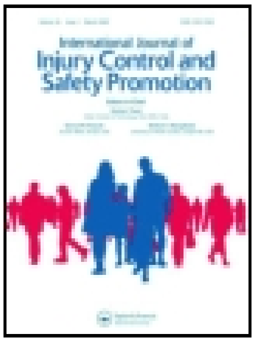

International Journal of Injury Control and Safety Promotion

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/nics20

Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study

Arshad Jamal, Muhammad Zahid, Muhammad Tauhidur Rahman, Hassan M. Al-Ahmadi, Meshal Almoshaogeh, Danish Farooq & Mahmood Ahmad

INTERNATIONAL JOURNAL OF INJURY CONTROL AND SAFETY PROMOTION
https://doi.org/10.1080/17457300.2021.1928233

Taylor & Francis
Taylor & Francis Group

Check for updates

RESEARCH ARTICLE

Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study

Arshad Jamal[a], Muhammad Zahid[b], Muhammad Tauhidur Rahman[c], Hassan M. Al-Ahmadi[a], Meshal Almoshaogeh[d], Danish Farooq[e,f] and Mahmood Ahmad[g]

[a]Department of Civil and Environmental Engineering, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia; [b]College of Metropolitan Transportation, Beijing University of Technology, Beijing, China; [c]Department of City and Regional Planning, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia; [d]Department of Civil Engineering, College of Engineering, Qassim University, Buraydah, Qassim, Saudi Arabia; [e]Department of Transport Technology and Economics, Budapest University of Technology and Economics, Budapest, Hungary; [f]Department of Civil Engineering, University of Engineering and Technology Peshawar (Bannu Campus), Peshawar, Pakistan; [g]Department of Civil Engineering, University of Engineering and Technology Peshawar (Bannu Campus), Peshawar, Pakistan

ABSTRACT

A better understanding of injury severity risk factors is fundamental to improving crash prediction and effective implementation of appropriate mitigation strategies. Traditional statistical models widely used in this regard have predefined correlation and intrinsic assumptions, which, if flouted, may yield biased predictions. The present study investigates the possibility of using the eXtreme Gradient Boosting (XGBoost) model compared with few traditional machine learning algorithms (logistic regression, random forest, and decision tree) for crash injury severity analysis. The data used in this study was obtained from the traffic safety department, ministry of transport (MOT) at Riyadh, KSA, and contains 13,546 motor vehicle collisions along 15 rural highways reported between January 2017 to December 2019. Empirical results obtained using k-fold (k = 10) for various performance metrics showed that the XGBoost technique outperformed other models in terms of the collective predictive performance as well as injury severity individual class accuracies. XGBoost feature importance analysis indicated that collision type, weather status, road surface conditions, on-site damage type, lighting conditions, and vehicle type are the few sensitive variables in predicting the crash injury severity outcome. Finally, a comparative analysis of XGBoost based on different performance statistics showed that our model outperformed most previous studies.

ARTICLE HISTORY
Received 21 January 2021
Revised 25 March 2021
Accepted 5 May 2021

KEYWORDS
Traffic crashes; crash injury severity prediction; machine learning; XGBoost; feature sensitivity analysis

Introduction

Road traffic crashes are one of the leading causes of human casualties and injuries. According to a recent report of the world health organization (WHO), over 1.35 million people are annually killed in traffic crashes, and approximately 50 million are injured worldwide (World Health Organization, 2019). They are the leading cause of deaths among young people (aged between 15-29 years) and are predicted to be the leading cause of fatalities across all age groups by the year 2030 (World Health Organization, 2019). Interestingly, the low and middle-income countries carry a disproportionately high number of fatalities even though they share only around 48% of the registered vehicles in the world (Kopits & Cropper, 2005). A rapid rise in population growth and fast economic growth in these countries has been associated with a surge in motorization and road traffic injuries (Bhalla et al., 2020; Wahab & Jiang, 2019a).

Growing traffic fatalities in the Kingdom of Saudi Arabia (KSA) have become a primary concern for government and local authorities. Traffic collisions in KSA account for around 4.7% of total mortalities, while it is below 2% in developed countries like Australia, the UK, and the United States (S. M. Lee & Al-Mansour, 2020). Similarly, over the decade, traffic-related fatalities have fluctuated between 22-27 per 100,000 persons, which is high compared to the global average (17.4), and worse compared to 3 in the UK and 10 in the United States. Due to traffic crashes in KSA, the economic losses are estimated to be approximately 4.3% of GDP (Mohamed, 2015). The average crash-to-injury ratios of 8:4 and 8:6 reported for different regions in KSA are also significantly high compared to the global ratio of 8:1 (Ansari et al., 2000; Jamal et al., 2019). In recent years, several studies have investigated the crash causation and injury severity risk factors in KSA and neighbouring gulf countries. Existing literature suggests that factors such as distracted driving, non-compliance with traffic rules, overspeeding and aggressive driving are the few predominant factors responsible for increased crash incidence and severity (Al-Tit et al., 2020; Bendak, 2005; DeNicola et al., 2016; Jamal et al., 2019; 2021; Tauhidur Rahman et al., 2020; Zahid, Chen, Khan, et al., 2020). In recent years, the

CONTACT Arshad Jamal arshad.jamal@kfupm.edu.sa

government has initiated few efforts like installing the SAHER scheme, imposition of hefty fines, strict enforcement of traffic rules, road safety education, campaigns, etc. However, the safety situation on national roads has been marginally improved. In addition to challenging road safety concerns, the lack of public transportation and increased auto-ownership in KSA have also led to alarming traffic congestion and environmental issues (Al-Turki et al., 2020; Jamal et al., 2020).

Crash injury severity prediction is an important and promising research target in road safety studies. A better understanding of factors contributing to crash injury severity is fundamental to implement appropriate mitigation strategies proactively. The outcome of crash injury severity is influenced by several concuring humans, environments, roadway, traffic, vehicle, crash characteristics, etc. However, traffic crashes are random events and exhibit spatial heterogeneity, meaning that severity contributing factors may vary from one location to another. Therefore, a local investigation of the risk factors is a must for reliable crash prediction and effective deployment of suitable counter-measures.

In literature, different statistical-based regression has been widely used for modelling crash injury severity. Although statistical models have good mathematical interpretation and provide a better understanding on the role of individual predictor variables, they do have some limitations. First, they are built on a number of underlying assumptions (regarding linear link functions and error distribution term) and the predefined association between the variables; if flouted, they may yield bias model estimation (Ullah et al., 2021; Zahid, Chen, Jamal, Al-Ahmadi, et al. 2020; Zahid, Chen, Jamal, and Memon 2020). Second, they suffer from poor prediction accuracies, which are not highly reliable as well. In recent years, with rapid advances in soft computing methods, machine learning-based models have emerged as promising tools in road safety research to overcome the limitations of statistical methods. However, most of these studies have focussed on improving the overall prediction accuracies that do not advance researchers' perception of the individual role of severity risk factors. Feature-based sensitivity analysis is essential to address this issue.

The contributions of this paper are three-fold. First, we compared the performance of the eXtreme Gradient Boosting (XGBoost) technique with three traditional algorithms (logistic regression, random forest, decision tree) for crash injury severity modelling. XGBoost is a relatively new applied machine learning technique, initially proposed in 2016 by Chen and Guestrin (T. Chen & Guestrin, 2016). XGBoost is an ensemble machine learning method that utilizes an implementation of a gradient boosted decision trees framework. The method provides high prediction accuracy with fast processing time and is computationally less costly and complex (Mousa et al., 2019; Parsa et al., 2020). Second, we use the XGBoost classifier feature importance technique for sensitivity analysis of predictor variables. Finally, we provide various performance statistics to compare our model with literature in the subject domain in terms of overall severity prediction and the model's predictive performance by different injury severity categories.

Literature review

A better understanding of circumstances under which drivers or vehicle occupants are more likely to sustain severe or fatal injuries in motor vehicle crashes is vital to improving highway safety situations. The primary objective of crash severity analysis is to uncover the latent relationships between crash injury severity and various contributing factors such as driver and passenger characteristics, roadway geometric design features, vehicle attributes, traffic attributes, crash characteristics, environmental conditions, and features of the built environment (Ijaz et al., 2021). In literature, crash injury severity analysis has been investigated using statistical and machine learning methods. A brief description of previous studies focused on modelling of crash injury severity from various methodological fronts is provided in the following passages.

Crash severity modeling using statistical methods

Crash injury severity is generally represented by discrete categories such as property damage, possible injuries, capacitating injury, incapacitating injury, and fatal etc. Many studies have used three injury severity categories(C. Ma et al., 2018; Mesa-Arango et al., 2018): property damage only (PDO)/no injury, injury, and fatal. Owing to the discrete nature of crash injury severity levels, discrete outcome models like binary or multinomial logit/probit models have been widely used (Azimi et al., 2020; Rifaat & Chin, 2007; Shankar & Mannering, 1996; Yu & Abdel-Aty, 2014a). To account for heterogeneity and endogeneity, and ordinal nature within-crash correlation, a number of advanced models as Bayesian hierarchical (Huang et al., 2008; Li et al. 2018), ordered logit models (Azimi et al., 2020; C. Chen et al., 2016; Khattak et al., 1998; O'Donnell & Connor, 1996), bivariate/multivariate models (Aguero-Valverde & Jovanis, 2009; C. Lee & Abdel-Aty, 2008; Russo et al., 2014; Zeng et al., 2017), nested logit model (Osman et al., 2016; Shankar et al., 1996), random parameter model (Milton et al., 2008; J. Wang et al., 2020), Markov switching multinomial model (Malyshkina & Mannering, 2009; Xiong et al., 2014), and their mixed versions (Christoforou et al., 2010; Eluru & Bhat, 2007; Huang et al., 2011; Li et al. 2019), were investigated.

Application of logit/probit models

Different types of logit and probit and their hybrid versions were initially used in crash injury severity analysis. For example, Garrido et al. used an ordered probit model to investigate injury severity prognosis of occupants in motor vehicle collisions (Garrido et al., 2014). Model estimation results showed that occupants in light vehicles (cars) on two-way highways, dry road surfaces, and traveling in rural areas tend to carry more severe injuries. Kockelman and

Kweon applied ordered probit models to examine risks of injury severity levels for single and bi-vehicle crashes considering different crash types (Kockelman & Kweon, 2002). The results indicated that occupants in sport utility vehicles and pickups suffered severe injuries compared to those riding cars under single-vehicle crash circumstances. Similarly, ordered logit models were used by other researchers to incorporate the ordinal nature of crash injury severity (Haghighi et al., 2018; Kamruzzaman et al., 2014; Yasmin et al., 2014).

Mohamed Abdel-Aty utilized multinomial logit models and ordered probit models to analyze driver injury severity (M. Abdel-Aty, 2003). Models' estimation results indicated that drivers' sociodemographic attributes (age and gender), speed ratio, dark lighting conditions, seat belt usage, vehicle type, point of impact, and presence of curves aggravate crash severity. Comparing the analysis methods, the ordered probit method was deemed more promising. Chang et al. utilized mixed ordered logit modes for crash severity analysis of motorcyclists using one-year crash data from Hunan, China (F. Chang et al., 2016). The study reported that factors associated with severe and fatal crashes include elderly riders (age> 60 years), the absence of helmets, collisions with heavy vehicles, and night-time travel conditions. Kim et al. also utilized mixed logit models to study the injury severity of pedestrians in motor vehicle crashes (Kim et al., 2010). Model estimation results revealed that chances of fatal crashes were increased significantly with speeding, absence of street lights, drunk driving, collisions with trucks, and crashes along freeways.

Application of random parameter models

In recent years, random parameters are increasingly used in the severity prediction of traffic crashes. In their study, Rusli et al. employed random parameter logit models to explore factors contributing to crash injury severity along rural mountainous terrain in Malaysia (Rusli et al., 2018). The study showed that the proposed method was more robust than standard scobit and logit models in capturing unobserved heterogeneity. Wang et al. also utilized the random parameter probit model and showed that factors including rainy weather, involvement of pedestrians, and crashes involving buses had a higher probability of severe injuries (J. Wang et al., 2020). To account for unobserved heterogeneity, Tulu et al. applied random-parameters logistic regression for crash injury severity analysis in Ethiopia (Tulu et al., 2017). The study reported that factors such as night-time driving, collisions involving heavy vehicles, and overspeeding resulted in severe and fatal crashes. Yu and Abdel-Aty used two statistical modelling techniques (fixed and random parameter logit models) to investigate the influence of real-time traffic and weather data on injury severity of crashes on mountainous terrain (Yu & Abdel-Aty, 2014b). Experimental results indicated that random parameters model outperformed the fixed-parameter logit model in terms of AUC values. Sensitivity analysis for parameters showed that large speed variations, snowy season, low temperature, and presence of steep grades increased the probability of having more severe crashes. Another recent study also proved the robustness of random parameter models for crash injury severity modelling (F. Chang et al., 2019).

Application of Bayesian hierarchical models

Bayesian Hierarchical models were also used by numerous previous studies in road safety research. For example, Huang et al. employed the Bayesian hierarchical binomial logistic model to explore factors causing driver injury severity level at signalized intersections in Singapore (Huang et al., 2008). It was found that intersection type, presence of camera at signals, street lighting, nature of lanes, the involvement of pedestrians, and time of the day were observed to be more significantly associated with crash injury severities. Haque et al., in their study, also proposed Bayesian hierarchical models to investigate motorcyclist injury severity at signalized intersections (Haque et al., 2010) and found that the presence of red-light cameras significantly reduces the risk of motorcycle crashes at both four-legged and T-intersections. Meng et al. proposed a Bayesian hierarchical logistic modelling approach for injury severity analysis of taxis occupants (Meng et al., 2017). The findings indicated that elderly drivers, fatigue driving, usage of seatbelt were significantly associated with taxi occupant's injury severity.

Limitations of statistical modeling methods

Existing literature is abundant, with research studies focussing on crash severity modelling from the diverse statistical methodological fronts. Even though statistical models have good theoretical interpretability, they do suffer from several limitations. For example, traditional statistical models are based on a certain distribution of crash data and assume linear function forms to link the dependent variable with predictor variables. However, such assumptions may not always remain valid; and if violated, it could easily lead to erroneous estimations and biased model inferences (Li et al. 2012; X. Wang & Kim, 2019; J. Zhang et al., 2018; Zahid, Chen, Jamal, and Mamadou, 2020). Further, unobserved heterogeneity and multi-collinearity within crash data could also affect model estimation (P. Savolainen & Mannering, 2007; Washington et al., 2020; Xu & Huang, 2015). To neutralize the negative impacts of mentioned issues, complex frameworks are often needed, which make the statistical models computationally exhaustive and challenging to use in practice (P. T. Savolainen et al., 2011). Alternatively, non-parametric methods and artificial intelligence models, particularly machine learning and deep learning, have emerged as promising road safety research tools.

Crash severity modeling using machine learning methods

To overcome the drawbacks of statistical methods, different machine learning (ML) models are increasingly explored for modelling the potentially nonlinear relationships between crash contributing factors and injury severity outcomes (Abdel-Aty & Abdelwahab, 2004; Iranitalab & Khattak, 2017; Li et al. 2012; Pradhan & Sameen, 2020; Sameen & Pradhan, 2017; Sarkar et al., 2020; Tang et al., 2019). Machine learning models have the advantage that they have higher adaptability to process outliers, noisy or missing data, are more flexible with no or very few pre assumptions for input variables. ML methods are reported to have better fitting compared to statistical methods. Some widely used ML algorithms in crash severity modelling domain are: Artificial Neural Networks (ANN) (Abdelwahab & Abdel-Aty, 2001; Amiri et al., 2020; Zeng & Huang, 2014), Support Vector Machines (SVM) (Dong et al., 2015; Mokhtarimousavi et al., 2019; Zhibin Li et al. 2012), Decision Trees (DT) (Abellán et al., 2013; Oña et al., 2013; P. Lu et al., 2020), K-means Clustering (KC) (Anderson, 2009; Fiorentini & Losa, 2020; Mauro et al., 2013), Random Forest (Iranitalab & Khattak, 2017; Mondal et al., 2020; J. Zhang et al., 2018), and Naïve Bayes (Arhin & Gatiba, 2020; Budiawan et al., 2019; C. Chen et al., 2016). A major criticism of ML models is that they are operated in a black box, giving no explicit correlation between the dependent variable and explanatory variables. Previously ML methods were mostly used as prediction tools; however, in recent years, the issue of the black box phenomenon is tackled using sensitivity analysis to identify the impact of individual predictor variables (Li et al. 2012; Y. Zhang & Xie, 2007). Sensitivity analysis may be employed for feature extraction and determining relative variable importance towards the target variable. It has greatly expanded the potential of adopting ML models in road safety studies.

Application of artificial neural network (ANN) models

Different ANN architectures were explored in earlier studies among the artificial intelligence-based framework for crash severity prediction. For example, Abdelwahab and Abdel-Aty used a multi-layer perceptron (MLP) framework to classify vehicular crash severity at signalized intersections using one year crash data from Central Florida, in the US (Abdelwahab & Abdel-Aty, 2001). It was reported that MLP, with an average classification accuracy of 65.6% for the test dataset, outperformed the ordered logit model. . Sarkar and Sarkar compared the performance of ANN and multiple logistic regression (MLRG) for accident severity forecasting on urban roads (A. Sarkar & Sarkar, 2020). To identify significant features influencing crash injury severity, the researchers utilized three different filter-based ranking methods such as Gain Ratio, Information Gain, and Symmetric Uncertainty. Results indicated that ANN is more reliable to cater for unobserved heterogeneity in predicting crash severity. Kunt et al. attempted to predict injury severity of moto vehicle crashes using ANN, pattern search, genetic algorithm (GA) modelling methods (Kunt et al., 2012). The models were constructed and trained with twelve input predictor variables using one-year crash data from Tehran–Ghom Freeway, in Iran. The best fit model was selected according to R-value, mean absolute errors (MAE), and root-mean-square errors (RMSE). ANN, with an R-value of approximately 0.87, demonstrated superior performance compared to other models. Similarly, reported values for MAE (.16), RMSE (0.22), and SSE (123.43) for ANN were better compared to other data mining techniques. Jamal et al. developed a feedforward neural network (FFNN) model to explore the injury severity risk factors in fatal crashes in KSA (Jamal & Umer, 2020). FFNN architecture with logistic as activation function, back-propagation (BP) as a training algorithm, and six number of hidden neurons generated the best model performance. Feature-based sensitivity analysis of optimized neural network (NN) was conducted to examine the relative importance of each predictor variable towards the injury severity occurrence. Results showed that weather conditions, crash type, on-site damage conditions, road and vehicle type, pedestrians' involvement, travel speeds, and road surface conditions all increased the crash injury severity.

Application of support vector machine (SVM) models

Support vector machine (SVM) is another widely used machine method in crash severity prediction. Li et al. applied SVM to predict crash injury severity based on data collected for 326 freeway diverge segments in the State of Florida, US (Li et al. 2012). The researchers compared the performance of the developed SVM model with the ordered logit model using the same dataset. Results showed that SVM achieved better prediction performance than the logit model. Mokhtarimousav et al. compared random parameter mixed logit models for the severity prediction of crashes occurring near work zones (Mokhtarimousavi et al., 2019). Three optimization algorithms (i.e., particle swarm optimization, harmony search, and whale optimization) were employed to enhance the predictive performance of SVM. Empirical findings showed that SVM had a better predictive performance than the mixed logit models. Analyzing the parameters' sensitivity, variables such as highway types, morning peaks, nature of termination areas in the work zones, type of activities, and rear-end crashes all had a positive influence on crash severity.

Delen et al. compared four different different machine methods for predicting the injury severity of traffic crashes as a function of different input variables like human attributes, vehicle, and roadway characteristics (Delen et al., 2017). The four methods used were: logistic regression, decision trees, NN, and SVM. The study results reveal that the support vector machine was the best in predicting the crash severity, while logistic regression was the least accurate. Non-compliance with seat belt, collision type, and drug involvement were noted to be some of the predominant predictors for crash injury severity. Wang et al. compared MLP and SVM for injury severity prediction of traffic crashes and found that SVM achieved better prediction accuracy (W. Wang et al., 2011).

Application of decision tree (DT) models

In recent years, the application of different types of Decision Trees (DT) models has also gained rapid popularity in road safety research, particularly for injury severity prediction and classification of traffic crashes. Chong et al. developed DT-based models for predicting automobile crash injury severity classified by KABCO (Chong et al., 2004). KABCO is an injury severity classification scale rated in descending order. For example, K indicates a fatal crash, O represents a crash involving property damage only, whereas the letters A, B and C corresponds to decreasing crash injury severities, respectively. Experimental results from this study showed that DT models achieved better prediction results compared to NN. Abellán et al. proposed an interesting concept based on decision rules via DT for crash severity prediction (Abellán et al., 2013). This study was based on seven years (2003-2009) crash data from Granada, Spain. Findings indicated that proposed methods could extract a high number of relevant rules for accurate classification of crash injury severity groups. Similarly, in another study, Griselda et al. demonstrated that DT could be successfully used to extract rules to define and accurately classify injury severity into crash distinct categories (Griselda et al., 2012). Ali and Mohamed proposed a novel probit DT-based approach to examine the impact of adverse weather on injury severity of work zones crashes (Ghasemzadeh & Ahmed, 2017). Experimental results were compared with the conventional probit model. It was concluded that the proposed methods outperformed the probit model in terms of estimation accuracy, reliability, and robustness.

Application of random forest (RF) models

Random Forest (RF) is yet another promising ML tool widely for severity prediction and sensitivity analysis of crash contributing factors. RF is an ensemble learning method that combines several decision trees via bootstrap sampling to improve predictive performance of model. Zhou et al. studied crash severity prediction of crashes at Highway-Rail Grade Crossings using RF and DT (Zhou et al., 2020). Results showed that RF performed better than DT based on the considered eight classification evaluation metrics. ML algorithms were validated using a k-fold cross-validation technique. Model's prediction showed that RF was the best classifier out of the proposed three algorithms. Zhang et al. also compared the crash severity predictive performance among different machine learning (DT, RF, SVM, K-Nearest Neighbor) and two statistical modelling (ordered probit and MNL) approaches (J. Zhang et al., 2018). The study used crash data near freeway diverge ramps in the state of Florida in the US. Results showed that the RF method had the best overall model's performance as well as by individual crash severity groups, while ordered probit models were the least accurate. The authors argued that variable importance for severity prediction was not consistent among different methods, which is a critical issue. In another study, Dadashova et al. compared the performance of RF and various discrete-choice models for crash injury severity prediction and noted that roadway design factors as super elevation, vertical and horizontal curvature, shoulder and lane width are among the few sensitive variables influencing the crash injury severity (Dadashova et al., 2020).

Princess and Rajsingh also performed a comparative study of RF, K-nearest neighbor DT, and Adaboost to classify crash injury severity using imbalanced crash datasets. The results showed that RF with an AUC value of 0.97 outperformed other models. In their study, Mondal et al. proposed RF and Bayesian additive regression trees (BART) models to classify and predict bi-level crash severity (i.e., severe and non-severe) using four years (2015-2018) crash data from Connecticut, US (Mondal et al., 2020). Sensitivity analysis was also conducted to establish the relative importance of crash contributing factors. Experimental findings suggested that RF had a better severity prediction capability for weather-related crashes. RF model with a skill score of 0.73 outperformed the BART model having a corresponding score of 0.60. Crash manner and weather conditions were identified as the two most important variables that impact the severity of crashes.

Application of eXtreme gradient boosting (XGBoost) models

XGBoost is an integrated ensemble machine learning model based on gradient boosted decision trees. It is a relatively new method designed to improve model predictive performance, avoid overfitting issues, and to reduce computational efforts (T. Chen & Guestrin, 2016). XGBoost is reported to have achieved better predictive performance compared to traditional ML algorithms across various domains (Duan et al., 2020; Fauzan & Murfi, 2018; Y. Lu et al., 2021; Qu et al., 2019; C. Wang et al., 2020). However, its application in road safety, particularly crash injury severity analysis, is relatively scarce. For example, two previous studies have shown that XGBoost performs better than conventional machine learning techniques (such as logistic regression, neural network, support vector machine, Bayesian network, and gradient boosting) in predicting the likelihood of traffic crashes (Mousa et al., 2019; Schlögl et al., 2019). Parsa et al. used XGBoost for feature analysis and real-time detection of traffic accidents (Parsa et al., 2020). Results showed that the method could robustly detect traffic accidents with a detection rate and an accuracy of 79% and 99%, respectively. A recent study conducted by Pradhan and Sameen has also demonstrated the robust performance of XGBoost, and deep neural networks (DNN) in traffic crashes severity prediction (Pradhan & Sameen, 2020). Malkoatle et al. utilized multivariate logistic regression and XGBoost for road accident severity classification in South Africa and concluded that XGBoost outperforms the multivariate regression (Mokoatle et al., 2019). Chen et al. employed XGBoost and Classification and Regression Tree (CART) models to examine factors contributing to the severity of automated vehicle crashes (H. Chen et al., 2020). Crash data from AV-related crash reports were extracted from the data originally procured from the California department of motor vehicles for the year 2019. Study results showed that XGBoost performs better in identifying crash injury severity. Feature importance analysis showed that predictor variables such as weather, accident location, degree of vehicle damage, and crash type were the significant crash contributing factors. Ma et al. compared the performance of XGBoost and grid analysis with five traditional algorithms (RF, LR, MLP, SVM, and RF) to analyze factors contributing to fatal crashes in Los Angeles (J. Ma et al., 2019). XGBoost, with a modelling accuracy of 86.73%, outperformed other methods. Eight variables, including drunk-driving, lighting condition, pedestrian and motorcycle involvement, the number of parties involved, rear-end crashes, time of the day, day of the week, were reported to be highly influential in determining the injury severity outcome. Guo et al. examined the crash injury severity factors in traffic crashes involving elderly pedestrians (age >65 years) using the emerging XGBoost model (Guo et al., 2021). Crash data (2006-2016) containing13,856 pedestrian traffic crashes in Colorado, the USA, was used in this research. Crash injury severity was classified into three-levels, i.e., fatal, injury, and property damage only (PDO). The study reported that factors contributing significantly to elderly pedestrian crashes include driving speed, lighting conditions, time of the day, road alignment, and driver age.

Data description

The traffic crash dataset used in this research was procured from the traffic safety department, Ministry of Transport (MOT), in KSA. The data covers all types of motorized vehicle crashes from January 2017 to December 2019 that occurred along 15 major inter-cities rural highways (shown in Figure 1) in the country. The selected highways mostly pass through desert and plain terrain, having warm-to-high temperatures during most part of the year. The traffic safety department is mainly responsible for collecting crash data on rural highways under MOT jurisdiction, whereas local traffic police departments maintain crash records within the cities. The on-site emergency response expert crew prepares a crash report file which is then used to extract and compile the crash database. The collected data was cleaned and preprocessed by deleting the duplicate records, outliers, and crash record with missing information. The final dataset had a total of 13,546 valid crash records that contained 5,559 injuries instances and 1,320 fatalities. The data had six explanatory variables under different categories (environmental, traffic, roadway, vehicle, temporal, and crash) with 59 sub-levels (child features) for categorical variables and seven-levels for continuous/numeric predictor variables. Temporal characteristics include the time of the crash (peak or off-peak hours), day of the week (weekday or weekend), and season (Winter, Spring, Summer, or Autumn), in which the crashes occurred. Variables under environmental

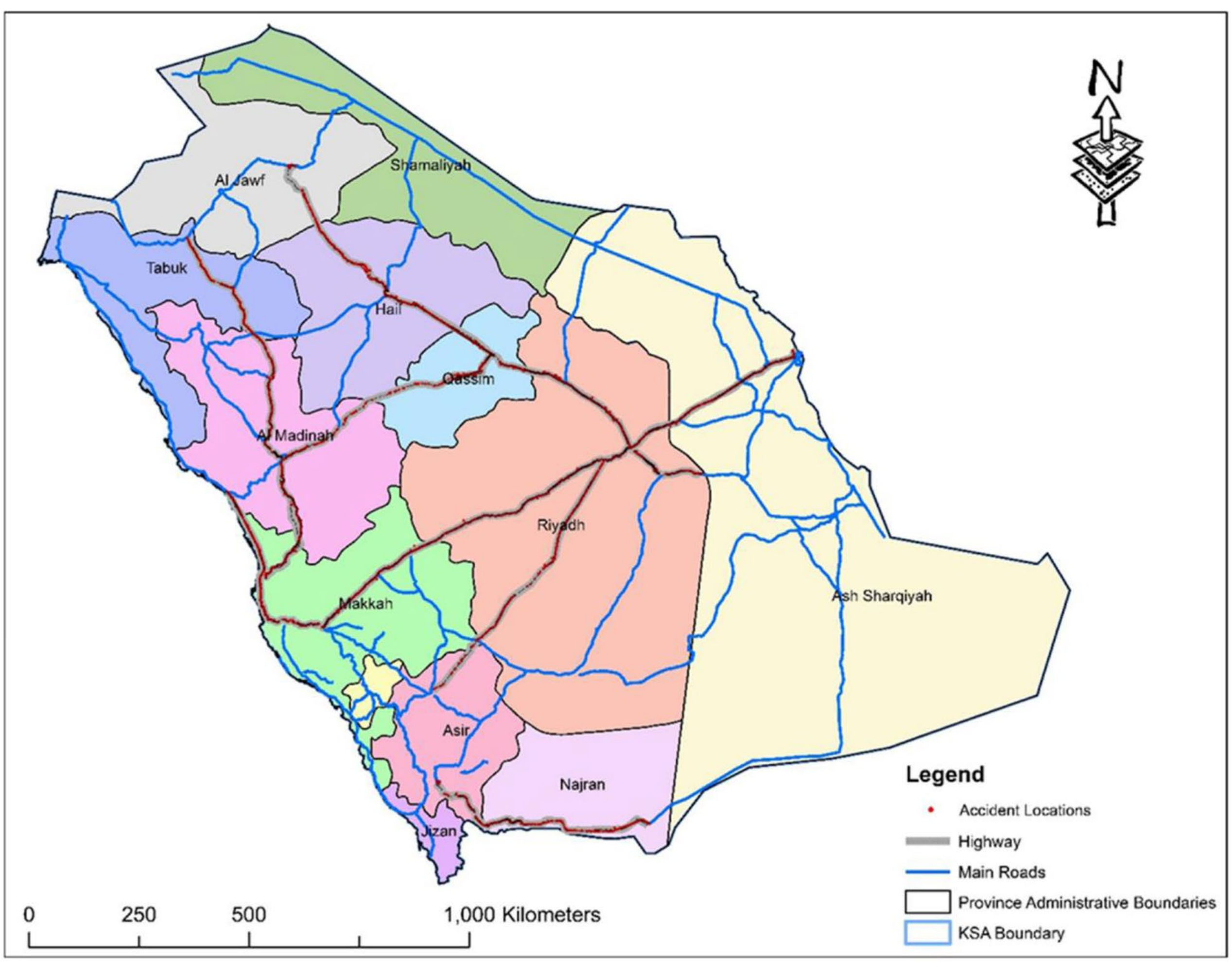

Figure 1. Study area and crash locations.

characteristics comprised of lighting and weather conditions. The explanatory variables under roadway attributes include highway type, alignment type, road surface conditions, on-site damaged road type, dimensions of roadway cross-sectional elements at crash locations, and presence or absence of road markings and road cat-eyes at crash sites. Category, traffic characteristics provided information about Average Annual Daily Traffic (AADT) on the link, the percentage of trucks in the AADT, and average stream speed. The vehicle attributes consisted of detailed information about the type of vehicle at-fault and the number of vehicles involved. Likewise, crash characteristics include collision type, crash contributing circumstance. Over 98% of the drivers involved were males. Unfortunately, information about other sociodemographic attributes for drivers was not available from crash records. Crash injury severity classified in three levels (fatal, injury, and property damage only) was the dependent variable. The crash data recorded collected from MOT three types of crash injury severity categories: PDO that do not involve any injury, injury that includes both capacitating and incapacitating injuries, and fatal that indicates that at least one person was killed during the crash. Traffic Data and other essential road inventory data that were also acquired from MOT.

Table 1 summarizes descriptive statistics of explanatory variables. Approximately 71% of the total crashes were reported to have occurred during the weekdays, mostly during peak periods (57%) and summer seasons (43%). The largest percentage (about 76%) of crashes were recorded on expressways. Similarly, crashes were more prevalent along tangent segments accounting for over 64% of total crashes, while approximately 4% occurred near-horizontal curves. A vast majority of crashes occurred during daytime (60%) and clear weather conditions (around 87%), which is consistent with previous studies that suggest that higher speeds are common during such environmental conditions. Collisions between motor vehicles were the most prevailing crash types, followed by roller over and run-off crashes. Similarly, drivers related factors (distractions, speeding, violations, fatigue) accounted for three-fourth of total crashes. Over 50% of the crashes that were reported during the study period involved at least one car, while almost 20% of the crashes involved trucks. Table 2 provides the distribution by injury severity category across different years. Out of the total 13,546 crashes, 961 (7%) are classified as fatal crashes, 5,559 (41%) as injury crashes, and the remaining 7,026 (52%) as PDO crashes.

Methodology

This section reviews the machine learning classification algorithms and discusses their application for crash severity prediction. Machine learning classifiers are supervised training algorithms used in classifying datasets that can produce promising results due to their multi-dimensional data processing capability, flexibility in implementation, versatility, and superior predictive capabilities. In this research, the target (dependent) variable (crash severity level) is a variable with three possible outcomes (fatal, injury, and PDO). Three algorithms including XGBoost, Random Forest, Decision Tree, and Logistic Regression, were implemented using stratified 10-fold cross-validation. All the algorithms for current study were implemented on Python analytic platform. To examine the individual role of injury severity risk factors, XGBoost feature/variable sensitivity analysis was also performed. The detailed methodology is further discussed in the following passages.

Logistic regression

Linear regression is not suitable for analysis crash severity classification since it might generate probabilities greater than one or less than zero. It is, therefore, more appropriate to use logistic regression (LR). In addition, logistic regression helps to prevent certain assumptions made by the linear regression model. LR is not a classifier but creates probabilities between one and zero., In this study, we used LR to predict the probability of crash severity like injury, fatal and PDO from available vehicle crash accident data. LR model establishes the relationship between the target class $y = (y_1, y_n)$ given $p = (p_1, p_n)$ and set of j predictors $X = (x_1, \ldots\ldots, x_j)$. The method attempts to model the relationship f between predictors or class variables y and a number of predictors or independent variables x. The dependent/target variable was programmed to have three possible outcomes: $\{y_1 = \mathit{Property\ damage};\ y_2 = \mathit{Injury};\ y_3 = \mathit{Fatal}\}$, which can be coded as $\{y_1 = 0;\ y_2 = 1,\ y_3 = 2\}$. The modelling function for LR describes the relationship between the probability of a specific class, e.g., $y = 1$ and the set of independent or predictors variables. A typical LR model equation for current problem may be expressed below (shown in equation 1 and 2):

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

$$z=\beta_{\circ}+\beta_1 x_1+\ldots\ldots+\beta_n x_n=x\beta \tag{2}$$

Where $x\beta$ represents the sigmoid S-shaped function. When the probability greater than 0.5, then the dataset classified as a fatal, injury and property damage. The parameters include in logistic regression were number of iteration, epsilon, learning rate strategy, step size and regularization. Moreover, the learning rate strategy and regularization were considered fixed and uniform.

Random forest

Random Forest (RF) rely on the concept of a Classification and Regression Tree (CART), there by constructing a large number of probable trees consisting of different sets of independent or predictors variables to prevent overfitting issues. A drawback with CART model is that it is very sensitive to new input data and lack generalization. In

Table 1. Descriptive statistics of explanatory variables.

Attribute	Type	Description	Mean	SD	Min	Max	Percent (%)
Temporal Attributes							
Time of crash	Categorical	1: Peak/rush hours; 2: Off-peak	1.436	0.496	1	2	57.40/42.60
Day	Categorical	1: Weekday (Sunday-Thursday); 2: Weekend (Friday and Saturday)	1.283	0.450	1	2	70.84/29.16
Season	Categorical	1: Winter (Nov. to Feb.); 2: Spring (March, Apr.); 3: Summer (May to Aug.); 4: Autumn (Sept. and Oct.)	2.604	1.038	1	4	21.90/15.79/42.2/ 20.05
Environmental Features							
Lighting Condition	Categorical	1: Daytime driving; 2: Night-time driving	1.395	0.488	1	2	60.40/40.60
Weather	Categorical	1: Clear; 2: Rain; 3: Cloudy; 4: Fog/Sand storm; 5: others	1.144	0.632	1	5	86.92/4.26/1.89/ 2.94/3.99
Roadway Characteristics							
Highway Type	Categorical	1: Divided Highway; 2: Expressway; 3: Single Highway	1.796	0.449	1	3	22.36/75.62/2.02
Alignment Type	Categorical	1: Tangent/Straight Segments; 2: Horizontal curve; 3: Vertical curve; 4: near intersection; 5: others	2.356	1.872	1	5	64.20/4.04/1.66/ 1.02/29.08
Surface Conditions	Categorical	1: Good; 2: Cracks; 3: Debris; 4: wet/slippery; 5: others	2.570	1.953	1	5	56.58/10.0/3.59/ 2.10/27.72
Damage at Site	Categorical	1: Fence damaged; 2: Barrier damaged; 3: Pole damaged; 4: Signpost damaged; 5: others	4.225	1.50	1	5	20.74/10.16/3.98/ 2.44/62.68
Carriageway width (m)	Continuous	[7.30, 13.20]	10.139	1.511	7.30	13.20	–
Shoulder width (m)	Continuous	[0, 4]	3.204	0.460	0	4	–
Median width (m)	Continuous	[0, 17.25]	13.552	4.468	0	17.25	–
Road Markings	Categorical	1: Present at crash site; 2: Absent at crash site	1.12	0.096	1	2	97.48/2.52
Road Cateyes	Categorical	1: Present at crash site; 2: Absent at crash site	1.09	0.078	1	2	98.61/1.39
Traffic Characteristics							
AADT	Continuous	[1198, 27546]	11362	6124	1198	27546	–
% of Trucks in ADDT	Continuous	[0, 30]	14.707	7.246	0	30	–
Average Speed (kmph)	Continuous	[60, 120]	101.263	9.60	60	120	–
Vehicle Characteristics							
Type of Vehicle at Fault	Categorical	1: Car; 2: Bus; 3: Small truck; 4: Big truck; 5: others	1.816	1.208	1	5	59.68/7.55/9.96/ 16.78/6.05
No. of vehicles involved	Continuous	[1,26]	2.182	0.750	1	26	–
Crash Characteristics-							
Collision Type	Categorical	1: Vehicle Collisions; 2: Hit Animal; 3: Hit Pedestrian; 4: Rollover; 5: Run-off the road; 6: Skidding; 7: Vehicle Burnt; 8: others	2.90	2.236	1	8	53.46/1.35/0.48/ 24.10/10.44/0.65/2.26/7.30
Contributing Circumstance	Categorical	1: Driver (distractions, fatigue driving, disregard traffic rules and TCD); 2: Animal; 3: Faulty vehicle component, 4: Poor roadway; 5: others	1.669	1.219	1	5	74.79/1.60/15.84/ 1.63/6.14

Table 2. Frequency and Percentage Distribution of Dependent Variable (Injury Severity Categories).

Year	Injury Severity Category	Frequency	Percent (%)
2017	PDO	2534	52.06%
	Injury	1969	40.47%
	Fatal	364	7.45%
2018	PDO	2573	52.46%
	Injury	2024	41.26%
	Fatal	308	6.28%
2019	PDO	1919	50.85%
	Injury	1566	41.49%
	Fatal	289	7.66%

general, this problem is referred to as "over-fitting". Random forests help in reducing the overfitting problem by constructing a large number of separate decision trees, created with various subsets of predictor variables. The basic framework of the RF utilizing decision trees is defined in three steps. 1) Generate a N_c size bootstrap sample from the overall N data to grow a tree by randomly selecting predictors $X = \{x_i, i= 1, ...p\}$. 2) Using the predictor x_i at different tree node n to vote for class label y in same node. The sample is further adjusted at each node, before the best predictor for the split is obtained. 3) To get the misclassification score, run the out-of-bag (OOB) data ($N - N_c$) down the tree, and OOBER is chosen. Whereas is the minimum out-of-bag error rate OOBEFor a large number of trees, repeating step (1–2–3) until OOBER is achieved. Assign each observation by majority vote to a final class y through averaging over the series of trees. Moreover, the split criterion used in this method is the information gain ratio, which is can be calculated from the below equation,

$$Information\ Gain\ Ratio\ (IGR) = \frac{Information\ Gain(X)}{Split\ Info(X)} \quad (3)$$

Where X is the randomly chosen example in the training set. Split info, defined as the information required to determine the branch to which the example or instance belongs.

Decision tree

A decision tree has a flow chart-like tree framework, whereas an internal node denotes a feature or attribute, the branch denotes a decision rule, and each leaf node denotes the outcome. The top node is called the root node in a decision tree. It intends to partition by the value of the attribute. It further divides the tree recursively by naming recursive partitioning. For instance, if a dataset D contains instances or examples from n classes. The Gini Index (Gini (D)) is defined as follows, where py is the relative frequency of class y, i.e. (fatal, injury, and property damage) in dataset D,

$$gini(D) = 1 - \sum y = 1\ to\ n\ py^2 \quad (4)$$

If Dataset D is divided into subsets D_1 and D_2 with sizes N_1 and N_2, the Gini index of the split data contains instances from n classes i.e. (fatal, injury and property damage), Gini (D) is defined as,

$$gini\ \text{split}(D) = \frac{N_1}{N} gini(D_1) + \frac{N_2}{N} gini(D_2) \quad (5)$$

The attribute providing the smallest gini split (T) is chosen to split the node. CART recursively expands the tree from a root node and then gradually prunes back to form the large tree. Other important DT parameters considered include minimum records per node and number of threads or average split point and nominal binary split.

eXtreme gradient boosting (XGBoost)

XGBoost is an ensemble technique developed on the basis of Friedman's proposed Gradient Boosting approach (Friedman, 2001). Chen and Guestrin added some improvement to the original Gradient Boosting Decision Tree (GBDT) algorithm in 2016, and it was then named as the XGBoost (T. Chen & Guestrin, 2016). XGBoost and Gradient Boosting (GB) are both tree-based ensemble algorithms. Practically, all the tree-based ensemble algorithms usually consist of several trees that further combine to improve classification accuracy. Therefore, for all ensemble models, a general model ($\hat{y}$) can be written as a summation of all classification scores for samples from all trees (x) It learns a series of classification and regression trees (CARTs) in parallel and receives the outcome by summing each CART score. The overall formulation of gradient boosting models is provided in the following relation (equation 1).

$$\check{y}_i(x) = \sum_{T=1}^{T} f_T(x_i), (f_T \in F) \quad (6)$$

Where T is the number of trees, and F is space for all the trees. This equation is further optimized for the objective function,

$$\text{Obj}(\theta) = \sum_{i=1}^{n} l(y_i, \check{y}_i) + \sum_{T=1}^{T} \Omega(f_T) \quad (7)$$

The first term is the loss function that helps to differentiate between (y_i) target and ($\hat{y}_i$) prediction. The latter term represents the regularisation that controls the complexity of the model and avoid overfitting. In contrast to the GB, which only takes into account a learning rate (L) for the regularization term that further decreases the impact of each successive tree, XGB introduces an incremental regularization term as below.

$$\Omega(f_T) = \gamma \text{t} + \frac{1}{2} \lambda \sum_{\text{j}=1}^{\text{t}} \text{C}_{\text{q(x)}^2 \text{j}} \quad (8)$$

Where t is the total number of leaves $C_{q(x)^2 j}$ is the score for $j$thleaf. λ and γ are regularization parameters. In addition, XGBoost is an enhanced GB implementation that is widely recognized for its accuracy, effectiveness, and ease of implementation, among other machine learning algorithms. These benefits allow XGBoost to produce better results than traditional GBDT. According to the results of the experiments in our case study, it also outperforms other commonly used machine learning algorithms.

Model evaluation

In this research, the most commonly known performance metrics were utilized to test the efficiency of the various techniques. These include confusion metrics, precision, recall, accuracy, and F-1 score. For classification problems, the confusion matrix is made up of four possible scenarios, i.e., true (TP) positive rate, true negative (TN) rate, false positive (FP) rate, and false negatives (FN) rate, that are shown in Table 3.

Accuracy is the percentage of the sample correctly classified/predicted to the total number of samples and is given by equation 9. The metric sensitivity is calculated using equation 10. Similarly, other performance measures like precision, F-measure, geometric mean, specificity, and Cohen's kappa can be calculated from equations 11, 12, 13, 14, and 15.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Geometric\ Mean\ (G.M) = (x_1, x_2, x_3, \ldots\ldots\ldots, x_n)^{1/n} \tag{12}$$

$$F\text{ - }measure = \frac{1}{Precision} + \frac{1}{Recall} \tag{13}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{14}$$

$$Cohen's\ kappa = \frac{p_o - p_e}{1 - p_e} \tag{15}$$

Where n represent the number of periods.

Table 3. Confusion matrix for evaluating model's performance.

	Predicted Condition	
Actual Condition	Positive	Negative
Positive	True Positives (TPs)	False Negatives (FNs)
Negative	False positives (FPs)	True Negatives (TNs)

Results and discussion

Feature importance analysis

One of the main criticism of machine learning models is the input information processing in a black box without explicitly emphasizing the contribution/role of individual predictor variables towards the target variable. Mere accurate and better prediction of the machine learning model does improve researchers' perception of variables importance and injury severity risk factors. One of the objectives of this research was to establish the relative importance of predictor variables/features in predicting crash severity. Feature importance is a technique that assigns a score to input features based on how important they are at predicting a target variable. The gain ratio feature evaluation method was adopted to explore the relative significance of each factor in predicting crash injury severity. It is an extension of the information gain, a measure frequently employed in the decision tree-based learning algorithm for finding the attribute importance. By applying stepwise normalization, the gain ratio metric subjugates the bias of information gain towards features with a large number of values (Han et al., 2011). For the current study, the XGBoost classifier features importance technique via scikit-learn was implemented in Python to establish the variables relative importance based on quantitative gain ratio scores.

Figure 2 shows the relative importance ranking of each attribute based on the gain ratio evaluator obtained using XGBoost. A high value of the evaluator implies comparatively greater feature importance. The results suggest that explanatory variables that heavily the crash injury severity outcome are collision type, weather status, road surface conditions, on-site damage road type, vehicle type, crash cause, number of lanes in each direction, lighting conditions, and road type. While the least sensitive variables are the day of the week, the hour of the day, crash season, AADT, shoulder and median width, and road alignment. Variable importance results showed herein pertaining to crash injury severity are mostly in agreement with number of previous studies from different regions (Abdel-Aty, 2003; Al-Ghamdi, 2002; C. Chen et al., 2016; L.-Y. Chang & Chien, 2013; Donnell & Mason, 2004; Fiorentini & Losa, 2020; Huang et al., 2011; Jamal & Umer, 2020; Liu et al., 2015; Z. Ma et al., 2015; Razi-Ardakani et al., 2019; Theofilatos & Yannis, 2014). However, few contributing factors, such as on-site damage road type, road surface conditions, road type, and the number of lanes, type are not commonly investigated in the literature, particularly in KSA and Gulf region context.

Models performance evaluation

Table 4 provides the model's predictive performance expressed in terms of respective confusion matrices for each classification algorithm which were obtained using the K-fold cross-validation (with k = 10) method. The confusion matrices given in Table 4 show the discrepancy between the predicted and actual observations for individual injury

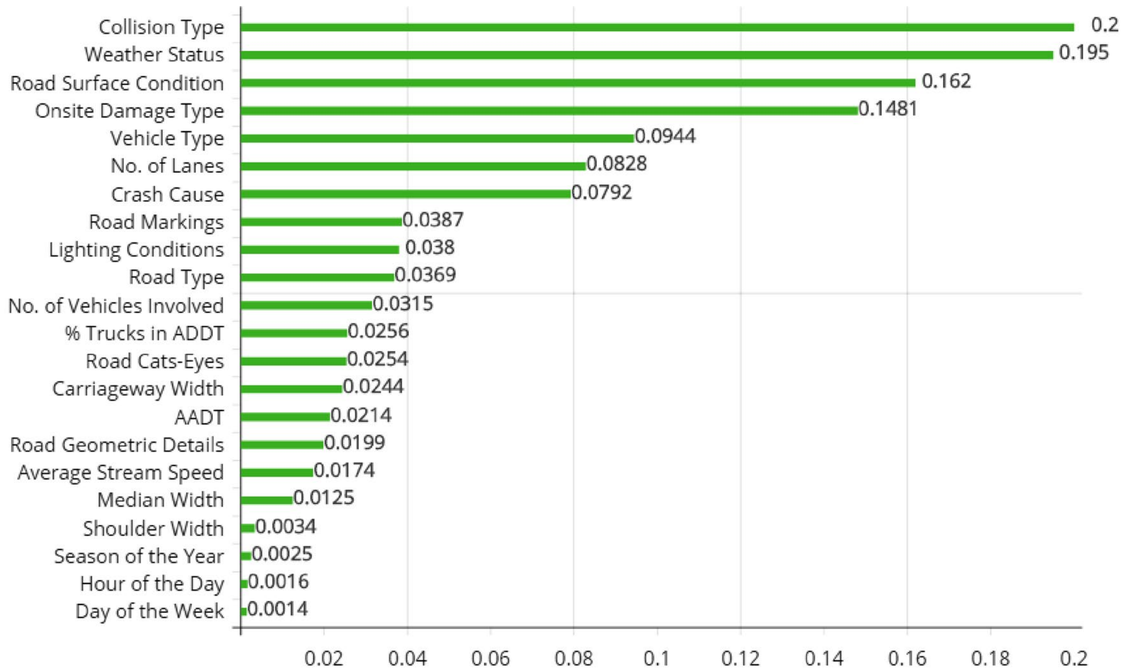

Figure 2. Variable importance based on XGBoost sensitivity analysis.

Table 4. Confusion matrix generated for different classifiers using k-fold cross-validation.

Classifier	Actual Severity Class	Predicted Severity Class		
		PDO	Injury	Fatal
Logistic Regression	PDO	5225 (74.37%)	702 (9.99%)	1099 (15.64%)
	Injury	1538 (27.67%)	3258 (58.61%)	763 (13.73%)
	Fatal	396 (41.21%)	456 (47.45%)	109 (11.34%)
Random Forest		PDO	Injury	Fatal
	PDO	7024 (99.97%)	1 (0.01%)	1 (0.01%)
	Injury	1181 (21.24%)	4375 (78.70%)	3 (0.05%)
	Fatal	490 (50.90%)	446 (46.41%)	25 (2.60%)
Decision Tree		PDO	Injury	Fatal
	PDO	6748 (96.04%)	0 (0.00%)	278 (3.96%)
	Injury	0 (0.00%)	5136 (92.39%)	423 (7.61%)
	Fatal	340 (35.38%)	438 (45.58%)	183 (19.04%)
XGBoost		PDO	Injury	Fatal
	PDO	6997 (99.59%)	0 (0.00%)	29 (0.41%)
	Injury	0 (0.00%)	5511 (99.14%)	48 (0.86%)
	Fatal	373 (38.81%)	501 (52.13%)	87 (9.05%)

severity classes in the dataset. In the contingency table shown below, each row in the matrix represents the actual number of observations for a specific crash category, while the columns indicate the predicted number of observations for a particular crash injury severity class. The cells' values across the diagonal imitate the accurate predictions, whereas the off-diagonal cells are indicative of misclassifications resulting in an over or underestimation of the model. The percentage (%) values in the brackets with each confusion matrix represents the ratio of the predicted number of crashes with severity level "i" (where i = PDO, Injury, and fatal injury) divided by the actual number of crashes in a specific crash category. As shown in Table 4, the XGBoost model yielded better severity prediction results compared to other machine learning algorithms. For the XGBoost model, 6,997 (99.60%) of PDO instances are correctly classified as PDO crashes. Only 29 cases of PDO crashes are misclassified as fatal. Similarly, crashes involving at least one injury, about 5,511 out of 5,559 crashes belonging to this category are correctly classified. In contrast, only 0.86% of the injury crashes are wrongly predicted as fatal crashes. Considering the fatal injury group, approximately 9% of the cases are correctly classified. The lower prediction performance of the model for fatal crashes may be attributed to the data imbalance issue that produces biased model estimations towards the majority class. This observation is intuitive and supported by studies in the existing literature (Bejjanki et al., 2020; Hasanin et al., 2019; Jiang et al., 2020; Pradhan & Sameen, 2020). The overall prediction accuracy of the XGBoost model is around 93%, which indicates an excellent model performance. The confusion matrices produced by Logistic Regression, Random Forest, and Decision Tree Trees can be interpreted in the same fashion. It is worth noting that all the algorithms produce an acceptable

Table 5. Summary of Models Classification Performance Using K-Fold Cross-validation.

Algorithms to classify crash injury severity	Classification results from stratified KFold cross-validation				
	Average accuracy (%)	Macro Average Precision	Macro Average Recall	F-measure	Specificity
XGBoost	0.95	0.81	0.70	0.65	0.97
Decision Tree	0.92	0.70	0.69	0.62	0.94
Random Forest	0.89	0.85	0.60	0.52	0.89
Logistic Regression	0.72	0.51	0.48	0.39	0.76

overall model performance and are particularly good in predicting the PDO and injury-related crashes.

Table 5 summarizes the predictive performance of all the four classification algorithms model performance using different classification evaluation metrics such as average accuracy, macro-average precision-recall, macro-average recall, F1 measure, and specificity. Average accuracy is the mean of class accuracies. Precision measures the exactness of a classifier. A low value of precision is indicative of several False Positives (FP). Recall also known as sensitivity, is a measure that determines the completeness of a classification algorithm. A low recall indicates many False Negatives (FN). The measure F-1 is determined by taking the harmonic mean of precision and recall values. Specificity measure the model's capability to predict true negatives for each category available in the dataset. As shown in Table 5, XGBost, with an average accuracy of approximately 95%, outperformed other models. On the other hand, logistic regression yielded the lowest average accuracy (72%) followed by Random Forest (89%). Similarly, collective model's predictive performance considering all other evaluation metrics, XGBoost provided better prediction.

Figure 3 presents a more illustrative plot for comparing the collective predictive performance of classification algorithms using a couple of other important classification evaluation measures, i.e., overall accuracy and Cohen's Kappa. Model overall accuracy is the ratio of the frequency of correctly predicted instances to the total of cases to predict. Literature suggests that it a better measure of model prediction instead of average model accuracy. Cohen's kappa statics is also frequently used in classification problems that measure the agreement between two raters that classify N number of items into C mutually exclusive groups/categories. Its value range from 0 (indicating random agreement) to 1 (showing complete agreement).

The accuracy and Cohen's Kappa were obtained for XGBoost using booster, tree method, and grow policy. The booster includes parameters such as maximum depth, minimum child node, column sampling rate by a tree, column sampling rate by level, subsampling rate, gamma, alpha, beta, and Lambda. The values of these parameters were 15, 1, 1, 1, 0.1, 1, 1 and 0.3, respectively. The tree method further includes two parameters, such as sketch epsilon, scale positive weight. The values for these parameters were 0.03 and 1. The grow policy was considered depth-wise, and the parameters for grow policy was the maximum number of leaves and the maximum number of bins. The values for these parameters were 2 and 256. Besides that, the objective function considered in XGBoost was a soft probe. The achieved XGBoost accuracy and Cohen's Kappa for crash severity were 0.93 and 0.87. Models comparison showed that XGBoost received high overall accuracy and Cohen's Kappa compared to Random Forest, Decision Tree, and Logistic Regression. The accuracy and Cohen's kappa obtained for Logistic Regression were 0.63 and 0.37. For DT and FR models, these values were 0.88, 0.79, and .84, 0.69 for overall accuracy and Cohen's Kappa, respectively.

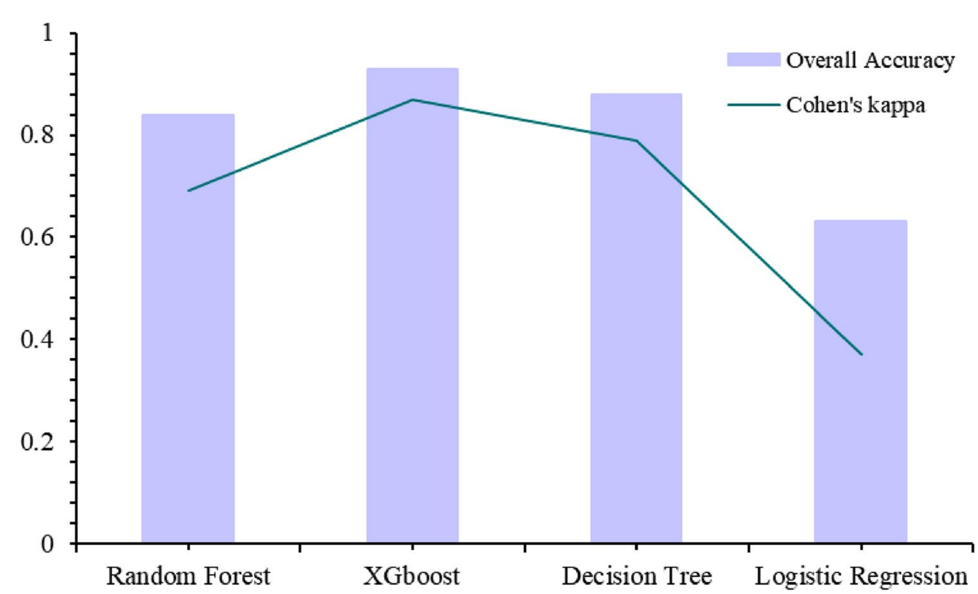

Figure 3. Models performance comparison based on overall accuracy and Cohen's Kappa.

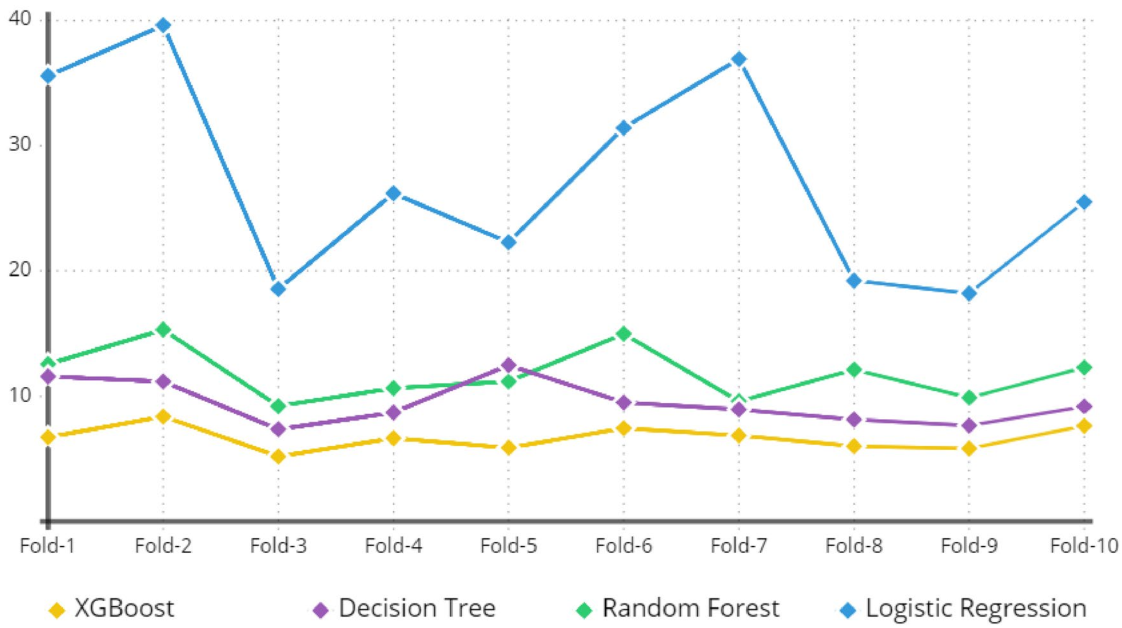

Figure 4. Error rate (%) per fold by 10-fold cross-validation method.

As mentioned, the model's prediction results were obtained using K-fold (using k = 10) cross-validation. In Figure 4, we showed an error rate per fold (expressed in %) for different models. The corresponding model prediction accuracy at each fold can be obtained by subtracting the respective values from 100. It may be noted from the figure that the error rates for all models fluctuating between the lowest values achieved at fold three, eight, and nine for different models. It may be argued from the results above that all the classifiers were found robust in predicting injury severity within the available dataset. However, XGBoost showed better predictive performance.

While it is important to achieve an acceptable overall model performance, knowing the predictive performance by each injury severity class is fundamental for prioritizing and selecting appropriate treatment and mitigation measures. Table 6 summarises the predictive accuracies by crash injury severity class for all the four classification algorithms. Explicitly, it presents the precision, specificity, average accuracy, and F-measure obtained using 10-fold cross-validation. The XGBoost achieved an accuracy of 95%, with a precision of 0.95, 0.92, 0.53 for PDO, injury, and fatal crashes, respectively. For the Decision Tree, the average accuracy achieved was around 92%, with a precision of 0.95, 0.92, 0.21 for PDO, injury, and fatal crashes, respectively. For Random Forest, the accuracy was 88%, with precision values of 0.81, 0.91, and 0.86 for the corresponding classes. Finally, Logistic Regression achieved an accuracy of 72% with precision values of 0.73, 0.74, and .06 for PDO, injury, and fatal crashes, respectively. The results reported in Table 6 suggest that all the classification algorithms have better performance in predicting PDO and injury crashes. The low prediction accuracy of the fatal crash injury severity category across all models may be attributed to their relatively small number of observations in the dataset. Considering all the performance metrics, for individual crash severity groups, the XGBoost model again outperformed other models.

XGBoost performance comparison with previous studies

In Table 7, we showed the injury severity prediction from the best performing model (XGBoost), with a few recent

Table 6. Performance metrics by individual severity class for different models.

Classifier	Class	Precision	Specificity	Average Accuracy	F-measure
XGBoost	Property Damage Only	0.95	0.94	0.97	0.94
	Injury	0.92	0.93	0.96	0.91
	Fatal	0.53	0.99	0.93	0.08
DT	Property Damage Only	0.95	0.93	0.95	0.91
	Injury	0.92	0.94	0.93	0.85
	Fatal	0.21	0.94	0.89	0.11
RF	Property Damage Only	0.81	0.72	0.87	0.81
	Injury	0.91	0.94	0.87	0.73
	Fatal	0.86	0.99	0.92	0.03
LR	Property Damage Only	0.73	0.63	0.70	0.58
	Injury	0.74	0.82	0.71	0.48
	Fatal	0.06	0.82	0.76	0.04

Table 7. Comparison of injury severity prediction performance of XGBoost with previous studies.

Crash Data Description	Number of classes/(breakdown, %)	Classification Algorithm	Overall Accuracy (%)	Algorithm Classification Results (ordered from least to most severe)						Studies
				Class 1	Class 2	Class 3	Class4	Class 5	Geometric Mean	
Five-year (2011-2015) of the data from Ghana (N = 8,516)	Four (Damage, 5.6)/ (Minor Injury, 29.4)/ (Major Injury, 42.1)/ (Fatal, 22.9)	Random Forest (RF)	73.91	47.4	N/A	77.2	79.8	58.2	64.20	(Wahab & Jiang, 2019a)
Three-years (2004-2006) of the data from Florida (N= 5,538)	Five (No Injury, 52.5)/ (Possible/Invisible, 26.4)/ No-capacitating Injury, 15.1)/ Incapacitating Injury, 5.1)/ (Fatal Injury, 0.9)	Support Vector Machine (SVM)	48.8	77.0	25.4	10.1	2.3	1.7	9.50	(Li et al., 2012)
Five-year of the crash data from Washington (N = 308,641)	Four (PDO, 82.6)/ (Possible Injury, 13.1)/ (Evident Injury, 3.6)/ (Severe and Fatal Injury, 0.8)	Gradient Boost (GB)	82.4	90.5	1.5	1.22	N/A	2.17	4.35	(Jiang et al., 2019)
Onr-year (2007) data from Tehran (N = 1,063)	Three (No Injury, 29.4)/ (Evident Injury, 42.1)/ (Fatality, 22.9)	Multilayer Perceptron (MLP)	77.37	83.72	N/A	69.36	N/A	75.87	76.58	(Kunt et al., 2012)
Six-year (2008-2013) of data from Abu Dhabi (N = 5,973)	Four (Minor, 59)/ (Moderate, 31)/ (Severe, 7)/ (Death, 3)	Multilayer Perceptron (MLP)	74.6	N/A	98.9	97.2	73.5	60.0	80.69	(Alkheder et al., 2017)
Three-year of the crash data from Florida (N = 5,538)	Four (No Injury, 52.5)/ (Possible/Invisible, 26.4)/ No-capacitating Injury, 15.1)/ Incapacitating Injury, 5.1)/ (Fatal Injury, 0.9)	Nearest Neighbor (KNN)	80.5	94.7	69.3	60.7	55.5	69.7	68.79	(Zhang et al., 2018)
Four-year (2011-2014) of data from Sydney (N = 42,102)	Three (Class 1, 20.22)/ (Class 2, 47.02)/ (Class 3, 32.76)	Decision tree (C4.5)	90.2	80.3	86.9	87.9	N/A	N/A	84.96	(Nguyen et al., 2017)
Eight-year (2009-2016) of data from Leeds (N = 21,436)	Three (Slight, 87.68)/ (Serious, 11.62)/ (Fatal, 0.70)	Convolutional Neural Network (CNN)	90.2	N/A	N/A	91.20	19.96	6.30	22.55	(Zheng et al., 2019)
Five-year (2011-2015) of the data from Ghana (N = 8,516)	Four (Damage, 5.6)/ (Minor Injury, 29.4)/ (Major Injury, 42.1)/ (Fatal, 22.9)	Classification and Regression Tree (CART)	73.81	43.90	N/A	77.15	80.02	57.38	62.80	(Wahab & Jiang, 2019b)
Two-years (2016-2017) of the data from Michigan (N= 297,113)	Five (No Injury, 84.54)/ (Possible/Invisible, 9.11)/ No-capacitating Injury, 4.53)/ Incapacitating Injury, 1.48)/ (Fatal Injury, 0.34)	Gradient Boost (GB)	54.19	52.90	24.41	9.14	41.38	58.55	30.99	(Jeong et al., 2018)

Two-years (2005) of the data from Taiwan (N= 1,620)	Three (No Injury, 60.2)/ Injury, 35.0)/ (Fatal Injury, 4.80)	Classification and Regression Tree (CART)	67.7	N/A	84.7	N/A	42.6	47.8	55.7	(Chang & Chien, 2013)
Two-years (2005) of the data from Mexico (N= 23,433)	Three (No Injury, 62.8)/ Injury, 29.0)/ (Fatal Injury, 8.2)	Bayesian Network(BN)	65.8	N/A	85.2	N/A	33.2	27.3	42.6	(Chen et al., 2015)
Two-years (2005) of the data from Mexico (N= 23,433)	Two (PDO, 91.24)/ (Fatal and Injury, 8.76)	Logistic Regression (LR)	62.53	92.3	N/A	N/A	N/A	0.00	–	(Fiorentini & Losa, 2020)
Two-years (2005) of the data from Mexico (N= 23,433)	Three (No Injury, 62.8)/ Injury, 29.0)/ (Fatal Injury, 8.2)	Bayesian Network(BN)	65.8	N/A	85.2	N/A	33.2	27.3	42.6	(Chen et al., 2015)
Three-years (2017-2019) of the data from KSA (N= 13,546)	Three (PDO, 51.87)/ Injury, 41.04)/ (Fatal, 7.09)	eXtreme Gradient Boosting (XGBoost)	93.10	94.1	N/A	N/A	91.24	8.04	41.02	Our Study

studies on crash severity prediction using machine learning methods. A brief description of data utilized, number and breakdown of injury severity categories, prediction model employed, along with overall prediction accuracies and per class accuracies are shown. Most of these studies have considered the data imbalance issue, while others have mentioned it as a prospect for future work. It is worth noting that the overall accuracy for the XGBoost achieved during the current study outperformed all of the existing literature reported herein. Similarly, the prediction accuracies by individual severity categories are also comparable with previous studies.

It is obvious that data with binary injury severity categories will have high overall and class prediction accuracy compared to the multi-class problem. Similarly, balanced data (having an equal proportion of injury severity categories) will exhibit better predictive performance compared to imbalanced data. Past studies have used different dataset, (specific to some jurisdiction) that contains information on diverse predictor variables with a common goal of improving the injury severity prediction using available data. Predictor variables present in one dataset from one location/country may be missing in others. Additionally, it is well-known that crashes exhibit spatial heterogeneity. Similarly, different jurisdictions report crash severities on different injury classes, usually ranging between two to five. All these factors account for a significant deviation in reported average injury severity accuracies reported in previous studies.

Conclusions

The analysis of crash injury severity is a promising research target in road safety studies. The present study is focussed on comparing the predictive performance of the XGBoost technique with traditional machine algorithms for crash injury severity modelling. The study also employed XGBoost based feature importance analysis to explore the key risk factors influencing the crash injury severity in motor vehicle collisions. Based on three years of crash data (2017-2019, N = 13546) collected along rural highways in KSA, four machine learning models were developed for predicting the crash injury severity associated with individual crashes. Various performance statistics (such as overall accuracy, average accuracy, Cohen's Kappa, precision, recall, F-measure, class accuracy, etc.) were used to compare the injury severity prediction performance among proposed methods. According to empirical cross-validation results, the XGBoost model with an overall prediction of 93% outperformed decision trees (88%), random forest (84%), and logistic regression (63%). Comparison by performance measures demonstrated the importance of determining the class prediction accuracy besides model collective performance. Variable importance analysis using XGBoost revealed that collision type, weather status, road surface conditions, on-site damage type, vehicle type, number of lanes, and crash cause are the most sensitive variables in predicting the crash injury severity outcome. Additionally, the proposed XGBoost outperformed most of the existing models by achieving high overall accuracy and better predictive performance among different injury severity classes.

This study does have some limitations that may be addressed in future studies. The current study utilized three years of limited crash data with no detailed drivers' detailed sociodemographic attributes. In the future, detailed datasets over prolonged periods covering expanded explanatory variables may be considered. The issue of lower prediction accuracy of fatal crash category arising due to data imbalance may be addressed by adopting appropriate over-and-under sampling strategies. Similarly, this study aimed at the severity prediction of all motor vehicle collisions from an aggregated viewpoint. Instead, severity prediction focussing on specific road users or crash types could reveal interesting and more useful insights, which may be vital for implementing improvement strategies. In conclusion, it may be argued XGBoost is promising and rapid-useful for crash severity forecasting that can help the agencies to allocate the safety improvement budget effectively.

Acknowledgments

The authors appreciate and gratefully acknowledge the support of KFUPM for conducting this study. The authors also acknowledge the support of the traffic safety center and traffic safety department at the ministry of transport (MOT), Riyadh, for providing the data used in this research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, *34*(5), 597–603. https://doi.org/10.1016/j.jsr.2003.05.009

Abdel-Aty, M. A., & Abdelwahab, H. T. (2004). Predicting injury severity levels in traffic crashes: a modeling comparison. *Journal of Transportation Engineering*, *130*(2), 204–210. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:2(204)

Abdelwahab, H. T., & Abdel-Aty, M. A. (2001). Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, *1746*(1), 6–13. https://doi.org/10.3141/1746-02

Abellán, J., López, G., & de Oña, J. (2013). Analysis of traffic accident severity using decision rules via decision trees. *Expert Systems with Applications*, *40*(15), 6047–6054. https://doi.org/10.1016/j.eswa.2013.05.027

Aguero-Valverde, J., & Jovanis, P. P. (2009). Bayesian multivariate poisson lognormal models for crash severity modeling and site ranking. *Transportation Research Record: Journal of the Transportation Research Board*, *2136*(1), 82–91. https://doi.org/10.3141/2136-10

Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident; Analysis and Prevention*, *34*(6), 729–741. https://doi.org/10.1016/S0001-4575(01)00073-2

Alkheder, S., Taamneh, M., & Taamneh, S. (2017). Severity prediction of traffic accident using an artificial neural network: Traffic accident severity prediction using artificial neural network. *Journal of Forecasting*, *36*(1), 100–108. https://doi.org/10.1002/for.2425

Al-Tit, A. A., Dhaou, I. B., Albejaidi, F. M., & Alshitawi, M. S. (2020). Traffic safety factors in the Qassim region of Saudi Arabia. *SAGE Open*, *10*(2), 215824402091950. 0. https://doi.org/10.1177/2158244020919500

Al-Turki, M., Jamal, A., Al-Ahmadi, H. M., Al-Sughaiyer, M. A., & Zahid, M. (2020). On the potential impacts of smart traffic control for delay, fuel energy consumption, and emissions: An NSGA-II-based optimization case study from Dhahran, Saudi Arabia. *Sustainability*, *12*(18), 7394. https://doi.org/10.3390/su12187394

Amiri, A. M., Sadri, A., Nadimi, N., & Shams, M. (2020). A comparison between artificial neural network and hybrid intelligent genetic algorithm in predicting the severity of fixed object crashes among elderly drivers. *Accident; Analysis and Prevention*, *138*, 105468. https://doi.org/10.1016/j.aap.2020.105468

Anderson, T. K. (2009). Kernel density estimation and K-Means clustering to profile road accident hotspots. *Accident; Analysis and Prevention*, *41*(3), 359–364. https://doi.org/10.1016/j.aap.2008.12.014

Ansari, S., Akhdar, F., Mandoorah, M., & Moutaery, K. (2000). Causes and Effects of Road Traffic Accidents in Saudi Arabia. *Public Health*, *114*(1), 37–39. https://doi.org/10.1016/S0033-3506(00)00306-1

Arhin, S. A., & Gatiba, A. (2020). Predicting crash injury severity at unsignalized intersections using support vector machines and naïve bayes classifiers. *Transportation Safety and Environment*, *2*(2), 120–132. https://doi.org/10.1093/tse/tdaa012

Azimi, G., Rahimi, A., Asgari, H., & Jin, X. (2020). Severity analysis for large truck rollover crashes using a random parameter ordered logit model. *Accident; Analysis and Prevention*, *135*, 105355. https://doi.org/10.1016/j.aap.2019.105355

Bejjanki, K. K., Gyani, J., & Gugulothu, N. (2020). Class imbalance reduction (CIR): A novel approach to software defect prediction in the presence of class imbalance. *Symmetry*, *12*(3), 407. https://doi.org/10.3390/sym12030407

Bendak, S. (2005). Seat belt utilization in saudi arabia and its impact on road accident injuries. *Accident; Analysis and Prevention*, *37*(2), 367–371. https://doi.org/10.1016/j.aap.2004.10.007

Bhalla, K., Mohan, D., & O'Neill, B. (2020). How much would low- and middle-income countries benefit from addressing the key risk factors of road traffic injuries?*International Journal of Injury Control and Safety Promotion*, *27*(1), 83–90. https://doi.org/10.1080/17457300.2019.1708411

Budiawan, W., Saptadi, S., Tjioe, C., & Phommachak, T. (2019). Traffic accident severity prediction using naive Bayes algorithm-a case study of Semarang toll road. *IOP Conference Series: Materials Science and Engineering*, *598*, 012089. IOP Publishing. https://doi.org/10.1088/1757-899X/598/1/012089

Chang, L.-Y., & Chien, J.-T. (2013). Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. *Safety Science*, *51*(1), 17–22. https://doi.org/10.1016/j.ssci.2012.06.017

Chang, F., Li, M., Xu, P., Zhou, H., Haque, M., & Huang, H. (2016). Injury severity of motorcycle riders involved in traffic crashes in Hunan, China: A mixed ordered logit approach. *International Journal of Environmental Research and Public Health*, *13*(7), 714. https://doi.org/10.3390/ijerph13070714

Chang, F., Xu, P., Zhou, H., Chan, A. H. S., & Huang, H. (2019, October). Investigating injury severities of motorcycle riders: A two-step method integrating latent class cluster analysis and random parameters logit model. *Accident; Analysis and Prevention*, *131*, 316–326. https://doi.org/10.1016/j.aap.2019.07.012

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, (pp. 785–94).

Chen, H., Chen, H., Liu, Z., Sun, X., & Zhou, R. (2020). Analysis of factors affecting the severity of automated vehicle crashes using XGBoost model combining POI data. *Journal of Advanced Transportation*, *2020*, 1–12. https://doi.org/10.1155/2020/8881545

Chen, C., Zhang, G., Huang, H., Wang, J., & Tarefder, R. A. (2016). Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. *Accident; Analysis and Prevention*, *96*, 79–87. https://doi.org/10.1016/j.aap.2016.06.015

Chen, C., Zhang, G., Qian, Z., Tarefder, R. A., & Tian, Z. (2016, May). Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accident; Analysis and Prevention*, *90*, 128–139. https://doi.org/10.1016/j.aap.2016.02.011

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., & Guan, H. (2015). A multinomial logit model-bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. *Accident; Analysis and Prevention*, *80*, 76–88. https://doi.org/10.1016/j.aap.2015.03.036

Chen, C., Zhang, G., Yang, J., Milton, J. C., & Alcántara, A. D. (2016). An explanatory analysis of driver injury severity in rear-end crashes using a decision table/naïve Bayes (DTNB) hybrid classifier. *Accident; Analysis and Prevention*, *90*, 95–107. https://doi.org/10.1016/j.aap.2016.02.002

Chong, M. M., Abraham, A., & Paprzycki, M. (2004). Traffic accident analysis using decision trees and neural networks. ArXiv Preprint Cs/0405050

Christoforou, Z., Cohen, S., & Karlaftis, M. G. (2010). Vehicle occupant injury severity on highways: An empirical investigation. *Accident Analysis and Prevention*, *42*(6), 1606–1620. https://doi.org/10.1016/j.aap.2010.03.019

Dadashova, B., Arenas-Ramires, B., Mira-McWillaims, J., Dixon, K., & Lord, D. (2020). Analysis of crash injury severity on two trans-European transport network corridors in Spain using discrete-choice models and random forests. *Traffic Injury Prevention*, *21*(3), 228–233. https://doi.org/10.1080/15389588.2020.1733539

Delen, D., Tomak, L., Topuz, K., & Eryarsoy, E. (2017, March). Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. *Journal of Transport & Health*, *4*, 118–131. https://doi.org/10.1016/j.jth.2017.01.009

DeNicola, E., Aburizaize, O. S., Siddique, A., Khwaja, H., & Carpenter, D. O. (2016). Road traffic injury as a major public health issue in the Kingdom of Saudi Arabia: A review. *Frontiers in Public Health*, *4*, 215. https://doi.org/10.3389/fpubh.2016.00215

Dong, N., Huang, H., & Zheng, L. (2015). Support vector machine in crash prediction at the level of traffic analysis zones: Assessing the spatial proximity effects. *Accident; Analysis and Prevention*, *82*, 192–198. (September): . https://doi.org/10.1016/j.aap.2015.05.018

Donnell, E. T., & Mason, J. M.Jr. (2004). Predicting the severity of median-related crashes in Pennsylvania by using logistic regression. *Transportation Research Record: Journal of the Transportation Research Board*, *1897*(1), 55–63. https://doi.org/10.3141/1897-08

Duan, J., Asteris, P. G., Nguyen, H., Bui, X.-N., & Moayedi, H. (2020). A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model. *Engineering with Computers*, 1–18. https://doi.org/10.1007/s00366-020-01003-0

Eluru, N., & Bhat, C. R. (2007). A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis and Prevention*, *39*(5), 1037–1049. https://doi.org/10.1016/j.aap.2007.02.001

Fauzan, M. A., & Murfi, H. (2018). The accuracy of XGBoost for insurance claim prediction. *International Journal of Advances in Soft Computing and its Applications*, *10*(2), 159–171.

Fiorentini, N., & Losa, M. (2020). Handling imbalanced data in road crash severity prediction by machine learning algorithms. *Infrastructures*, *5*(7), 61. https://doi.org/10.3390/infrastructures5070061

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232.

Garrido, R., Bastos, A., de Almeida, A., & Elvas, J. P. (2014). Prediction of road accident severity using the ordered probit model. *Transportation Research Procedia*, *3*, 214–223. https://doi.org/10.1016/j.trpro.2014.10.107

Ghasemzadeh, A., & Ahmed, M. M. (2017). 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.

Griselda, L., Juan, D. O., & Joaquín, A. (2012). Using decision trees to extract decision rules from police reports on road accidents. *Procedia - Social and Behavioral Sciences*, *53*, 106–114. https://doi.org/10.1016/j.sbspro.2012.09.864

Guo, M., Yuan, Z., Janson, B., Peng, Y., Yang, Y., & Wang, W. (2021). Older pedestrian traffic crashes severity analysis based on an emerging machine learning XGBoost. *Sustainability*, *13*(2), 926. https://doi.org/10.3390/su13020926

Haghighi, N., Liu, X. C., Zhang, G., & Porter, R. J. (2018). Impact of roadway geometric features on crash severity on rural two-lane highways. *Accident; Analysis and Prevention*, *111*, 34–42. https://doi.org/10.1016/j.aap.2017.11.014

Han, J., Pei, J., & M. Kamber (2011). *Data Mining: Concepts and techniques*. Elsevier.

Haque, M. M., Chin, H. C., & Huang, H. (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident; Analysis and Prevention*, *42*(1), 203–212. https://doi.org/10.1016/j.aap.2009.07.022

Hasanin, T., Khoshgoftaar, T. M., Leevy, J. L., & Bauder, R. A. (2019). Severely imbalanced big data challenges: investigating data sampling approaches. *Journal of Big Data*, *6*(1), 1–25. https://doi.org/10.1186/s40537-019-0274-4

Huang, H., Chin, H. C., & Haque, M. M. (2008). Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. *Accident; Analysis and Prevention*, *40*(1), 45–54. https://doi.org/10.1016/j.aap.2007.04.002

Huang, H., Siddiqui, C., & Abdel-Aty, M. (2011). Indexing crash worthiness and crash aggressivity by vehicle type. *Accident; Analysis and Prevention*, *43*(4), 1364–1370. https://doi.org/10.1016/j.aap.2011.02.010

Ijaz, M., Lan, L., Zahid, M., & Jamal, A. (2021, May). A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw. *Accident; Analysis and Prevention*, *154*, 106094. https://doi.org/10.1016/j.aap.2021.106094

Iranitalab, A., & Khattak, A. (2017). Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis and Prevention*, *108*, 27–36. https://doi.org/10.1016/j.aap.2017.08.008

Jamal, A., & Umer, W. (2020). Exploring the injury severity risk factors in fatal crashes with neural network. *International Journal of Environmental Research and Public Health*, *17*(20), 7466. https://doi.org/10.3390/ijerph17207466

Jamal, A., Mahmood, T., Riaz, M., & Al-Ahmadi, H. M. (2021). GLM-based flexible monitoring methods: An application to real-time highway safety surveillance. *Symmetry*, *13*(2), 362. https://doi.org/10.3390/sym13020362

Jamal, A., Rahman, M. T., Al-Ahmadi, H. M., & Mansoor, U. (2019). The dilemma of road safety in the eastern province of Saudi Arabia: Consequences and prevention strategies. *International Journal of Environmental Research and Public Health*, *17* (1), 157. https://doi.org/10.3390/ijerph17010157

Jamal, A., Rahman, M. T., Al-Ahmadi, H. M., Ullah, I. M., & Zahid, M. (2020). Intelligent intersection control for delay optimization: Using meta-heuristic search algorithms. *Sustainability*, *12*(5), 1896. https://doi.org/10.3390/su12051896

Jeong, H., Jang, Y., Bowman, P. J., & Masoud, N. (2018). Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. *Accident; Analysis and Prevention*, *120*, 250–261. https://doi.org/10.1016/j.aap.2018.08.025

Jiang, L., Xie, Y., & Ren, T. (2019). Modelling highly unbalanced crash injury severity data by ensemble methods and global sensitivity analysis. In Proceedings of the Transportation Research Board 98th Annual Meeting, Washington, DC, USA (pp. 13–17).

Jiang, L., Xie, Y., Wen, X., & Ren, T. (2020). Modeling highly imbalanced crash severity data by ensemble methods and global sensitivity analysis. *Journal of Transportation Safety & Security*, 1–23. https://doi.org/10.1080/19439962.2020.1796863

Kamruzzaman, M., Haque, M. M., & Washington, S. (2014). Analysis of traffic injury severity in Dhaka, Bangladesh. *Transportation*

Research Record: Journal of the Transportation Research Board, 2451(1), 121–130. https://doi.org/10.3141/2451-14
Khattak, A. J., Kantor, P., & Council, F. M. (1998). Role of adverse weather in key crash types on limited-access: Roadways implications for advanced weather systems. *Transportation Research Record: Journal of the Transportation Research Board*, *1621*(1), 10–19. https://doi.org/10.3141/1621-02
Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident; Analysis and Prevention*, *42*(6), 1751–1758. https://doi.org/10.1016/j.aap.2010.04.016
Kockelman, K. M., & Kweon, Y.-J. (2002). Driver injury severity: An application of ordered probit models. *Accident; Analysis and Prevention*, *34*(3), 313–321. https://doi.org/10.1016/S0001-4575(01)00028-8
Kopits, E., & Cropper, M. (2005). Traffic fatalities and economic growth. *Accident; Analysis and Prevention*, *37*(1), 169–178. https://doi.org/10.1016/j.aap.2004.04.006
Kunt, M. M., Aghayan, I., & Noii, N. (2012). Prediction for traffic accident severity: Comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. *Transport*, *26*(4), 353–366. https://doi.org/10.3846/16484142.2011.635465
Lee, C., & Abdel-Aty, M. (2008). Presence of passengers: Does it increase or reduce driver's crash potential?*Accident; Analysis and Prevention*, *40*(5), 1703–1712. https://doi.org/10.1016/j.aap.2008.06.006
Lee, S. M., & Al-Mansour, A. I. (2020). Development of a new traffic safety education material for the future drivers in the Kingdom of Saudi Arabia. *Journal of King Saud University - Engineering Sciences*, *32*(1), 19–26. https://doi.org/10.1016/j.jksues.2018.11.003
Li, Z., Chen, C., Ci, Y., Zhang, G., Wu, Q., Liu, C., & Qian, Z. S. (2018). Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical bayesian models. *Accident; Analysis and Prevention*, *120*, 139–151. https://doi.org/10.1016/j.aap.2018.08.009
Li, Z., Liu, P., Wang, W., & Xu, C. (2012). Using support vector machine models for crash injury severity analysis. *Accident; Analysis and Prevention*, *45*, 478–486. https://doi.org/10.1016/j.aap.2011.08.016
Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X., & Zhang, G. (2019). Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accident; Analysis and Prevention*, *129*, 230–240. https://doi.org/10.1016/j.aap.2019.04.001
Liu, D., Solomon, D., & Hardy, L. (2015). Investigating injury occurrence in motor vehicle collision using artificial neural networks: Investigating injury occurrence in motor vehicle collision. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *25*(3), 294–303. https://doi.org/10.1002/hfm.20543
Lu, P., Zheng, Z., Ren, Y., Zhou, X., Keramati, A., Tolliver, D., & Huang, Y. (2020). A gradient boosting crash prediction approach for highway-rail grade crossing crash analysis. *Journal of Advanced Transportation*, *2020*, 1–10. https://doi.org/10.1155/2020/6751728
Lu, Y., Fu, X., Guo, E., & Tang, F. (2021). XGBoost algorithm-based monitoring model for urban driving stress: Combining driving behaviour, driving environment, and route familiarity. *IEEE Access.*, *9*, 21921–21938. https://doi.org/10.1109/ACCESS.2021.3055551
Ma, C., Hao, W., Xiang, W., & Yan, W. (2018). The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents. *Journal of Advanced Transportation*, *2018*, 1–10. https://doi.org/10.1155/2018/9841498
Ma, J., Ding, Y., Cheng, J. C., Tan, Y., Gan, V. J., & Zhang, J. (2019). Analyzing the leading causes of traffic fatalities using XGBOOST and grid-based analysis: A city management perspective. *IEEE Access.*, *7*, 148059–148072. https://doi.org/10.1109/ACCESS.2019.2946401
Ma, Z., Zhao, W., Steven, I., Chien, J., & Dong, C. (2015). Exploring Factors Contributing to Crash Injury Severity on Rural Two-Lane Highways. *Journal of Safety Research*, *55*, 171–176. https://doi.org/10.1016/j.jsr.2015.09.003
Malyshkina, N. V., & Mannering, F. L. (2009). Markov switching multinomial logit model: An application to accident-injury severities. *Accident; Analysis and Prevention*, *41*(4), 829–838. https://doi.org/10.1016/j.aap.2009.04.006
Mauro, R., De Luca, M., & Dell'Acqua, G. (2013). Using a K-means clustering algorithm to examine patterns of vehicle crashes in before-after analysis. *Modern Applied Science*, *7*(10), 11. https://doi.org/10.5539/mas.v7n10p11
Meng, F., Xu, P., Wong, S. C., Huang, H., & Li, Y. C. (2017, November). Occupant-level injury severity analyses for taxis in Hong Kong: A Bayesian space-time logistic model. *Accident; Analysis and Prevention*, *108*, 297–307. https://doi.org/10.1016/j.aap.2017.08.010
Mesa-Arango, R., Valencia-Alaix, V. G., Pineda-Mendez, R. A., & Eissa, T. (2018). Influence of socioeconomic conditions on crash injury severity for an urban area in a developing country. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(31), 41–53. https://doi.org/10.1177/0361198118758684
Milton, J. C., Shankar, V. N., & Mannering, F. L. (2008). Highway accident severities and the mixed logit model: An exploratory empirical analysis. *Accident; Analysis and Prevention*, *40*(1), 260–266. https://doi.org/10.1016/j.aap.2007.06.006
Mohamed, H. A. (2015). Estimation of socio-economic cost of road accidents in Saudi Arabia: willingness-to-pay approach (WTP). *Advances in Management and Applied Economics*, *5*(3), 43.
Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2019). Improved support vector machine models for work zone crash injury severity prediction and analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(11), 680–692. https://doi.org/10.1177/0361198119845899
Mokoatle, M., Marivate, V., & Bukohwo, M. E. (2019). Predicting road traffic accident severity using accident report data in South Africa. In *Proceedings of the 20th Annual International Conference on Digital Government Research* (pp. 11–17).
Mondal, A. R., Bhuiyan, M. A. E., & Yang, F. (2020). Advancement of weather-related crash prediction model using nonparametric machine learning algorithms. *SN Applied Sciences*, *2*(8), 1–11.
Mousa, S. R., Bakhit, P. R., & Ishak, S. (2019). An extreme gradient boosting method for identifying the factors contributing to crash/near-crash events: A naturalistic driving study. *Canadian Journal of Civil Engineering*, *46*(8), 712–721. https://doi.org/10.1139/cjce-2018-0117
Nguyen, H., Cai, C., & Chen, F. (2017). Automatic classification of traffic incident's severity using machine learning approaches. *IET Intelligent Transport Systems*, *11*(10), 615–623. https://doi.org/10.1049/iet-its.2017.0051
O'Donnell, C. J., & Connor, D. H. (1996). Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. *Accident Analysis & Prevention*, *28*(6), 739–753. https://doi.org/10.1016/S0001-4575(96)00050-4
Oña, J. d., López, G., & Abellán, J. (2013). Extracting decision rules from police accident reports through decision trees. *Accident; Analysis and Prevention*, *50*, 1151–1160. https://doi.org/10.1016/j.aap.2012.09.006
Osman, M., Paleti, R., Mishra, S., & Golias, M. M. (2016). Analysis of injury severity of large truck crashes in work zones. *Accident; Analysis and Prevention*, *97*, 261–273. https://doi.org/10.1016/j.aap.2016.10.020
Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. K. (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident; Analysis and Prevention*, *136*, 105405. https://doi.org/10.1016/j.aap.2019.105405
Pradhan, B., & Sameen, M. I. (2020). Predicting injury severity of road traffic accidents using a hybrid extreme gradient boosting and deep neural network approach. In *Laser scanning systems in highway and safety assessment* (pp. 119–127). Springer.
Qu, Y., Lin, Z., Li, H., & Zhang, X. (2019). Feature recognition of urban road traffic accidents based on GA-XGBoost in the context of big data. *IEEE Access.*, *7*, 170106–170115. https://doi.org/10.1109/ACCESS.2019.2952655

Razi-Ardakani, H., Mahmoudzadeh, A., & Kermanshah, M. (2019). What factors results in having a severe crash? A closer look on distraction-related factors. *Cogent Engineering*, *6*(1), 1708652. https://doi.org/10.1080/23311916.2019.1708652

Rifaat, S. M., & Chin, H. C. (2007). Accident severity analysis using ordered probit model. *Journal of Advanced Transportation*, *41*(1), 91–114. https://doi.org/10.1002/atr.5670410107

Rusli, R., Haque, M. M., Saifuzzaman, M., & King, M. (2018). Crash severity along rural mountainous highways in Malaysia: An application of a combined decision tree and logistic regression model. *Traffic Injury Prevention*, *19*(7), 741–748. https://doi.org/10.1080/15389588.2018.1482537

Russo, B. J., Savolainen, P. T., Schneider, W. H., IV, & Anastasopoulos, P. C. (2014). Comparison of factors affecting injury severity in angle collisions by fault status using a random parameters bivariate ordered probit model. *Analytic Methods in Accident Research*, *2*, 21–29. https://doi.org/10.1016/j.amar.2014.03.001

Sameen, M. I., & Pradhan, B. (2017). Severity prediction of traffic accidents with recurrent neural networks. *Applied Sciences*, *7*(6), 476. https://doi.org/10.3390/app7060476

Sarkar, A., & Sarkar, S. (2020). Comparative assessment between statistical and soft computing methods for accident severity classification. *Journal of the Institution of Engineers (India): Series A*, *101*(1), 27–40. https://doi.org/10.1007/s40030-019-00422-7

Sarkar, S., Pramanik, A., Maiti, J., & Reniers, G. (2020). Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data. *Safety Science*, *125*, 104616. https://doi.org/10.1016/j.ssci.2020.104616

Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident; Analysis and Prevention*, *43*(5), 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025

Savolainen, P., & Mannering, F. (2007). Probabilistic Models of motorcyclists' injury severities in single- and multi-vehicle crashes . *Accident; Analysis and Prevention*, *39*(5), 955–963. https://doi.org/10.1016/j.aap.2006.12.016

Schlögl, M., Stütz, R., Laaha, G., & Melcher, M. (2019). A comparison of statistical learning methods for deriving determining factors of accident occurrence from an imbalanced high resolution dataset. *Accident; Analysis and Prevention*, *127*, 134–149. https://doi.org/10.1016/j.aap.2019.02.008

Shankar, V., & Mannering, F. (1996). An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of Safety Research*, *27*(3), 183–194. https://doi.org/10.1016/0022-4375(96)00010-2

Shankar, V., Mannering, F., & Barfield, W. (1996). Statistical Analysis of accident severity on rural freeways. *Accident Analysis & Prevention*, *28*(3), 391–401. https://doi.org/10.1016/0001-4575(96)00009-7

Tang, J., Liang, J., Han, C., Li, Z., & Huang, H. (2019). Crash injury severity analysis using a two-layer stacking framework. *Accident Analysis and Prevention*, *122*, 226–238. https://doi.org/10.1016/j.aap.2018.10.016

Tauhidur Rahman, M., Jamal, A., & Al-Ahmadi, H. M. (2020). Examining hotspots of traffic collisions and their spatial relationships with land use: A GIS-based geographically weighted regression approach for Dammam, Saudi Arabia. *ISPRS International Journal of Geo-Information*, *9*(9), 540–522. https://doi.org/10.3390/ijgi9090540

Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident; Analysis and Prevention*, *72*, 244–256. https://doi.org/10.1016/j.aap.2014.06.017

Tulu, G. S., Washington, S., Haque, M. M., & King, M. J. (2017). Injury severity of pedestrians involved in road traffic crashes in Addis Ababa, Ethiopia. *Journal of Transportation Safety & Security*, *9*(sup1), 47–66. https://doi.org/10.1080/19439962.2016.1199622

Ullah, I., Liu, K., Yamamoto, T., Zahid, M., & Jamal, A. (2021). Electric vehicle energy consumption prediction using stacked generalization: An ensemble learning approach. *International Journal of Green Energy*, 1–14. https://doi.org/10.1080/19439962.2020.1796863

Wahab, L., & Jiang, H. (2019a). A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. *PLoS One*, *14*(4), e0214966. https://doi.org/10.1371/journal.pone.0214966

Wahab, L., & Jiang, H. (2019b). Severity prediction of motorcycle crashes with machine learning methods. *International Journal of Crashworthiness*, 25(5), 1–8.

Wang, C., Deng, C., & Wang, S. (2020). Imbalance-XGBoost: Leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern Recognition Letters*, *136*, 190–197. https://doi.org/10.1016/j.patrec.2020.05.035

Wang, J., Huang, H., Xu, P., Xie, S., & Wong, S. C. (2020). Random parameter probit models to analyze pedestrian red-light violations and injury severity in pedestrian–motor vehicle crashes at signalized crossings. *Journal of Transportation Safety & Security*, *12*(6), 818–837. https://doi.org/10.1080/19439962.2018.1551257

Wang, W., Liu, C., & Chen, D. (2011). Predicting driver injury severity in freeway rear-end crashes by support vector machine. In Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE) (pp. 1800–1803). IEEE. https://doi.org/10.1109/TMEE.2011.6199563

Wang, X., & Kim, S. H. (2019). Prediction and factor identification for crash severity: Comparison of discrete choice and tree-based models. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(9), 640–653. https://doi.org/10.1177/0361198119844456

Washington, S., Karlaftis, M. G., Mannering, F., & Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*. CRC press.

World Health Organization. (2019). *Global Status Report on Road Safety*. World Health Organization.

Xiong, Y., Tobias, J. L., & Mannering, F. L. (2014, September). The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity. *Transportation Research Part B: Methodological*, *67*, 109–128. https://doi.org/10.1016/j.trb.2014.04.007

Xu, P., & Huang, H. (2015). Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis and Prevention*, *75*, 16–25. https://doi.org/10.1016/j.aap.2014.10.020

Yasmin, S., Eluru, N., & Ukkusuri, S. V. (2014). Alternative ordered response frameworks for examining pedestrian injury severity in New York City. *Journal of Transportation Safety & Security*, *6* (4), 275–300. https://doi.org/10.1080/19439962.2013.839590

Yu, R., & Abdel-Aty, M. (2014a, January). Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. *Accident Analysis & Prevention*, *62*, 161–167. https://doi.org/10.1016/j.aap.2013.08.009

Yu, R., & Abdel-Aty, M. (2014b, March). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science*, *63*, 50–56. https://doi.org/10.1016/j.ssci.2013.10.012

Zahid, M., Chen, Y., Jamal, A., & Mamadou, C. Z. (2020). Freeway short-term travel speed prediction based on data collection time-horizons: A fast forest quantile regression approach. *Sustainability*, *12*(2), 646. https://doi.org/10.3390/su12020646

Zahid, M., Chen, Y., Jamal, A., & Memon, Q. M. (2020). Short Term traffic state prediction via hyperparameter optimization based classifiers. *Sensors*, *20*(3), 685. https://doi.org/10.3390/s20030685

Zahid, M., Chen, Y., Jamal, A., Al-Ahmadi, H. M., & Khalaf Al-Ofi, A. (2020). Adopting machine learning and spatial analysis techniques for driver risk assessment: Insights from a case study. *International Journal of Environmental Research and Public Health*, *17*(14), 5193–5115. https://doi.org/10.3390/ijerph17145193

Zahid, M., Chen, Y., Khan, S., Jamal, A., Ijaz, M., & Ahmed, T. (2020). Predicting risky and aggressive driving behavior among taxi drivers: Do spatio-temporal attributes matter?*International Journal of Environmental Research and Public Health*, *17*(11), 3937. https://doi.org/10.3390/ijerph17113937

Zeng, Q., & Huang, H. (2014, December). A stable and optimized neural network model for crash injury severity prediction. *Accident;*

Analysis and Prevention, *73*, 351–358. https://doi.org/10.1016/j.aap.2014.09.006

Zeng, Q., Wen, H., Huang, H., Pei, X., & Wong, S. C. (2017). A multivariate random-parameters tobit model for analyzing highway crash rates by injury severity. *Accident; Analysis and Prevention*, *99*(Pt A), 184–191. https://doi.org/10.1016/j.aap.2016.11.018

Zhang, J., Li, Z., Pu, Z., & Xu, C. (2018). Comparing prediction performance for crash injury severity among various machine learning and statistical methods. *IEEE Access.*, *6*, 60079–60087. https://doi.org/10.1109/ACCESS.2018.2874979

Zhang, Y., & Xie, Y. (2007). Forecasting of short-term freeway volume with v-support vector machines. *Transportation Research Record: Journal of the Transportation Research Board*, *2024*(1), 92–99. https://doi.org/10.3141/2024-11

Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z., & Wang, Z. (2019). Traffic accident's severity prediction: A deep-learning approach-based CNN network. *IEEE Access.*, *7*, 39897–39910. https://doi.org/10.1109/ACCESS.2019.2903319

Zhou, X., Lu, P., Zheng, Z., Tolliver, D., & Keramati, A. (2020, August). Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliability Engineering & System Safety*, *200*, 106931. https://doi.org/10.1016/j.ress.2020.106931