

Article

Severity Prediction of Highway Crashes in Saudi Arabia Using Machine Learning Techniques

Ibrahim Aldhari ¹, Meshal Almoshaogeh ^{1,*}, Arshad Jamal ², Fawaz Alharbi ¹, Majed Alinizzi ¹
and Husnain Haider ¹

¹ Department of Civil Engineering, College of Engineering, Qassim University, Buraydah 51452, Saudi Arabia

² Transportation and Traffic Engineering Department, College of Engineering, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31451, Saudi Arabia

* Correspondence: m.moshaogeh@qu.edu.sa

Abstract: Kingdom of Saudi Arabia (KSA) is facing alarming traffic safety issues compared to other G-20 countries. Mitigating the burden of traffic accidents has been identified as a primary focus as part of vision 2030 goals. Driver distraction is the primary cause of increased severity traffic accidents in KSA. In this study, three different machine learning-based severity prediction models were developed and implemented for accident data from the Qassim Province, KSA. Traffic accident data for January 2017 to December 2019 assessment period were obtained from the Ministry of Transport and Logistics Services. Three classifiers, two of which are ensemble machine learning methods, namely random forest, XGBoost, and logistic regression, were used for crash injury severity classification. A resampling technique was used to deal with the problem of bias due to data imbalance issue. SHapley Additive exPlanations (SHAP) analysis interpreted and ranked the factors contributing to crash injury. Two forms of modeling were adopted: multi and binary classification. Among the three models, XGBoost achieved the highest classification accuracy (71%), precision (70%), recall (71%), F1-scores (70%), and area curve (AUC) (0.87) of receiver operating characteristic (ROC) curve when used for multi-category classifications. While adopting the target as a binary classification, XGBoost again outperformed the other classifiers with an accuracy of 94% and an AUC of 0.98. The SHAP results from both global and local interpretations illustrated that the accidents classified under property damage only were primarily categorized by their consequences and the number of vehicles involved. The type of road and lighting conditions were among the other influential factors affecting injury severity outcome. The death class was classified with respect to temporal parameters, including month and day of the week, as well as road type. Assessing the factors associated with the severe injuries caused by road traffic accidents will assist policymakers in developing safety mitigation strategies in the Qassim Region and other regions of Saudi Arabia.



Citation: Aldhari, I.; Almoshaogeh, M.; Jamal, A.; Alharbi, F.; Alinizzi, M.; Haider, H. Severity Prediction of Highway Crashes in Saudi Arabia Using Machine Learning Techniques. *Appl. Sci.* **2023**, *13*, 233. <https://doi.org/10.3390/app13010233>

Academic Editors: Filipe Moura and Manuel Marques

Received: 18 November 2022

Revised: 10 December 2022

Accepted: 16 December 2022

Published: 24 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: traffic safety; severity prediction; machine learning; SHapley Additive exPlanations; SHAP; XGBoost; random forest; regression analysis

1. Introduction

Every year, traffic accidents cause a huge number of injuries and casualties worldwide, and the socioeconomic and emotional consequences suffered are enormous. It is important to understand what precedes a road traffic accident (RTA) and the resulting injury severity so that a reduction in road trauma is given top priority. Although traffic volumes declined during the COVID-19 pandemic worldwide, the average annual reduction in the number of deaths due to RTAs was well below the target (50% by 2020) established by the United Nations Decade of Action for Road Safety [1]. As per the latest global road status report on road safety published by World Health Organization (WHO), traffic accidents are responsible for over 1.3 million mortalities worldwide. Motor vehicle accidents are the leading cause of death among teenagers and young people globally [2]. Over three-quarters,

(80%) of all deaths in developing countries are from traffic accidents involving men. The Kingdom of Saudi Arabia (KSA) has faced serious traffic safety issues, particularly since the oil boom in the early 1970s. With over 300,000 road accidents occurring annually [3], the KSA surpasses all the G20 countries [3,4]. The KSA suffers through a loss of SAR 13 billion annually associated with RTAs [4], as 30% of hospital capacity is affected. Therefore, it is critical to understand what precedes a road traffic accident so that a reduction in road trauma prevalence and severity can be achieved with effective interventions.

The top five provinces in Saudi Arabia in terms of RTA frequency are Riyadh, Jeddah, Makkah, Madinah, and Qassim [5]. With the lowest population among these top five provinces, the Traffic Police Department in Qassim recorded more than 18,000 accidents in 2010. Around 23,000 people were involved in these accidents, leading to 2000 people being injured and the deaths of nearly 370 people [5]. There are few studies available in the current literature exploring crash severity in different cities in the KSA. Furthermore, Qassim Province is one of the 13 administrative provinces of the KSA, and has a high road traffic accident rate, similar to other provinces in the Kingdom [5]. Despite the growing traffic safety situation, no comprehensive study has been undertaken in recent times to assess the causes of these accidents and their severity in Qassim Province. To fill this gap, this research aims to develop a model for traffic accident severity in Qassim Province and identify the factors contributing to severity. To model crash injury severity, two ensemble machine learning (ML) methods, random forest and XGBoost, as well as logistic regression are proposed. Further, to overcome the interpretability issue of the ML methods, this study also proposes the application of the Shapley Additive exPlanations (SHAP) analysis for model interpretation and ranking of crash injury severity contributing factors. The findings of the study are expected to provide useful guidance to traffic engineers and safety practitioners for the proactive deployment of policy-related and engineering countermeasures to mitigate the burden of losses caused by RTAs.

2. Related Works

2.1. Crash Injury Severity Categorization

The injury severity of traffic crashes is typically classified as discrete outcome categories, such as fatal injury, severe injury, minor injury, and no injury. Several countries have established their own labeling system; for instance, the well-known system of KABCO developed by the National Safety Council (NSC), and adopted by the Federal Highway Administration (FHWA) [6], states: K = killed, A = incapacitating injuries, B = evident injuries, C = possible injuries, and O = no apparent injuries. The most common types of injury severity fall under three to five classes [7–11]. Severity can also be classified using a binary system [12–14]. Three or more classes represent the degree of severity in descending manner, with death or fatality indicating the worst case, severe or incapacitating referring to the victim being clearly injured, minor or slight indicating the victim is injured but not hospitalized, and the absences of injury or as loss of property as the last severity class. On the other hand, the binary system has also been used for studies with specific objectives, such as the scope of the paper, for instance, fatal/non-fatal, injury/non-injury, serious/minor accident, or property damage only (PDO)injury [14–22].

2.2. Modeling Approaches

Machine learning (ML) has been frequently used for severity prediction over the last decade due to its ability to capture complex relations and produce higher accuracy. The statistical means of severity prediction has shown some limitations, such as low accuracy and lack of reality due to the underlying assumptions [23,24]. ML does not require prior assumptions about the variables or the stochastic process that generates them, thus producing more reliable predictions. A number of ML approaches to severity prediction were implemented at the beginning of the millennium.

2.3. Previous Studies

The first study in the KSA used categorical types of attributes in a decision tree (CHAID and J48) and a probabilistic model, achieving 98% accuracy [25]. However, the classifiers were insufficient to represent the minority groups, i.e., injury and death. Driver distraction was found to be an important cause of injury severities. Two studies in the United Arab Emirates (UAE) used the same data and applied a multilayer perceptron (MLP) and ordered probit, with four classes targeted [26,27]. The overall prediction accuracy of the decision tree (J48) model after resampling the training set was 88.08% and 0.93 area under (AUC) of the ROC curve. The studies found that age between 18 and 30 years, gender (male), and collision type were the most important factors associated with fatal severity. The literature identifies the need for advanced approach applications on severity prediction modeling in the Gulf region.

Breiman et al. [28] developed a popular ML algorithm known as the random forest (RF) classifier, which has been frequently used for predicting the severity of RTAs. The random forest always outperformed the decision trees and other algorithms [20,29–33]. Random Forest is an effective approach for predicting road traffic crash injury severity [34]. It was also adopted by Khan et al. in its latest application in the field of engineering [35]. Rezapour et al. [36] studied motorcycle crashes on two-lane highways with four-level classes of data, and the study relied on the literature to construct a binary category response. They found that the posted speed limit is the most important predictor, followed by age, highway functional class, and speed compliance. However, the authors stated that there was a possibility of bias identification, as the posted speed limit was the only continuous variable. Mokoatle et al. [37] were among the first to apply extreme gradient boosting trees (XGBoost), a relatively new classifier, to severity prediction [38]. The data was imbalanced, so the synthetic minority oversampling technique (SMOTE), a popular technique developed by [39] to handle imbalanced datasets, was used for the minority classes. Despite utilizing SMOTE, the four-class accuracy of the classifiers did not meet expectations, and the multi-class label was switched to binary classes.

Over the last five years, crash severity research has witnessed a revolution in terms of ML techniques due to the development of data science approaches. Rahim and Hassan [40] developed a framework based on a deep neural network (DNN) and applied ImageNet competition to improve the accuracy of the predictions using a stacking technique, a general procedure developed by [41] to enhance predictive accuracy by combining several predictive algorithms. The stacked model outperformed the other classifiers [10,22]. Tang et al. used Shapley Additive exPlanations (SHAP) to interpolate feature importance, which is a recent approach developed by Lundberg et al. [42]. The SHAP value impact on a model gives a high value for collision type, person count, and vehicle count. The authors in [43] adopted a light gradient boosting machine (LightGBM), novel natural gradient boosting (NGBoost), categorical boosting (CatBoost), and AdaBoost to predict two class responses. LightGBM outperformed the others regarding accuracy and ROC curve. Feature importance was determined via SHAP. Injury severity was found to be affected by month, age, cause, and collision type.

Random forest (RF) is an effective approach for predicting RTA injury severity, followed by support vector machine, decision tree, and k-nearest neighbor [34]. However, DNN has been adopted in recent years and its results show promising accuracy [40,44]. Furthermore, the stacking and boosting techniques [45,46] are also accurate and distinguishable from a label perspective. Recently, extreme gradient boosting (XGBoost) outperformed random forest in a study conducted by Jamal et al. [47–50]. Other studies also found XGBoost to be highly reliable for severity prediction [51–53].

Table of the Appendix A summarizes some of the past studies on severity prediction discussed earlier. The table contains detailed information on each research study, including the study theme, study area, duration, and size (i.e., number of crashes) of the study, approaches adopted, best approach, accident categories, and the most significant factors. It may be observed from the table that previous studies have adopted different types of

classification techniques (e.g., KNN, SVM, MLP, and decision trees) for severity prediction. In literature, several studies have also used boosting-based methodologies for severity prediction modeling. Similarly, the application of various statistical modeling techniques (e.g., logistic regression and ordered probit) has also been proposed.

3. Materials and Methods

3.1. Proposed Severity Prediction Framework

Figure 1 presents the severity prediction framework adopted in the present study. The framework starts by defining the accident data for the 2017–2019 period, collected from relevant agencies and authorities, and preprocessed for analysis. Parallel paths for binary and multi-responses were followed. Each path adopted the same steps for data splitting built using 10-fold cross-validation. Injury severity modeling was accomplished via logistic regression (LR) along with two machine ensemble learning algorithms, XGBoost and RF. Once the optimal model with the best performance was identified based on the evaluation metrics, an additive attribute was established using the SHAP approach to determine how important the variables were for injury severity and how each variable contributed to each severity mode. The following passages provide a detailed description of the methodology.

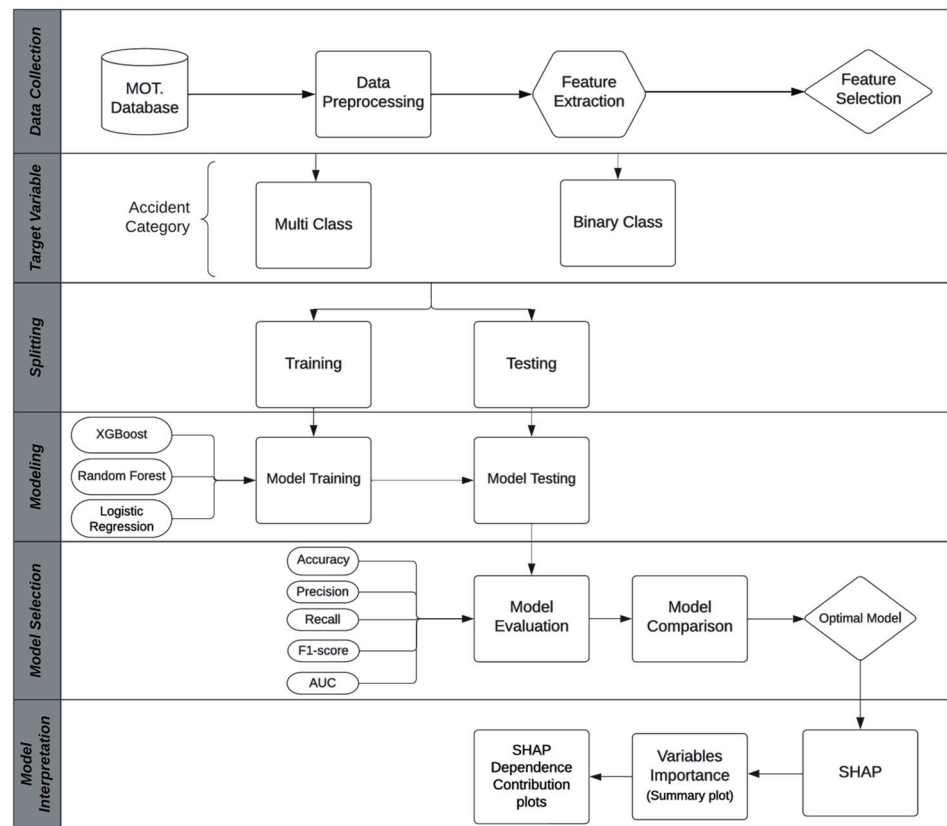


Figure 1. Severity prediction framework.

3.2. Study Area and Collection of Accident Data

Out of 13 provinces in Saudi Arabia, Qassim Province (shown in Figure 2B) was selected as the study area [54]. In geographical terms, Qassim Province lies between 40°000 E and 45°000 E longitude and 23°300 N and 28°000 N latitude, with 10 sub-provinces and 155 localities, divided into 13 governorates, namely Buraidah, Unaizah, AlAsyah, Uyun Al-Jiwa, Al-Badaya’a, Al Bakiriyah, Daria, Al Mithnab, Al Nabhanya, Ar Rass, Riyadh Al-Khabra, Al Shammasiya, and Oklat AlSkoor [55]. There are 1.3 million people living in Qassim, and the province covers 58,046 square kilometers [55]. Approximately 49% of Qassim’s population resides in Buraydah, the province’s capital city. Located in the

middle of the KSA with extensive agricultural activities, it links the north with the south and the west with the east, hence heavy carriers travel on its roads for different agricultural, commercial, religious, and cultural purposes. With 205,000 tons of dates produced annually (both for local consumption and export), the region has been able to enhance its economic value, being one of the largest producers of luxury dates in the Middle East. In addition, various types of fruits and vegetables are extensively produced.

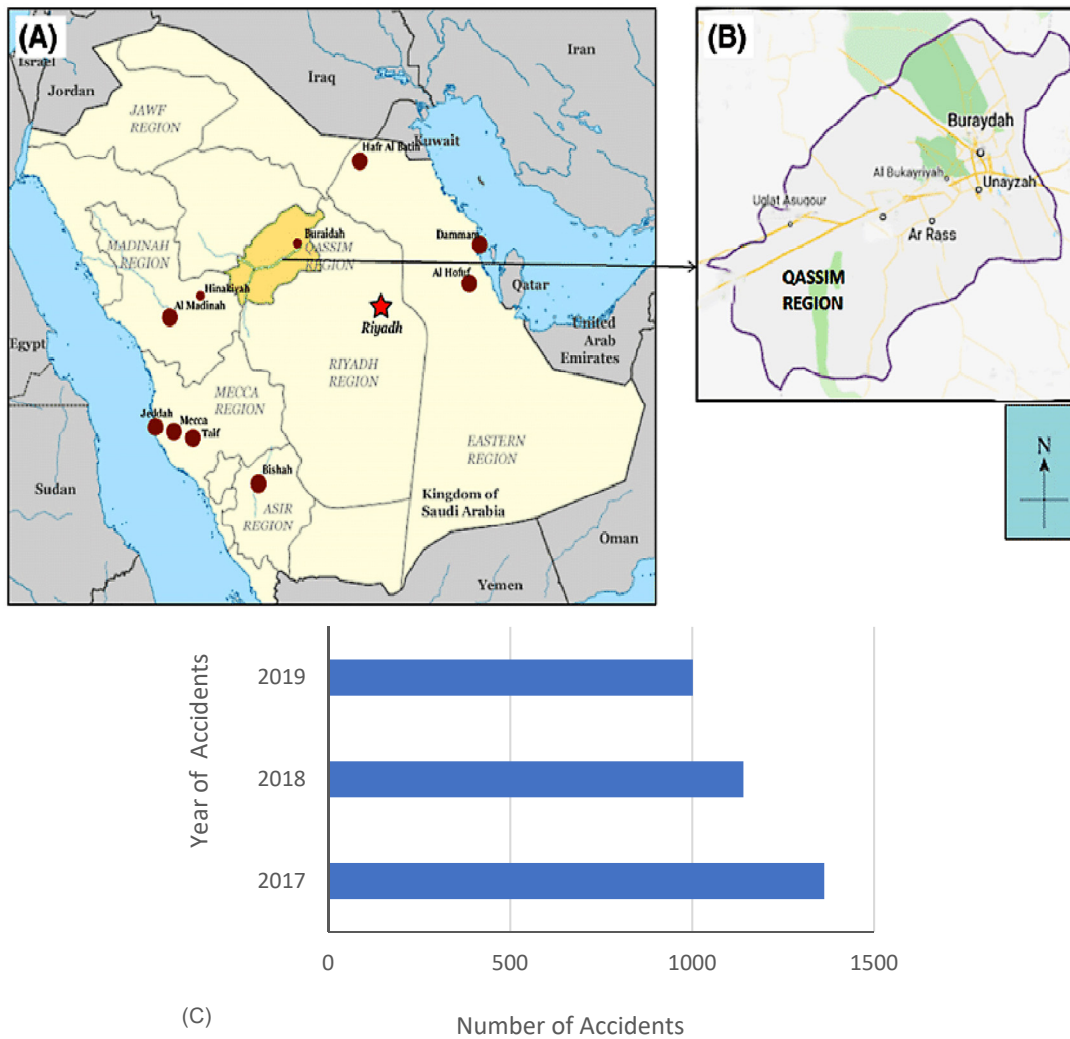


Figure 2. Description of study area: (A) Qassim Province highlighted on a map of Saudi Arabia; (B) study area boundaries; (C) yearly distribution of data collected.

It is worth mentioning that the number of traffic accidents has witnessed a decreasing trend in recent years for many reasons, mainly due to the implementation of mitigation strategies [56]. In 2010, more than 4232 injuries and 1054 deaths were recorded in Qassim [5], reducing periodically until 2016 [57]. Figure 2 illustrates the distribution of RTAs for the given assessment period and clearly shows a continuous reduction in the number of accidents. Al-Atit et al. [57] found that speeding, irregular overtaking, irregular turning, failing to prioritize other vehicles, irregular stops, lack of road readiness, driver carelessness, cell phone usage while driving, and disobeying traffic laws were the most important causes of RTAs. Apart from not wearing a seat belt, these are the ten most common causes of RTAs in Qassim Province.

This analysis is based on traffic crash data collected from the Ministry of Transport and Logistic Services. The data contains information about accidents that have occurred on the main highways under the jurisdiction of Ministry of Transport. During the period of

data collection, between January 2017 and December 2019, 3506 accidents were reported. The data also consists of detailed information on various explanatory variables, as shown in in Table 1.

Table 1. Descriptive statistics of research data explanatory variables *.

Variables	Description	Variable Categories	Frequencies
Dependent Variable			
Accident_categories	Severity of crash	Property damage only (PDO)/Injury/Death	49.2%, 43.3%, 7.5%
Independent Variables			
Temporal Attributes			
Day	Day of the week	Sunday/Monday/Tuesday/ Wednesday/Thursday/ Friday/Saturday	15%, 15%, 15%, 16%, 15%, 11%, 13%
Day_Type	Weekday or weekend	Weekday/Weekend	76%, 24%
Rush_Hour	Time of the day (TOD)	Peak hours/Non-peak hours	59%, 41%
Season	Season of the year	Summer/Autumn/Winter/Spring	62%, 11%, 15%, 12%
Month	Month of the year	January/February/March/ April/May/June/July/ August/September/October /November/December	4.7%, 4.7%, 6.5%, 6%, 6.5%, 5%, 30%, 15%, 5%, 5.7%, 5.4%, 5.5%
Quarter	Quarter of the year	Q1/Q2/Q3/Q4	15.8%, 17.5%, 50.1%, 16.6%
Environmental Factors			
Lighting Condition	Light condition at time of accident	Twilight/Daylight/Darkness/Dusk	3%, 57%, 36%, 4%
Weather condition	Weather status at time of the accident	Good/Rainy/Dusty/Other	95.9%, 1.9%, 1.9%, 0.3%
Roadway Characteristics			
Road_Status	Condition of the road	Good/Road Works/Other	99.5%, 0.1%, 0.4%
Road_Type	Type of road the accident occurred in	Single Carriageway/Dual Carriageways/Highway	19.9%, 28.6%, 51.5%
Geometric_Road_Type	Road geometric characteristics	Straight road/Vertical curve/Horizontal curve/Intersection/U Turn	96.2%, 0.4%, 2.5%, 0.6%, 0.3%
Paints	Existence of road paint marking	True/False	99%, 1%
Eyes	Existing of cat's eyes on road	True/False	99%, 1%
Vehicle Characteristics			
Vehicle_Type	Type of vehicles in accident	Private car/Light truck/Heavy truck/Bus	81%, 5.3%, 13.4%, 0.3%
No_Vehicles_Involved	Number of vehicles in accident	1–17	
Crash Characteristics			
Accident_Type	Collision type	Swerving/Burning/Collision/ Rollover/Rear-end collision/Head-on collision/ Faulty tire/Animal runover/Human runover	12.1%, 3.8%, 50.5%, 24.9%, 6%, 0.5%, 0.1%, 1.7%, 0.4%
ACC_Cause	Cause of accident	Driver/Vehicle/Driver + Vehicle/Road	83.8%, 13.2%, 2.6%, 0.4%
Main_Cause	Root cause of accident	Speeding/Inattention/Tire explosion/Obstacles/Vehicle malfunction/Traffic violation	31.8%, 49.3%, 11.8%, 1.2%, 2.9%, 3%
Damage_Road_Type	Type of damage to accident surroundings	No damage/Flexible barrier/Fixed barrier/Road sign/Road surface/Light post	75%, 18.4%, 2%, 2.5%, 0.6%, 1.5%
Consequence	What are the consequences of that accident?	Vehicle/Vehicles damage/Vehicle and infrastructure damage/Rollover/Run off the road/Animal runover/Human runover	40.3%, 11.7%, 36.9%, 9.2%, 1.6%, 0.3%

* All the explanatory variables are categorical type, except "No_Vehicles_Involved".

3.3. Methods

3.3.1. Platform

The most popular options for classification problems are the Python analytic platform and Waikato Environment for Knowledge Analysis (WEKA). Python is a well-known programming language due to its flexibility and it is an open-source platform [58]. WEKA is a free software developed at the University of Waikato, New Zealand, which is a companion software to the book *Data Mining: Practical Machine Learning Tools and Techniques* [59]. WEKA is being used extensively for severity predictions [7,11,26,60–62]. In 2017, Mitranont et al. [63] compared Python with other machine-learning algorithms and found Python to be the best performer in terms of precision, recall, and correct/incorrect instances. Hence, Python was adopted in this research.

3.3.2. Response Process

The dataset is considered imbalanced if one label/class of the data appears significantly more often than others. Imbalanced classifications can be slight or severe; for example, the difference ratio between PDO, injury, and death was found to be 1726:1518:262 in the present research. Figure 3 illustrates death as a minority classification, while PDO and injury are majority classifications. The ML models face difficulty with regard to extracting accurate information from imbalanced data and producing biased results. There are various methods to overcome this problem, for instance, resampling techniques and the synthetic minority oversampling technique (SMOTE).

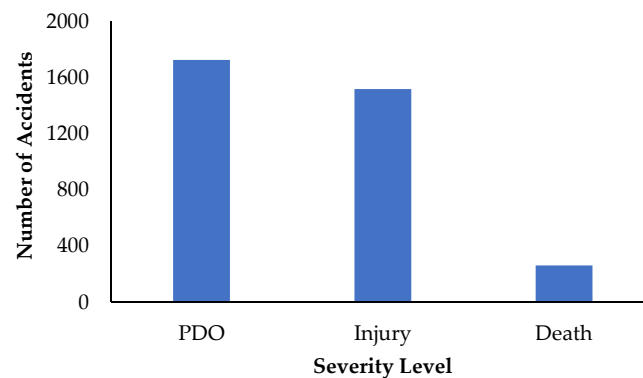


Figure 3. Number of accidents under each injury severity classification for highway crashes in Qassim Region between January 2017 and December 2019.

Resampling can be done through either over-sampling or under-sampling. While performing under-sampling, the run time and storage problems can be improved by reducing the amount of training data, but this might result in a loss of potentially important information. Over-sampling, however, does not lead to information loss but does increase the likelihood of overfitting since it replicates minority class events. Using SMOTE, overfitting is mitigated as synthetic examples are generated instead of replicating instances. In addition, no useful information is lost.

SMOTE, however, does not consider neighboring examples from other classes when generating synthetic examples. Consequently, classes can overlap, and noise can be introduced. Furthermore, SMOTE does not support categorical data in Python yet; even if categorical data are transformed into numerical; SMOTE would generate data that did not make sense. Therefore, a resampling technique with a typical random over-sampling was adopted in this research to overcome the ML bias problem.

3.3.3. Logistic Regression

The logistic regression (LR) or logit model is commonly used for classification and prediction. The method creates probabilities between one and zero. Using a given dataset of independent variables, LR estimates the probability of the occurrence of an event. In this

study, LR predicted the probability of injury, death, and PDO with three possible outcomes: y_1 = property damage; y_2 = injury, and y_3 = death (coded as $y_1 = 0$; $y_2 = 1$; $y_3 = 2$). For LR, the modeling function described the relationship between a class’s probability and the set of independent or predictor variables. For the current problem, a typical LR model equation may be expressed as:

$$P(y = 1|x) = \frac{1}{1 + e^{(-z)}} = \frac{e^{(z)}}{1 + e^{(z)}} \in [0, 1] \tag{1}$$

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n = x\beta \tag{2}$$

where $x\beta$ represents the sigmoid S-shaped function. In the event that the probability is greater than 0.5, the dataset is classified as a death, an injury, or property damage. The parameters included in LR were the number of iterations, epsilon, learning rate strategy, step size, and regularization. Moreover, the learning rate strategy and regularization were considered fixed and uniform.

3.3.4. Extreme Gradient Boosting (XGBoost)

XGBoost was used to solve problems requiring supervised learning, where multiple features were used to predict the target variable from the training data. XGBoost is based on the gradient boosting technique, which is a term coined by [64], in a paper titled “Greedy Function Approximation: A Gradient Boosting Machine”. The gradient-boosting decision tree (GBDT) algorithm was improved by Chen and Guestrin in 2016, resulting in the XGBoost algorithm [38]. This method helps models to predict more accurately, reduce computational efforts, and avoid overfitting issues. In addition, it has been reported that XGBoost demonstrates better predictive performance compared with traditional machine learning algorithms. XGBoost provides a parallel tree boosting and is a leading machine-learning library for regression, classification, and ranking. XGBoost is an ensemble algorithm, which refers to the use of multiple learning algorithms to achieve better predictive performance than any single algorithm.

The general form of the XGBoost objective function is presented in Equation (3), where l is the training loss function, and Ω is the regularization term. It uses the training data (with multiple features) x_i to predict a target variable y_i , and \hat{y}_i is the prediction; k refers to the number of trees; and f_k represents a function in the functional space. However, the XGBoost objective function at iteration t needs to be minimized using Equation (4):

$$Objective\ Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \tag{3}$$

$$obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) + \Omega(f_t) \tag{4}$$

In the regularization term in Equation (5), XGBoost uses two parameters to define the complexity of a tree, . . . as the number of leaves and L_2 as the norm of leaf scores, while γ and λ are hyperparameters. The overall complexity is defined as the sum of the number of leaves weighted by γ and a λ weighted L_2 norm of the leaf score-leaf weight:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \tag{5}$$

The second part is the training loss, which can be represented as an additive model since it is under the boosting space as given in Equation (6). This is the final model \hat{y} of t composed of the previous model, which is $\hat{y}_i^{(t-1)}$ plus the new model that we want to learn:

$$Final\ model\ (t) = \hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{6}$$

To use traditional optimization techniques, the objective function must be transformed into a Euclidean domain via Taylor approximation using Equation (7). To simplify the objective function’s training loss, Taylor’s equation is applied on the two terms in Equation (8), which

will introduce two new terms g_i and h_i in Equation (9), representing the values of each leaf in the tree. Making the original function in quadratic form:

$$Taylor\ approximation = f(x + \Delta x) \simeq f(x) + f'(x)\Delta x + \frac{1}{2}f''(x)\Delta x^2 \tag{7}$$

XGBoost Simplified with Taylor Approximation:

$$obj^{(t)} \simeq \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{8}$$

$$\begin{aligned} g_i &= \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \\ h_i &= \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) \end{aligned} \tag{9}$$

where the g_i term indicates the first-order differentiation, and h_i is the second order differentiation for approximating the function. The new form is when it is put together, i.e., training loss and regularization, and it is simplified in Equation (10) by using G and H to refer to the summation of $g_{(s)}$ and $h_{(s)}$, respectively. G and H now represent the entire tree structure for this formulation:

XGBoost Objective Function Representing Trees:

$$obj^{(t)} = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \tag{10}$$

where, $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$.

The equation above is now becoming sum T quadratic equations. For each quadratic function, it is now possible to drive optimal weight which is w_j the purpose to drive its simplified, as in Equation (11):

XGBoost Objective Function Optimal Weight:

$$Optimal\ weight = w_j^* = -\frac{G_j}{H_j + \lambda} \tag{11}$$

The final step is to substitute w_j back into the original objective function. At this point, a new objective function without any $w_{(s)}$ is Equation (12). This objective function is known as the minimum objective function, and this is the most simplified form of quadratic approximation of the original objective function [65]:

XGBoost Minimum Objective Function:

$$MinObj = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \tag{12}$$

3.3.5. Random Forest

To implement the RF utilizing decision trees, three steps must be followed. Start by generating an NC size bootstrap sample from the overall N data to grow a tree by randomly selecting predictors $X = (x_i, i = 1, \dots, p)$. Then use the predictors' X_i at different nodes n of the tree to vote for the class label y for the same node. Further adjustments are made at each node before the best predictor for the split is obtained. As a last step, the out-of-bag (OOB) data (N–Nc) are run down the tree to determine the misclassification error. Repeat these procedures until the minimum out-of-bag error rate (OOBER) is reached for a large number of trees. Averaging the series of trees allows each observation to be assigned to a final class 'y' by majority vote. Moreover, this method uses the information gain ratio as its split criterion, which may be calculated from Equation (13):

$$Information\ Gain\ Ration\ (IGR) = \frac{Information\ Gain\ (x)}{Split\ info\ (x)} \tag{13}$$

where x represents the randomly chosen example in the training set. In short, split info refers to information needed to determine which branch an instance or example belongs to.

3.3.6. Hyperparameter Tuning

Hyperparameter tuning plays a crucial role in the outcome of ML algorithms by identifying the optimal parameters for an algorithm. A hyperparameter is a set of values tuned before the learning process begins to reduce model complexity, improve generalization performance, and avoid overfitting. Iteratively traversing the entire space of available hyperparameter values is time-consuming for larger parameter spaces. It is possible to tune hyperparameters using several techniques, such as grid search and random search. Grid search was used to tune the hyperparameters and the performance metrics based on the model’s classification accuracy. Table 2 lists the hyperparameter range used for XGBoost. Several parameters were tuned, and the accuracy of the model was found to be affected by them. Table 2 also includes the RF and LR hyperparameter range. The parameter values as stated, were tuned for optimal accuracy value via grid search.

Table 2. Hyperparameter optimization off ML models.

XGBoost Hyperparameter Optimization		
Parameter	Typical Value	Values
learning_rate	0.01 to 0.2	0.05, 0.10, 0.15, 0.20, 0.25, 0.30
max_depth	3 to 10	3, 4, 5, 6, 8, 10, 12, 15
min_child_weight	N/A	1, 3, 5, 7
gamma	N/A	0.0, 0.1, 0.2, 0.3, 0.4
n_estimators	N/A	10, 50, 100, 150, 200, 250, 300
colsample_bytree	0.5–0.9	0.3, 0.4, 0.5, 0.7
Random Forest Hyperparameter Optimization		
max_depth	3 to 10	3, 4, 5, 6, 8, 10, 12, 15
min_samples_leaf		1, 5, 10
min_samples_split	N/A	2, 4, 10, 12, 16
n_estimators		10, 50, 100, 150, 200, 250, 300
criterion	Gini or Entropy	Gini or Entropy
Logistic Regression Hyperparameter Optimization		
c_values		100, 10, 1.0, 0.1, 0.01
penalty	N/A	l1 and l2
solvers		Newton-cg, lbfgs, or liblinear

3.3.7. Model Interpretation

SHapley Additive exPlanations (SHAP) is a method to explain the individual predictions of any machine learning model. SHAP is based on the game theoretically optimal Shapley values [66]. SHAP aims to explain the prediction of instance x by computing the contribution of each feature [67]. It is frequently adopted due to its durability, solid theoretical foundation, fair distribution of prediction among features, fast implementation for tree-based models, and global model interpretations. It includes feature importance, feature dependence, interactions, clustering, and summary plots.

Equation (14) generates the Shapley value for a feature i . For a specific feature value (float number), the inputs for the blackbox model f , and the input datapoint x , this datapoint will be a singular row in the dataset. Iterate the overall possible subset prime combination of features to ensure that the interaction between individual feature values is

accounted for. One of the subsets could be any other feature, and the remaining are treated as unknown values. By this, consideration is left only to the selected feature:

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z'/i)] \tag{14}$$

where $\phi_i(f, x)$ is the Shapley value for a feature i , f corresponds to the blackbox model, x refers to the input data point, z' represents the subset feature, x' stands for the simplified data input, and M denotes the total number of features.

The core step is to get the blackbox model f output for the subset with $f_x(z')$ and without $f_x(z'/i)$ for the feature desired (i.e., interested in i). The difference in these two terms reveals how much i contributes to the prediction in the subset known as the marginal value. Repeating these for all the possible combinations, each permutation of subsets for each of those is additionally weighted according to how many players in that collation.

3.3.8. Model Evaluation

In modeling a classification problem, the metrics to test the model’s performance are the same for statistical and machine learning. Figure 4 presents the four types of outcomes that can occur when performing classification predictions.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figure 4. Typical confusion matrix scenario for classification problems.

When an observation is predicted to belong to a class and does belong to that class, it is considered a true positive (TP). True negatives (TN) occur when a prediction of an observation does not belong to a class, and it actually does not belong to one. A false positive (FP) occurs when an observation is predicted to belong to a class when it actually does not. In false negative (FN) situations, the prediction of an observation does not belong to a particular class, but in reality, it does. Figure 4 presents a confusion matrix with a commonly used binary classification system. The true positive ratio (TPR) quantifies the proportion of positives that are correctly identified, while the false positive ratio (FPR), using Equation (15), quantifies the proportion of negatives that are incorrectly classified as positives:

$$False\ Positive\ Ratio\ (FPR) = \frac{FP}{FP + TN} \tag{15}$$

Accuracy, precision, F1-score, and recall were used to evaluate the performance of the classification model. Accuracy is the percentage of correct predictions for the test data using the following equation:

$$Accuracy = \frac{Coorect\ predictions}{All\ predictions} = \frac{TP + TN}{TP + FN + TN + FP} \tag{16}$$

The precision is the percentage of true positive examples among all the examples predicted to belong to a particular class using Equation (17). Low precision indicates a high number of false positives (FPs):

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

The sensitivity/recall of an algorithm determines its completeness using Equation (18). A low recall indicates the presence of a large number of FNs:

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

By taking the harmonic mean of a classifier's precision and recall, the F1-score is calculated. The F1-score comes into play when one classifier has a high recall, while the other has high precision. The F1-scores of both classifiers can be used in this case to determine which classifier produces better results via Equation (19):

$$F1 - score = \frac{2(Precision \times Recall)}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (19)$$

Moreover, the area under the receiver operating characteristic (AUC-ROC) curve is used for evaluating the model performance and is one of the most important metrics, indicating how well the model distinguishes between injury classes. If the AUC is high, the model is more accurate in predicting severity classes to label injury as injury and death as death. The ROC curve is plotted with TPR on the y-axis and FPR on the x-axis.

4. Results and Discussion

4.1. Multi-Classification

We investigated the model performance prediction using the evaluation metrics of accuracy, precision, recall, F1-score, and AUC. It is worth mentioning that non-resampling models were employed, and their performance was poor and biased toward the majority classes, as predicted earlier. The combined classifiers could not achieve more than 0.57 accuracy. Even if the accuracy was ignored, the F1 the score was also very poor. Therefore, that eliminated the need to check the recall and precision scores.

The importance of balancing the data is illustrated in Figure 5. Clearly, the ensemble models were capable of learning and predicting, not only in terms of accuracy but also in terms of sensitivity and precision (Table 3). It should be noted that the results demonstrated in Table 3 are the optimal models via a grid search process and by taking their macro average. Apart from LR, the other two models performed spectacularly in Figure 5. RF and XGBoost were able to distinguish between the classes by seeing the AUC reach 0.87. Furthermore, the XGBoost precision, recall, and F1-score were 0.7, 0.71, and 0.71, respectively. XGBoost gained an edge over RF in all evaluation metrics. In addition, RF had a low observation to be predicted to belong to a class when it actually did not (i.e., FP). These finding supports the claim of the ability of RF to generate desirable outcomes with accuracies of 0.69 and 0.87 AUC.

Table 3. Models' multi-class performances.

Multi-Classes					
Classifier	Accuracy	Precision	Recall	F1-Score	AUC
XGBoost	0.71	0.7	0.71	0.7	0.87
Random Forest	0.69	0.68	0.69	0.68	0.87
Logistic Regression	0.43	0.43	0.43	0.43	0.62

The confusion matrix of both RF and XGBoost in Figure 5a,b were analyzed along with their ROC curves. Previously, the ensembled algorithm was not able to identity the minority classes, while in Figure 5a,b, it is vice versa. From their confusion matrices in Figure 5a, XGBoost was better able to correctly classify the minority classes than RF. The ROC curves complete the story by showing the class curves, and that the minority class area is much greater than the majority class area. Moreover, the outcomes are within other researchers' range of outcomes [10,43,62,68,69].

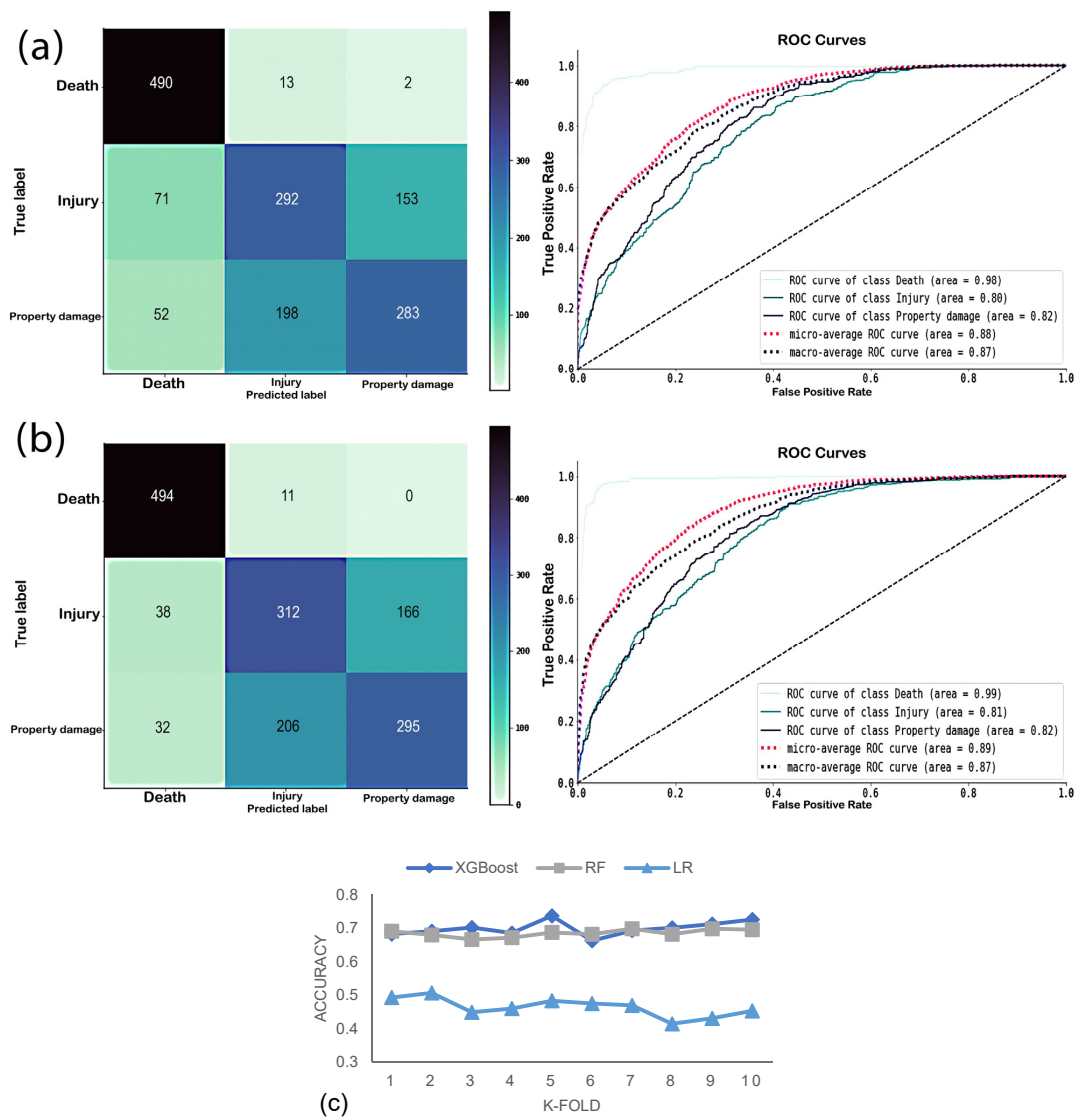


Figure 5. Multi-class model confusion matrix and ROC curves: (a) XGBoost model; (b) random forest model; (c) classifier accuracies by 10-fold multi-class.

4.2. Binary Classification

The outcome of the previous models’ performance is more than satisfying if it is compared with previous studies in the literature. The Pillajo et al. [69] approach achieved more than 70% accuracy, with 77% precision, an AUC of 0.8. was obtained by Tang et al. [10]; and an accuracy and ROC curve value of 0.73 and 0.71, respectively, was attained by Dong et al. [43]. Further, some studies treated the target as a binary classification [37,70]. To maximize the potential of the dataset, the present study adopted the same approach by merging PDO and injury in a non-fatal class and using the minority class as the fatal class. By merging them, the accuracy was increased; however, the ability to distinguish the class remained an issue. Table 4 shows the performance of the three models when they were balanced to prevent the misclassification of minority classes.

Table 4. Models’ binary-class performances.

Classifier	Binary-Classes				
	Accuracy	Precision	Recall	F1-Score	AUC
XGBoost	0.94	0.94	0.94	0.94	0.98
Random Forest	0.9	0.91	0.9	0.9	0.97
Logistic Regression	0.65	0.65	0.65	0.65	0.70

In Table 4 and Figure 6, the accuracies are self-explanatory in their increments. Even RF and LR generated a high precision of 0.91 and 0.65, respectively. XGBoost improved in terms of accuracy, precision, recall, and F1-score (all 0.94), making its prediction reliable and truly representative of the minority class (Figure 6). The XGBoost area under the ROC curve was 0.98, which indicates its ability to classify the injury severity classes correctly. Meanwhile, RF precision, recall, and F1-score improved significantly with 0.91, 0.9, and 0.9, respectively. Moreover, the TP and low FP of the confusion matrix enhanced the AUC to achieve 0.97 (Figure 6a). These outcomes also comply with those of other studies [11,25,70–75]. The average fold accuracy of LR is not competitive (Figure 6b), as in the previous situation; however, the evaluation metrics further improved.

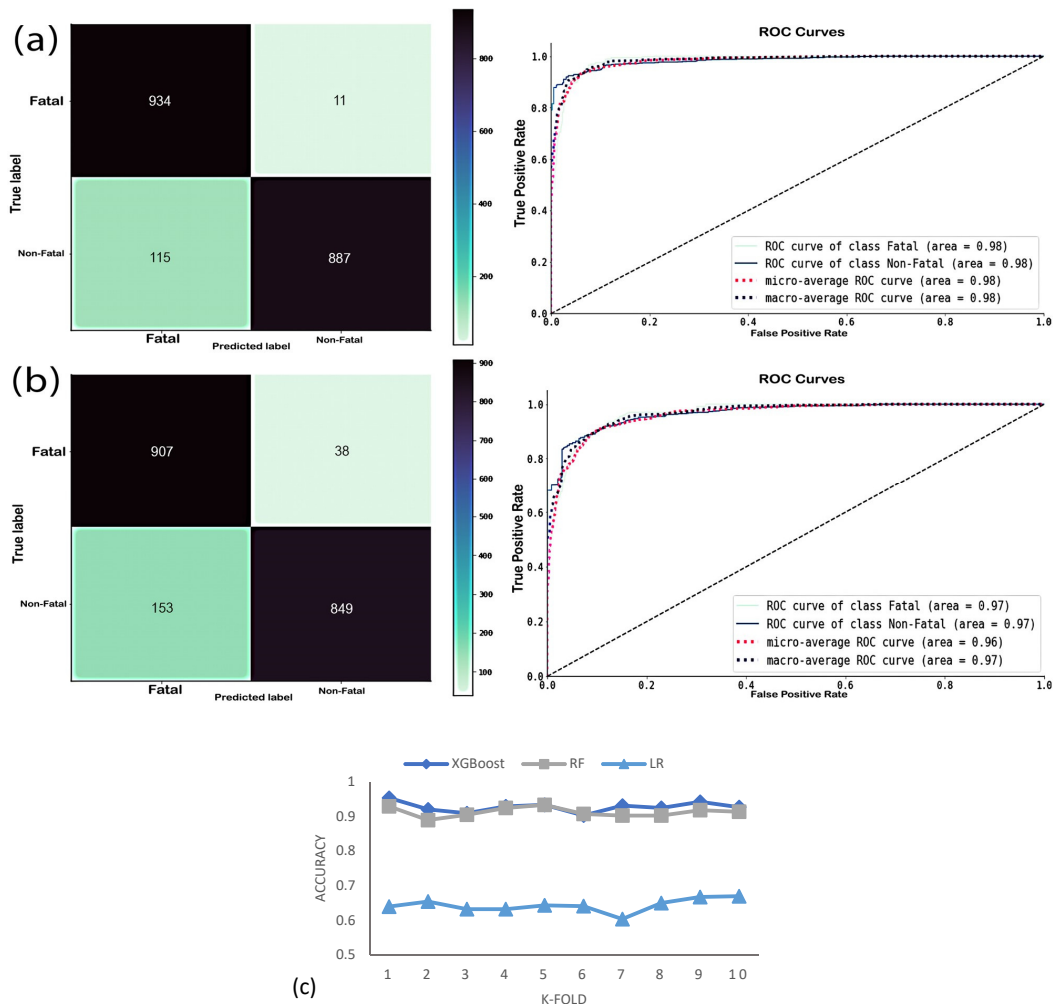


Figure 6. Binary-class model confusion matrix and ROC curves: (a) XGBoost model; (b) random forest model; (c) classifier accuracies by 10-fold binary-class.

4.3. Model Interpretations

An enhanced model analysis was conducted through a SHAP summary assessment. The calculated quantitative value that sums the Shapley values by each severity class and

represents the contribution of each variable to the model can be seen in Figure 7. The input variables are arranged vertically according to their influence, beginning with the most influential variable. In this figure, the horizontal axis represents the SHAP value, and the color scales indicate the significance level of each variable, with blue indicating low significance and pinkish-red indicating high significance. Data points within a range of SHAP values with more data points show a stronger correlation between input variables and injury severity. Figure 7a is the summary plot of the PDO class, which identifies that PDO accidents were mainly classified depending on the consequences and the number of vehicles involved. Moreover, the main cause of the accident is a factor in the higher severity level. Figure 7b identifies the most important parameters for classifying the injury class: main cause followed by road type and lighting condition. The most severe class (i.e., death class) is affected mostly by temporal parameters: month and day of the week, and, again, road type (Figure 7c). Based on the summary plots in Figure 7, it can be seen that a feature’s value directly correlates with its impact on prediction. However, only SHAP dependence plots can reveal the exact shape of the relationships.

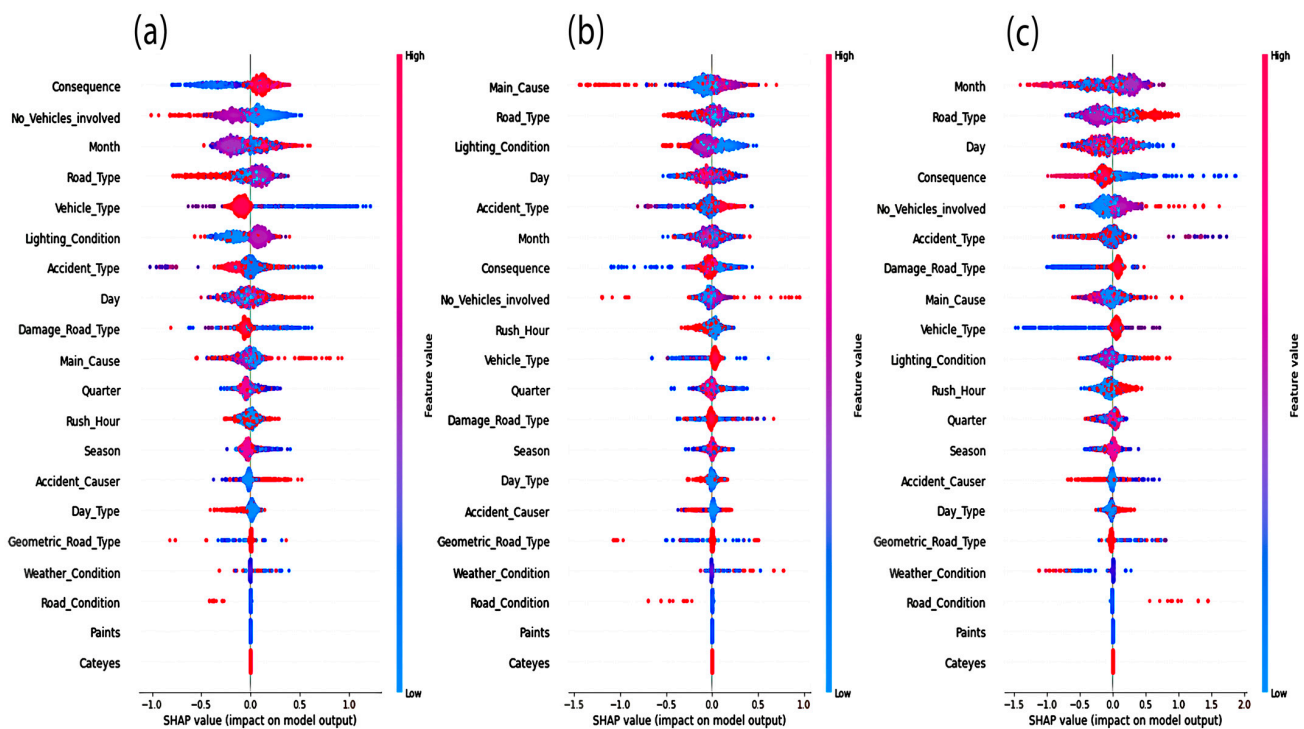


Figure 7. SHAP plots for variable importance per class: (a) PDO class; (b) injury class; (c) death class.

4.4. Dependence Plot

A SHAP dependence contribution plot provides similar insights to a partial dependence plot (PDP) but yields much more information. Dots represent the data rows. The horizontal location indicates the actual value in the dataset, while the vertical location indicates how it affected the prediction. In the interaction analysis, trends were examined in terms of month of the year, day of the week, consequences, main causes of accidents, lighting conditions, accident types, and vehicle types. It is also possible, however, to evaluate other variable interactions.

The SHAP value in Figure 8a, lighting condition, is greater when PDO accidents occur in daylight and at dusk. It can be seen that the vehicles involved were private cars and light trucks, while the heavy trucks were mostly quarantined during the peak hours, occupying the road areas far less than at dusk. When a vehicle malfunctions, it is more likely to interact with other vehicles or infrastructure. This type of impact positively favors PDO severity (Figure 8b). In Figure 8c, it can be seen that rear-end collision, rollover, and swerving were

positively categorized as injury severity. For instance, in rear-end collisions, there is a higher risk that the driver will suffer a spinal cord injury, and such a scenario is positively classified as injury severity.

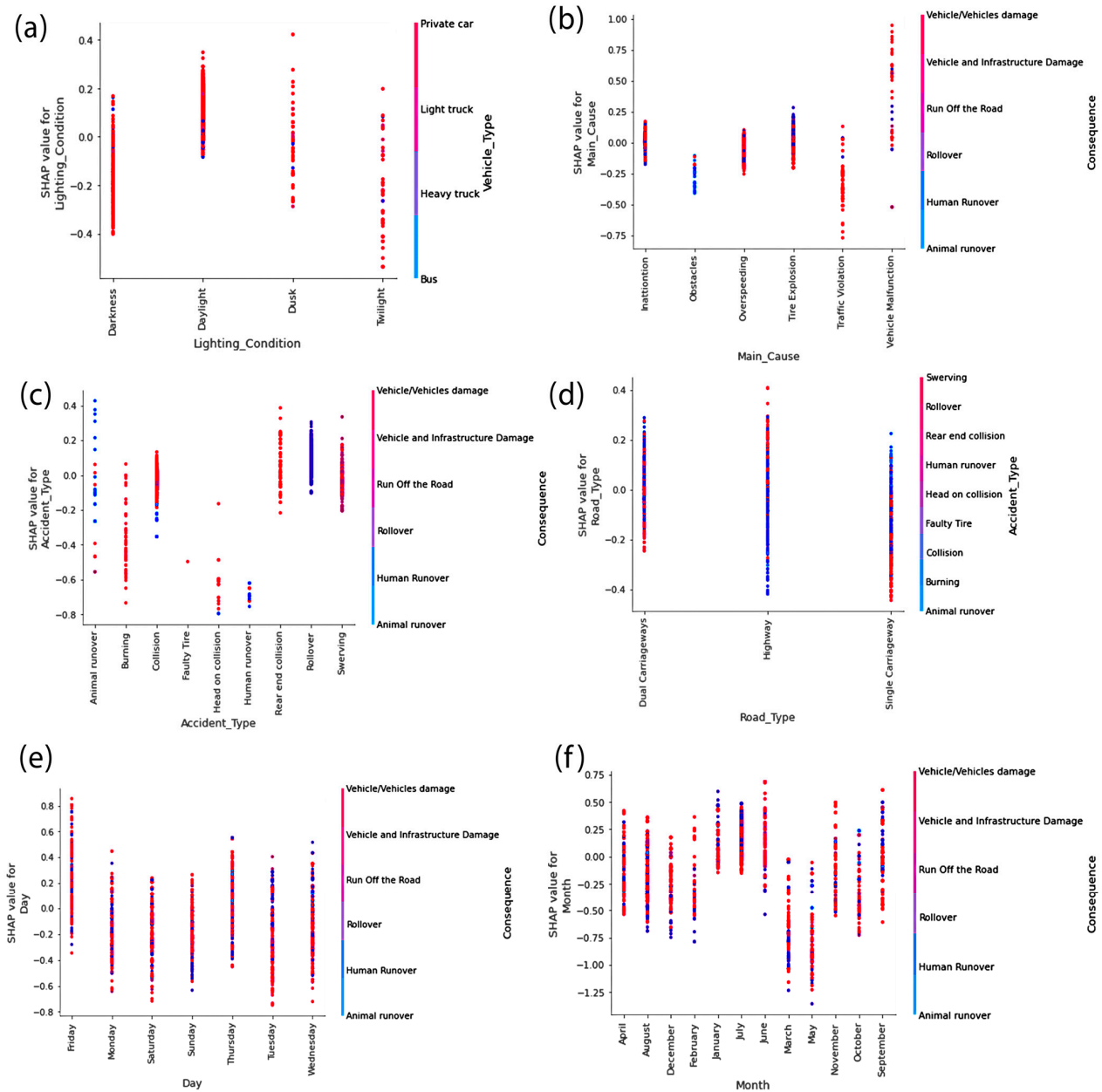


Figure 8. SHAP dependence plot for variable importance per class: (a,b) PDO class; (c,d) injury class; (e,f) death class.

The national weekend in Saudi Arabia falls on Friday and Saturday, and as the weekend approaches, accidents are more likely to cause death (Figure 8e). As head-on collisions, burns, and human runovers have a higher likelihood of resulting in fatal injuries, they do not positively qualify as injury severity. The national summer holiday in Saudi Arabia is from June to late August, so in Figure 8f, it can be seen that June and July had the highest SHAP value. On the other hand, in March and May, death severity was found to be less likely.

5. Conclusions

Traffic accidents cause a large number of fatalities and enormous socioeconomic consequences every year in KSA. The current study proposes an application of machine learning-based algorithms and SHAP for modeling and interpreting traffic accident injury severity in the Al-Qassim province of KSA. Using two ensemble machine learning methods, random forest and XGBoost, as well as logistic regression, revealed that XGBoost outperformed the others by achieving an accuracy of 0.71 and an AUC of 0.87. Further analysis conducted by treating the target variable as a binary classification revealed that XGBoost achieved an accuracy of 0.94 and distinguished the class with an AUC of 0.98. SHapley Additive exPlanations can effectively extract the feature importance and explore the relationships with the help of dependence plots.

Results revealed that the following factors primarily impact the accident injury severity: month of the year, road type, consequences of the accident, and day of the week. Furthermore, the presence of paint, cateyes and road conditions have the least impact on the severity of the injury. A precise analysis for each injury severity class revealed that PDO accidents mainly depend on the consequences and the number of vehicles involved. The main cause of accidents is the most important parameter for classifying an injury class, followed by road type and lighting conditions. For the death class, month of the year, day of the week, and road type were the most significant variables.

The nature of the available data resulted in the present study having some limitations. The explanatory variables were categorical, except the number of vehicles involved. There was also an absence of some important variables, such as the physiology of the driver, age, and gender. The authors suggest the adoption of XGBoost and SHapley Additive exPlanations in injury severity classification studies and recommend investigating the implementation of DNN or hybrid neural networks in future research.

Author Contributions: I.A. was involved in data analysis, methodology, software, and manuscript writing. M.A. (Meshal Almoshaogeh) was involved in conceptualization, data collection, supervision, and paper review. H.H. was involved in conceptualization, supervision, and manuscript writing. F.A., M.A. (Majed Alinizzi) and A.J. were involved in conceptualization and paper review. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Highly confidential data that cannot be publicly shared due to the confidentiality agreement between the research and data sharing organization.

Acknowledgments: The researchers would like to thank the Deanship of Scientific Research, Qassim University for funding the publication of this project. The authors also acknowledge the Ministry of Transport for data sharing.

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

Appendix A. Summary of Past Studies on Injury Severity Prediction

No.	Year	Country	Duration	Size	Injury Severity Classes	Approach	Best Approach	Significant Factors	Reference
1	2022	Pakistan	2015–2019	1784	- Fatal - Non-Fatal	- Naive Bayes (NB) - K-Nearest Neighbor (K-NN) - Binary Logistic Regression (BLR) - Random Forest (RF) - Extreme Gradient Boosting (XGBoost)	- XGBoost	- Vehicle type - Month of the year - Driver’s age - Alignment of the road	S. Zhang, Khattak, Matara, Hussain, & Farooq (2022) [55]
2	2022	China	2018	567	- Property Loss - Fatal - Injury	- XGBoost - Bayesian network	- Bayesian Network	- Roadside protection facilities - Roadway type - Central isolation facilities - Light condition - Crash time - Weather conditions	Yang, Wang, Yuan, & Liu (2022) [75]
3	2021	Saudi Arabia	Jan 2017–Dec 2019	13,546	- Fatal - Injury - PDO	- XGBoost - Logistic regression - Random Forest - Decision Tree	- XGBoost	- Collision type - Weather condition - Road surface - On-site damage - Vehicle type - No. of lanes - Crash cause	Jamal et al. (2021) [47]
4	2021	US	2004–2021	204,758	- PDO - Injury	- Gradient Boosting (GB) - AdaBoost - MLP - Stacking Ensemble	- Stacking Ensemble	- Collision type - Person count - Vehicle count	Niyogisubizo et al. (2021) [19]

No.	Year	Country	Duration	Size	Injury Severity Classes	Approach	Best Approach	Significant Factors	Reference
5	2020	US	2010–2018	8859	<ul style="list-style-type: none"> - Suspected Serious Injury - Non-Incapacitating Injury - Possible Injury - Killed 	<ul style="list-style-type: none"> - Random Forest - XGBoost 	<ul style="list-style-type: none"> - XGBoost 	<ul style="list-style-type: none"> - Road class - Speed limit - First harmful event 	Lin, Wu, Liu, Xia, & Bhattarai (2020) [51]
6	2020	India	2016–2018	7654	<ul style="list-style-type: none"> - No Injury - Simple Injury - Grievous Injury - Fatal Injury 	<ul style="list-style-type: none"> - Ordered Probit - Random Forest - Conditional Inference Forest 	<ul style="list-style-type: none"> - Conditional Inference Forest 	<ul style="list-style-type: none"> - Type of colliding vehicle - Collision type - Driver age - Visibility of the road. 	Panicker & Ramadurai (2022) [9]
7	2019	US	2017	201,581	<ul style="list-style-type: none"> - Property Damage - Injury - Fatal 	<ul style="list-style-type: none"> - Multinomial Logit - Random Forest 	<ul style="list-style-type: none"> - Random Forest 	<ul style="list-style-type: none"> - Collision type - Occupant age - Speed limit 	Wang & Kim (2019) [32]
8	2019	South Africa	2015–2017	1525	<ul style="list-style-type: none"> - No Injury - Serious - Slight - Killed 	<ul style="list-style-type: none"> - Multivariate Logistic Regression - XGBoost 	<ul style="list-style-type: none"> - XGBoost 	<ul style="list-style-type: none"> - Truck license - Light motor duty license - Vehicle type 	Mokoatle et al. (2019) [37]

No.	Year	Country	Duration	Size	Injury Severity Classes	Approach	Best Approach	Significant Factors	Reference
9	2018	US	2008–2012	32,730	- No Injury - Minor Injury - Severe Injury	- Decision Tree C4.5 - Nearest-Neighbor - Instance-Based - Random Forest	- Random Forest	- Age/gender groups - Area type - Median width - Crash contributing factors	Mafi et al. (2018) [30]
10	2018	US	2012–2015	15,164	- No Injury - Non-Incapacitating Injury - Incapacitating/Fatal	- SVM - Back Propagation Neural Network - Ordered Logit	- SVM	- Relative speed - Gross vehicle weight	Liao et al. (2018) [74]
11	2017	Malaysia	2009–2015	1130	- PDO - Evident Injury - Disabling Injury	- Recurrent Neural Network - Bayesian Logistic Regression - Multilayer Perceptron	- Recurrent Neural Network	- Dry surface - Lighting conditions (dark with and without street lights)	Sameen and Pradhan (2017) [75]
12	2017	UAE	2008–2013	5973	- Minor - Moderate - Severe - Death	- Decision Tree (J48) - Naive Bayes - Rule Induction (PART) - Multilayer Perceptron	- Decision Tree (J48)	- Nationality - Collision type (pedestrian–vehicle) - Age (18–30) - Male	Taamneh et al. (2017) [25]
13	2016	Saudi Arabia	2014–2015	85,605	- Injury - No Injury - Deaths	- Decision Tree (CHAID) - Decision Tree (J48) - Naive Bayes	- Decision Tree	- Distraction while driving	Al-Turaiki et al. (2016) [24]

No.	Year	Country	Duration	Size	Injury Severity Classes	Approach	Best Approach	Significant Factors	Reference
14	2015	Iran	2007	1063	<ul style="list-style-type: none">- Fatality- Evident Injury- No Injury	<ul style="list-style-type: none">- SVM, Kernels- Multilayer Perceptron (MLP)- Genetic Algorithm- Combined Genetic Algorithm and Pattern Search	- MLP	- Speed of impact distance	Aghayan et al. (2015) [72]

References

1. International Transport Forum (ITF). *Road Safety Annual Report 2021: The Impact of COVID-19*; ITF: Paris, France, 2021.
2. World Health Organization (WHO). *Global Status Report on Road Safety*; WHO: Geneva, Switzerland, 2018.
3. Al-Atawi, A.M.; Kumar, R.; Saleh, W. A framework for accident reduction and risk identification and assessment in Saudi Arabia. *World J. Sci. Technol. Sustain. Dev.* **2014**, *11*, 214–223. [[CrossRef](#)]
4. Memish, Z.A.; Jaber, S.; Mokdad, A.H.; AlMazroa, M.A.; Murray, C.J.; Al Rabeeah, A.A.; Saudi Burden of Disease Collaborators. Peer reviewed: Burden of disease, injuries, and risk factors in the Kingdom of Saudi Arabia, 1990–2010. *Prev. Chronic Dis.* **2014**, *11*, E169. [[CrossRef](#)]
5. 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.* **2012**, *6*, 31. [[CrossRef](#)] [[PubMed](#)]
6. FHWA. *Highway Safety Manual*; American Association of State Highway and Transportation Officials: Washington, DC, USA, 2010; Volume 19192.
7. Hosseinzadeh, A.; Moeinaddini, A.; Ghasemzadeh, A. Investigating factors affecting severity of large truck-involved crashes: Comparison of the SVM and random parameter logit model. *J. Saf. Res.* **2021**, *77*, 151–160. [[CrossRef](#)]
8. Al-Moqri, T.; Haijun, X.; Namahoro, J.P.; Alfalahi, E.N.; Alwesabi, I. Exploiting Machine Learning Algorithms for Predicting Crash Injury Severity in Yemen: Hospital Case Study. *Appl. Comput. Math* **2020**, *9*, 155–164. [[CrossRef](#)]
9. Panicker, A.K.; Ramadurai, G. Injury severity prediction model for two-wheeler crashes at mid-block road sections. *Int. J. Crashworthiness* **2022**, *27*, 328–336. [[CrossRef](#)]
10. Tang, J.; Liang, J.; Han, C.; Li, Z.; Huang, H. Crash injury severity analysis using a two-layer Stacking framework. *Accid. Anal. Prev.* **2019**, *122*, 226–238. [[CrossRef](#)]
11. Bahiru, T.K.; Singh, D.K.; Tessfaw, E.A. Comparative study on data mining classification algorithms for predicting road traffic accident severity. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018.
12. Prati, G.; Pietrantonio, L.; Fraboni, F. Using data mining techniques to predict the severity of bicycle crashes. *Accid. Anal. Prev.* **2017**, *101*, 44–54. [[CrossRef](#)]
13. Özden, C.; Acı, Ç. Analysis of injury traffic accidents with machine learning methods: Adana case. *Pamukkale Univ. J. Eng. Sci.* **2018**, *24*, 266–275. [[CrossRef](#)]
14. Beshah, T.; Ejigu, D.; Abraham, A.; Snasel, V.; Kromer, P. Mining Pattern from Road Accident Data: Role of Road User's Behaviour and Implications for improving road safety. *Int. J. Tomogr. Simul.* **2013**, *22*, 73–86.
15. Zhang, S.; Khattak, A.; Matara, C.M.; Hussain, A.; Farooq, A. Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents. *PLoS ONE* **2022**, *17*, e0262941. [[CrossRef](#)] [[PubMed](#)]
16. Arhin, S.A.; Gatiba, A. Predicting crash injury severity at unsignalized intersections using support vector machines and naïve Bayes classifiers. *Transp. Saf. Environ.* **2020**, *2*, 120–132. [[CrossRef](#)]
17. Candefjord, S.; Muhammad, A.S.; Bangalore, P.; Buendia, R. On Scene Injury Severity Prediction (OSISP) machine learning algorithms for motor vehicle crash occupants in US. *J. Transp. Health* **2021**, *22*, 101124. [[CrossRef](#)]
18. 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.* **2019**, *2673*, 680–692. [[CrossRef](#)]
19. 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.* **2021**, *160*, 106322. [[CrossRef](#)] [[PubMed](#)]
20. AlMamlook, R.E.; Kwayu, K.M.; Alkasisbeh, M.R.; Frefer, A.A. Comparison of machine learning algorithms for predicting traffic accident severity. In Proceedings of the 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Amman, Jordan, 9–11 April 2019.
21. Chen, M.-M.; Chen, M.-C. Modeling road accident severity with comparisons of logistic regression, decision tree and random forest. *Information* **2020**, *11*, 270. [[CrossRef](#)]
22. Niyogisubizo, J.; Liao, L.; Lin, Y.; Luo, L.; Nziyumva, E.; Murwanashyaka, E. A Novel Stacking Framework Based On Hybrid of Gradient Boosting-Adaptive Boosting-Multilayer Perceptron for Crash Injury Severity Prediction and Analysis. In Proceedings of the 2021 IEEE 4th International Conference on Electronics and Communication Engineering (ICECE), Xi'an, China, 17–19 December 2021.
23. Shibata, A.; Fukuda, K. Risk factors of fatality in motor vehicle traffic accidents. *Accid. Anal. Prev.* **1994**, *26*, 391–397. [[CrossRef](#)]
24. Duncan, C.S.; Khattak, A.J.; Council, F.M. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. *Transp. Res. Rec.* **1998**, *1635*, 63–71. [[CrossRef](#)]
25. Al-Turaiki, I.; Aloumi, M.; Aloumi, N.; Alghamdi, K. Modeling traffic accidents in Saudi Arabia using classification techniques. In Proceedings of the 2016 4th Saudi International Conference on Information Technology (Big Data Analysis)(KACSTIT), Riyadh, Saudi Arabia, 6–9 November 2016.
26. Taamneh, M.; Alkheder, S.; Taamneh, S. Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. *J. Transp. Saf. Secur.* **2017**, *9*, 146–166. [[CrossRef](#)]
27. Alkheder, S.; Taamneh, M.; Taamneh, S. Severity prediction of traffic accident using an artificial neural network. *J. Forecast.* **2017**, *36*, 100–108. [[CrossRef](#)]

28. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
29. Zhang, J.; Li, Z.; Pu, Z.; Xu, C. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. *IEEE Access* **2018**, *6*, 60079–60087. [[CrossRef](#)]
30. Krishnaveni, S.; Hemalatha, M. A perspective analysis of traffic accident using data mining techniques. *Int. J. Comput. Appl.* **2011**, *23*, 40–48. [[CrossRef](#)]
31. Jiang, H. A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. *PLoS ONE* **2019**, *14*, e0214966.
32. Mafi, S.; Abdelrazig, Y.; Doczy, R. Machine learning methods to analyze injury severity of drivers from different age and gender groups. *Transp. Res. Rec.* **2018**, *2672*, 171–183. [[CrossRef](#)]
33. Wang, X.; Kim, S.H. Prediction and factor identification for crash severity: Comparison of discrete choice and tree-based models. *Transp. Res. Rec.* **2019**, *2673*, 640–653. [[CrossRef](#)]
34. Santos, K.; Dias, J.P.; Amado, C. A literature review of machine learning algorithms for crash injury severity prediction. *J. Saf. Res.* **2022**, *80*, 254–269. [[CrossRef](#)] [[PubMed](#)]
35. Khan, M.U.A.; Shukla, S.K.; Raja, M.N.A. Load-settlement response of a footing over buried conduit in a sloping terrain: A numerical experiment-based artificial intelligent approach. *Soft Comput.* **2022**, *26*, 6839–6856. [[CrossRef](#)]
36. Rezapour, M.; Molan, A.M.; Ksaibati, K. Analyzing injury severity of motorcycle at-fault crashes using machine learning techniques, decision tree and logistic regression models. *Int. J. Transp. Sci. Technol.* **2019**, *9*, 89–99. [[CrossRef](#)]
37. Mokoatle, M.; Vukosi Marivate, D.; Michael Esiefarienne Bukohwo, P. Predicting road traffic accident severity using accident report data in South Africa. In Proceedings of the 20th Annual International Conference on Digital Government Research, Dubai, United Arab Emirates, 18–20 June 2019.
38. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
39. Hawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357.
40. Rahim, M.A.; Hassan, H.M. A deep learning based traffic crash severity prediction framework. *Accid. Anal. Prev.* **2021**, *154*, 106090. [[CrossRef](#)] [[PubMed](#)]
41. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [[CrossRef](#)]
42. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 2493.
43. Dong, S.; Khattak, A.; Ullah, I.; Zhou, J.; Hussain, A. Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHAPley Additive exPlanations. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2925. [[CrossRef](#)] [[PubMed](#)]
44. Chakraborty, M.; Gates, T.; Sinha, S. Causal Analysis and Classification of Traffic Crash Injury Severity Using Machine Learning Algorithms. *arXiv* **2021**, arXiv:2112.03407.
45. Jiang, L.; Xie, Y.; Wen, X.; Ren, T. Modeling highly imbalanced crash severity data by ensemble methods and global sensitivity analysis. *J. Transp. Saf. Secur.* **2022**, *14*, 562–584. [[CrossRef](#)]
46. Labib, M.F.; Rifat, A.S.; Hossain, M.M.; Das, A.K.; Nawrine, F. Road accident analysis and prediction of accident severity by using machine learning in Bangladesh. In Proceedings of the 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, 28–30 June 2019.
47. Jamal, A.; Zahid, M.; Rahman, M.T.; Al-Ahmadi, H.M.; Almoshaogeh, M.; Farooq, D.; Ahmad, M. Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study. *Int. J. Inj. Control Saf. Promot.* **2021**, *28*, 408–427. [[CrossRef](#)]
48. Sattar, K.; Oughali, F.C.; Assi, K.; Ratrou, N.; Jamal, A.; Rahman, S.M. Prediction of electric vehicle charging duration time using ensemble machine learning algorithm and Shapley additive explanations. *Int. J. Energy Res.* **2022**, *46*, 15211–15230.
49. 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.* **2023**, *31*, 78–92. [[CrossRef](#)]
50. Sattar, K.; Chikh Oughali, F.; Assi, K.; Ratrou, N.; Jamal, A.; Masiur Rahman, S. Transparent deep machine learning framework for predicting traffic crash severity. *Neural Comput. Appl.* **2022**, *34*, 1–13. [[CrossRef](#)]
51. Lin, C.; Wu, D.; Liu, H.; Xia, X.; Bhattarai, N. Factor identification and prediction for teen driver crash severity using machine learning: A case study. *Appl. Sci.* **2020**, *10*, 1675. [[CrossRef](#)]
52. Wu, S.; Yuan, Q.; Yan, Z.; Xu, Q. Analyzing accident injury severity via an extreme gradient boosting (XGBoost) model. *J. Adv. Transp.* **2021**, *2021*, 3771640.
53. Zhu, S.; Wang, K.; Li, C. Crash injury severity prediction using an ordinal classification machine learning approach. *Int. J. Environ. Res. Public Health* **2021**, *18*, 11564. [[CrossRef](#)]
54. Aldhari, I.; Bakri, M.; Alfawzan, M.S. Prediction of California Bearing Ratio of Granular Soil by Multivariate Regression and Gene Expression Programming. *Adv. Civ. Eng.* **2022**, *2022*, 7426962.
55. Ministry of Interior (MOI). Emirate of Al-Qasim Province. National Information Center “Absher”, Saudi Arabia 2022; 2022. Available online: <https://www.moi.gov.sa/wps/portal/Home/emirates/qasim/> (accessed on 8 October 2022).
56. Almoshaogeh, M.; Abdulrehman, R.; Haider, H.; Alharbi, F.; Jamal, A.; Alarifi, S. Traffic accident risk assessment framework for qassim, saudi arabia: Evaluating the impact of speed cameras. *Appl. Sci.* **2021**, *11*, 6682. [[CrossRef](#)]

57. 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* **2020**, *10*, 2158244020919500. [[CrossRef](#)]
58. Witten, I.H.; Frank, E.; Trigg, L.E.; Hall, M.A.; Holmes, G.; Cunningham, S.J. *Weka: Practical Machine Learning Tools and Techniques with Java Implementations*; University of Waikato, Department of Computer Science: Hamilton, New Zealand, 1999.
59. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2005; Volume 2.
60. Beshah, T.; Hill, S. Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia. In 2010 AAAI Spring symposium series. 2010. Available online: <https://www.aaai.org/ocs/index.php/SSS/SSS10/paper/view/1173/1343> (accessed on 17 November 2022).
61. Khera, D.; Singh, W. Prediction and analysis of injury severity in traffic system using data mining techniques. *Int. J. Comput. Applic* **2015**, *7*, 1–7.
62. Castro, Y.; Kim, Y.J. Data mining on road safety: Factor assessment on vehicle accidents using classification models. *Int. J. Crashworthiness* **2016**, *21*, 104–111. [[CrossRef](#)]
63. Mitranont, J.; Sawangphol, W.; Vithantirawat, T.; Paengkaew, S.; Suwannasing, P.; Daramas, A.; Chen, Y.C. A study on using Python vs Weka on dialysis data analysis. In Proceedings of the 2017 2nd International Conference on Information Technology (INCIT), Nakhonpathom, Thailand, 2–3 November 2017.
64. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
65. Xgboost Developers. Revision 62ed8b5f, Introduction/XGBoost Objective Function. 2022. Available online: <https://xgboost.readthedocs.io/en/stable/tutorials/model.html> (accessed on 23 August 2022).
66. Shapley, L. *Quota Solutions op n-person Games1*; Artin, E., Morse, M., Eds.; Princeton University Press: Princeton, NJ, USA, 1953; p. 343.
67. SHAP. SHAP Documentation. SHAP 2017. Available online: <https://shap.readthedocs.io/en/latest/index.html> (accessed on 3 September 2022).
68. Yu, R.; Abdel-Aty, M. Utilizing support vector machine in real-time crash risk evaluation. *Accid. Anal. Prev.* **2013**, *51*, 252–259. [[CrossRef](#)]
69. Pillajo-Quijia, G.; Arenas-Ramírez, B.; González-Fernández, C.; Aparicio-Izquierdo, F. Influential factors on injury severity for drivers of light trucks and vans with machine learning methods. *Sustainability* **2020**, *12*, 1324. [[CrossRef](#)]
70. Aghayan, I.; Hosseinlou, M.H.; Kunt, M.M. Application of support vector machine for crash injury severity prediction: A model comparison approach. *J. Civ. Eng. Urban.* **2015**, *5*, 193–199.
71. Shanthi, S.; Ramani, R.G. Feature relevance analysis and classification of road traffic accident data through data mining techniques. In Proceedings of the World Congress on Engineering and Computer Science, San Francisco, CA, USA, 24–26 October 2012.
72. Kunt, M.M.; Aghayan, I.; Noii, N. Prediction for traffic accident severity: Comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. *Transport* **2011**, *26*, 353–366. [[CrossRef](#)]
73. Yang, Y.; Wang, K.; Yuan, Z.; Liu, D. Predicting freeway traffic crash severity using XGBoost-Bayesian network model with consideration of features interaction. *J. Adv. Transp.* **2022**, *2022*, 4257865. [[CrossRef](#)]
74. Liao, Y.; Zhang, J.; Wang, S.; Li, S.; Han, J. Study on crash injury severity prediction of autonomous vehicles for different emergency decisions based on support vector machine model. *Electronics* **2018**, *7*, 381. [[CrossRef](#)]
75. Sameen, M.I.; Pradhan, B. Severity prediction of traffic accidents with recurrent neural networks. *Appl. Sci.* **2017**, *7*, 476. [[CrossRef](#)]

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