proportional-odds or generalized ordered-logit models to estimate threshold-specific relationships. Spatio-temporal modeling approaches, real-time traffic data, and explainable artificial intelligence techniques should also be considered to improve predictive performance and model interpretability. Subject to the availability of detailed traffic-demand, signal-control, vehicle-trajectory, and traffic-conflict data, calibrated microscopic simulation platforms such as SUMO or AnyLogic could be applied to selected

high-risk road segments to examine how driver behavior and traffic conditions influence surrogate safety indicators and their relationship with observed crash severity.

Author Contributions: Conceptualization, P.N. and G.M.; methodology, P.N. and G.M.; software, P.N. and G.M.; validation, P.N. and G.M.; formal analysis, P.N. and G.M.; investigation, P.N.; resources, P.N.; data curation, P.N.; writing—original draft preparation, P.N.; writing—review and editing, G.M.; visualization, P.N.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from Cyprus' Police Traffic Department and are not available.

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

References

1. World Health Organization. Road Traffic Injuries. 2024. Available online: https://www.who.int/health-topics/road-safety#tab=tab_1 (accessed on 5 May 2026).
2. European Commission. EU Road Safety: Towards “Vision Zero”. Contributions of Horizon 2020 Projects Managed by CINEA. 2022. Available online: <https://cinea.ec.europa.eu/system/files/2023-02/H2020%20Transport-Road%20Safety%202022-web.pdf> (accessed on 5 May 2026).
3. United Nations. The 17 Sustainable Development Goals. 2015. Available online: <https://sdgs.un.org/goals> (accessed on 5 May 2026).
4. Mokarami, H.; Alizadeh, S.S.; Rahimi Pordanjani, T.; Varmazyar, S. The relationship between organizational safety culture and unsafe behaviors, and accidents among public transport bus drivers using structural equation modeling. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *65*, 46–55. [\[CrossRef\]](#)
5. Bekelcho, T.; Baru, A.; Woldemeskel, A.; Alemayehu, M.; Guta, G. Identification of determinant factors for crash severity levels occurred in Addis Ababa City, Ethiopia, from 2017 to 2020: Using ordinal logistic regression model approach. *BMC Public Health* **2023**, *23*, 1884. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Shaik, M.E.; Hassan, A.M.H.A.; Rahman, S.M. Predicting road traffic crash severity using machine learning models: A comparative study. *Afr. Transp. Stud.* **2026**, *4*, 100100. [\[CrossRef\]](#)
7. Silva, P.B.; Andrade, M.; Ferreira, S. Machine learning applied to road safety modeling: A systematic literature review. *J. Traffic Transp. Eng. (Engl. Ed.)* **2020**, *7*, 775–790. [\[CrossRef\]](#)
8. 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 (KDD '16)*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [\[CrossRef\]](#)
9. Politi, R. Prediction of signalized intersection delays using genetic programming, XGBoost, and clustering approach. *Transp. Lett.* **2026**, *18*, 1173–1193. [\[CrossRef\]](#)
10. Markou, G.; Bakas, N.P.; Chatzichristofis, S.A.; Papadarakakis, M. A general framework of high-performance machine learning algorithms: Application in structural mechanics. *Comput. Mech.* **2024**, *73*, 705–729. [\[CrossRef\]](#)
11. Liu, T.; Chong, T.; Xu, G.; Wang, X.; Peng, P.; Ma, J. XGBoost-LR: A method for network traffic anomaly detection. *Int. J. Cogn. Comput. Eng.* **2025**, *7*, 325–333. [\[CrossRef\]](#)
12. Parsa, A.B.; Movahedi, A.; Taghipour, H.; Derrible, S.; Mohammadian, A. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accid. Anal. Prev.* **2020**, *136*, 105405. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Al-Sheikh, O.; Alfari, R.; Jalayer, M. ‘How we roll’: The effects of human, vehicle, and environmental contributors to the occurrence and severity of rollover crashes. *J. Saf. Res.* **2026**, *97*, 183–196. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18)*; Curran Associates Inc.: Red Hook, NY, USA, 2018; pp. 6639–6649.
15. Jabeur, S.B.; Gharib, C.; Mefteh-Wali, S.; Arfi, W.B. CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technol. Forecast. Soc. Change* **2021**, *166*, 120658. [\[CrossRef\]](#)
16. Zhou, G.; Mao, L.; Bao, T.; Zhuang, F. Machine learning-driven CO2 emission forecasting for light-duty vehicles in China. *Transp. Res. Part D Transp. Environ.* **2024**, *137*, 104502. [\[CrossRef\]](#)

17. Zahid, M.; Habib, M.F.; Ijaz, M.; Ameer, I.; Ullah, I.; Ahmed, T.; He, Z. Factors affecting injury severity in motorcycle crashes: Different age groups analysis using Catboost and SHAP techniques. *Traffic Inj. Prev.* **2024**, *25*, 472–481. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Agheli, A.; Aghabayk, K. How does distraction affect cyclists' severe crashes? A hybrid CatBoost-SHAP and random parameters binary logit approach. *Accid. Anal. Prev.* **2025**, *211*, 107896. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Junaid, M.; Bendak, S.; Luo, S.; Jiang, C.; Pervez, A. Understanding Motorcycle Crash Severity in Pakistan: Insights from Machine Learning and Interaction Effects. *KSCE J. Civ. Eng.* **2026**, *30*, 100569. [\[CrossRef\]](#)
20. Akopov, A.S.; Beklaryan, L.A. Simulation of rates of traffic accidents involving unmanned ground vehicles within a transportation system for the 'smart city'. *Bus. Inform.* **2022**, *16*, 19–35. [\[CrossRef\]](#)
21. Smith, D.; Djahel, S.; Murphy, J. A SUMO based evaluation of road incidents' impact on traffic congestion level in smart cities. In Proceedings of the 39th Annual IEEE Conference on Local Computer Networks Workshops, Edmonton, AB, Canada, 8–11 September 2014; pp. 702–710. [\[CrossRef\]](#)
22. Young, W.; Sobhani, A.; Lenné, M.G.; Sarvi, M. Simulation of safety: A review of the state of the art in road safety simulation modelling. *Accid. Anal. Prev.* **2014**, *66*, 89–103. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Gettman, D.; Head, L. Surrogate Safety Measures from Traffic Simulation Models. *Transp. Res. Rec. J. Transp. Res. Board* **2003**, *1840*, 104–115. [\[CrossRef\]](#)
24. Savolainen, P.T.; Mannering, F.L.; Lord, D.; Quddus, M.A. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accid. Anal. Prev.* **2011**, *43*, 1666–1676. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Iranitalab, A.; Khattak, A. Comparison of four statistical and machine learning methods for crash severity prediction. *Accid. Anal. Prev.* **2017**, *108*, 27–36. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Tukey, J.W. *Exploratory Data Analysis*; Pearson: Hoboken, NJ, USA, 1977.
27. Dodge, Y. *The Concise Encyclopedia of Statistics*; Springer: New York, NY, USA, 2008.
28. Meulien, J.-F.; Xiong, R.; Findlay, C.J. *Multivariate and Probabilistic Analyses of Sensory Science Problems*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
29. Bender, R.; Grouven, U. Ordinal Logistic Regression in Medical Research. *J. R. Coll. Physicians Lond.* **1997**, *31*, 546–551. [\[CrossRef\]](#)
30. Korstanje, J. *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR*; Apress: Berkeley, CA, USA, 2021.
31. Brenndorfer, M. CatBoost: Complete Guide to Categorical Boosting with Target Encoding, Symmetric Trees & Python Implementation. Mbrenndorfer.com. 2025. Available online: <https://mbrenndorfer.com/writing/catboost-categorical-boosting-complete-guide-target-encoding-symmetric-trees-python-implementation> (accessed on 5 May 2026).
32. Shen, R.; Ma, L.; Yan, X.; Li, X. The nonlinear impact of road safety policy implementation on the severity of road traffic crashes: A fusion of deep learning and Bayesian random parameter methods. *Accid. Anal. Prev.* **2026**, *232*, 108503. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Theofilatos, A.; Yannis, G.; Kopelias, P.; Papadimitriou, F. Predicting Road Accidents: A Rare-events Modeling Approach. *Transp. Res. Procedia* **2016**, *14*, 3399–3405. [\[CrossRef\]](#)
34. European Commission. Mobility & Transport—Road Safety. Speed and Injury Severity. 2023. Available online: https://road-safety.transport.ec.europa.eu/eu-road-safety-policy/priorities/safe-road-use/safe-speed/archive/speeding/speed-central-issue-road-safety/speed-and-injury-severity_en (accessed on 5 May 2026).
35. Gbey, E.K.; Atombo, C.; Adanu, E.K.; Agyemang, W. Predicting higher-risk crash factors and patterns using machine learning-association rule mining. *IATSS Res.* **2025**, *49*, 565–579. [\[CrossRef\]](#)
36. Rezapour, M.; Moomen, M.; Ksaibati, K. Ordered logistic models of influencing factors on crash injury severity of single and multiple-vehicle downgrade crashes: A case study in Wyoming. *J. Saf. Res.* **2019**, *68*, 107–118. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Alrejail, A.; Moomen, M.; Ksaibati, K. Evaluating the impact of traffic violations on crash injury severity on Wyoming interstates: An investigation with a random parameters model with heterogeneity in means approach. *J. Traffic Transp. Eng. (Engl. Ed.)* **2022**, *9*, 654–665. [\[CrossRef\]](#)
38. Assi, K.; Rahman, S.M.; Mansoor, U.; Ratrou, N. Predicting Crash Injury Severity with Machine Learning Algorithm Synergized with Clustering Technique: A Promising Protocol. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5497. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Wisutwattanasak, P.; Se, C.; Sum, S.; Champahom, T.; Ratanavaraha, V.; Jomnonkwao, S. Temporal instability of highway pedestrian crash severity: Comparative analysis of machine learning models. *Transp. Res. Interdiscip. Perspect.* **2026**, *37*, 101935. [\[CrossRef\]](#)
40. Yannis, G.; Michalaraki, E. Effectiveness of 30 km/h speed limit—A literature review. *J. Saf. Res.* **2025**, *92*, 490–503. [\[CrossRef\]](#) [\[PubMed\]](#)

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