ARTICLE | Article in Press

Severity prediction of traffic accidents with recent machine learning paradigms and SHAP

Meshal Almoshaogeh · Fawaz Alharbi

Received: 26 December 2025 / Accepted: 21 September 2026
© The Author(s) 2026

Abstract

Traffic-related fatalities are among few leading causes of death worldwide. Crash severity prediction of traffic crashes has become a widespread research subject in recent years. Severity outcomes of crashes are usually influenced by factors such as humans, vehicles, roadway, weather, environment, and their interactions. A better understanding of crash-contributing circumstances and crash severity risk predictors is vital for the timely and effective implementation of appropriate mitigation measures. In the literature, various statistical methods have been widely adopted for crash severity modeling; however, such approaches are based on several underlying unrealistic assumptions, which, if violated, may lead to biased predictions. The current study aims to investigate the applicability of the state-of-the-art gradient boosting algorithm "XGBoost" for crash severity prediction of traffic crashes in the Al-Qassim region, Saudi Arabia. Crash severity classification performance of XGBoost was compared with two traditional machine learning algorithms, including Random Forest (RF) and Logistic Regression (LR), using different statistical measures including accuracy, recall, precision, F-1 scores, receiver operating characteristic (ROC) curves, and area under the curve (AUC). Empirical results revealed that all the models yielded acceptable predictive performance; however, XGBoost achieved the highest accuracy (90.5%), outperforming both the LR and RF models. Models' comparison using performance metrics also demonstrated superior predictive performance of the XGBoost model. To address the infamous "lack of interpretation" issue frequently raised against the application of machine learning, this study also presents XGBoost-based feature sensitivity and ranking analysis. The results showed that predictors including accident type, weather conditions, number of vehicles involved, lighting conditions, accident cause, and time of the day (TOD) have a strong influence on the crash severity outcome of crashes. SHAP analysis was also performed to assess the consistency of severity risk factors with the XGBoost feature importance technique. The findings of this study are expected to provide crucial insights to safety practitioners and policy-makers for improving highway safety.

Keywords Traffic crashes · Risk factors · Injury prevention · Machine learning · Feature sensitivity analysis · SHAP

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



https://doi.org/10.1038/s41598-026-73109-2

Published online: 29 September 2026

scientific reports

Introduction

Traffic injuries are among the leading causes of human casualties worldwide. The recent World Health Organization (WHO) global road safety report suggests that the global annual fatalities due to traffic crashes have exceeded 1.19 million fatalities and 50 million injuries[1,2]. The resulting socio-economic losses are enormous and are estimated to be around USD 520 billion globally[1,3]. The Kingdom of Saudi Arabia (KSA) has witnessed a rapid surge in traffic fatalities and injuries during the past four decades, particularly after the oil boom in the early 1970s. The annual traffic fatality index (deaths/100,000 pop.) is around 27, which is significantly high compared to other countries in the Arabian Gulf region[4]. Traffic injuries in KSA are around 4.7% of total mortalities[5,6]. Similarly, the average crash-to-injury ratio in different regions of KSA varies from 8:4 to 8:6, which is also significantly high compared to a global ratio of 8:1[2,7]. In recent years, several studies have been undertaken to examine the influential factors of traffic crashes and injury severity in different regions of KSA and reported that predominant causes accounting for high crash incidence and increased crash severity include aggressive driving, overspeeding, non-compliance with traffic rules, distracted driving, sudden lane deviation, etc[2,6–10]. Recently, the Ministry of Transport (MOT) is undertaking some serious efforts to negate and mitigate the growing road safety issues in all regions of the country. Notable actions in this regard include the installation of speed cameras (SAHER), the installation of appropriate signage including Variable Message Signs (VMS), lane-marking, periodic maintenance of national highways, and strict enforcement of traffic regulations. Such efforts have resulted in noticeable safety improvements in some regions[11].

An important goal of highway safety studies is to accurately model the crash severity outcomes and to identify the contributing factors. Accurate prediction of crash injury is essential for the adoption of appropriate crash severity remedial measures. However, it is often very challenging to fully understand and precisely predict crash patterns since they are complex events that usually involve many multi-faceted interactions among contributing factors rather than one factor. The crash severity outcomes of traffic crashes are influenced by several factors like driver attributes, vehicle factors, weather and lighting conditions, traffic attributes, roadway geometric characteristics (highway types and alignment), features of the surrounding built environment, and so on[4,12–19]. The main aim of crash severity studies is to analyze the latent correlations between the crash severity outcomes and the accompanying crash risk predictors[20]. In the literature, different statistical analyses and modeling methods have been widely explored for crash severity. While such models have a sound theoretical basis and clear mathematical formulations to examine the individual role of crash predictors on the crash severity outcome, they have a few serious drawbacks, such as unrealistic linear link function assumptions and predefined associations between the variables, and lower prediction accuracies[21–26]. Recently, with rapid advances in data mining and computational methods, machine learning and deep learning-based models have emerged as promising alternatives to address the aforementioned limitations of their counterparts (statistical methods). However, these data mining-based methods are mainly criticized for their infamous back-box nature and lack of interpretation[27–29]. Recent studies have employed various interpretation techniques such as feature importance analysis and SHAP analysis to overcome this issue[30,31]. Recent studies have reported that ensemble learning approaches, particularly gradient-boosting models, can substantially enhance crash severity prediction by efficiently capturing complex non-linear relationships among the crash risk factors[32,33]. However, despite their superior predictive capabilities, such methods are often criticized for their limited explainability, hence restricting adoption by traffic safety practitioners' policymakers. Consequently, explainable machine learning (XML), specifically SHAP and LIME have become increasingly important for model interpretation and identifying underlying factors contributing to crash severity outcomes[34,35].

Despite the growing application of XGBoost and SHAP in recent traffic safety and crash severity studies, several important research gaps remain. First, the majority of existing studies have primarily focused on enhancing the predictive performance of various ML models, while comparatively limited attention has been devoted to exploring the consistency of risk factors between different explainability techniques. Second, there is a dearth of

literature linking the model interpretations into practical safety interventions and policy recommendations. Third, many existing studies have been conducted in the context of specific roadway environments or highway facilities such as rural freeways, highways, or urban intersections, making the transferability of identified crash severity predictors uncertain across heterogeneous traffic environments.

To address the identified gaps, the current study aims to pursue four main research objectives and contributions. First, it proposes a unified modeling framework for comprehensive comparative assessment of three widely used classification algorithms, namely XGBoost, Random Forest RF), and Logistic regression for crash severity classification. Second, the study employed XGBoost feature importance analysis with SHAP-based explainability to investigate the consistency and robustness of influential crash severity predictors rather than relying on a single interpretability tool. Third, the study aims to translate explainability outputs into practical traffic safety and policy recommendations. Finally, the analysis is conducted using a crash dataset representing both rural and urban roadway environments of the Qassim region in KSA, thereby offering valuable insights into crash severity prediction across diverse and heterogeneous traffic conditions.

The remainder of this paper is organized as follows. Section 2 provides a detailed overview of different modeling approaches used for crash severity prediction and a summary of relevant previous studies. Section 3 introduces the study area and describes the data utilized for the current research. Section 4 describes the proposed machine modeling approaches and classifier evaluation metrics. Section 5 presents the experimental results and related discussion. Finally, Sect. 6 concludes the paper and provides an outlook for future studies.

Related works

Modeling crash severity is a promising research target in traffic safety studies, and it has been the subject of widespread attention during the past few decades. The outcome of crash severity is usually represented by discrete categories, like no injury/property damage, possible injury, minor injury, severe injury, and fatal injury. Most studies have considered three crash severity classes[36,37]: crashes involving property damage only, at least one injury, and/or a fatality. Due to the discrete nature of crash severity outcomes, studies have opted for discrete outcome models such as multinomial logit (MNL) and probit models for modeling crash severity[38–41]. For example, Mohamed Abdel-Aty proposed the application of MNL and ordered probit models for studying driver's injury severity[42]. The estimation analysis showed that driver age and gender variables, dark lighting conditions, non-compliance with seat-belt usage, speed ratio, point of impact, presence of curves, and vehicle are likely to intensify the resulting crash severity. A comparative analysis of modeling methods suggested that the ordered probit models were more accurate and robust. In their study, Hu et al.[43] adopted conventional logit models for crash severity modeling of crashes that occurred near/at railroad crossings. Variables identified to have a significant statistical influence on the output crash severity included obstacle detection devices near the crossing zones, daily trips, and the presence of markings at approach roads. In another study, Chen et al.[44] applied probit models to investigate the crash severity risk factors in crashes involving heavy vehicles and found that predictors including drivers' sociodemographic attributes, wet and slippery surface conditions, time of the day (peak/off-peak) and adverse weather conditions will aggravate the crash severity.

To capture the ordinal nature of crash data and to account for heterogeneity and endogeneity within-crash correlation, studies have proposed other advanced statistical models like ordered logit models[38,45–48], nested logit model[49,50], Bayesian hierarchical[51–53], bivariate/multivariate models[54–57], latent class models[58–61], random parameter model[62–65], and their hybrid versions[12,66–68]. In a recent research study, Kim et al.[69] attempted to investigate the significant crash risk factors in hit pedestrian crashes. They reported that the likelihood of fatal crashes is expected to increase under circumstances such as driving under the influence of drugs, collisions involving trucks, no street lighting, overspeeding, and crashes that occurred along freeways. Another study undertaken by Rusli et al. examined the applicability of random parameter logit models to uncover crash severity risk factors for crashes alongside hilly areas in Malaysia[70]. Comparative analysis with traditional Scobit and logit models

revealed that the random parameter model was more robust in capturing unobserved heterogeneity. Wang et al.[63] also employed a random parameter probit model to study the influential factors that could lead to severe crashes. Model estimation results showed that variables such as bus and pedestrian involvement and rainy weather conditions increase the probability of severe injury outcomes. Random parameter-based models have been increasingly used for crash severity analysis in recent years[63,70–74]. Bayesian hierarchical models have also been employed widely for crash severity investigation. For example, a recent study by Meng et al. applied Bayesian hierarchical logistic models to explore the crash severity patterns of taxi occupants[75]. It was observed that factors such as non-compliance with a seatbelt, driving under fatigue, and elderly drivers would amplify the odds of severe injuries among the taxi occupants.

In recent years, various data mining and artificial intelligence-based techniques have gained widespread attention on the horizon of crash severity modeling. Such methods have emerged as promising alternatives to overcome the primary drawbacks of statistical analytic[76]. The main advantage of these artificial intelligence-based methods is that, without any prior assumptions, they can accurately capture the latent and non-linear correlations between crash severity outcomes and contributing factors. Few widely used machine learning techniques for crash severity prediction are: Artificial Neural Networks, Naïve Bayes, K-means Clustering, Random Forest, Support Vector Machines, and Decision Trees algorithms. Though several studies have focused on the application of machine learning for the prediction and modeling of crash severity, only a few have investigated the potential of deep learning models[77–80]. For instance, a prior research study proposed hybrid deep learning architectures integrating temporal and spatial learning mechanisms to enhance road accident prediction performance, thus highlighting the continuing development of data-driven traffic safety modeling[81].

Among the machine learning methods, artificial neural networks (ANNs) are perhaps the most widely used paradigm for crash injury prognosis. For example, Al-Kheder et al.[82] adopted ANN models for predicting the crash severity of crashes using six years of data for the state of Abu Dhabi in the United Arab Emirates. A systematic k-means clustering was first employed to split the data into three clusters to improve the models' predictive performance. Crash severity (the output variable) had four possible outcomes: fatal injury, severe injury, moderate injury, and minor injury. ANN prediction results were also compared with statistically ordered probit models. The results indicated that ANN, with a classification accuracy of 74.6%, outperformed the ordered probit model, having a corresponding value of approximately 59%. Abdelwahab and Abdel-Aty, in their study, also reported the superior crash severity predictive performance of the multilayer perceptron (MLP) framework compared to the standard ordered logit model in classifying vehicular crash record of signalized intersections in Central Florida, US[83]. Jamal and Waleed[63] recently introduced an improved feedforward neural network (FFNN) framework to examine the significant crash severity risk factors associated with fatal crashes along interstate rural highways in KSA. The optimized FFNN structure with back-propagation as a training algorithm, logistic as an activation function, and six hidden-layer neurons resulted in overall prediction accuracy of around 78%. Feature sensitivity of the optimized FFNN structure revealed that the presence of factors that could intensify crash severity include pedestrian involvement, adverse weather conditions, vehicle and highway type, average travel speeds, traffic volume, and on-site damage characteristics.

Support vector machine (SVM) is another commonly utilized machine learning technique for traffic crash severity analysis. For example, Li et al.[84], using data of 326 diverge segments in Florida, developed SVM-based predictive analytics for classifying crash severity records, and the results were compared with the ordered logit model. Modeling results indicated that SVM performed better than the logit model. In another similar study, the authors[85] compared the crash severity classification performance of SVM and random parameter mixed logit models for traffic crashes that occurred near the work zones and noted that SVM results are more accurate, consistent, and robust. The study also performed a feature sensitivity-based analysis, which showed that the predictors, such as activity type in the work zone, morning off-peak periods, highway types, nature, and termination areas on the approach segments, all had a significantly positive effect on the severity of both drivers and occupants. A study undertaken by Wang et al.[86] also showed that the SVM model achieved better injury severity prediction performance compared to a multilayer perceptron (MLP).

In recent years, Decision Trees (DTs) and Random Forest (RF) have also been increasingly investigated for modeling and prediction of crash severity. A recent study demonstrated that the applicability of DT models is preferred for crash severity classification and analysis, as they can be used to extract rules for assigning injury severity into distinct severity groups under specific circumstances[87]. Bernard et al.[88] adopted DT models for evaluating the injury severity prognosis of motor vehicle crashes using ten years of data from Missouri State in the United States. It was found that explanatory variables such as drunk driving, overspeeding, and failing to yield are likely to increase the chances of fatal crash occurrence. In their study, Ijaz et al.[89] compared the injury severity classification performance of three DT-based algorithms (RF, DT, and Decision Jungle) using three years (2017 to 2019) of data for traffic crashes involving three-wheeled motorized rickshaws (3 W-MR). Empirical results revealed that Decision Jungle, having an overall accuracy of 83.7%, outperformed DT and RF models. The study proposed the extraction of relevant decision rules based on RF in order to categorize crash injuries into distinct severity groups according to the prevailing contributing circumstances. Further, the feature sensitivity analysis indicated that predictors such as drivers aged between 20 and 30 years, nighttime driving conditions, facilities with comparatively high-speed limits (over 60 mph), off-peak, weekdays, and crashes that occurred under shiny/clear weather led to a high probability of severe and fatal injuries. Princess and Rajsingh[90] also performed a comparative study analysis of four machine learning algorithms such as RF, K-nearest neighbor, Adaboost, and DT for injury severity classification using imbalanced crash datasets. Experimental results suggested that the RF model achieved an excellent AUC value of 0.97, significantly higher compared to other algorithms.

XGBoost is another newly introduced machine learning technique used for crash severity analysis; however, there are very limited applications of this algorithm in highway safety studies. A study conducted by Malkoatle et al.[91] evaluated the application of XGBoost and multivariate logistic regression for traffic crash severity identification in South Africa. It was concluded that XGBoost produced better injury severity classification results. In recent research, Jamal et al. conducted a comparative study to evaluate the applicability of XGBoost against traditional machine learning techniques (RF, DT, and LR) for modeling crash severity[20]. Proposed models were developed using three years (2017 to 2019) of rural interstate highway crash data in KSA. The output injury severity outcome had three possible outcomes, i.e., property damage, injury crashes, and fatal crashes. Empirical results obtained via k-fold cross-validation for various metrics (precision, recall, F-scores) indicated that XGBoost outperformed other models both in terms of aggregate model performance as well as for class injury severities. Feature importance analysis using XGBoost showed that variables such as weather conditions, collision type, on-scene damage type, road surface conditions, vehicle type, and lighting conditions were all sensitive in predicting crash severity outcomes. Several recent studies have focused on applying different feature engineering techniques and new ensemble methods to improve the predictive performance of existing machine learning algorithms in crash severity analysis[92,93]. It is worth noting that the present work has some key differences from the previous study conducted locally[20]. First, this previous study is focused on the accident dataset that occurred only on multi-lane rural interstate highways in KSA, while the current study utilized a different dataset encompassing accident data in the Al-Qassim region for both rural and urban highways and those that occurred within the city and its outskirts. It is established from the literature that crash characteristics within (urban) and outside (rural) cities are different, which implies that it is likely that injury severity predictors are likely to be different between both jurisdictions and worthy of investigation. Second, model' predictive performance comparisons are also based on additional ROC and AUC metrics which is not the case in the previous study. Third, the data also contain a few new variables. Lastly and more importantly, the specific contributions of the current work lie in the application of advanced interpretation analysis (SHAP summary plots and main effects plots for sensitive variables) instead of simple feature importance analysis as provided in the previously published study.

More recently, researchers have shifted their attention from emphasizing predictive accuracy alone toward integrating XML frameworks that combine high-performance models with transparent model interpretation[34,94]. Several prior studies have employed ensemble learning algorithms together with SHAP-based explanations to identify influential predictors of crash severity, thereby improving transparency and practical applicability of these frameworks[35,95,96]. Increasing application of explainable AI-based models in traffic safety research

highlights the growing importance of considering interpretability and predictive performance for supporting evidence-based road safety decision-making. However, only a few studies have been devoted to evaluating the consistency between different explainability approaches and translating these findings into practical safety recommendations[94,97,98]. The present study aims to address this aspect by proposing a unified modelling and interpretability framework, providing a robust and reliable comparison between different classifiers and explainability techniques.

Study area and data description

Qassim region in KSA was selected as the study area for this research. It is one of the thirteen administrative regions of KSA and is located between N23° 30', N28° 00' latitude and E40° 00', E45° 00' longitude, at the center of the country. Qassim province has an area of approximately 58,000 km² and a population of 1.3 million people, almost half of whom live in the capital city of Buraidah. The entire Qassim region consists of 13 administrative local governorates, including Buraidah, Al-Badaya'a, Al-Asyah, Al Bakiriyah, Uyun Al-Jiwa, Daria, Al Mithnab, Ar Rass, Al Nabhanya, Riyadh Al-Khabra, Oklat AlSkoor, Al Shammasiya, and Unaizah[99]. Figure 1 shows the location of the study area in its entirety, including its roadnetwork and the spatial distribution of major urban districts. The Qassim province is well known for its agricultural resources, such as the production of dates, vegetables, and various other types of fruits. The province capital (Buraidah) is also the hub of various public and commercial businesses, public offices, academic institutions, and health facilities. The region has witnessed a significant increase in motorization rate in recent years. The province has witnessed an unusually high number of traffic accidents compared to other regions in KSA[100,101]. Increasing traffic safety issues have become key

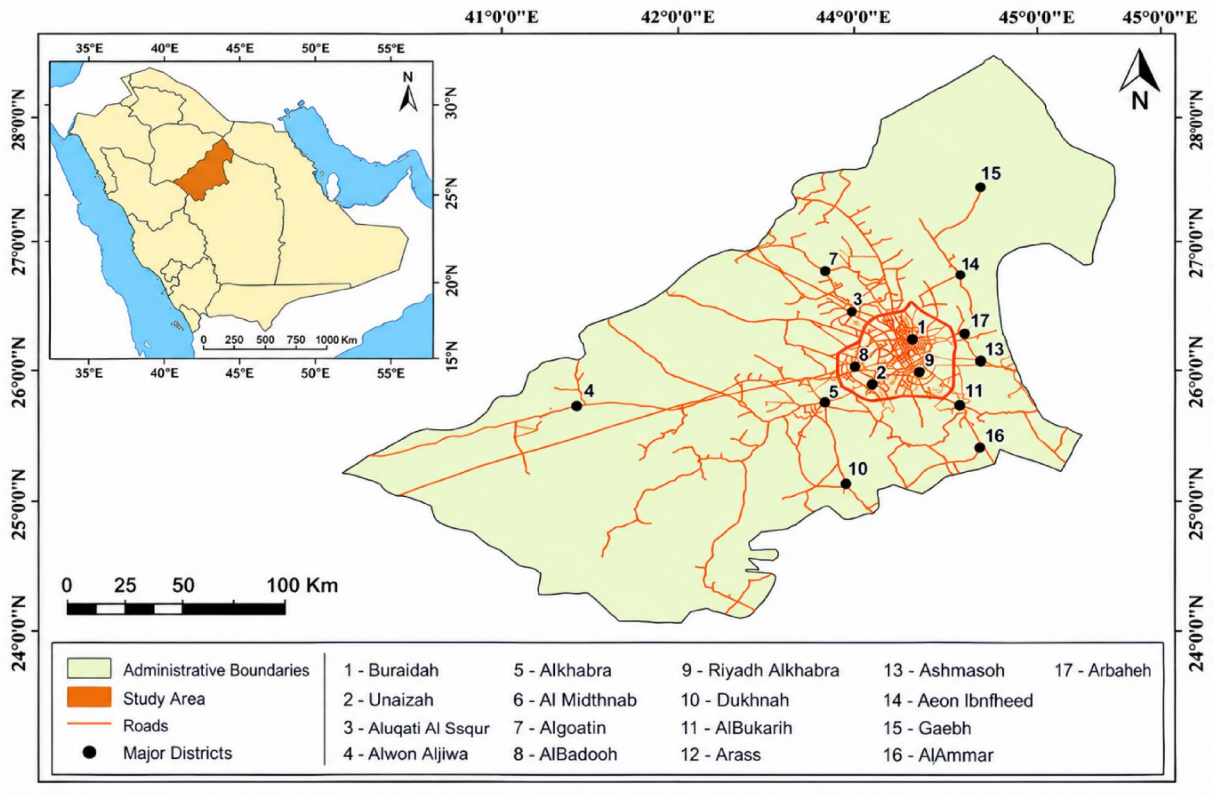


Fig. 1 Location of the study area: Al-Qassim region. Prepared by the authors using QGIS (Version 3.42.0; https://qgis.org) with OpenStreetMap basemap data (© OpenStreetMap contributors).

concerns for the local authorities. A recent study[102] attempted to identify the significant causes of increased traffic crashes in the Al-Qassim region and found that overspeeding, irregular stops, irregular bypasses, lack of road readiness, careless driving, disobedience of traffic laws, and cell phone use while driving are the few predominant crash risk factors.

The crash dataset was obtained from the traffic safety department at the Ministry of Transport (MOT), in Riyadh, KSA, with authorization for their use for research purposes. Since the study involved only secondary analysis of anonymized administrative records, with no direct involvement of human participants or collection of personally identifiable information, institutional ethics review was not required. Crash data for the study area (Qassim region) was then extracted and preprocessed by removing crash records with missing information, duplicate records, and outliers to make it ready for the subsequent analysis. The analysis was based on one year (2018) of historical crash records for the study area. The final data contained 1,142 valid crash instances over the study period. A total of 948 injuries and 170 fatalities were reported. The variables in crash data were divided into several main categories (such as temporal attributes, roadway characteristics, vehicle features, crash characteristics, etc.). The data type and the number of sub-categories under each main attribute are shown in Table 1. Crash injury severity, the target variable, was classified into two levels, i.e., fatal and non-fatal. The fatal crash instances, however, had limited frequency compared to non-fatal crash records. The number of fatal injury instances was 170, whereas the remaining 948 were non-fatal. Table 1 also shows the sub-classification of the main explanatory variables as well as their frequency share.

Although the final dataset consists of one year of crash data, comprising 1,142 valid crash observations, this sample size is considered appropriate for the present binary classification problem involving a moderate number of explanatory variables and structured tabular data. Similar sample sizes have been reported in previous traffic safety crash severity modelling literature, employing both statistical and ML approaches[89,103]. Furthermore, to improve the model robustness and reduce the likelihood of overfitting, the proposed modelling framework incorporates various procedures, including independent training-testing datasets, model regularization, Bayesian hyperparameter optimization, etc. Therefore, the available dataset provides a suitable basis for comparative performance assessment of selected ML paradigms while recognising the inherent limitations associated with temporal coverage.

Methods

This section presents a brief description of machine learning modeling approaches used for crash severity analysis. The proposed machine learning techniques are Logistic Regression (LR), Random Forest (RF), and the extreme gradient boosting algorithm (XGBoost), which were implemented in Python (version 3.9.12). Different Python libraries such as pandas for Tabular data reading, preprocessing, and data analysis, ImbalancedLearn for data imbalance problems, Scikit-learn for model implementation, and Matplotlib for results visualization were utilized. Hyperparameter tuning was accomplished through Bayesian optimization within predefined search ranges. This approach has a number of inherent advantages over the grid and random search techniques[104]. To enhance the transparency and reproducibility of the developed ML model, the final optimized hyperparameter settings obtained through Bayesian optimization are given in Table 2. The final/optimized parameter configuration shown in Table 2 corresponded to the combination that achieved the highest validation performance on the training dataset while reducing the risk of overfitting.

The original crash dataset exhibited a significant class imbalance, with approximately 85% non-fatal crashes and 15% fatal crashes. Analysis based on such an imbalanced dataset may yield biased training and high prediction accuracy toward the majority class. To minimize the possible bias in predictions toward the majority class, the Synthetic Minority Oversampling Technique (SMOTE) was systematically applied. To avoid data leakage, the dataset was first randomly partitioned into independent training (70%) and testing (30%) subsets using a stratified approach to preserve the distribution of each severity class. The stratified sampling approach for a 70%/30%

Table 1 Descriptive statistics of explanatory variables related to crash data.

Input variables	No of categories	Variable nature	Frequency distribution
Temporal attributes			
Period of the day	2	Categorical	Off-Peak = 723 Peak = 418
Day of the week	2	Categorical	Weekday = 862 Weekend = 279
Environmental attributes			
Lighting conditions	*2*	Categorical	Dayligh = 627 Nighttime = 514
Weather status	*4*	Categorical	Good = 846 Dusty = 168 Rainy = 95 Others = 32
Vehicle attributes			
Number of vehicles involved	3	Numeric	Single vehicle = 673 Two Vehicles − 408 Multiple Vehicles = 60
Vehicle type	5	Categorical	Private car = 659 Light Truck = 268 Light Truck = 162 Buses = 39 Others = 51
Roadway attributes			
Road type	3	Categorical	Expressway = 529 Divided Highway = 400 Single Highway = 212
Road alignment details	4	Categorical	Straight = 923 Horizontal Curve = 88 Vertical Curve = 39 Near Inersection = 91
Road status (surface conditions)	3	Categorical	Good = 688 Cracking/Disressed Pavement = 278 Others= 175
Damage type	8	Categorical	No Damage = 831 Cable Barrier = 46 Concrete Barrier = 34 Culvert Fence = 62 Light Post = 17 Metal Barrier = 69 Road Sign = 32 Road Surface = 50
Road markings	2	Categorical	Present = 992 Absent = 149
Road cateyes	2	Categorical	Present = 1037 Absent = 104
Area type	2	Categorical	Urban = 781 Rural = 360
Speed limit	-		-
Crash locations (coordinates)	-	Numeric	-
Crash attributes			
Number of causalities	-	Numeric	-
Accident type	7	Categorical	Vehicle Collision = 600 Vehicle Burnt = 40 Hit Animal = 20 Hit Pedetrian = 13 Rollover 300 Swerving = 111 Others = 55

Table 1 (continued)

Input variables	No of categories	Variable nature	Frequency distribution
Accident cause	4	Categorical	Driver = 915 Vehicle = 164 Roadway = 43 Others = 19
Crash severity	2	Categorical	Fatal = 170 Non-Fatal = 948

Table 2 Hyperparameter search settings used for selected ML models.

Algorithm	Parameters description	Search range/setting	Optimized value
Logistic regression (LR)	Penalty	L1 / L2	L2
	Regularization (C)	0.01–100 (log scale)	1.0
	Maximum iterations	100–1500	800
	Solver	lbfgs / liblinear	Lbfgs
Random forest (RF)	Number of trees	100–1000	500
	Minimum samples leaf	1–10	3
	Minimum samples split	2–20	5
	Maximum tree depth	5–30	18
	Maximum features	sqrt, log2	Sqrt
XGBoost	Learning rate	0.01–0.3	0.11
	Maximum depth	3–10	7
	Number of estimator	100–1000	400
	Subsample	0.5–1.00	0.90
	Minimum child weight	1–10	3
	Gamma	0–5	0.20

train-test partition ensured that both subsets preserved the original class distribution before SMOTE application. SMOTE was then applied exclusively to the training subset to synthetically generate minority-class (fatal crash) observations until a balanced class distribution was attained. The testing subset remained unchanged throughout the model development and was used only for final model evaluation. This procedure follows standard and established best practices for handling imbalanced classification problems using SMOTE while ensuring unbiased estimation of predictive performance[105]. It should be mentioned that class imbalance was addressed exclusively using the SMOTE technique, and no additional class-weighting strategy was deployed during the model training.

Predictive performance of machine learning models has been assessed in terms of different evaluation metrics such as confusion matrices, accuracy, precision, recall, F-1 scores, the area under the curve (AUC), and receiver operating characteristic (ROC) curves. The analysis was based on crash data described in the preceding section. Finally, a trained XGBoost model was used to conduct feature importance analysis for quantifying the relative contribution of each predictor to crash severity outcomes. Subsequently, XGBoost feature importance results were interpreted together with SHAP analysis to provide complementary global and local explanations of the developed prediction model. Detailed methodology is presented in the following sections.

Prediction models

Logistic regression

Logistic regression is the most widely used method in machine learning for solving classification problems that have discrete choice categories, like the mode choice outputs for the current study. The name logistic is taken after the function which is employed at the core of this technique, “the logistic function.” The method is not very good at performing regression tasks despite its name. The method is commonly used when the target variable has binary output, although it can be extended as multinomial logistic regression for multi-class problems

as well. The main advantage of using logistic regression over discriminant analysis is that it does not assume a normal distribution of independent variables. The method can also provide ROC curves and confidence intervals on predicted values that help identify the best cut-off points in classification. The LR model aims to establish the relation between the target class y= (y1, y2) and the given set of j predictors x = (x1,…, xj) and p = (p1, pn). The approach seeks to represent the relationship f between several predictors or independent variables x and predictors or class variables y. For the present study, two possible outcomes were encoded into the target variable (fatal and non-fatal crash injury severity). The modeling function of LR describes a correlation between a specific class against a set of independent variables. The general formulation of LR is presented in the following relations (Eqs. 1 and 2).

$$P(y=1 \mid x) = \frac{1}{1+e^{-\left(\beta_0+\sum_{j=1}^{n}\beta_j x_j\right)}} \tag{1}$$

$$\log\left(\frac{P(y=1|x)}{1-P(y=1|x)}\right) = \beta_0 + \sum_{j=1}^{n}\beta_j x_j \tag{2}$$

Where $P(y=1 \mid x)$denotes the probability of a fatal crash given the predictor vector x, $\beta_0$is the intercept, $\beta_j$is the regression coefficient for predictor x_j, and $n$is the total number of predictors. The logistic regression model assigns a probability value to each severity class based on the input features. By comparing these probability values with a predefined threshold, the predicted class for each sample is determined. The threshold is typically set at 0.5, meaning that if the predicted probability for a severity class is greater than or equal to 0.5, the corresponding class is assigned to that sample.

Random forest

Random Forests (RF) is another commonly used classification algorithm that was initially proposed by Leo Breiman in 2001[106]. Unlike logistic regression, RF can be used effectively for both classification and regression tasks depending on whether the target variable has a categorical nature or continuous response. In predicting the output, RF forms a family of tree predictors (ensemble trees) with the same distribution of trees throughout the model[26]. The values at each tree are extracted from a random vector that is sampled independently. As the number of trees becomes large, the generalization error in the RF model converges and is a function of the strength of the individual tree and the correlation between them. The algorithm starts with the selection of bootstrap samples from the input data. Observations in the actual dataset that do not belong to the bootstrap sample are referred to as out-of-bag (OOB) observations. This is followed by fitting an ensemble classification tree to each bootstrap sample by random selection of a small number of variables. The fully grown trees are then used to predict OOB observations. The majority vote method is subsequently used for the class prediction of that observation. Error rates and accuracies are calculated by averaging the out-of-bag predictions for each observation. RF model has some advantages for solving classification tasks, that include their ability to be used for high-dimensional problems, the requirement of only one or two tuning parameters, the ability to train fast and predict, and a built-in estimate for providing the generalization error[69]. The typical classification model of RF is shown in Eq. 3.

$$\hat{y} = \frac{1}{T}\sum_{i=1}^{T} h_i(x) \tag{3}$$

Where $\hat{y}$ is the final prediction, T is the number of trees, $h_i(x)$ represents the prediction from ith tree. It can be observed from the above relation that the prediction/output of RF model is the mean of prediction values of all the trees.

XGBoost

The extreme gradient boosting algorithm (XGBoost) is a newly developed scalable machine learning technique based on a gradient boosting strategy to improve the predictive model performance. The algorithm is originally based on work conducted by Friedman in early 2000[107], and later on, Chen and Guestrin extended this work and made some improvements, and it was named XGBoost[108]. In recent years, the algorithm has been applied across diverse research domains[109–112]. The prediction of XGBoost is based on the "boosting" mechanism, which implies an additive strategy is employed to combine the predictions of weak learners to produce strong learners in an iterative fashion. To optimize the loss function, the residuals are used to correct the preceding predictor for individual iterations of gradient boosting. To improve the algorithm's predictive performance and avoid overfitting, random regularization is added to the loss function. The general ensemble model of XGBoost may be expressed as a summation of the sample's classification scores from all trees. The typical formulation of the model is shown in Eq. (4).

$$\overset{'}{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in \mathcal{F} \tag{4}$$

In the above equation, K denotes the number of regression trees, f_k is the k^{th} regression tree, and $\mathcal{F}$ is the space of all regression trees. For optimizing the objective function, Eq. 4 can be further optimized as follows: (Eq. 5):

$$\text{Obj} = \sum_{i=1}^{n} l\left(y_i, \overset{'}{y}_i\right) + \sum_{k=1}^{K} \Omega\ (f_k) \tag{5}$$

Where $l(y_i, \hat{y}_i)$ denotes the loss function measuring the difference between the predicted and observed values, while $\Omega\ (f_k)$ is the regularization term used to control model complexity and reduce overfitting. The first term in the equation denotes the loss function to help in differentiating between the prediction and the target. The second term in the same expression shows the regularization terms, which are aimed at controlling the overfitting issue and reducing the model complexity. Unlike conventional gradient boosting approaches, XGBoost adopts an incremental regularization term to account for a learning rate (L) (shown in Eq. 6), which helps to reduce the impact of successive trees.

$$\Omega\ (f) = \gamma K + \frac{1}{2}\lambda \sum_{j=1}^{K} w_j^2 \tag{6}$$

Where γ controls tree complexity, λ is the L2 regularization parameter, K denotes the number of leaves in the tree, and w_j is the score assigned to the j^{th} leaf.

Model's evaluation

Different performance metrics were used to evaluate the predictive efficiency of the proposed machine learning models. These include confusion matrices, accuracy, recall, precision, AUC, and ROC. For classification analysis, the confusion matrix is composed of four possible outcomes, i.e., true negatives (TNs), true positives (TPs), false negatives (FNs), and false positives (FPs). Accuracy is the rate of correctly predicted observations over the total number of observations in the dataset. It can be determined using the expression shown in Eq. 7. Accuracy ranges from zero to one, with zero showing the worst prediction and vice versa. An accuracy of 0.5 indicates random classification. For imbalanced datasets, it frequently indicates a biased tendency toward the majority class, and hence other performance metrics should be reported as well to have a reliable comparison between different machine learning modeling approaches. The metric recall is the ratio of correctly classified/predicted observations of a class divided by the total actual number of observations of that class. Recall may be computed using the relation shown in Eq. 8. Similarly, precision is another important evaluation metric, which may be obtained by dividing the total accurately predicted observations of a given class/label by the total instances predicted as that

class/label (Eq. 9). Although both recall and precision can provide some useful insights about model performance in predicting individual classes, there is a trade-off between the two and hence do not provide a complete picture of the model. The F-1 score is another useful metric commonly used for evaluating the classification performance of machine learning models. It is calculated by taking the harmonic mean of both recall and precision values (Eq. 10) and is considered a better representative of the performance of individual classes. ROC is yet another important performance criterion that indicates the diagnostic ability of adopted machine learning models. ROC depicts the relationship between the true positive rate and the false-positive rate. AUC is also reported alongside the ROC, which represents the area under its curve (Eq. 11). An AUC value of 0.7 and above is regarded as acceptable for machine learning classification tasks.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$F - 1\ Score = 2 * \frac{Precsion\ .\ Recall}{Precision + Recall} \tag{10}$$

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(x)\ d\mathrm{FPR}(x) \tag{11}$$

Since the available crash dataset exhibits class imbalance, model performance was evaluated using several complementary metrics rather than relying on overall accuracy alone. Accuracy provides an overall measure of correct classifications; however, it may overestimate predictive capability for the majority class under imbalanced dataset situations. Hence, other complementary class-sensitive metrics such as recall, precision, F1-score, ROC curves, and AUC were jointly considered to provide a more comprehensive and reliable assessment of the model's predictive performance.

SHAP analysis

The model's interpretability is the main concern when applying machine learning models. Model interpretability is essential to build confidence in their application and understanding of the role of individual predictors on the target variable. To address these black box mechanisms of machine learning models, various interpretation techniques have recently been increasingly employed in various applications. The current investigation employed one such recently proposed and well-known interpretation technique known as SHapley Additive exPlanations (SHAP) to examine the role of independent variables on the resulting crash severity outcomes. In order to improve the interpretability of machine learning models, Lundberg & Lee (2017) developed SHAP. Using game theory and local explanation, this approach quantifies the contribution of each feature to model prediction based on SHAP values. The SHAP values of each feature show the anticipated deviation from the base value in the model forecast. For ranking feature importance using the SHAP analysis method, the mean SHAP values of each instance of the dataset for that feature are calculated. Thus, the average importance for feature j is the mean of SHAP values for all of the samples, calculated as Eq. 11.

$$I_j = \frac{1}{N} \sum_{i=1}^{N} \left| \phi_j^{(i)} \right| \tag{12}$$

where I_j denotes the global importance of feature j, N is the total number of observations, and $\phi_j^{(i)}$ is the SHAP value of feature j for observation i.

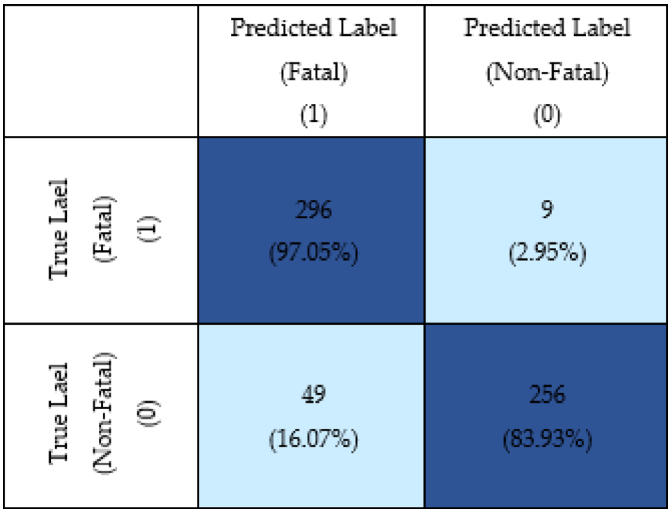


Fig. 2 Confusion matrix obtained using XGBoost for crash severity classification.

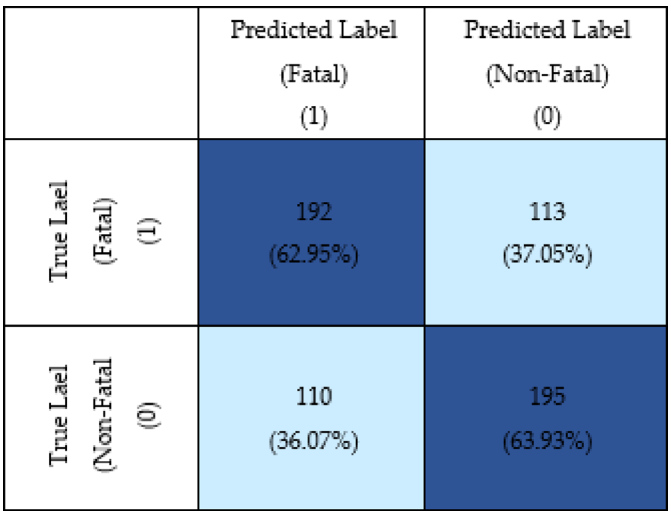


Fig. 3 Confusion matrix obtained using logistic regression for crash severity classification.

Results and discussion

Modeling results

The confusion matrices for the test dataset of the developed models are presented in Figs. 2 and 3, and 4. Confusion matrix results are vital in understanding the interactions between the true and wrong classification of target labels. Rows in the confusion contingency table represent the actual frequencies of target labels in the original dataset, while the columns in this matrix show the frequency/percentage of predicted observations of each class. Further, the values in the diagonal cells show accurate classification/predictions. In comparison, the off-diagonal elements represent misclassified observations, which is indicative of model under- and/or overestimation. The overall prediction accuracy is obtained by taking the sum of correctly predicted instances and the total number of input observations.

As shown in Fig. 2, the confusion matrix for XGBoost yielded better crash severity classification results compared to logistic (LR) regression and random forest (RF) models. Out of 305 instances taken for the test dataset, 296 (approximately 97%) are correctly classified under injury severity "class 1" (fatal crashes) by the XGBoost algorithm. However, the model also resulted in wrong predictions or misclassified instances. For example, 9 (about

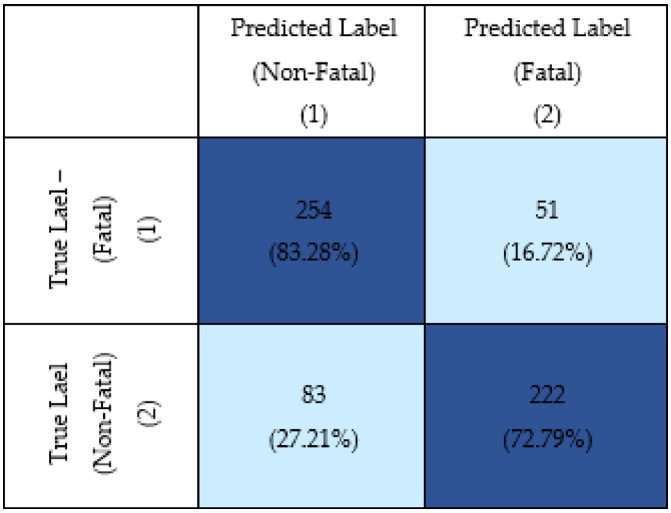


Fig. 4 Confusion matrix obtained using the random forest for crash severity classification.

Table 3 Comparison of confusion matrix results for selected ML classifiers.

Classifier	True positives (TPs)	False negatives (FNs)	False positives (FPs)	True negatives (TNs)	Accuracy
Logistic regression (LR)	192 (62.95%)	113 (37.05%)	110 (36.07%)	195 (63.93%)	63.4%
Random forest (RF)	254 (83.28%)	51 (16.72%)	83 (27.21%)	222 (72.79%)	78.0%
XGBoost	296 (97.05%)	9 (2.95%)	49 (16.07%)	256 (83.93%)	90.5%

3%) observations of "class 1" (fatal crashes) are wrongly predicted to belong to "class 2" (non-fatal crashes). Approximately 84% of "class 2" (non-fatal crashes) are correctly predicted under the same class. Similarly, 49 (about 16%) instances of "class 2" are misclassified as "class 1" (fatal crashes). Nevertheless, the summation of correct model classifications (XGBoost model) achieved an overall prediction accuracy of approximately 90%, which is indicative of favorable model performance. In contrast, the LR model showed a relatively higher rate of misclassification than both the ensemble-based classifiers (RF and XGBoost), especially for fatal crashes (shown in Fig. 3). This is reflected by the larger number of false positives (113) and false negatives (110) observed in the confusion matrix for LR, indicating that the linear decision boundary of the classifier was less effective in capturing the complex non-linear relationships between the predictors and crash severity outcomes. On the other hand, RF demonstrated a noticeable improvement over LR by correctly classifying a larger proportion of both non-fatal and fatal collisions (shown in Fig. 4). However, the remaining false positive (51) and false negative (83) predictions indicate that its performance was still inferior to that of the XGBoost model.

For ease of comparison, Table 3 provides a consolidated summary of confusion matrix results for the evaluated models, reporting both the absolute frequencies and their corresponding percentages for correct classification (represented by TPs and TNs) and wrong classification (represented by FNs and FPs). The summarized results confirm that XGBoost achieved the highest classification performance by correctly identifying the largest proportion of both fatal and non-fatal crashes. LR achieved the lowest predictive performance, particularly for fatal crash observations, while the RF model exhibited intermediate performance.

Table 4 summarizes a comparative analysis of models' injury severity classification/predictive performance based on various evaluation metrics, including accuracy, precision, recall, and F-1 scores. All reported performance metrics shown in the table below were calculated using predictions on the original independent test dataset. SMOTE was used exclusively during the model training phase and was not included in model evaluation, ensuring that the reported model prediction results improve the generalization performance on unseen crash records. The experimental results shown suggest that the developed/proposed model performed well on the input crash data. LR achieved an acceptable score for the considered metrics despite its simplicity and shallow architecture. However, it is the worst-performing model with an average accuracy of 63.4%. The RF model obtained

Table 4 Comparative assessment of crash severity classification for selected ML models.

ML model	Accuracy	Precision	Recall	F1 score	AUC
Logistic regression (LR)	63.4%	63.0%	64.0%	63.0	0.71
Random forest (RF)	78.0%	78.0%	91.0%	79.0%	0.84
XGBoost	90.5%	91.0%	90.0%	90.0	0.92

*Metrics were calculated using on independent hold-out test dataset (n=305; 30% of the dataset). The training dataset (n=713; 70%) was balanced using SMOTE to address data imbalance and avoid any data leakage. F1-score was computed as the harmonic mean of precision and recall.

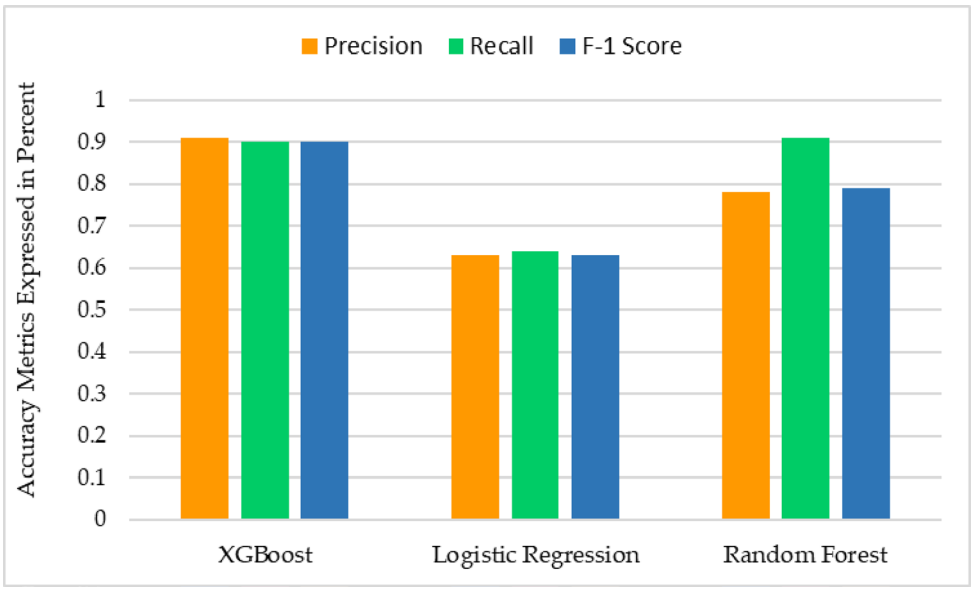


Fig. 5 Injury severity classification comparison of proposed models.

accuracy, precision, recall, and F-1 scores values of 0.78, 0.78, 0.91, and 0.79, respectively. The improved prediction performance of the RF model against LR may be attributed to its ensemble tree structure, where a bootstrap sample of a small number of variables is randomly selected at each iteration, which enhances its performance. Several previous studies also demonstrated the superior performance of RF compared to LR in crash severity applications[20,113,114]. Overall, the results indicated XGBoost achieved the highest overall prediction accuracy (90.5%) and superiority as consistently reflected by higher precision, recall, F1-score, and AUC values. Such complementary metrics are principally important for evaluating the classifier's predictive performance under imbalanced crash datasets because they are designed to assess its ability to correctly identify both minority and majority severity classes. Therefore, the conclusion regarding the superior performance of the XGBoost model is based on the collective evidence provided by multiple evaluation measures rather than solely overall accuracy. Similarly, it also had a better performance in terms of other evaluation metrics compared to other models. However, the XGBoost model required a relatively longer training run compared to other models, which may be attributed to the complex model structure and exhaustive tuning of the number of hyperparameters to optimize the model's performance. XGBoost is based on a gradient boosting technique at the core of its architecture, where the weak learners are iteratively combined to form a strong classifier while making the prediction, and hence usually high performance may be expected[115]. Since computational efficiency was beyond the scope of this study, a quantitative runtime comparison of selected ML models was not included because it is dependent on the hardware resources and computing environment. Figure 5 presents more illustrative bar charts showing the comparative prediction efficiency of proposed machine learning models.

The comparative analysis of the developed models was also evaluated by Receiver Operating Characteristic (ROC) curves as shown in Fig. 6. ROC curve depicts the interactions between false positive rate and true positive rate in two-dimensional space. Points located on a hypothetical diagonal plane are indicative of random

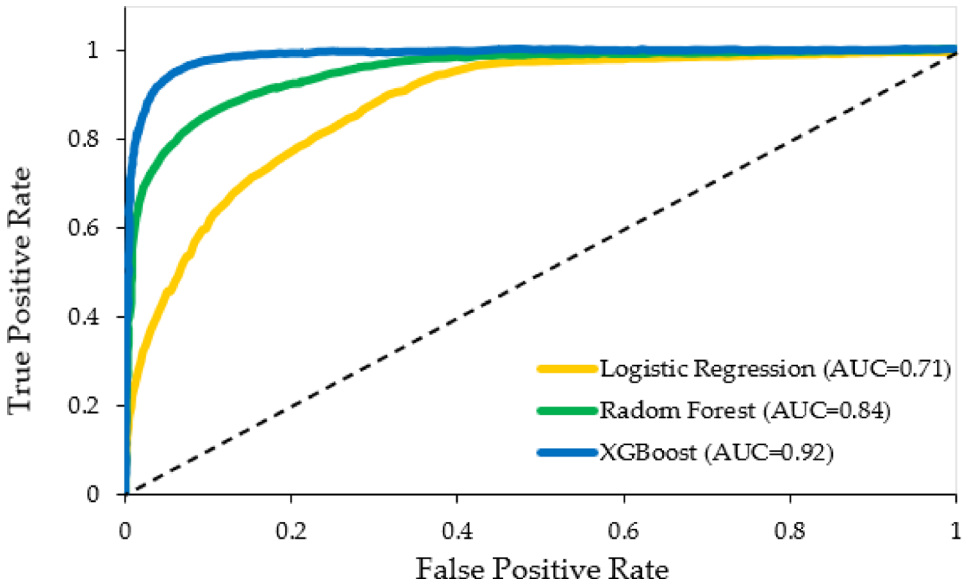


Fig. 6 ROC for the model's predictive performance evaluation.

prediction/classification. Similarly, points lying on this plane imply a worse prediction than a random prediction, while those lying above the diagonal show that a model performs better than a random prediction. The metric area under the curve (AUC) may be quantitatively estimated by taking the integral of the ROC curve. An AUC value of 1 implies that an algorithm has achieved a perfect classification for all the classes/labels. Model predictions are considered random if the AUC value is near 0.5, while a value of 0.7 and above is considered adequate for classification tasks[69]. As shown in Fig. 6, the AUC metric for all the classifiers is above 0.7, indicating that the models performed reasonably well on the input crash dataset. A comparative overview of different models reveals that XGBoost has the highest AUC value of 0.92, followed by RF (0.84) and LR (0.71). These results further reinforce that the XGBoost algorithm is more robust, suggesting that it can better capture the patterns in the dataset compared to other classifiers.

The predictive performance results of selected ML models demonstrate that the available crash dataset contains sufficient information for distinguishing between fatal and non-fatal injury outcomes. Nevertheless, model performance should be interpreted within the context of the available data, as crash characteristics may evolve because of changes in driver behaviour, traffic demand, roadway infrastructure, vehicle technologies, and enforcement strategies. Consequently, the reported results reflect the predictive capability of the selected ML models for the analysed dataset and should not be understood as universally transferable without supplementary external validation.

XGBoost feature importance analysis

The primary drawback and frequently encountered criticism against the application of machine learning models is that the obtained results are not explicitly interpretable. For applications such as crash severity analysis, it is often very important to understand the impact of individual risk factors and contributing circumstances on the severity outcome. Mere accurate classification/prediction of crash severity instances does not provide crucial insights about the relative role of significant severity risk variables. Accurate information about both (accurate severity prediction and variable importance) is helpful to researchers and practitioners for proposing and prioritizing suitable policy implications within the limited available highway safety budget. To enhance the interpretability of the developed ML framework, feature importance analysis was performed using the trained XGBoost model, which achieved the highest predictive performance compared to other evaluated classifiers. Selecting the best-performing model for interpretation ensures consistency between model prediction and interpretation, thereby

providing more reliable and meaningful insights into factors influencing crash severity outcomes. XGBoost was preferred for feature importance analysis due to its gradient boosting framework which can capture complex non-linear relationships and variable interactions more effectively than traditional ensemble learning methods. Subsequently, XGBoost feature importance results were independently examined using SHAP analysis, allowing to evaluate the consistency of the crash severity predictors through two complementary different explainability techniques.

Feature ranking analysis using XGBoost for the top six crash severity predictors is presented in Fig. 7. High scores of the evaluator are indicative of relatively greater feature importance. Labels for input features are shown on the y-axis, while the corresponding feature scores are shown on the x-axis. It is evident from the results presented in Fig. 7 that crash severity is significantly associated with explanatory variables such as accident type, weather conditions, number of vehicles involved, lighting conditions, accident cause, and time of the day (peak/off-peak periods). Other predictors (not shown in the figure) had comparatively less significance on the crash severity outcome. The feature ranking analysis given below is mostly intuitive and consistent with recent previous studies in the KSA context, Gulf countries, and worldwide[20,42,67,116–125]. However, it is well established that traffic crashes usually exhibit unobserved endogeneity heterogeneity, which means that the contributing crash injury risk factors may vary from one geographic location to another as well as under different time intervals. Hence, it is also quite possible that the significant severity risk factors of traffic crashes may be different under different spatiotemporal domains.

SHAP interpretation analysis

Unlike traditional feature importance analysis, SHAP interpretation analysis provides both global and local explanation insights by quantifying the contribution of each predictor to individual model predictions, thereby improving the transparency and interpretability of the optimized XGBoost model. The results of the SHAP interpretation analysis are presented in Figs. 8, 9, 10, 11 and 12. Figure 8 shows the global SHAP feature importance obtained from the optimized XGBoost model. SHAP values were derived directly from the trained XGBoost classifier to quantify the contribution of each predictor to the predicted crash severity outcomes. The effect of explanatory

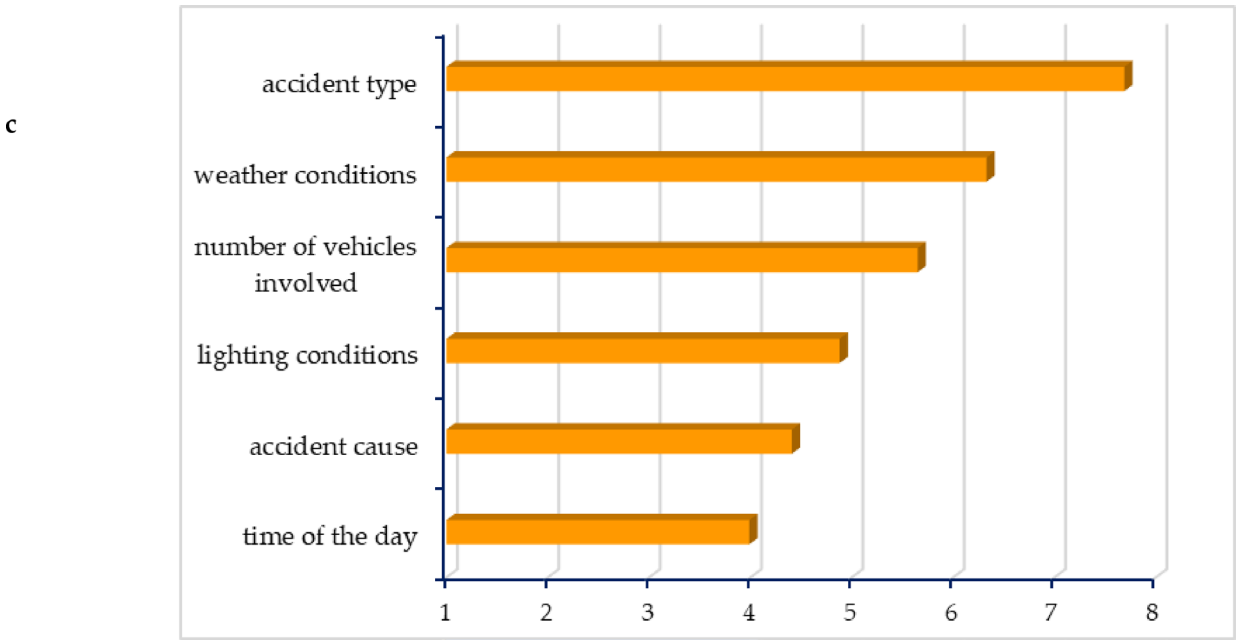


Fig. 7 Features' importance analysis using XGBoost.

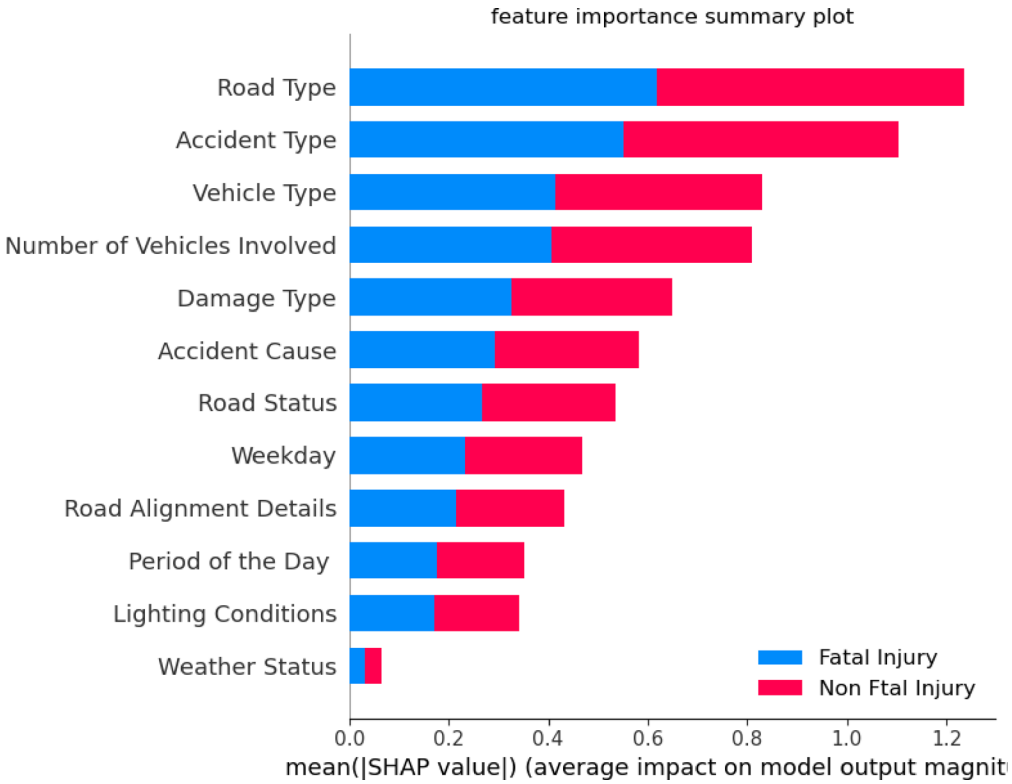


Fig. 8 Bar plots for ranking of significant crash severity risk factors obtained using SHAP.

variables on each injury severity outcome (fatal, non-fatal) is shown. The horizontal axis of the SHAP summary plots indicates the SHAP value, which quantifies the magnitude and direction of each predictor's contribution toward the model prediction. Larger absolute SHAP values indicate a greater influence on crash injury severity prediction.

For fatal crashes, the SHAP summary plots indicate that variables including accident type, vehicle type, road type, number of vehicles involved, damage type, accident cause, and road status exert the greatest influence on the model prediction. These predictors generally exhibit larger SHAP values, signifying a stronger contribution and increasing the probability of fatal crash outcomes. The results shown are intuitive and support the findings of several previous studies[94,98,126].

Similarly, for non-fatal crashes, few variables such as accident type, weather conditions, number of vehicles involved, vehicle type, and damage type emerged as the most influential predictors, although their relative contributions differ from those observed for fatal crashes. These findings imply that the influence of individual predictors varies according to the modeled crash severity outcome, highlighting the importance of separate interpretation of the SHAP results for each severity class. Prior studies have also highlighted the importance of studying the crash severity factors separately for individual severity outcomes[95,127].

Figures 9 and 10 present the SHAP summary plots for the identification and ranking of the crash severity risk factors for fatal and non-fatal crashes, respectively. Interpreting the SHAP summary plots, a feature/variable is more important in model prediction if it has more skewness along the horizontal plane(x-axis). Similarly, variables lying higher on the vertical plane refer to greater feature importance (y-axis). It is clear from the results shown in the Figure below that the features including road type, accident type, vehicle type, number of vehicles involved, damage type, crash cause, and road status are the most significant factors influencing the nature/severity of fatal crashes. Similarly, the likelihood of non-fatal crashes is expected to increase with variables such as accident type, weather status, vehicle type, number of vehicles involved, damage type, etc. It is worth noting

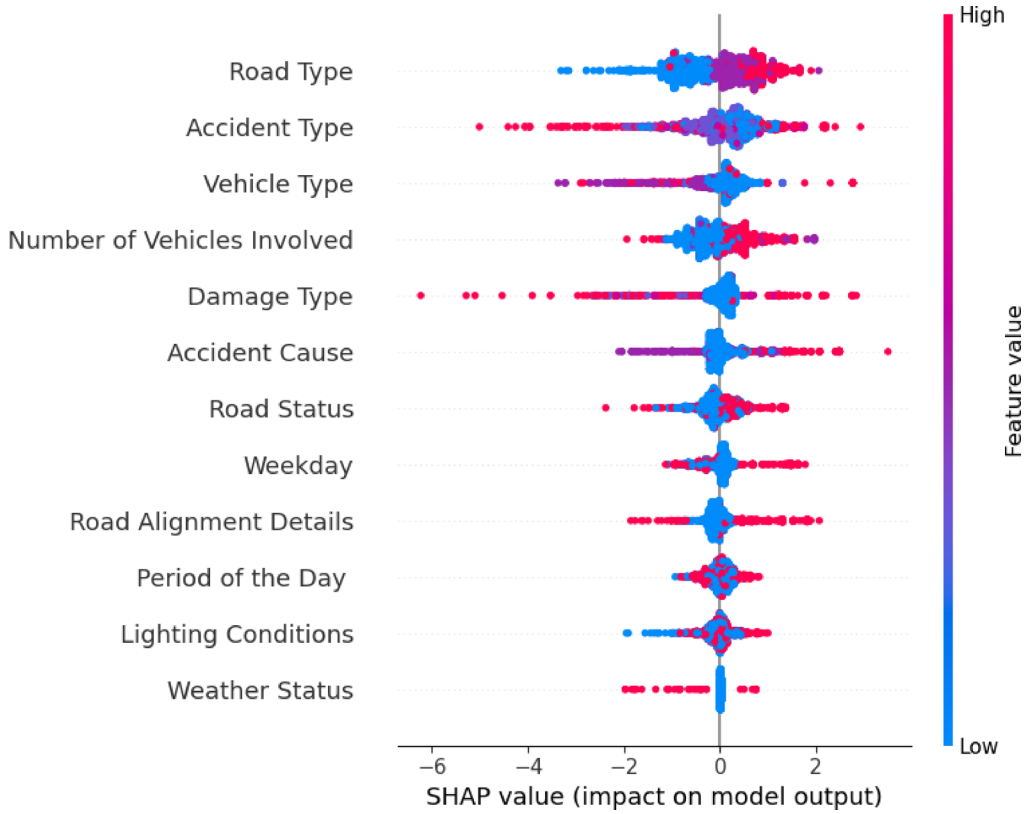


Fig. 9 SHAP summary plots for identification of severity risk factors in fatal crashes.

that high injury severity risk factors identified through SHAP summary analysis are mostly consistent with those obtained through the SHAP method.

Beyond the identification of influential crash severity predictors, the SHAP analysis provides insights into the relative contribution of each factor toward the model predictions. Table 3 compares the influential crash severity predictors identified by the XGBoost feature importance and SHAP analyses to examine the consistency between the two interpretability approaches. It shows the influential crash severity predictors identified by XGBoost feature importance with their corresponding SHAP ranking, and highlights additional predictors identified exclusively by SHAP. Overall, both methods demonstrated substantial agreement for several key predictors, including accident type, number of vehicles involved, and accident cause, indicating that these variables induce consistent and stable contributions to crash severity outcomes prediction regardless of the employed interpretability method. This consistency strengthens confidence in the robustness of the identified crash severity risk factors, suggesting that these variables contribute reliably to crash severity prediction across different interpretation methods/techniques. Variables such as time of day exhibited partial agreement, whereas some predictors like lighting and weather conditions exhibited comparatively larger differences in their rankings. Further, road type and damage type were identified as top predictors only by SHAP, highlighting their localized non-linear effects not emphasized by global feature importance. These differences are expected since both methods quantify feature importance from different perspectives. XGBoost feature importance provides a global ranking based on each variable's contribution to the tree-splitting process during model construction. In contrast, SHAP explains individual predictions by quantifying each feature's marginal contribution while accounting for feature interactions and nonlinear relationship effects. Subsequently, SHAP may identify features with considerable local influence that are not necessarily among the highest-ranked variables in the global XGBoost-based feature importance

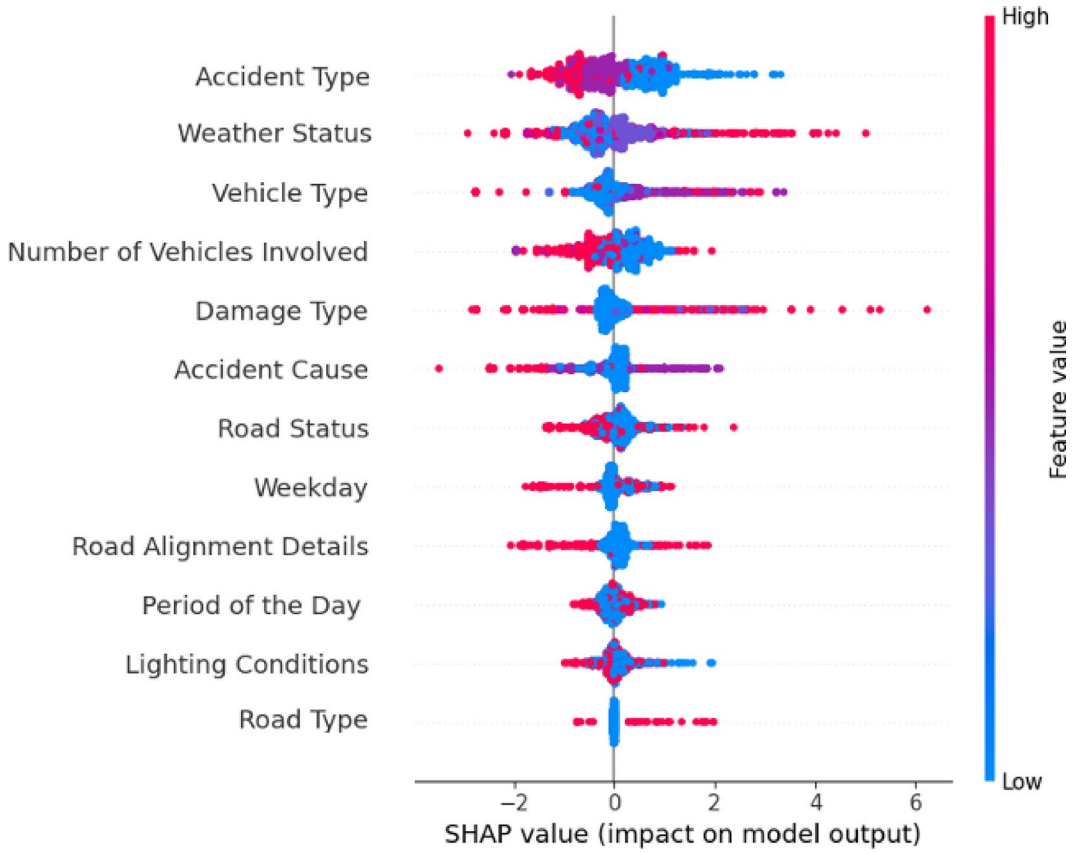


Fig. 10 SHAP summary plots for identification of severity risk factors in non-fatal crashes.

analysis. The two approaches should be viewed as complementary, jointly providing both global and local perspectives on crash severity prediction.

Findings of the current study are generally consistent with recent XML studies for crash severity prediction[34,128,129], which similarly reported that accident characteristics, environmental factors, human-related variables, and roadway conditions constitute the dominant predictors of crash severity. Comparative analysis of XGBoost feature importance analysis and SHAP interpretation observed in this study further supports recent evidence that combining ensemble learning with explainability improves both predictive performance and model transparency, hence facilitating targeted safety interventions and more reliable policy-oriented decision-making (Table 5).

Figures 11 and 12 show the SHAP main-effects interaction plots for the two sensitive features, including accident type and weather status. Main-effects interaction plots are shown for both fatal and non-fatal injury accidents. Main-effects SHAP interaction plots can provide better insights into individuals' effects of sub-predictors (child variables) instead of main variables (parent variables). A better understanding of the role of specific sub-features on crash severity outcomes is important for the selection and suggestion of appropriate countermeasures. Interpreting the main-effects interactions plots, it is worth mentioning that points located above the hypothetical horizontal plane above zero imply greater feature importance and vice versa.

Figure 11a shows the SHAP main-effects plot for fatal crashes. Higher positive SHAP values suggest that certain crash types contribute more strongly in predicting fatal injury outcomes. Further, the relatively wide spread of SHAP values also implies that the contribution of variable accident type varies across individual crash observations, reflecting the heterogeneous nature of fatal crash severity prediction. Considering the main- interaction plot for the feature accident type, it is worth noting that human-runover crashes, rollover crashes, and those that

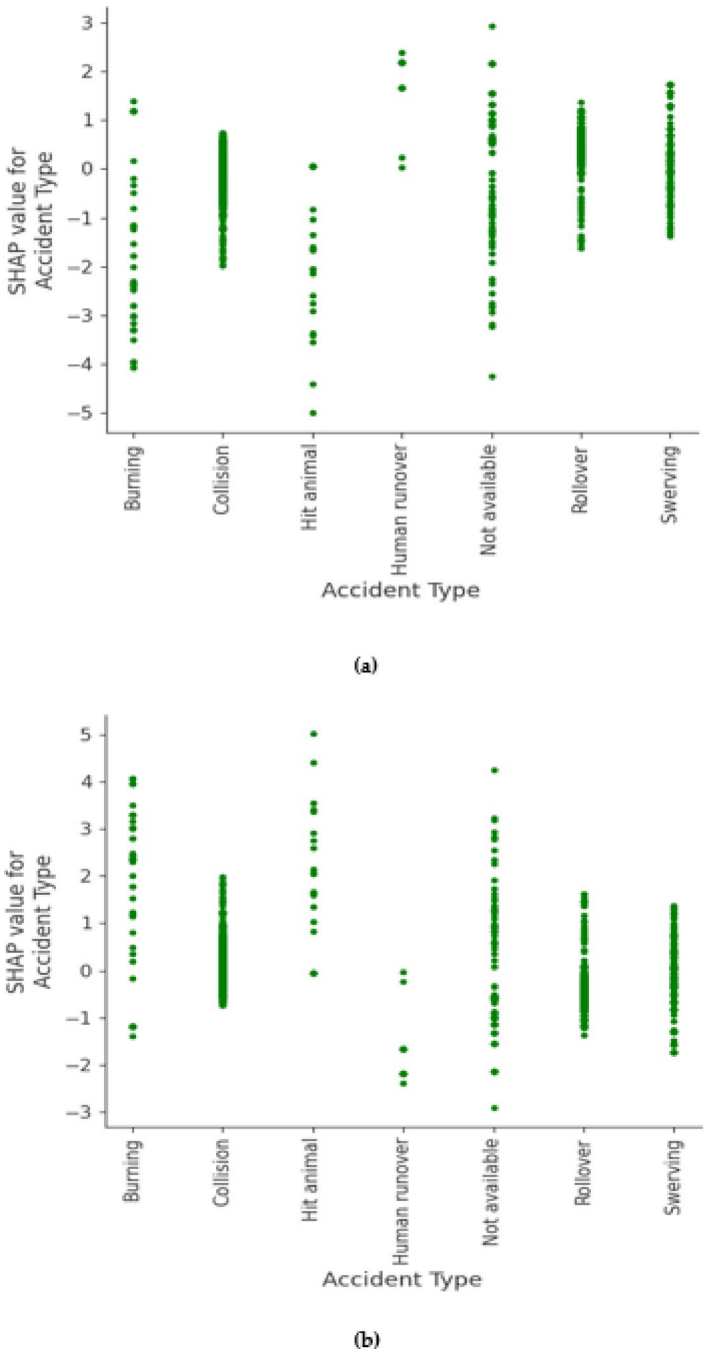


Fig. 11 SHAP main-effects interaction plot showing the contribution of accident type to predictions of fatal crash severity (**a**); and non-fatal crash severity (**b**).

occurred due to sudden swerving are expected to increase the probability of fatal injuries. Similarly, considering the main-effects SHAP interaction plot for feature weather status, it is revealed that adverse weather such as rain and poor visibility conditions during dust storms are associated with increased probability of fatal crashes. The reported findings are in line with recent related literature[69,124].

Figure 11b presents the corresponding SHAP main-effects plot for non-fatal crashes. Unlike the fatal crash analysis, the SHAP values distribution exhibits comparatively lower positive contributions, demonstrating that the same accident categories may have different predictive influences depending on the injury severity outcome. This highlights the importance of separately examining the predictor effects for fatal and non-fatal crash classifications.

https://doi.org/10.1038/s41598-026-73109-2

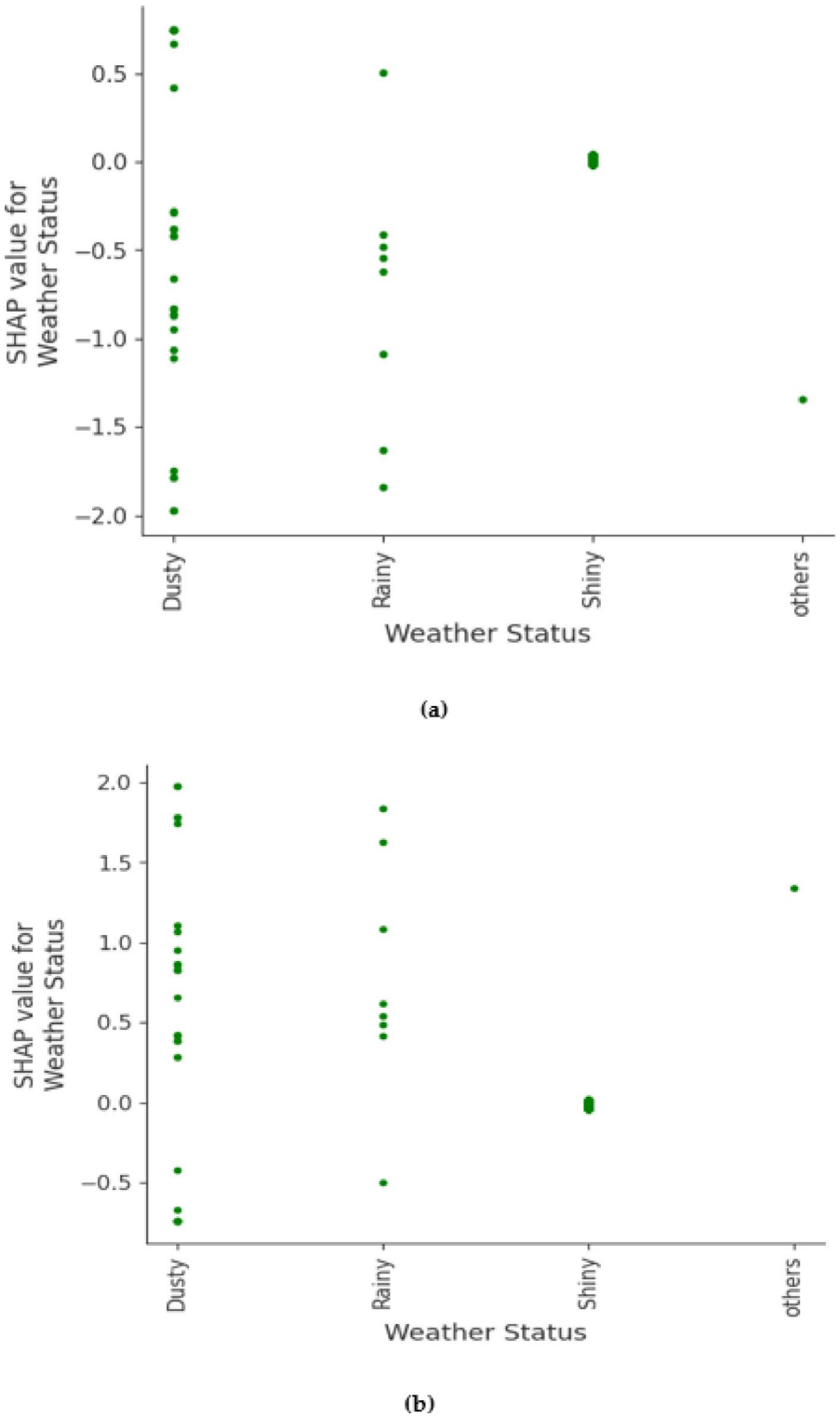


Fig. 12 SHAP main interaction plot showing the contribution of weather conditions to predictions of fatal crash severity (**a**); and non-fatal crash severity (**b**).

Table 5 Comparison of influential crash severity predictors identified using XGBoost feature importance and SHAP and analyses.

Predictor	XGBoost rank	Corresponding SHAP Rank	Agreement	Interpretation
Accident type	1	2	High	Consistently ranked as most influential predictors, suggesting its robust effect on crash severity across both methods.
Weather condition	2	12	Low	Discrepancy in ranking by both methods, indicating its influence is primarily reflected in the model's global structure rather than local predictions.
Number of vehicles	3	4	High	Similar rankings across both methods indicating a consistent and stable contribution to crash severity outcomes.
Lighting condition	4	11	Low	Higher global importance but weaker local contribution, potentially implying its interaction effects with other crash predictors.
Accident cause	5	6	High	Closely aligned variable rankings demonstrating consistent contribution to crash severity prediction.
Time of the day	6	10	Partial	Identified by both methods, although SHAP indicates relatively weaker influence on individual predictions, possibly due to local interaction effects with other crash predictors
Road type	-	1	SHAP only	Identified primarily by SHAP, highlighting the localized nonlinear effects not emphasized by global feature importance.
Damage type	-	5	SHAP only	Identified only by SHAP, showing its important role in explaining individual crash severity outcomes rather than overall model structure.

Policy implications

The findings of the current study could provide valuable insights and essential guidance for effective policymaking to improve highway safety. The explainability analysis outputs yielded by SHAP and XGBoost feature importance analysis provide valuable guidance for practical traffic safety improvement and management. For instance, the high contribution of accident type suggests that safety programmes should focus on prioritising crash types often associated with fatal severity outcomes through engineering improvements, public awareness campaigns, and targeted enforcement. Similarly, the importance of adverse weather and lighting conditions advocates the need for enhanced warning systems during poor weather, weather-responsive traffic management strategies, and improved roadway illumination, especially during the night. The high and consistent contribution of the number of vehicles involved indicator and various crash causes further highlights the importance of strengthening driver behaviour interventions, safety awareness campaigns, speed management, and strict enforcement of traffic regulations.

Conclusions and future works

Traffic crashes remain a major public health challenge worldwide. Analysis of crash severity is an important aspect of highway safety research. Severity outcome of traffic crashes is influenced by complex interactions between various predictors such as humans, vehicles, roadway, weather, and features of the built environment. A better understanding of crash severity risk and crash-contributing circumstances is important for the timely and effective implementation of suitable countermeasures. In the literature, crash severity is frequently studied through the application of parametric (statistical) and conventional non-parametric (machine learning); however, such methods have several inherent discrepancies. The present study is aimed at investigating the applicability

of the state-of-the-art ensemble XGBoost algorithm for crash severity prediction of traffic crashes in the Qassim region of KSA. The predictive performance of XGBoost was evaluated using different statistics (such as accuracy, recall, precision, F-1 scores, ROC, AUC, etc.) in a comparative study with two traditional and well-known machine learning techniques, including LR and RF.

Experimental results revealed that all the developed models yielded an acceptable crash severity classification performance. However, model comparison analysis showed that XGBoost, with an average accuracy of 90.50%, outperformed both the LR and RF models. Similarly, it also achieved superior values for other considered evaluation metrics, indicating that it is more robust and better explains/captures the patterns in the input crash data compared to other classifiers. To overcome the blackbox criticism of machine learning models, the current study also employed XGBoost-based feature ranking/sensitivity analysis. The results identified the top six predictors that can significantly impact the crash severity outcome. These top sensitive variables in this regard include accident type, weather conditions, number of vehicles involved, lighting conditions, accident cause, and time of the day (TOD). Crash severity risk factor evaluation was also conducted through the SHAP approach. Rather than serving only as an interpretation tool for feature ranking, SHAP analysis provides transparent insights for translating the ML and explainable AI predictions into resource prioritisation and evidence-based safety policies. SHAP results revealed that variables including road type, accident type, number of vehicles involved, damage characteristics at the site, and accident cause are expected to increase the likelihood of fatal injury crashes. The explainability analysis demonstrates that the proposed ML framework can provide not only accurate crash severity prediction but also transparent decision support for transportation officials and agencies. Overall, the XGBoost feature importance and SHAP analyses showed substantial agreement in identifying the dominant crash severity predictors, while their differences provide complementary insights into both the global importance and localized nonlinear effects of the explanatory variables. The consistency of risk factors observed between SHAP results and XGBoost feature importance analysis increases confidence in the identified risk factors and enables policymakers to prioritise interventions toward crash scenarios that contribute most to severe crash outcomes. Further study findings demonstrate that explainable AI frameworks can serve as a practical tool for developing targeted and data-driven road safety strategies.

The growing adoption of XML in transportation safety research demonstrates that future crash severity prediction frameworks should simultaneously emphasize improving predictive accuracy, model interpretability, and practical applicability. The framework proposed in this research contributes to this growing research direction by integrating high-performing ensemble ML learning models with transparent model interpretation to support evidence-based road safety interventions and policy-oriented decision-making.

This research is not without a few limitations, which may be covered in forthcoming studies. This study used crash data from a single year in Qassim province of KSA. While adequate for model development and evaluation, the limited geographical and temporal coverage may reduce the generalizability of the findings and transferability to other regions. Future studies should validate the proposed framework using multi-region, multi-year, and external datasets to assess its temporal stability, robustness, and transferability. Similarly, future studies should utilize detailed datasets containing information on other missing explanatory variables and information about drivers/occupants' socio-demographics. Similarly, the present analysis was focused on severity modeling of all crash types; future studies could investigate targeted road user groups such as vulnerable road users (VRUs) or specific crashes, which may reveal more interesting insights. In addition, although SMOTE was employed to address class imbalance, no sensitivity analysis or comparison with alternative imbalance-handling methods (ADASYN, Borderline-SMOTE, or cost-sensitive learning) was conducted. Future work should evaluate different class-balancing procedures to assess their impact on model predictive capabilities and robustness. Further, the present study employed a single hold-out validation framework rather than repeated cross-validation or multiple experimental runs. Although this approach provides an unbiased assessment using an independent test dataset, future studies may employ repeated stratified cross-validation or bootstrap resampling to further quantify the robustness and uncertainty associated with model performance estimates. Future studies can also focus on evaluating computational efficiency under different hardware configurations to investigate the robustness of the

proposed framework using publicly available benchmark datasets and repeated validation procedures. Finally, further research is needed to examine the applicability of various other ensemble techniques, deep learning approaches, and different interpretability analyses such as IME, SAGE (Shapley Additive Global Explanations), partial dependence plots (PDP), Accumulated Local Effects (ALE), permutation feature importance, and counterfactual explanations, to improve the transparency, robustness, and practical interpretability of crash severity prediction models.

Acknowledgements The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University (https://www.qu.edu.sa) for financial support (QU-APC-2026).

Author contributions M.A.: Conceptualization, Formal Analysis, Methodology, Software, Writing Original DraftA.A. Data Curation, Investigation, Visualization, Funding, Writing Draft_review and editing.

Data availability The crash dataset used in this research study was provided by the Ministry of Transport and Logistics Services (MOT), Saudi Arabia, with authorization for academic and research purposes. The underlying data are not publicly available because the authors are not authorized to transfer, share, redistribute, or deposit the data in an open access public repository without prior written authorization from the MOT. To access the relevant data, interested researchers may submit an official data-access request directly to the MOT through its designated communication channels published on their website. all such requests will be independently reviewed by the MOT and remain subject to its applicable requirements, policies, confidentiality conditions, and necessary written approval, as appropriate.

Code availability The custom code used for the data analysis including preprocessing, model development and optimization, and SHAP-based interpretability analysis is publically avilable in the following GitHub repository: https://github.com/moshaogeh87/Qassim-Crash-Analysis. The code was developed specifically for structure and variables of the dataset used in this study. Full reproduction of the reported results and findings requires authorized access to the underlying data directly from the original data provider (MOT).

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Organization, W.H. *Global Status Report on Road Safety 2018 (2018)*. (WHO, 2019).
2. Jamal, A., Rahman, M. T., Al-Ahmadi, H. M. & Mansoor, U. The dilemma of road safety in the Eastern Province of Saudi Arabia: Consequences and prevention strategies. *Int. J. Environ. Res. Public Health*. **17**, 157 (2020).
3. OECD/ITF Road Safety Annual Report 2016.
4. Tauhidur Rahman, M., Jamal, A. & Al-Ahmadi, H. M. Examining hotspots of traffic collisions and their spatial relationships with land use: A GIS-based geographically weighted regression approach for Dammam, Saudi Arabia. *ISPRS Int. J. Geo-Inf.* **9**, 1–22. https://doi.org/10.3390/ijgi9090540 (2020).
5. Lee, S. M. & Al-Mansour, A. I. Development of a new traffic safety education material for the future drivers in the Kingdom of Saudi Arabia. *J. King Saud Univ.-Eng. Sci.* **32**, 19–26 (2020).
6. Mansuri, F. A., Al-Zalabani, A. H., Zalat, M. M. & Qabshawi, R. I. Road safety and road traffic accidents in Saudi Arabia: A systematic review of existing evidence. *Saudi Med. J.* **36**, 418 (2015).
7. Ansari, S., Akhdar, F., Mandoorah, M. & Moutaery, K. Causes and effects of road traffic accidents in Saudi Arabia. *Public. Health*. **114**, 37–39 (2000).

8. Al Turki, Y. A. How can Saudi Arabia use the decade of action for road safety to catalyse road traffic injury prevention policy and interventions? *Int. J. Injury Control Saf. Promot.* **21**, 397–402 (2014).
9. Jamal, A., Mahmood, T., Riaz, M. & Al-Ahmadi, H. M. GLM-based flexible monitoring methods: An application to real-time highway safety surveillance. *Symmetry* **13**, 362 (2021).
10. Al-Ahmadi, H. M. et al. Calibrating the highway safety manual predictive models for multilane rural highway segments in Saudi Arabia. *Arab. J. Sci. Eng.* https://doi.org/10.1007/s13369-021-05944-6 (2021).
11. Abdulrahman, R., Almoshaogeh, M., Haider, H., Alharbi, F. & Jamal, A. Development and application of a risk analysis methodology for road traffic accidents. *Alexandria Eng. J.* **111**, 293–305 (2025).
12. Christoforou, Z., Cohen, S. & Karlaftis, M. G. Vehicle occupant injury severity on highways: An empirical investigation. *Accid. Anal. Prev.* **42**, 1606–1620 (2010).
13. Iranitalab, A. & Khattak, A. Comparison of four statistical and machine learning methods for crash severity prediction. *Accid. Anal. Prev.* **108**, 27–36 (2017).
14. Jeong, H., Jang, Y., Bowman, P. J. & Masoud, N. Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. *Accid. Anal. Prev.* **120**, 250–261 (2018).
15. Ijaz, M., Lan, L., Usman, S. M., Zahid, M. & Jamal, A. Investigation of factors influencing motorcyclist injury severity using random parameters logit model with heterogeneity in means and variances. *Int. J. Crashworthiness* 1–11. (2021).
16. Zahid, M. et al. Predicting risky and aggressive driving behavior among taxi drivers: Do spatio-temporal attributes matter? *Int. J. Environ. Res. Public Health*. **17**, 3937 (2020).
17. Ahmed, T. et al. A new pedestrian crossing level of service (PCLOS) method for promoting safe pedestrian crossing in urban areas. *IJERPH* **18**, 8813. https://doi.org/10.3390/ijerph18168813 (2021).
18. Jamal, A. et al. Implementing the maximum likelihood method for critical gap estimation under heterogeneous traffic conditions. *Sustainability* **14**, 15888 (2022).
19. Ullah, I., Jamal, A. & Subhan, F. Public perception of autonomous car: A case study for Pakistan. *Advances Transp. Stud.* **49**. (2019).
20. Jamal, A. et al. Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study. *Int. J. Injury Control Saf. Promot.* 1–20. (2021).
21. Ullah, I., Liu, K., Yamamoto, T., Zahid, M. & Jamal, A. Electric vehicle energy consumption prediction using stacked generalization: An ensemble learning approach. *Int. J. Green. Energy* 1–14. (2021).
22. Zahid, M., Chen, Y., Jamal, A., Al-Ahmadi, H. M. & Al-Ofi, A. K. Adopting machine learning and spatial analysis techniques for driver risk assessment: Insights from a case study. *Int. J. Environ. Res. Public Health*. **17** (14), 1–15. https://doi.org/10.3390/ijerph17145193 (2020).
23. Zahid, M., Chen, Y., Jamal, A. & Memon, Q. M. *Short Term Traffic State Prediction via Hyperparameter Optimization Based Classifiers*. Vol. 20. (2020).
24. Zahid, M., Chen, Y. & Jamal, A. Freeway short-term travel speed prediction based on data collection time-horizons: A fast forest quantile regression approach. *Sustainability* **12**, 1–19. https://doi.org/10.3390/su12020646 (2020).
25. Jamal, A., Reza, I. & Shafiullah, M. Modeling retroreflectivity degradation of traffic signs using artificial neural networks. *IATSS Res.* S0386111222000462 https://doi.org/10.1016/j.iatssr.2022.08.003 (2022).
26. Zahid, M., Jamal, A., Chen, Y., Ahmed, T. & Ijaz, M. Predicting red light running violation using machine learning classifiers. In *Green Connected Automated Transportation and Safety* 137–148. (Springer, 2022).
27. Ullah, I., Liu, K., Yamamoto, T., Zahid, M. & Jamal, A. Modeling of machine learning with SHAP approach for electric vehicle charging station choice behavior prediction. *Travel Behav. Soc.* **31**, 78–92. https://doi.org/10.1016/j.tbs.2022.11.006 (2023).
28. Tamim Kashifi, M., Jamal, A., Samim Kashefi, M., Almoshaogeh, M. & Masiur Rahman, S. Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. *Travel Behav. Soc.* **29**, 279–296. https://doi.org/10.1016/j.tbs.2022.07.003 (2022).
29. Ihsanullah, I., Alam, G., Jamal, A. & Shaik, F. Recent advances in applications of artificial intelligence in solid waste management: A review. *Chemosphere* 136631. (2022).
30. Ullah, I., Liu, K., Yamamoto, T., Zahid, M. & Jamal, A. Prediction of electric vehicle charging duration time using ensemble machine learning algorithm and shapley additive explanations. *Int. J. Energy Research*.
31. Alkhulaifi, A., Jamal, A. & Ahmad, I. Predicting traffic sign retro-reflectivity degradation using deep neural networks. *Appl. Sci.* **11**, 11595. https://doi.org/10.3390/app112411595 (2021).
32. Butt, M. S., Shafique, M. A. A. & Literature review AI models for road safety for prediction of crash frequency and severity. *Discov Civ. Eng.* **2**, 99. https://doi.org/10.1007/s44290-025-00255-3 (2025).
33. Verma, R. & Agarwal, M. M. A Systematic review on road accident prediction: A special attention toward machine learning and deep learning approaches. *J. Circuit. Syst. Comp.* **35**, 2530009. https://doi.org/10.1142/S0218126625300090 (2026).
34. Abdulrashid, I., Chiang, W. C., Sheu, J. B. & Mammadov, S. An interpretable machine learning framework for enhancing road transportation safety. *Transp. Res. E.* **195**, 103969. https://doi.org/10.1016/j.tre.2025.103969 (2025).

35. Ahmed, A., Topuz, K., Moqbel, M. & Abdulrashid, I. What makes accidents severe! explainable analytics framework with parameter optimization. *Eur. J. Oper. Res.* **317**, 425–436. https://doi.org/10.1016/j.ejor.2023.11.013 (2024).
36. Ma, C., Hao, W., Xiang, W. & Yan, W. The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents. *J. Adv. Transportat. 2018*. (2018).
37. Mesa-Arango, R., Valencia-Alaix, V. G., Pineda-Mendez, R. A. & Eissa, T. Influence of socioeconomic conditions on crash injury severity for an urban area in a developing country. *Transp. Res. Rec.* **2672**, 41–53 (2018).
38. Azimi, G., Rahimi, A., Asgari, H. & Jin, X. Severity analysis for large truck rollover crashes using a random parameter ordered logit model. *Accid. Anal. Prev.* **135**, 105355 (2020).
39. Rifaat, S. M. & Chin, H. C. Accident severity analysis using ordered probit model. *J. Adv. Transp.* **41**, 91–114 (2007).
40. Shankar, V. & Mannering, F. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *J. Saf. Res.* **27**, 183–194. https://doi.org/10.1016/0022-4375(96)00010-2 (1996).
41. Yu, R. & Abdel-Aty, M. Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. *Accid. Anal. Prev.* **62**, 161–167. https://doi.org/10.1016/j.aap.2013.08.009 (2014).
42. Abdel-Aty, M. Analysis of driver injury severity levels at multiple locations using ordered probit models. *J. Saf. Res.* **34**, 597–603. https://doi.org/10.1016/j.jsr.2003.05.009 (2003).
43. Hu, S. R., Li, C. S. & Lee, C. K. Investigation of key factors for accident severity at railroad grade crossings by using a logit model. *Saf. Sci.* **48**, 186–194 (2010).
44. Chen, S., Zhang, S., Xing, Y. & Lu, J. Identifying the factors contributing to the severity of truck-involved crashes in Shanghai river-crossing tunnel. *Int. J. Environ. Res. Public Health*. **17**, 3155 (2020).
45. Khattak, A. J., Kantor, P. & Council, F. M. Role of adverse weather in key crash types on limited-access: Roadways implications for advanced weather systems. *Transp. Res. Rec.* **1621**, 10–19. https://doi.org/10.3141/1621-02 (1998).
46. Chen, C., Zhang, G., Huang, H., Wang, J. & Tarefder, R. A. Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. *Accid. Anal. Prev.* **96**, 79–87 (2016).
47. O'Donnell, C. J. & Connor, D. H. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. *Accid. Anal. Prev.* **28**, 739–753. https://doi.org/10.1016/S0001-4575(96)00050-4 (1996).
48. Rehman, Z. U. et al. Assessing heterogeneity in factors influencing three-wheeled motorized rickshaws crash outcomes between weekdays and weekends. *Sci. Rep.* **15**, 14164 (2025).
49. Shankar, V., Mannering, F. & Barfield, W. Statistical analysis of accident severity on rural freeways. *Accid. Anal. Prev.* **28**, 391–401. https://doi.org/10.1016/0001-4575(96)00009-7 (1996).
50. Osman, M., Paleti, R., Mishra, S. & Golias, M. M. Analysis of injury severity of large truck crashes in work zones. *Accid. Anal. Prev.* **97**, 261–273 (2016).
51. Huang, H., Chin, H. C. & Haque, M. M. Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. *Accid. Anal. Prev.* **40**, 45–54. https://doi.org/10.1016/j.aap.2007.04.002 (2008).
52. Li, Z. et al. Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical Bayesian models. *Accid. Anal. Prev.* **120**, 139–151 (2018).
53. Yin, J. et al. Fuzzy Bayesian quantitative risk assessment for language and communication induced accidents in maritime operations. *Ocean. Coastal. Manag.* **259**, 107449. https://doi.org/10.1016/j.ocecoaman.2024.107449 (2024).
54. Lee, C. & Abdel-Aty, M. Presence of passengers: Does it increase or reduce driver's crash potential? *Accid. Anal. Prev.* **40**, 1703–1712. https://doi.org/10.1016/j.aap.2008.06.006 (2008).
55. Aguero-Valverde, J. & Jovanis, P. P. Bayesian multivariate poisson lognormal models for crash severity modeling and site ranking. *Transp. Res. Rec.* **2136**, 82–91 (2009).
56. Russo, B. J., Savolainen, P. T., Schneider, I. V. & Anastasopoulos, W. H. Comparison of factors affecting injury severity in angle collisions by fault status using a random parameters bivariate ordered probit model. *Anal. Methods Accid. Res.* **2**, 21–29 (2014).
57. Zeng, Q., Wen, H., Huang, H., Pei, X. & Wong, S. C. A Multivariate random-parameters tobit model for analyzing highway crash rates by injury severity. *Accid. Anal. Prev.* **99**, 184–191 (2017).
58. Esmaili, A., Aghabayk, K. & Shiwakoti, N. Latent class cluster analysis and mixed logit model to investigate pedestrian crash injury severity. *Sustainability* **15**, 185 (2023).
59. Xiao, D., Šarić, Ž., Xu, X. & Yuan, Q. Investigating injury severity of pedestrian–vehicle crashes by integrating latent class cluster analysis and unbalanced panel mixed ordered probit model. *J. Transp. Saf. Secur.* **15**, 83–102 (2023).
60. Karabulut, N. C. & Ozen, M. Exploring driver injury severity using latent class ordered probit model: A case study of Turkey. *KSCE J. Civil Eng.* 1–11. (2023).
61. Almadi, A. I. et al. A fuzzy-logic approach based on driver decision-making behavior modeling and simulation. *Sustainability* **14**, 8874 (2022).
62. Milton, J. C., Shankar, V. N. & Mannering, F. L. Highway accident severities and the mixed logit model: An exploratory empirical analysis. *Accid. Anal. Prev.* **40**, 260–266. https://doi.org/10.1016/j.aap.2007.06.006 (2008).

63. Wang, J., Huang, H., Xu, P., Xie, S. & Wong, S. C. Random oarameter probit models to analyze pedestrian red-light violations and injury severity in Pedestrian–motor vehicle crashes at signalized crossings. *J. Transp. Saf. Secur.* **12**, 818–837 (2020).
64. Ijaz, M. et al. Temporal instability of factors affecting injury severity in Helmet-wearing and non-helmet-wearing motorcycle crashes: A random parameter approach with heterogeneity in means and variances. *Int. J. Environ. Res. Public Health*. **19**, 10526 (2022).
65. Rehman, Z. U., Chaozhe, J., Jamal, A. & Wang, C. Educational disparities shifts in motorcyclist injury severity: Accounting for unobserved heterogeneity. *J. Transp. Saf. Secur.* **17**, 715–756. https://doi.org/10.1080/19439962.2024.2449118 (2025).
66. Eluru, N. & Bhat, C. R. A Joint econometric analysis of seat belt use and crash-related injury severity. *Accid. Anal. Prev.* **39**, 1037–1049. https://doi.org/10.1016/j.aap.2007.02.001 (2007).
67. Huang, H., Siddiqui, C. & Abdel-Aty, M. Indexing crash worthiness and crash aggressivity by vehicle type. *Accid. Anal. Prev.* **43**, 1364–1370. https://doi.org/10.1016/j.aap.2011.02.010 (2011).
68. Li, Z. et al. Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accid. Anal. Prev.* **129**, 230–240 (2019).
69. Kim, J. K., Ulfarsson, G. F., Shankar, V. N. & Mannering, F. L. A Note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accid. Anal. Prev.* **42**, 1751–1758. https://doi.org/10.1016/j.aap.2010.04.016 (2010).
70. Rusli, R., Haque, M. M., Saifuzzaman, M. & King, M. Crash severity along rural mountainous highways in Malaysia: An application of a combined decision tree and logistic regression model. *Traffic Inj. Prev.* **19**, 741–748 (2018).
71. Chang, F., Xu, P., Zhou, H., Chan, A. H. S. & Huang, H. Investigating injury severities of motorcycle riders: A two-step method integrating latent class cluster analysis and random parameters logit model. *Accid. Anal. Prev.* **131**, 316–326. https://doi.org/10.1016/j.aap.2019.07.012 (2019).
72. Pervez, A., Jamal, A. & Khan, S. H. Analyzing injury severity of three-wheeler motorized rickshaws: A correlated random parameters approach with heterogeneity in means. *Accid. Anal. Prev.* **204**, 107651 (2024).
73. Ur Rehman, Z., Chaozhe, J., Adanu, E. K., Almarhabi, Y. & Jamal, A. Factors influencing injury severity in three-wheeled motorized rickshaw and motorcycle collisions. *Sci. Rep.* **15**, 18341. https://doi.org/10.1038/s41598-025-00145-9 (2025).
74. Zhang, P., Wang, C., Feng, C., Easa, S. M. & Jamal, A. Partially constrained temporal modelling injury severity of crashes caused by risk-taking behaviours. *Transportmet. A: Transp. Sci.* 1–50. https://doi.org/10.1080/23249935.2024.2406009 (2024).
75. Meng, F., Xu, P., Wong, S. C., Huang, H. & Li, Y. C. Occupant-level injury severity analyses for taxis in Hong Kong: A Bayesian space-time logistic model. *Accid. Anal. Prev.* **108**, 297–307. https://doi.org/10.1016/j.aap.2017.08.010 (2017).
76. Pervez, A., Jamal, A. & Exploring, E-S. Risk factors based on interpretable machine learning framework. *J. Saf. Res.* **94**, 128–140. https://doi.org/10.1016/j.jsr.2025.06.011 (2025).
77. Rahim, M. A. & Hassan, H. M. A Deep learning based traffic crash severity prediction framework. *Accid. Anal. Prev.* **154**, 106090 (2021).
78. Ma, Z., Mei, G. & Cuomo, S. An Analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. *Accid. Anal. Prev.* **160**, 106322 (2021).
79. Sattar, K. et al. Transparent deep machine learning framework for predicting traffic crash severity. *Neural Comput. Applic*. https://doi.org/10.1007/s00521-022-07769-2 (2022).
80. Rehman, Z. U. et al. Unveiling pedestrian injury severity risk factors: A comparative analysis of machine learning and deep learning methods. *Transp. Eng.* **21**, 100352. https://doi.org/10.1016/j.treng.2025.100352 (2025).
81. Verma, R., Agarwal, M. M. & Advancing traffic safety through caprnn: a hybrid capsule-recurrent neural network for accident forecasting. *Kufa J. Eng.* **16**, 335–352. https://doi.org/10.30572/2018/KJE/160419 (2025).
82. Alkheder, S., Taamneh, M. & Taamneh, S. Severity prediction of traffic accident using an artificial neural network: Traffic accident severity prediction using artificial neural network. *J. Forecast.* **36**, 100–108. https://doi.org/10.1002/for.2425 (2017).
83. Abdelwahab, H. T. & Abdel-Aty, M. A. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. *Transp. Res. Rec.* **1746**, 6–13. https://doi.org/10.3141/1746-02 (2001).
84. Li, Z., Liu, P., Wang, W. & Xu, C. Using support vector machine models for crash injury severity analysis. *Accid. Anal. Prev.* **45**, 478–486 (2012).
85. Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A. & Hadi, M. Improved support vector machine models for work zone crash injury severity prediction and analysis. *Transp. Res. Rec.* **2673**, 680–692 (2019).
86. Wang, W., Liu, C., Chen, D. & Engineering, E. Predicting driver injury severity in freeway rear-end crashes by support vector machine. In *Proceedings of the Proceedings 2011 International Conference on Transportation, Mechanical, and (TMEE)* pp. 1800–1803. (IEEE, 2011).

87. Ghasemzadeh, A. & Ahmed, M. M. *A Probit-Decision Tree Approach to Analyze Effects of Adverse Weather Conditions on Work Zone Crash Severity Using Second Strategic Highway Research* (Program Roadway Information Dataset, 2017).
88. Bernard, J. M. An Application of Decision Tree Models toExamine Motor Vehicle Crash Severity Outcomes. In *Proceedings of the Journal of the Transportation Research Forum* **56**. (2017).
89. Ijaz, M., lan, L., Zahid, M. & Jamal, A. A. Comparative Study of Machine Learning Classifiers for Injury Severity Prediction of Crashes Involving Three-Wheeled Motorized Rickshaw. *Accid. Anal. Prev.* **154**, 106094. https://doi.org/10.1016/j.aap.2021.106094 (2021).
90. Princess, P. J. B., Silas, S. & Rajsingh, E. B. Performance comparison of machine learning models for classification of traffic injury severity from imbalanced accident dataset. In *Intelligence in Big Data Technologies—Beyond the Hype* 361–369. (Springer, 2021).
91. Mokoatle, M., Vukosi Marivate, D. & Michael Esiefarienrhe Bukohwo, P. Predicting road traffic accident severity using accident report data in South Africa. In *Proceedings of the Proceedings of the 20th Annual International Conference on Digital Government Research* pp. 11–17. (2019).
92. Assi, K. Traffic crash severity prediction—A synergy by hybrid principal component analysis and machine learning models. *IJERPH* **17**, 7598. https://doi.org/10.3390/ijerph17207598 (2020).
93. Mansoor, U., Ratrout, N. T., Rahman, S. M. & Assi, K. Crash severity prediction using two-layer ensemble machine learning model for proactive emergency management. *IEEE Access.* **8**, 210750–210762. https://doi.org/10.1109/ACCESS.2020.3040165 (2020).
94. Ouyang, H., Liu, P. & Han, Y. Exploring factors contributing to pedestrian injury severity in Pedestrian–vehicle crashes: An integrated XGBoost–SHAP, Latent cluster, and mixed logit approach. *J. Transp. Eng. Part. A: Syst.* **151**, 04024115. https://doi.org/10.1061/JTEPBS.TEENG-8829 (2025).
95. Wang, C. & Serre, T. A. Hybrid approach to investigating factors associated with crash injury severity: Integrating interpretable machine learning with logit model. *Appl. Sci.* **15**, 10417. https://doi.org/10.3390/app151910417 (2025).
96. Amiri, M. A., Afshari, S. & Soltani, A. Machine learning approaches to traffic accident severity prediction: Addressing class imbalance. *Mach. Learn. Appl.* **22**, 100792. https://doi.org/10.1016/j.mlwa.2025.100792 (2025).
97. Chen, F. et al. Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model. *Sci. Rep.* **15**, 25729. https://doi.org/10.1038/s41598-025-00797-7 (2025).
98. Scarano, A. et al. Cyclist crash severity modeling: A hybrid approach of XGBoost-SHAP and random parameters logit with heterogeneity in means and variances. *J. Saf. Res.* **93**, 373–398. https://doi.org/10.1016/j.jsr.2025.04.003 (2025).
99. Almoshaogeh, M. et al. Traffic accident risk assessment framework for Qassim, Saudi Arabia: Evaluating the impact of speed cameras. *Appl. Sci.* **11**, 6682 (2021).
100. Barrimah, I., Midhet, F. & Sharaf, F. Epidemiology of road traffic Injuries in Qassim Region, Saudi Arabia: Consistency of police and health data. *Int. J. Health Sci.* **6**, 31 (2012).
101. Aldhari, I. et al. Severity prediction of highway crashes in Saudi Arabia using machine learning techniques. *Appl. Sci.* **13**, 233 (2023).
102. Al-Tit, A. A., Ben Dhaou, I., Albejaidi, F. M. & Alshitawi, M. S. Traffic safety factors in the Qassim Region of Saudi Arabia. *SAGE Open.* **10**, 2158244020919500 (2020).
103. Alanazi, F. & Okail, M. A. Predicting Traffic Crash Severity in Jeddah Using Machine Learning Techniques: A Data-Driven Approach to Road Safety. *Transp. Res. Record: J. Transp. Res. Board.* **2679**, 324–343. https://doi.org/10.1177/03611981251337460 (2025).
104. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms 2012.
105. Branco, P., Torgo, L. & Ribeiro, R. P. A Survey of predictive modeling on imbalanced domains. *ACM Comput. Surv.* **49**, 1–50. https://doi.org/10.1145/2907070 (2017).
106. Breiman, L., Random & Forests *Mach. Learn.* **45**, 5–32. (2001).
107. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 1189–1232. (2001).
108. 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*. 785–794. (2016).
109. Lim, S. & Chi, S. Xgboost application on bridge management systems for proactive damage estimation. *Adv. Eng. Inform.* **41**, 100922 (2019).
110. Jiang, J. & Liu, W. Application of XGBoost Algorithm in Manufacturing Quality Prediction. *Intell. Comput. Appl.* **6**, 61–63 (2017).
111. Zhang, L. & Zhan, C. Machine learning in rock facies classification: An application of XGBoost. In *Proceedings of the International Geophysical Conference, Qingdao, China*. Society of Exploration Geophysicists and Chinese Petroleum Society, pp. 1371–1374. (2017).
112. Ullah, I., Liu, K., Yamamoto, T., Al Mamlook, R. E. & Jamal, A. A. Comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability. *Energy & Environ.* 0958305X211044998. (2021).

113. Chen, M. M. & Chen, M. C. Modeling road accident severity with comparisons of logistic regression, decision tree and random forest. *Information* **11**, 270 (2020).
114. Mansoor, U., Jamal, A., Su, J., Sze, N. N. & Chen, A. Investigating the risk factors of motorcycle crash injury severity in Pakistan: Insights and policy recommendations. *Transport. Policy* (2023).
115. Devan, P. & Khare, N. An Efficient XGBoost–DNN-based classification model for network intrusion detection system. *Neural Comput. Appl.* **32**, 12499–12514 (2020).
116. Jamal, A. & Umer, W. Exploring the injury severity risk factors in fatal crashes with neural network. *IJERPH* **17**, 7466. https://doi.org/10.3390/ijerph17207466 (2020).
117. Fiorentini, N. & Losa, M. Handling imbalanced data in road crash severity prediction by machine learning algorithms. *Infrastructures* **5**, 61. https://doi.org/10.3390/infrastructures5070061 (2020).
118. Theofilatos, A. & Yannis, G. A. Review of the effect of traffic and weather characteristics on road safety. *Accid. Anal. Prev.* **72**, 244–256 (2014).
119. Ma, Z., Zhao, W., Steven, I., Chien, J. & Dong, C. Exploring factors contributing to crash injury severity on rural two-lane highways. *J. Saf. Res.* **55**, 171–176 (2015).
120. Razi-Ardakani, H., Mahmoudzadeh, A. & Kermanshah, M. What factors results in having a severe crash? A closer look on distraction-related factors. *Cogent Eng.* **6**, 1708652 (2019).
121. Liu, D., Solomon, D. & Hardy, L. Investigating injury occurrence in motor vehicle collision using artificial neural networks: Investigating injury occurrence in motor vehicle collision. *Hum. Factors Man.* **25**, 294–303. https://doi.org/10.1002/hfm.20543 (2015).
122. Donnell, E. T. & Mason, J. M. Jr Predicting the severity of median-related crashes in Pennsylvania by using logistic regression. *Transp. Res. Rec.* **1897**, 55–63 (2004).
123. Al-Ghamdi, A. S. Using logistic regression to estimate the influence of accident factors on accident severity. *Accid. Anal. Prev.* **34**, 729–741. https://doi.org/10.1016/S0001-4575(01)00073-2 (2002).
124. Chen, C., Zhang, G., Qian, Z., Tarefder, R. A. & Tian, Z. Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accid. Anal. Prev.* **90**, 128–139. https://doi.org/10.1016/j.aap.2016.02.011 (2016).
125. Chang, L. Y. & Chien, J. T. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. *Saf. Sci.* **51**, 17–22 (2013).
126. Mostafa, A. M. et al. AI-based prediction of traffic crash severity for improving road safety and transportation efficiency. *Sci. Rep.* **15**, 27468. https://doi.org/10.1038/s41598-025-10970-7 (2025).
127. Lee, E. J., Kim, S., Lee, H. J. & Jhong, J. H. Identifying policy-relevant traffic crash risk factors in Cheongju, South Korea using logistic regression and explainable machine learning. *PLoS One*. **21**, e0350616. https://doi.org/10.1371/journal.pone.0350616 (2026).
128. Verma, R. & Agarwal, M. M. Enhancing road accident classification with capsule recurrent neural network and improved reptile search algorithm. *IETE J. Res.* **71**, 1822–1841. https://doi.org/10.1080/03772063.2025.2464239 (2025).
129. Verma, R. & Agarwal, M. M. Road accident severity prediction using adaptive custom weight initialization and enhanced focal loss integration technique. *IETE J. Res.* **71**, 2489–2501. https://doi.org/10.1080/03772063.2025.2487628 (2025).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Meshal Almoshaogeh[1] · Fawaz Alharbi[1]

✉ Meshal Almoshaogeh
m.moshaogeh@qu.edu.sa

[1] Department of Civil Engineering, College of Engineering, Qassim University, Buraidah 51452, Saudi Arabia