

Full Length Article

Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers

J.P.S. Shashiprabha Madushani^a, R.M. Kelum Sandamal^a, D.P.P. Meddage^b, H.R. Pasindu^{c,*}, P.I. Ayantha Gomes^a^a Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Sri Lanka^b School of Engineering and Information Technology, University of New South Wales, Australia^c Department of Civil Engineering, Faculty of Engineering, University of Moratuwa, Sri Lanka

ARTICLE INFO

Keywords:

Explainable machine learning
Machine learning
Traffic crash severity
Expressways
Logistic regression

ABSTRACT

The number of expressway road accidents in Sri Lanka has significantly increased (by 20%) due to the expansion of the transport network and high traffic volume. It is crucial to identify the causes of these crashes for effective road safety management. However, traditional statistical methods may be insufficient due to their inherent assumptions. This study utilized explainable machine learning to investigate the factors that affect the severity of traffic crashes on expressways. The study evaluated two groups of traffic crashes: fatal or severe crashes, and other crashes that included non-severe injuries or only property damage. Five factors that contribute to crashes were analyzed: road surface condition, road alignment, location, weather condition, and lighting effect. Four machine learning models (Random Forest (RF), Decision Tree (DT), extreme gradient boosting (XGB), K-Nearest Neighbor (KNN)) were developed and compared with Logistic Regression (LR) using 223 training and 56 testing data instances. The study revealed that the machine learning algorithms provided more accurate predictions than the LR model. To explain the machine learning models, Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) were used. These methods revealed that all five features decreased the possibility of occurrence of fatal accidents. SHAP and LIME explanations confirmed the known interactions between factors influencing crash severity in expressway operational conditions. These explanations increase the trust of end-users and domain experts on machine learning models. Furthermore, the study concluded that using explainable machine learning methods is more effective than traditional regression analysis in evaluating safety performance. Additionally, the results of the study can be utilized to improve road safety by providing accurate explanations for decision-making processes for black-box models.

1. Introduction

Traffic crashes represent one of the most significant social problems globally, as reported by the World Health Organization (WHO) in 2022. Their data revealed that approximately 1.3 million individuals lose their lives each year whereas an additional 50 million suffer from non-fatal injuries on roads [1]. Surprisingly, despite low-middle-income countries possessing only 60% of the world's vehicles, they accounted for 93% of road traffic deaths [1]. It is crucial to identify the factors that contribute to the severity of these crashes, to effectively reduce traffic crash fatalities and serious injuries [2]. Several studies have been conducted to explore these contributing factors from various perspectives, including driver behavior, roadway issues, and environmental effects

[3–5]. When examining the trends and contributing factors in low and middle-income countries, it becomes evident that Sri Lanka carries a significant burden of traffic crash incidents and their consequences [1]. Addressing these issues in this context is of paramount importance to ensure improved road safety and a reduction in the overall impact of traffic crashes.

Expressways were first constructed in Sri Lanka in 2010 and currently span a total length of 310 km, according to the Road Development Authority (RDA) of Sri Lanka (2022) [6]. These expressways have experienced an average of 1000 traffic crashes per year with driver factors such as speeding, overtaking, and fatigue/drowsiness being the primary causes [7,8]. Additionally, environmental factors like weather conditions and poor road surface conditions have also contributed to

* Corresponding author.

E-mail address: pasindu@uom.lk (H.R. Pasindu).<https://doi.org/10.1016/j.treng.2023.100190>

Received 22 April 2023; Received in revised form 29 June 2023; Accepted 8 July 2023

Available online 9 July 2023

2666-691X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1

The summary of recent studies conducted on traffic crash severity analysis worldwide.

Casual factors	Key findings	Road network	References
Traffic condition	Low traffic volumes, higher percentages of heavy vehicles have increased the injury severity of drivers.	Crosstown roads in Spain	[20]
Road Geometry	The diverging diamond interchange (DDI) ramp terminals have 55%, and 31.4% lower fatal/injury crashes and property damage-only crashes respectively.	Freeways in Missouri	[23]
	Wider lanes, non-existence of lane markings, inattentive driving and exceeding speed have increased the injury severity of drivers.	Crosstown roads in Spain	[20]
	Straight road sections found to be one of highest contributing factor.	Lahore-Islamabad Highway M-2	[11]
Road surface	Wet roadway surfaces have increased the single-vehicle and multivehicle crash ratio by 8.87 and 2.82 times, respectively.	Freeways in Florida	[19]
	Road surface condition is significantly associated with the accident severity level.	Colombo, Sri Lanka	[10]
Weather condition	Precipitation, snow, temperature shown an impact on disability occurrence risk with a moderate degree.	Urban expressways in Central Shanghai City	[5]
	Rainy weather mostly affects the occurrence of accidents at identified prone locations.	Southern expressway, Sri Lanka	[24]
	Dry weather was found to be one of the critical contributing factors.	Lahore-Islamabad Highway M-2	[11]
Driver behavior	The leading cause found is the fault of the driver (77.1%).	Indian expressways	[25]

traffic crashes on the Southern Expressway [8].

Crashes are typically categorized based on severity levels, including serious or fatal, severe or minor injuries, and property damage [9]. In Sri Lanka, the same severity classification is adopted by RDA for national road crash analysis. However, there has been a lack of comprehensive analysis to evaluate the contribution of the aforementioned factors to crash severity.

Traditionally, crash data analysis has been conducted using statistical methods such as regression approaches [5,8,10–12]. Recently, Geographical Information System (GIS) based approaches and machine learning techniques have emerged as reliable methods for crash analysis worldwide [7,13–16].

This study aims to investigate the contribution of roadway and environmental factors to expressway crash severity using both Logistic Regression (LR) and explainable machine learning. Furthermore, it analyzes and compares the effectiveness of the explainable machine learning approach in safety analysis with the LR approach. Ultimately, the study recommends practices that should be emphasized on expressways under specific roadway and environmental conditions to reduce traffic crash fatalities and severe injuries.

2. Related work

2.1. Factors affecting road crashes in expressways

Expressways offer increased mobility with higher operating speeds. However, this also leads to a higher severity of road crashes. The risk of accidents rises with traffic volume and adverse weather conditions [17]. Therefore, it is crucial to identify the factors contributing to accidents related to roadway and weather conditions, despite the contingent, rare, and random nature of traffic crashes [17].

Geometric features of expressways, such as turning angles, downhill slopes, shoulder width, and the number of intersections, have a significant impact on crash severity [17–19]. Higher traffic volume and the presence of heavy vehicles are critical traffic factors associated with crash severity [17,19,20]. Driver-related factors, including age, gender, alcohol consumption, seat belt usage, and airbag deployment also play a substantial role in contributing to crash severity [21].

Weather conditions, such as wet road surfaces, fog, and gloomy environments, significantly affect expressway traffic crashes [19]. Expressways constructed in mountainous areas have a more pronounced impact on crash severity compared to those on flat terrains [17]. Additionally, vehicle speed is a critical consideration in crash analysis, since higher operating speeds than the posted speed limit on expressways dramatically increases the severity of crashes [22]. Recent studies have focused on evaluating the contribution of roadway and environmental factors to expressway crashes, as summarized in Table 1.

2.2. Accident risk and severity analysis

According to the WHO, unreliable and unsafe roads with defects in layout, design, and lack of maintenance are among the main causes of traffic crashes. Various road agencies have proposed safety evaluation criteria considering roadway, traffic data, environmental factors, and crash data, regardless of the functional class of the road [5,26–28].

Different indices and metrics have been adopted worldwide for safety analysis. Traffic crash analysis methods range from simple linear regression models to more advanced machine learning approaches [13]. LR is a highly effective method for traffic accident prediction due to its low data requirements and simple analysis structure.

Studies have used LR models to analyze accident data and identify key contributing factors. Paolo et al. [29] found heavy traffic in autumn and winter and speed limits to be key factors in an LR model analyzing accident data in Norway [29]. Dhananjaya & Alibuhitto [10] used a binomial LR model to explain variables affecting the probability of fatal accidents, including driver's age, lighting condition, vehicle condition, weather condition and location. Ma et al. [21] identified accident severity factors using an LR model, including accident location, road alignment, lighting condition and more [21].

In addition to LR models, various statistical approaches have been implemented in studies on accident severity analysis. Kodippili et al., [7] conducted a study on distinctive causes of traffic crashes on expressways using GIS-based interpolation approaches with different methods [7]. They found that uncontrolled speeding and unsuccessful overtaking were significant factors contributing to crash rates during evening hours.

However, standard regression models may not possess the necessary distributional characteristics to predict traffic crash causal factors [13]. Issues such as confounding variables and selection bias can affect the accuracy of regression analysis [30]. Overfitting or underfitting of regression models can also lead to unreliable estimates or miss important relationships. To address these problems, this study focused on incorporating machine learning approaches to analyze safety performance.

2.3. Machine learning approaches in safety performance analysis

In the field of safety and accident analysis, many researchers have employed machine learning-based classification methods. Chakraborty et al. [31] utilized tree-based classifiers and deep neural networks (DNN) to classify fatal accidents, non-fatal accidents, and property damage. They found that DNN performed better in predicting classes corresponding to fatal accidents while Decision Tree (DT) and Random Forest (RF) worked well for the remaining classes. They also used Granger causality to determine the feature importance of input parameters.

Table 2
Summary of crash analysis studies conducted in classifying road safety/accidents using machine learning.

Machine learning models	Road type	References
Decision Tree (DT)	Urban highways	[39,40,41]
	Two-way two-lane highways	[42]
	Highways	[43]
K-Nearest Neighbor (KNN)	Local, interstate and highway	[44]
	Freeway	[16]
	Urban highways	[40]
Random Forest (RF)	Local, interstate and highway	[44]
	Freeway	[16]
	Urban highways	[40,41,45]
Support Vector Machine (SVM)	Urban highways	[45,46]
	Local, interstate and highway	[44]
	Freeway	[16]
Classification and regression tree	Two-way two-lane highways	[42,47]
	Urban highways	[46]
Artificial Neural Network (ANN)	Urban highways	[39,45,48]
	Highways	[49,50]
Naïve Bayes	Highways	[43]
	Urban highways	[45]
Data fusion	Urban highways	[51]

Cigdem & Cevher [32] compared the performance of five machine learning classifiers (K-Nearest Neighbor (KNN), neural network, DT, Support Vector Machine (SVM), and Naïve Bayes) with traditional LR in classifying the severity of accidents. They observed that DT, KNN, and multilayer perception neural network achieved better performance than the other models. The presence of traffic control and ground surface temperature were identified as parameters that had a significant impact on the class prediction.

Yang et al. [33] developed a deep convolutional neural network to detect and classify freeway accidents, demonstrating that the model outperformed conventional methods. Das et al. [34] reported that the Extreme Gradient Boost (XGB) classifier showed good performance in classifying pedestrian crashes, achieving an accuracy of 77% in training and 72% in testing. Silva et al. [35] noted that DT, KNN, SVM, evolutionary algorithms, and Artificial Neural Networks (ANN) are commonly used machine learning algorithms in safety modeling.

The Least Square Support Vector Machine (LSSVM) is another effective method for small sample nonlinear problems [17]. Wang et al. [36] applied the LSSVM model and identified density, average annual daily traffic, and mean spacing distance as significant variables with an accuracy of 89% and an area under the curve (AUC) of 0.99. RF algorithm, which is a collection of classification trees, is also commonly used in safety analysis due to its ease of use [37,38]. Table 2 provides a summary of several studies that have employed machine learning techniques for classifying road safety and accidents.

In previous studies, the use of machine learning or deep learning methods has been prevalent, however the evaluation of these complex models through post-hoc explanation methods have been lacking. It is crucial to provide end-users with knowledge about how these machine learning models work, the importance and interaction of features, and dependencies among features to improve transparency. Statistical modeling techniques which rely on assumptions that may not always hold true in crash analysis due to issues like multicollinearity and unobserved heterogeneity can lead to prediction errors. On the other hand, machine learning models do not require pre-assumed relationships between variables.

Indeed, a significant research gap exists in providing proper explanations for machine learning-based classifiers, despite their higher accuracy compared to traditional methods. The transparency and interpretability of these classifiers are crucial, especially in the context of road safety, where the consequences of decisions can range from property damage to fatal accidents. Explaining the factors and features that contribute to the classification outcomes of machine learning models can greatly enhance the decision-making process and improve

the understanding of road safety issues. Furthermore, in the specific context of Sri Lanka, there is a lack of crash severity analysis conducted for expressways using data-driven methods. This highlights the need for comprehensive studies that specifically focus on analyzing the severity of accidents on expressways in Sri Lanka.

The objective of the present study is to investigate the performance of traditional regression models, and machine learning models with post hoc explanations to identify the crash severity of expressways in Sri Lankan context. This study is novel as it (a) uses accident data to be classified using supervised machine learning for the first time in an Asian context; (b) uses different machine learning algorithms to investigate the applicable algorithm in identifying crash severity (c) uses traditional regression models to compare their performance concerning machine learning models; (d) uses post-hoc methods to interpret the machine learning models and respective predicted classes (e) emphasizes the factors that govern the crash severity in the Sri Lankan context.

3. Materials & methods

3.1. Study area & data collection

In this study, the Southern Expressway (E01) has been selected as the study area for accident analysis. The Southern Expressway is a significant expressway in the Sri Lankan Road network, serving as a primary link. It is a four-lane expressway spanning a length of 126 km and designed for a speed of 120 km/h, with a speed limit of 100 km/h. The majority of vehicles on the expressway are passenger cars which is over 87%. The expressway mostly stretches over flat terrain, although the maximum longitudinal gradient observed is 7%.

Fig. 1 illustrates that the expressway can be divided into three segments based on their opening years: Kottawa Interchange (IC) to Pinnaduwa IC (2010), Pinnaduwa IC to Godagama IC (2014), and Godagama IC to Andarawewa IC (2020) [52]. Accident data for the study has been collected from RDA, Sri Lanka, covering the period from 2017 to 2021. However, it should be noted that there is limited data available for accident records specifically from the section between Godagama IC and Andarawewa IC as well as from the section between Kottawa IC and Godagama IC. The collected accident data includes information such as accident location, accident type, number and types of vehicles involved, number of fatalities and injuries, accident causes, road surface conditions, weather conditions, and lighting conditions.

Typically, accidents are categorized into four types such as property damage only, non-grievous, grievous and fatal based on the severity of the damage [8,54]. Grievous accidents are defined as accidents which consequence for serious injuries while non-grievous accidents account for only minor injuries. Moreover, Table 3 shows the accident distribution concerning severity levels on Southern Expressway from Kottawa IC to Godagama IC section (from Chainage 0 km to 125 km) from the year 2017 to 2021.

3.2. Data pre-processing

Data preprocessing will convert the data into a form which can be fed into the Machine Learning algorithms. In this study, the dependent variables and inputs variables were one-hot encoded using integer values. When we have non-numeric data, one-hot encoding is a fair way of converting them into numeric inputs. For example, the crashes are classified into two classes based on the severity of the crashes: fatal/grievous crashes as high severity and other crashes (non-grievous/property damage only crashes) as low severity. Fatal/grievous crashes were coded as 1 and the other crashes were coded as 0.

The roadway information and environmental conditions selected for this study include road surface condition (dry or wet), weather condition (clear, cloudy/foggy, or rainy), lighting condition (daylight, partial daylight, night with good street lighting or night with no street lighting), road alignment (curve or straight) and location (at an interchange or not

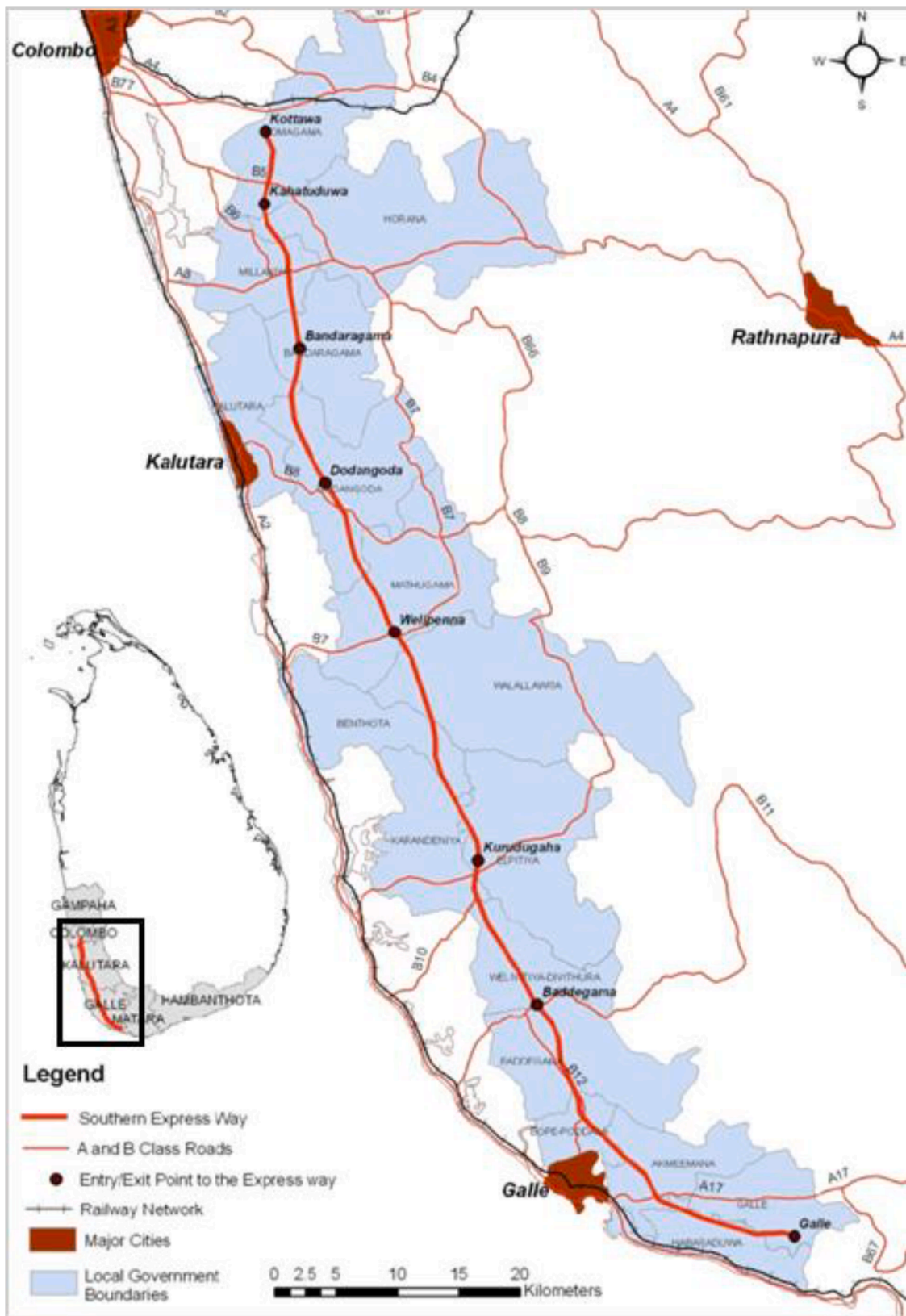


Fig. 1. Southern Expressway with interchanges [53].

at interchange). Thus, the analysis is further continued for identifying the contribution of the abovementioned factors on crash severity. Moreover, the numerical values for the input variables are listed in Table 4 (all the input variables are categorized into two or more levels,

and one-hot encoded as 0,1,2,3.

The total dataset initially consisted of 4720 data instances. However, during the survey, it was observed that some combinations of input variables were collected multiple times, resulting in replicates. These

Table 3
Accident distribution concerning severity levels on Southern expressway.

Year	Fatal crashes	Grievous crashes	Non-grievous crashes	Property damage only crashes	Total number of crashes per year
2017	5	13	69	802	889
2018	7	16	89	886	998
2019	10	7	92	904	1013
2020	5	14	49	735	803
2021	2	13	69	933	1017
Total number of crashes	29	63	368	4260	4720
Percentage	0.61	1.33	7.80	90.25	100.00

Table 4
Numerical values for explanatory variables.

Number	Variable	Codes/Values	Abbreviation
1	Location	0 = Not at an interchange 1 = At the interchange	LOC
2	Road alignment	0 = Straight 1 = Curve	ALIGN
3	Light condition	0 = Daylight 1 = Partial daylight 2 = Night with good street lighting 3 = Night with no street lighting	LIGHT
4	Road surface condition	0 = Dry 1 = Wet	SURF
5	Weather condition	0 = Clear 1 = Cloudy/Foggy 2 = Rainy	WEATH

Table 5
Descriptive statistics of five independent features.

Statistics	WEATH	LOC	ALIGN	LIGHT	SURF
Mean	0.47	0.27	0.73	1.06	0.25
Standard error	0.01	0.01	0.01	0.02	0.01
Standard deviation	0.80	0.44	0.44	1.38	0.43
Minimum	0.00	0.00	0.00	0.00	0.00
Maximum	2.00	1.00	1.00	3.00	1.00

replicates do not add any additional significance to the machine learning models and can potentially introduce bias or skewed results. Therefore, the authors decided to remove these replicates from the dataset. After removing the replicates, the final dataset comprised 279 unique instances. Among these instances, approximately 35% corresponded to fatal or grievous accidents, while the remaining 65% corresponded to non-grievous accidents and property damage accidents. This distribution of instances ensures a balanced representation of both classes in the dataset, allowing for a more accurate analysis and prediction of crash severity using machine learning models.

Moreover, the descriptive statistics of the variables is shown in Table 5.

3.3. Logistic regression model

A LR model is used to predict the accident type based on the severity of the crash in this study. Assume the severity of a crash as a random variable Y, follows a Bernoulli distribution with a success probability p (Eq. (1)).

$$Y \sim \text{Bernoulli}(p) \tag{1}$$

For LR, the log odds of a crash are a fatal or grievous crash is modeled

as in Eq. (2).

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 \tag{2}$$

Where, x_1 is the type of location, x_2 is the type of horizontal alignment, x_3 is the lighting condition, x_4 is the road surface condition and x_5 is the weather condition.

Thus, the probability of a crash is a fatal or grievous crash, p is Eqs. (3) & ((4)),

$$p = \frac{e^{g(x)}}{1 + e^{g(x)}} = \frac{1}{1 + e^{-g(x)}} \tag{3}$$

$$g(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 \tag{4}$$

3.4. Machine learning models

3.4.1. Decision tree classifier

DT analysis is a well-established analytical technique in data mining that creates a tree-based classification model that classifies cases or values of a dependent (target) variable based on the values of an independent (predictor) variable. DT analysis has strengths in identifying determinants compared to traditional regression modeling. First, it does not require specific rules for data format and can be used to study continuous independent, categorical and multivariate unordered categorical variables. Second, DT analysis is a non-parametric statistical method, with no specification of functional form and no hard assumptions about the distribution of the data. Third, it can explore interaction effects between independent variables and address multicollinearity issues [55].

Building a DT is a recurrent process involving two processes: tree building and tree pruning. The process of tree building is to classify the preliminary record level by level until it is no longer possible or necessary to split according to certain split criteria to properly generate the tree [56]. Specifically, in each classification, the model compares the differences for each branch obtained using different independent variables as classification variables and compares the independent variables that make the most significant differences to the classification variables of the nodes. The above process can produce huge trees that require pruning to reduce tree nodes, control tree complexity, and measure complexity by the number of leaf nodes in the tree.

Tree models can be divided into two categories based on the type of dependent variable: classification trees and regression trees. Commonly used algorithms include CHAID (Chi-square Automatic Interaction Detector), CRT (Classification and Regression Trees), and QUEST (Quick, Unbiased, Efficient, and Statistical Tree). CRT algorithm divides the data into segments and constructs this tree mode that attempts to maximize the homogeneity of the values of the dependent variable within the nodes [57].

3.4.2. K-Nearest Neighbor (KNN) classifier

KNN rule which is a distribution-free statistic pattern classification was introduced in 1951. However, it was not until the 1960s that the method became popular, and since then it has been widely used in pattern recognition and classification due to the advancement of computational power [58]. KNN is learned by comparing a given test tuple to a set of similar training tuples. K stands for the number of neighbors considered in determining the class [59]. The KNN simply stores a given training tuple, waits until it receives a test tuple, and performs a generalization to rank the tuples based on similarity or distance. They are often called "Lazy learner" or "instance-based learner" but do more work in the case of classification and prediction. Since unlike the other models, it uses the classification model before classifying the test tuples it received, it wants to classify all invisible tuples.

KNN classifiers typically enforce either Euclidean distance or cosine similarity between training and test tuples. We used the Euclidean

distance approach for this study [58].

Consider two tuples for example $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ and $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$ the Euclidean distance of two tuples can be obtained from the Eq. 5.

$$dist(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (5)$$

3.4.3. Random Forest classifier

The RF classifier is a DT algorithm-based meta-estimator [57]. RF classifier fits a series of DTs to different subsamples of a given dataset and uses averaging to improve the accuracy of prediction and reduce overfitting. It is therefore described as an ensemble learning method that can be used for both classification and regression [60]. RF classifiers are designed to reduce variance by averaging multiple deep DTs trained on different parts of the same training set, so they can be expected to consistently outperform DTs. Moreover, it makes accurate predictions and can handle large numbers of input variables. Perhaps the main drawback of RF is that the implementation of a large number of DTs causes the internal decision logic to be lost to the user. At the very least, the user should understand which variables lead to predictions. Moreover, the ensemble approach is computationally expensive and can make the algorithm less efficient when time is constrained. RF is extremely versatile and has been used in various areas [61].

3.4.4. Extreme gradient boosting classifier

XGB is a highly effective method for data classification. It is one of the highly scalable end-to-end tree-boosting systems used in machine learning [46]. The workflow of XGB can be explained as follows [62].

Consider a tree ensemble method of classification and regression trees (CARTs) with a set of K_k^i $i \in 1 \dots K$ nodes. For each tree k^{th} , total prediction scores at a leaf node f_k are predicted to obtain the final prediction output of class label \hat{y}_i , as expressed in Eq. (6).

$$\hat{y}_i = \varphi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (6)$$

Eq. (7) gives a regularization step that improves the validity of the results. Where the K score for all charts is represented by set F and x_i denotes the training set.

$$\mathcal{L}(\varphi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \quad (7)$$

Where l represents the differentiable loss function, defined by computing the error difference between the target y_i and the predicted class label \hat{y}_i . The second part performs a penalization Ω on model complexity to avoid overfitting problems. The penalty Ω function is calculated by Eq. (8).

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (8)$$

Here w stores the value of weights for each leaf whereas T represents the leaves in the tree. Where γ and λ are known as configurable parameters which control the level of regularization.

3.5. Explainable Artificial Intelligence

Explainable artificial intelligence (XAI) improves domain experts' and Machine Learning users' confidence by revealing the causality of machine learning outcomes [63–65]. For simple models such as DTs, intrinsic explanations are adequate. When models are complex, an extrinsic (post-hoc) interpretation is required (e.g. LIME [66], RISE [67], and SHAP [68], etc). These explanations are a necessary add-in to Machine Learning predictions by giving hidden reasoning. [64] conducted a comprehensive review of machine learning interpretability

Table 6

Pairwise correlation between independent features.

Variable	WEATH	LOC	ALIGN	LIGHT	SURF
WEATH	1				
LOC	-0.09	1			
ALIGN	-0.09	-0.19	1		
LIGHT	-0.01	-0.09	-0.21	1	
SURF	0.88	-0.08	-0.09	-0.01	1

methods. The authors used data-driven perturbation methods (LIME and SHAP) for the present study.

3.4.1. Local Interpretable Model-agnostic Explanations (LIME)

LIME is an instance-based explanation [69]. When alterations are given to the machine learning model, LIME conducts the change in predictions by observing the resulting impact on predictions. Therefore, a novel dataset is generated by disturbing the original dataset. Next, predictions (obtained from the perturbed inputs) are given weights based on their closeness to original predictions. The interpretation is automatically generated locally by replacing the model with an explainable one. Eq. 9 is used to express local surrogate models with interpretable constraints.

$$\text{Explanation}(x) = \operatorname{argmin}_{g \in M} L(f, g, \pi_x) + \omega(g) \quad (9)$$

f denotes the interpretable model for sample x . L is the loss term and $\omega(g)$ is the complexity. Term M stands for a collection of realizable explanations for a hypothetical case. The locality around sample x is presented by the closeness measure (π_x).

3.4.2. Shapley Additive Explanations (SHAP)

Lundberg & Lee, [68] introduced SHAP, which can explain a model in whole or instance wise. The core concept of SHAP is the game theory that relates the contribution of a player to the game [70]. SHAP is mostly identified as a method of obtaining unified feature importance. We used Tree-SHAP for the present study that uses following Eqs. 10-12 to compute the Shapley value.

$$f(y') = \phi_0 + \sum_{i=1}^N \phi_i y'_i \quad (10)$$

f - explanation model N - the maximum size of coalition and $\phi \in \mathbb{R}$ denote the feature attribution. Eqs. 11 and 12 are used to calculate feature attribution.

$$\phi_i = \sum_{S \subseteq \{1, \dots, p\} \setminus \{i\}} \frac{|S|!(p - |S| - 1)!}{p!} [g_x(S \cup \{i\}) - g_x(S)] \quad (11)$$

$$\text{Where; } g_x(S) = E[g(x)|x_S] \quad (12)$$

The term S denotes a subset of input features and x is a vector of feature values of instance (instance which needs to be interpreted). Shapley value is obtained through a value function (g_x). p is the number of features. $E[g(x)|x_k]$ expresses the expected value of the function on subset S .

4. Results and discussion

In the study, the authors conducted an analysis of the correlation between variables using the Pearson correlation coefficient. The results are presented in Table 6. It was observed that there is a strong correlation between weather condition and surface condition. This suggests that changes in weather, such as rain or fog, are associated with changes in the road surface condition, which is expected. On the other hand, the authors also identified variables that showed weak negative relationships based on the Pearson correlation coefficients. The coefficients ranged from -0.01 to -0.21, indicating a low level of correlation between these variables. This implies that these variables have little influence on

Table 7
Variables coefficients for the LR model.

Source	Coefficient	Standard Error	Pr > Ch ²	95% CI (Profile likelihood)		Odd ratio
				Lower bound	Upper bound	
Intercept	1.225	0.392	0.002	0.456	1.993	
WEATH	-1.048	0.234	<0.0001	-1.507	-0.589	0.35
LOC	-1.107	0.375	0.003	-1.841	-0.372	0.33
ALIGN	0.013	0.335	0.006	-0.643	0.670	1.01
LIGHT	-0.253	0.128	0.048	-0.503	-0.003	0.78
SURF	-1.156	0.396	0.004	-1.931	-0.380	0.32

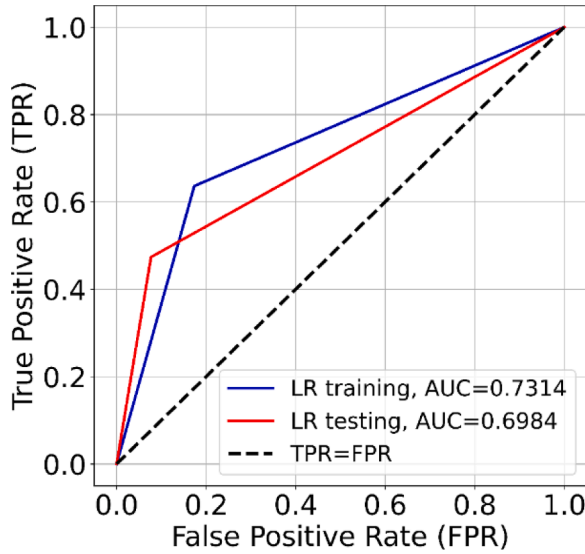


Fig. 2. ROC curve for the LR model.

each other and may have independent effects on crash severity. Correlation analysis provides insights into the relationships between different variables, helping to understand the interplay among them and identify potentially important factors contributing to crash severity.

4.1. Logistic regression model

Table 7 shows the variables’ coefficients for the LR model. The p-values for all the coefficients are less than 0.0005, which means the impacts of these dependent variables are significant. Therefore, the log odds of a crash being a fatal or serious injury crash can be calculated as Eq. 13.

$$\log\left(\frac{p}{1-p}\right) = 1.225 - 1.048WEATH - 1.107LOC + 0.013ALIGN - 0.253LIGHT - 1.156SURF \tag{13}$$

The probability of a crash being a fatal or serious injury crash can be calculated as Eq. 14.

$$P = \frac{1}{1 + e^{-(1.225 - 1.048WEATH - 1.107LOC + 0.013ALIGN - 0.253LIGHT - 1.156SURF)}} \tag{14}$$

The LR model used in the study revealed important insights into the relationship between the variables and the likelihood of a fatal or grievous accident. The coefficients of the model indicated the direction and strength of the association between each variable and the outcome. According to the model, variables such as weather condition (WEATH), location (LOC), lighting condition (LIGHT), and surface condition (SURF) showed negative coefficients, indicating that higher values of these variables are associated with a lower likelihood of a fatal or grievous accident. In other words, clear weather, dry surface, daylight, and mid-block sections were identified as less risky conditions for severe accidents.

On the other hand, the ALIGN variable, which represents road alignment (curve or straight), was found to be more critical in curve sections compared to straight sections when it comes to the likelihood of a fatal or grievous accident. This suggests that curves in the road pose a higher risk for severe accidents. The magnitude of the coefficients provides information about the strength of the association between each variable and the outcome. Larger magnitudes indicate a stronger impact, suggesting that these variables have a more significant influence on the likelihood of a fatal or grievous accident. Conversely, smaller magnitudes imply a relatively weaker effect.

The area under the receiver operating characteristics (ROC) curve is an aggregated metric that evaluates how well a LR model classifies positive and negative outcomes at all cutoffs. From the results it was found that the AUC is 0.73 for the training data set and 0.70 for the testing data set (Fig. 2). Moreover, the accuracy and precision are calculated by using the confusion matrix as shown in Fig. 3.

The LR model accurately distinguished between less severe accidents and fatal/grievous accidents in both the training and testing datasets. However, it had a tendency to misclassify fatal accidents as other types of accidents in the testing set. The study calculated the odds ratios for several factors, including location, road alignment, lighting condition, surface condition, and weather. The findings revealed that accidents occurring at intersections were 0.33 times less likely to result in fatal/grievous injuries compared to accidents at non-intersection locations. Road alignment did not have a significant impact on the severity of traffic accidents. Accidents during partial daylight had 0.78 times lower odds of being fatal/grievous than accidents during daylight. Interestingly, accidents during nighttime with good lighting or without street lighting had even lower odds of being fatal/grievous than accidents

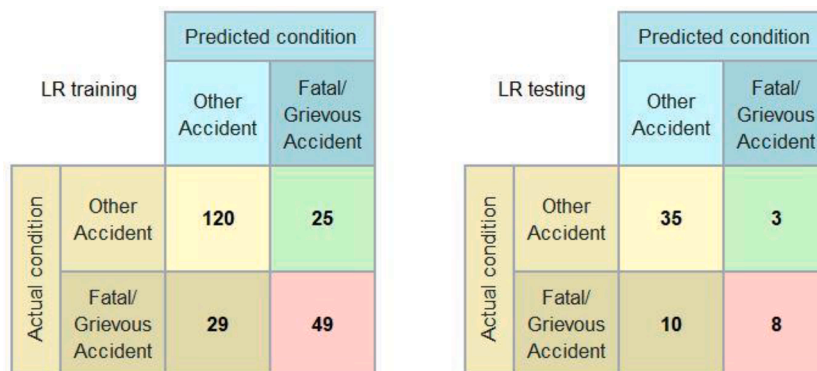


Fig. 3. Confusion matrices for training & testing for the LR model.

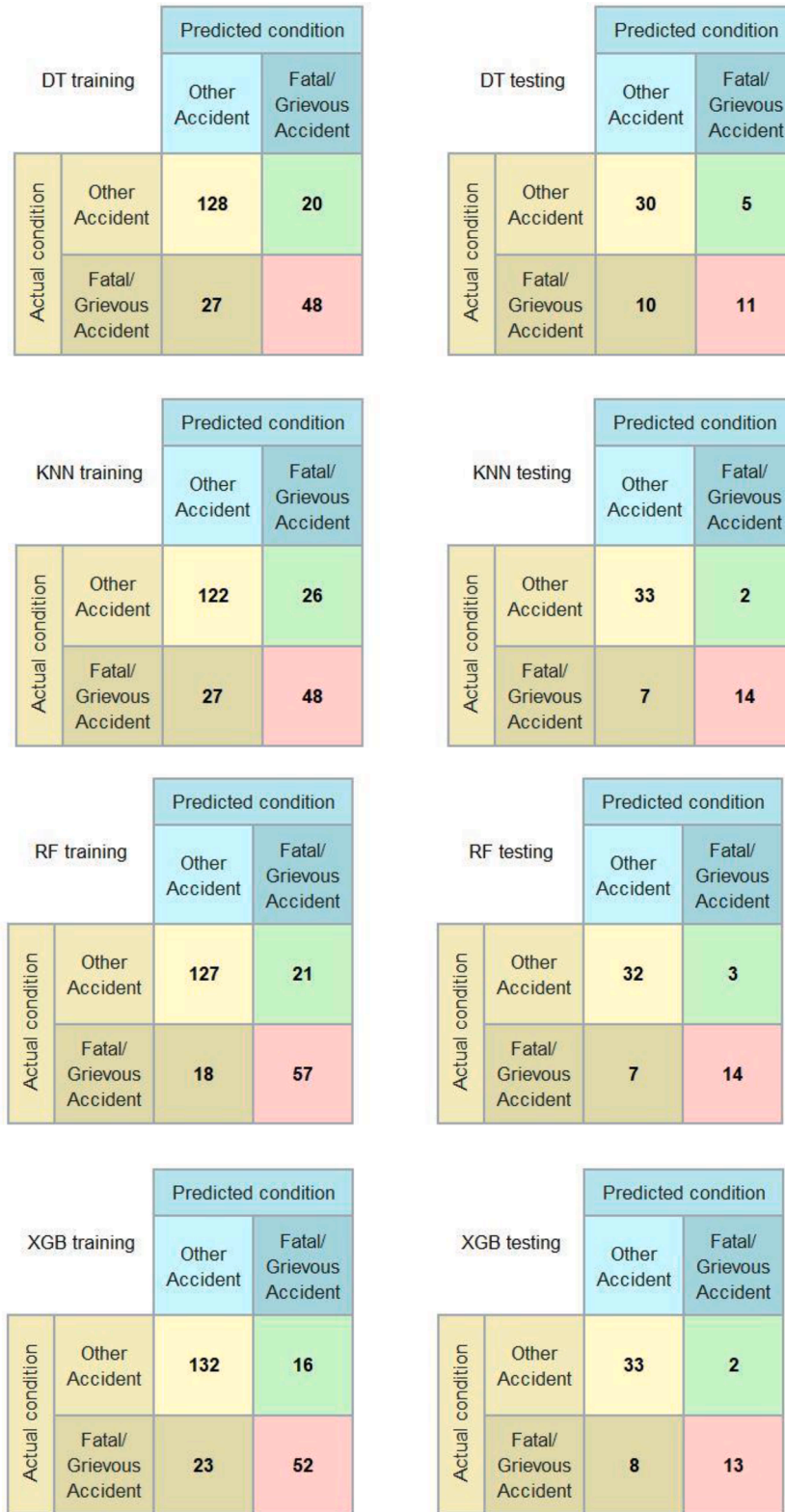


Fig. 4. Confusion matrices of four machine learning classifiers (Training and Testing).

during daylight. Accidents on wet pavement surfaces had 0.32 times lower odds of being fatal/grievous than accidents on dry pavement surfaces. Furthermore, accidents during clear weather conditions had a higher severity level.

4.2. Machine learning approach

The confusion matrix for each classifier is presented in Fig. 4 for both training and testing. The study used 223 instances for training and 56 for testing, with two crash severity levels: high-severe for fatal/grievous and low-severe for non-grievous/property damage only. In the training

Table 8
Performance evaluation parameters for LR & machine learning models.

Model	Data Category	Recall/ TPR / Sensitivity	Precision	Accuracy	F1 score	FPR
LR	Training	0.83	0.81	0.76	0.82	0.37
	Testing	0.92	0.78	0.77	0.84	0.56
DT	Training	0.86	0.83	0.79	0.84	0.36
	Testing	0.86	0.75	0.73	0.80	0.48
KNN	Training	0.82	0.82	0.76	0.82	0.36
	Testing	0.94	0.83	0.84	0.88	0.33
RF	Training	0.86	0.88	0.83	0.87	0.24
	Testing	0.91	0.82	0.82	0.86	0.33
XGB	Training	0.89	0.85	0.83	0.87	0.31
	Testing	0.94	0.80	0.82	0.87	0.38

process, the XGB model correctly classified the highest number of low-severe accidents (132), followed by DT (128) and RF (127). When classifying fatal or grievous accidents, both DT and RF showed comparable performance, while KNN assigned similar numbers for high-severe and low-severe categories. The XGB model misclassified 16 low-severe accidents as fatal/grievous and 23 fatal/grievous accidents as low-severe accidents. The RF model achieved the highest accurate classification of fatal accidents (57).

During the testing process, KNN and XGB accurately classified 33 low-severe accidents, while RF classified 32. KNN and RF both correctly classified 14 fatal/grievous accidents, but XGB only classified 13. DT had the lowest number of correctly classified accidents, with 30 low-severe and 11 fatal/grievous. The models were compared in terms of sensitivity, precision, accuracy, F1 score, and false positive rate using Eqs. 15-19, with Table 8 showing the results for each model, including LR and machine learning models. LR achieved an 83% recall score in training and 92% in testing, with an accuracy of 77% and F1 score of 84%. Among the machine learning models, XGB had the highest test accuracy and F1 score, slightly outperforming RF, while KNN and XGB had the highest true positive rate in testing. RF and KNN had comparable scores for precision matrices. Therefore, the machine learning models are versatile and efficient in classifying road safety-related accidents and outperform traditional LR methods.

$$\text{Sensitivity, Recall, True positive rate (TPR)} = \frac{TP}{TP + FN} \tag{15}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{16}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN} \tag{18}$$

$$\text{False positive rate (FPR)} = \frac{FP}{FP + TN} \tag{19}$$

The probabilities obtained by machine learning classifiers for each predicted class in the test set are shown in Fig. 5. KNN had the lowest count of zero probabilities, followed by RF and DT. The XGB model had a count near 25. KNN predicted classes near zero probability with reasonable accuracy. The distribution observed for RF and XGB was comparable, regardless of their frequencies. Each of the four methods had a distinct classification approach, resulting in unique predictions.

All of the machine learning models, including XGB, showed effectiveness in classifying accidents and fatal accidents, with XGB demonstrating the highest scores compared to other machine learning models and LR. However, before selecting XGB as the final classifier for crash severity, it was necessary to evaluate the ROC curves for each classifier. The ROC curve illustrates the trade-off between true positive and false positive rates for each classifier. Fig. 6 displays the ROC curve obtained for the machine learning classifiers in terms of training, testing, and the entire dataset. The AUC serves as a useful indicator for selecting the best model for the given task. The AUC values obtained were 0.86 for DT in both testing and training, 0.84 for KNN in training, and 0.87 for KNN in testing. While RF and XGB demonstrated similar performance in training and testing, XGB slightly outperformed RF in the testing phase. A higher AUC value signifies a better model, and in the present study, XGB was identified as the best model. However, the potential effectiveness of RF in performing a similar task should not be disregarded. Permutation-based feature importance was utilized to assess the impact of each feature for both RF and XGB.

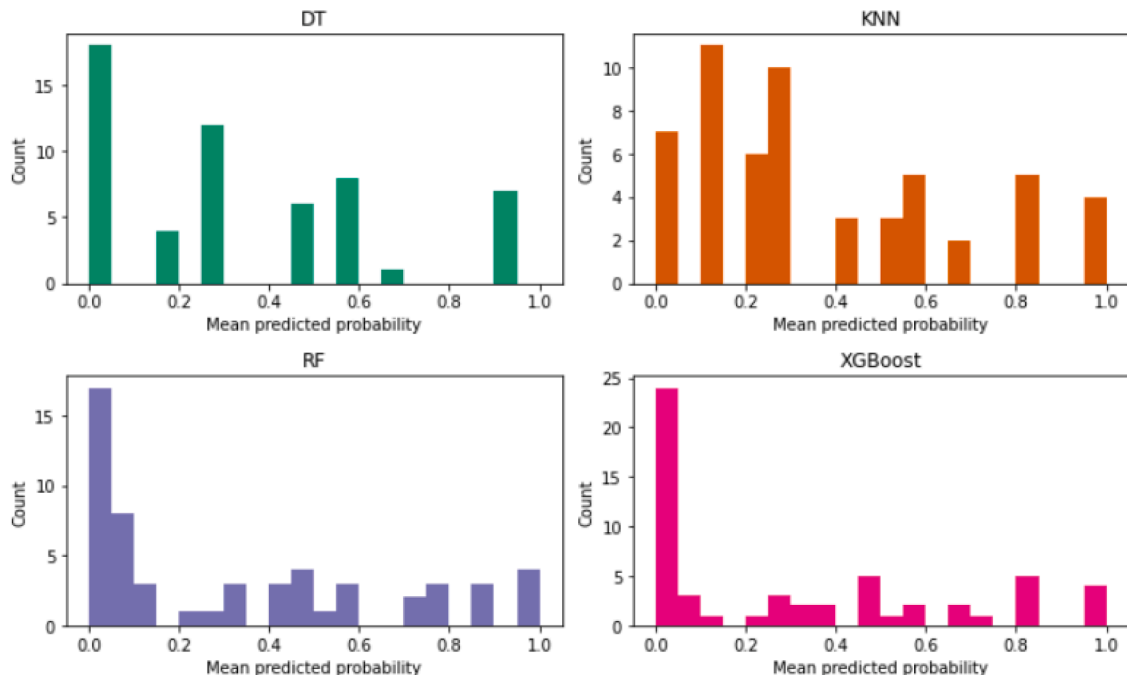


Fig. 5. Frequencies correspond to prediction probability for each machine learning classifier.

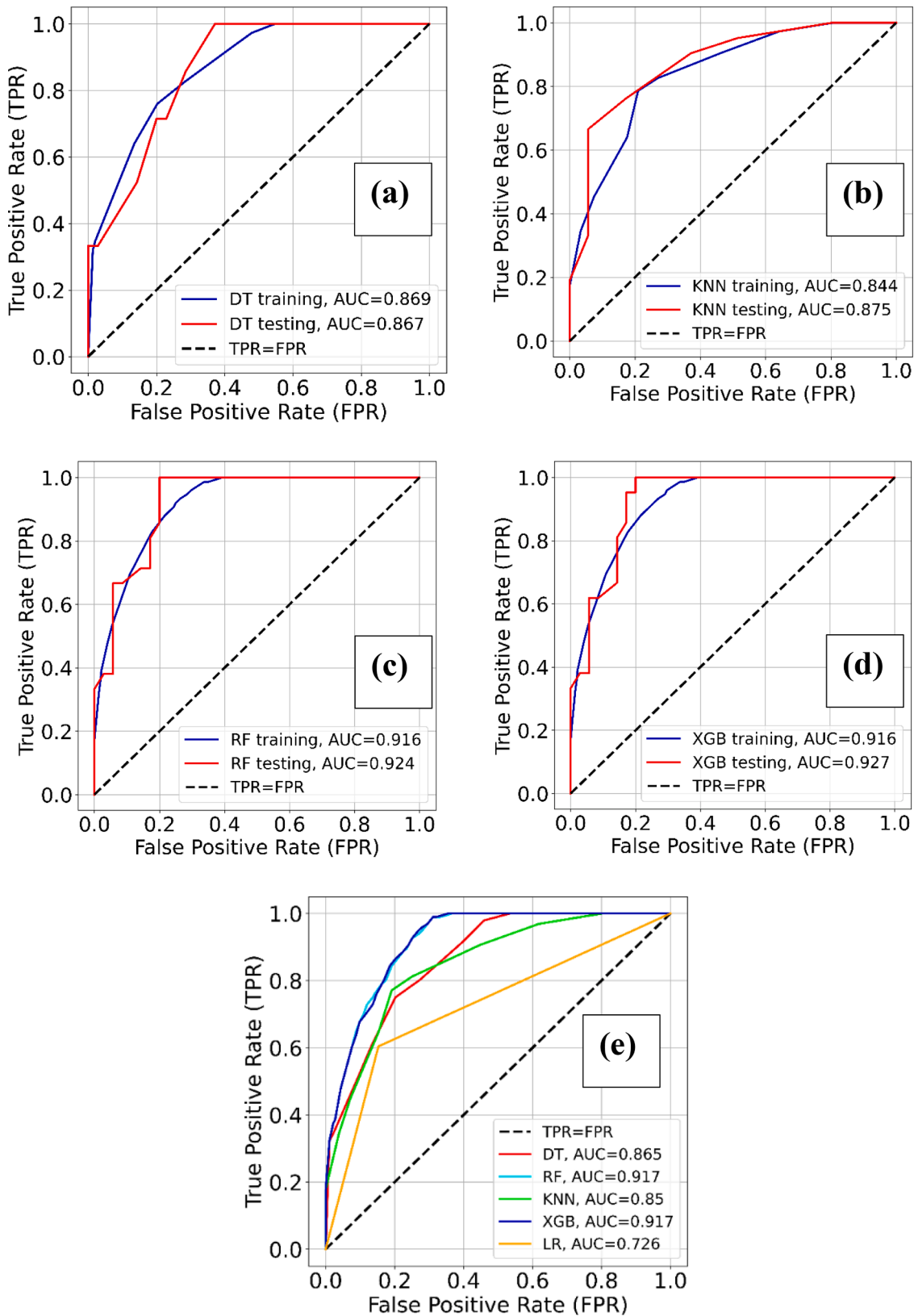


Fig. 6. ROC curves obtained for (a)-(d): machine learning models for training and testing; (e) all the models using the whole data set.

According to Fig. 7, weather has the strongest impact on the predicted class, with RF slightly overestimating its maximum feature importance compared to XGB. Similarly, the minimum feature importance of weather is higher in RF than in XGB. The minimum feature

importance for weather in XGB is less than 0.1. Lighting conditions and surface conditions are the dominant features for RF and XGB, respectively. However, XGB now overestimates the feature values of surface conditions compared to the feature importance observed in the RF

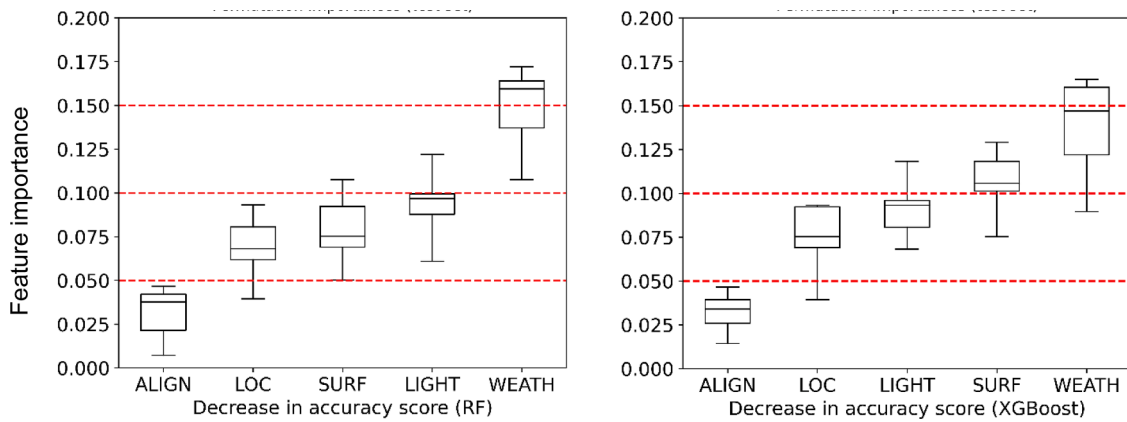


Fig. 7. Permutation-based feature importance.

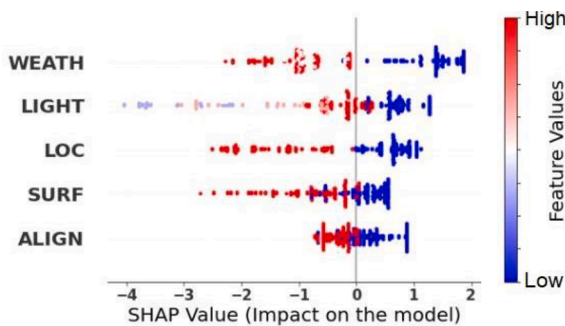


Fig. 8. Global explanation obtained from SHAP for XGB.

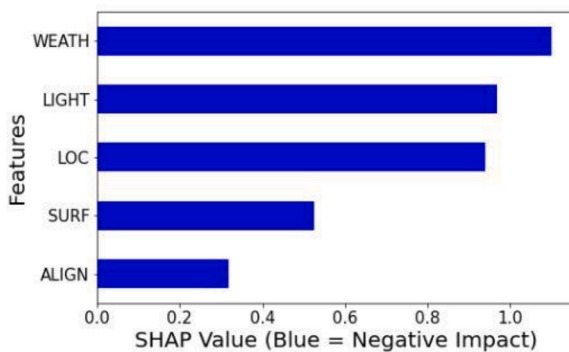


Fig. 9. SHAP average feature importance.

model. The median and both quartiles observed for surface conditions lie above 0.1 for XGB. Lighting conditions show comparable feature importance in both models. Alignment and location have the least impact on the predicted class for RF and XGB. The interquartile range of location parameters observed in XGB is higher than the corresponding range observed for the RF model. Therefore, permutation-based feature importance is unique for each model and may reflect the difference in decision-making in each algorithm. Despite the similar accuracies obtained for RF and XGB, permutation-based importance varies in terms of both magnitude and order of features. However, it is important to note that permutation-based feature importance has limitations as it overlooks important factors necessary for model explanation. To address these limitations, the authors employed an explainable machine learning approach, focusing on the superior model (XGB) for the analysis. However, for a complex tree-based model like XGB, implementing a post-hoc explanation is necessary to improve interpretability.

4.3. Machine learning explanation

4.3.1. Global explanations

The global explanation obtained from SHAP is presented in Fig. 8, where blue and red colors represent low and high values of each feature, respectively. The weather was found to have the most significant impact on the predicted class, with *clear weather* conditions influencing more on low severe accidents. *Rainy weather* conditions, on the other hand, increase the occurrence of fatal accidents. This suggests that the impact of *clear weather* on fatal accidents is comparable to the impact of *rainy weather* on low severe accidents.

Furthermore, Fig. 9 revealed that *bad lighting* conditions at nighttime may influence both fatal/grievous accidents and other accidents, while *partial daylight* and *nighttime good lighting* conditions reduced the probability of a fatal/grievous crash. The SHAP explanation of location indicated that fatal accidents were more likely to occur at mid-blocks than near interchanges. However, the SHAP importance of interchanges was more negative than the magnitude of feature importance obtained for mid-blocks. *Dry road surfaces* resulted in both fatal/grievous and other accidents, while *wet road surfaces* decreased the possibility of fatal accidents.

The explainable machine learning approach provided a deeper understanding of road safety modeling and revealed hidden sensitive factors. The five features (weather, lighting conditions, location, road surface, and alignment) all had an overall negative impact (decreasing the possibility of occurrence of fatal accidents) on the predicted class, with weather having the maximum impact, followed by lighting conditions and location. The overall impact of the road surface was 50% less than that of weather, and the alignment of the road had the least feature importance. The SHAP average values for the road alignment were around 0.3. These findings highlight the importance of considering multiple factors in road safety modeling and the potential of explainable machine learning to improve model explanations.

4.3.2. Dependency of features

The SHAP technique can detect how each explanatory variable interacts with the final class by using dependency maps as displayed in Fig. 10. The SHAP dependency plot offers an advanced approach compared to the traditional partial dependency plot because it not only illustrates the dependency but also identifies the feature that it interacts with. For instance, the features of weather, location, alignment, and road surface depict a decrease in SHAP value with the increase in feature value. Concerning weather, class value [0] had SHAP values ranging from 0 to 2, and class value [2] had SHAP values ranging from 0 to -2. Class value [1] also exhibited moderately negative SHAP values ranging from 0 to -1. An intriguing finding was that weather is primarily linked to road surface condition, which aligns with reality. It indicates that wet road conditions (red color) are associated with lower SHAP values for

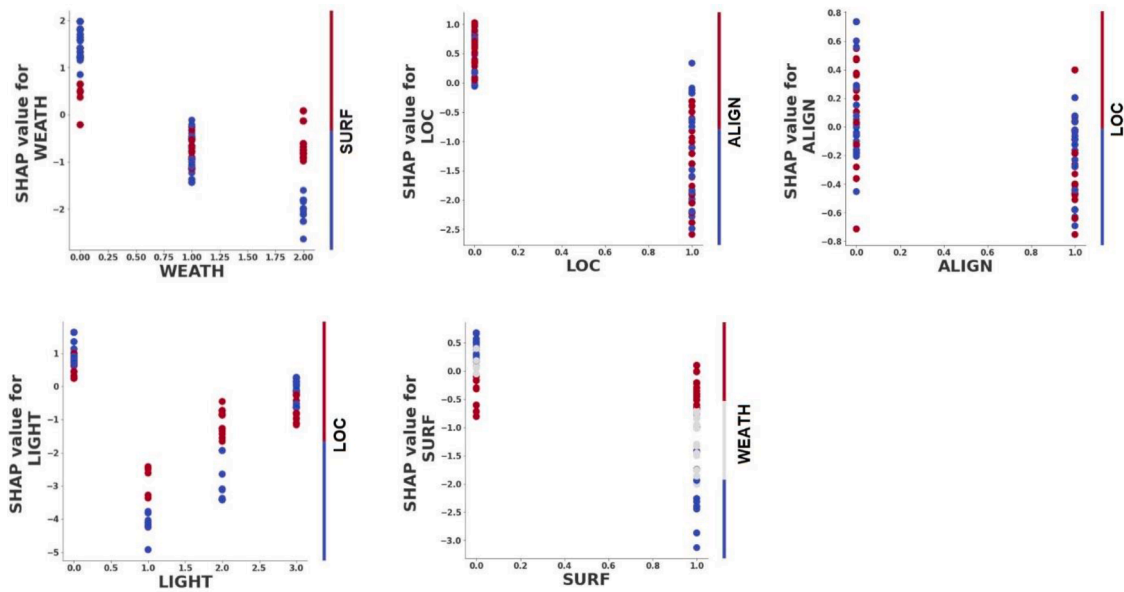


Fig. 10. Feature dependencies obtained from SHAP.

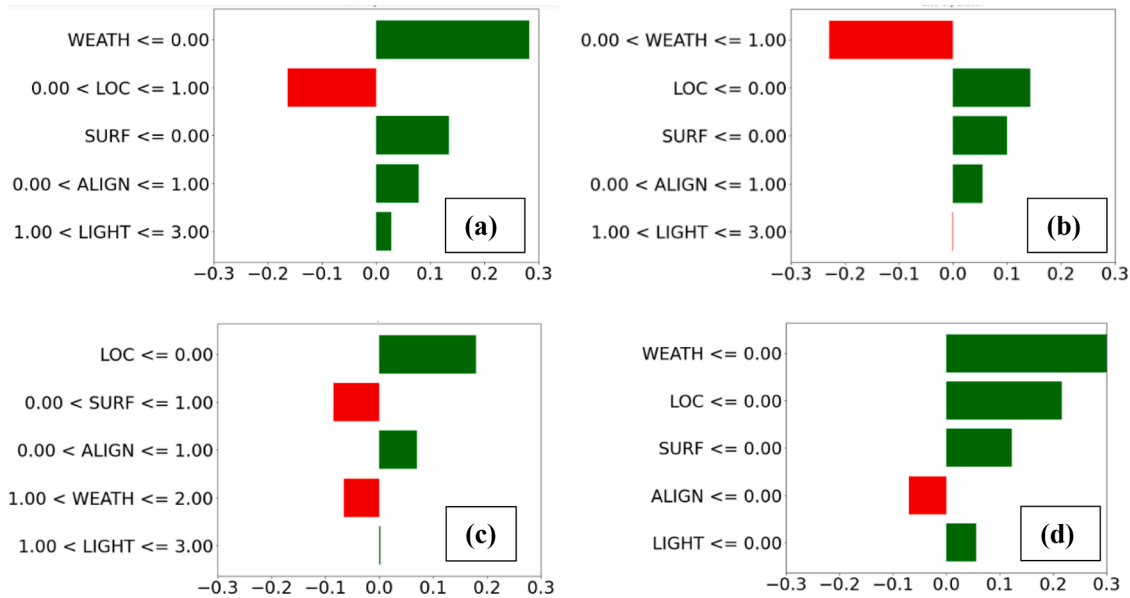


Fig. 11. Lime (local) explanations of four distinct fatal accidents.

clear weather.

While lighting conditions exhibited different behavior from other features, class value [0] displayed a large SHAP value and suddenly dropped at class value [1]. Subsequently, the SHAP value tended to increase with the increase in the class value of the feature. During daylight conditions, there was an increase of fatal accidents, and that impact dropped when the lighting became dark, likely due to higher vehicle speeds during the day. At *partial daylight* conditions, the impact of fatal accidents reduced, and interestingly, when the lighting conditions became dark (at nighttime), the impact increased again, possibly because of poor visibility at night. Additionally, lighting conditions were primarily associated with road alignment, as indicated by SHAP results.

Furthermore, a similar trend was observed in location and alignment, although the significance was not higher than in other scenarios. Explainable machine learning adds a crucial component to make the classification human-comprehensible beyond classification.

4.3.3. Local explanations

LIME was used for local interpretations as it is specifically designed for this purpose. Four instances of fatal accidents were selected for explanation, as shown in Fig. 11. The red color represents a negative impact and green represents a positive impact. The y-axis represents the range of the feature value. For example, in Fig. 11(a), it can be observed that a feature value of WEATH < 0 (which corresponds to a clear weather condition according to Table 4) has a positive impact (increased the possibility of occurrence) on the instance. It is important to note that local explanations may not always align with the global explanation. Among the four instances, weather emerged as the dominant feature in three instances, but its importance decreased in the remaining instance. Figs. 11(a) and 11(b) demonstrate that *clear weather* increased the possibility of fatal accidents while the feature values larger than 0 and less than or equal to 1 decreased the possibility of fatal accidents. Additionally, location had a negative impact in Fig. 11(a) and a positive impact in Fig. 11(b) on the occurrences, respectively. The other three

features showed comparable importance in both instances. In Fig. 11(c), location was found to have the greatest contribution, while weather conditions had a relatively less significant negative value. Lighting condition emerged as the second most influential feature but became the least significant feature across all four instances. Moreover, surface conditions had a negative impact that was comparable to the impact of weather conditions. Although local explanations may not always align with the global explanation, they provide critical factors to consider in different instances. This also highlights the independence of selected explanatory variables and their effectiveness in determining traffic crash severity on expressways.

As a limitation of the study, the current work proposes to accommodate critical factors such as environmental characteristics, traffic operation conditions, and driver behavior. In addition, the present work focused on expressways where fewer vehicle types are allowed to use the expressway. Therefore, the work can be extended to road types other than expressways by including one of the critical factors, vehicle type, into the analysis.

5. Conclusion

To evaluate road safety performance, a comprehensive examination of traffic crashes is necessary, considering their significant social and economic impacts. While statistical methods have been commonly employed in analyzing crash severity, the complex nature of machine learning techniques, often referred to as "black box" models has hindered their use in these studies. The authors overcome this limitation by employing explainable machine learning methods. Following are the remarks of this study.

- The study employed various machine learning techniques, including RF, DT, XGB, KNN, and a traditional LR model, to predict and classify the severity of expressway crashes. The crash severity was classified into two levels: fatal/grievous crashes and all other crashes. Performance metrics such as recall, precision, accuracy, F1 score, false positive rate (FPR), and AUC were used to evaluate the models. The LR model achieved a precision of 0.81, an accuracy of 0.76, and an AUC of 0.73 in predicting crash severities.
- The study revealed that machine learning methods outperformed the LR model in predicting and classifying crash severity. Specifically, the XGB model demonstrated the highest prediction performance, followed by the RF model.
- The study used explainable machine learning to interpret traffic crash severities. The global explanation of the data models was obtained using SHAP which revealed key insights. According to the findings, weather condition is found as the predominant factor on crash severity while wet weather condition has increased the occurrence of fatal/grievous accidents compared to dry weather.
- Additionally, the study employed local explanations using LIME to analyze four random instances of fatal accidents. The results demonstrated that local explanations often deviated from the global explanations, suggesting that the factors influencing crash occurrence are not necessarily dependent on each other and that crashes can happen randomly. This emphasizes the importance of considering both global and local perspectives when interpreting crash severity. In conclusion, explainable machine learning is a preferred approach for domain experts such as highway engineers when dealing with large volumes of environmental and roadway data. Given the high-stakes nature and the sensitivity of factors affecting crash severity, accurate prediction models are essential. Explainable machine learning provides a robust and accurate platform, surpassing the limitations of traditional statistical approaches in terms of analysis capabilities and offering a more comprehensive understanding of the underlying factors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The data collection was conducted with support given by the Planning Division, RDA, Sri Lanka.

References

- [1] World Health Organization, Road traffic injuries. Retrieved from World Health Organization, 2022. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>.
- [2] Y. Wang, W. Zhang, Analysis of roadway and environmental factors affecting traffic crash severities, in: *World Conference on Transport Research - WCTR*, Elsevier, Shanghai, 2016, pp. 2119–2125.
- [3] Guido, G., Haghshenas, S., Vitale, A., Astarita, V., Park, Y., & Geem, Z.W. (2022). Evaluation of contributing factors affecting number of vehicles involved in crashes using machine learning techniques in rural roads of Cosenza, Italy. *Safety*.
- [4] X. Wang, J. Liu, A.J. Khattak, D. Clarke, Non-crossing rail-trespassing crashes in the past decade: a spatial approach to analysis of injury severity, *Saf. Res.* (2016) 44–55.
- [5] J. Zhang, X. Chen, Y. Tu, Environmental and traffic effects on incident frequency occurred on urban expressways, in: *13th COTA International Conference of Transportation Professionals (CICTP 2013)*, Shanghai, Elsevier, 2013, pp. 1366–1377.
- [6] Road Development Authority, Safe Use of the Expressway, Expressway Operation Maintenance and Management Division, 2023. Retrieved from, http://www.exway.rda.gov.lk/index.php?page=user_guide#safe_use.
- [7] N.P. Kodippili, M. Akther, G. Naveendrakumar, Accident hotspots in southern expressway of Sri Lanka: interpolation evaluation using GIS, *Adv. Technol.* (2021) 191–206.
- [8] S. Kumara, C. Walgampaya, Identification of severity factors and risk areas of southern expressway accidents, *Engineer* (2021) 61–75.
- [9] M.S. Siraj, F. Rabbi, M.S. Asma, M.J. Islam, Road accident analysis: a case study Dhaka metropolitan area, *J. Transp. Syst.* (2021) 48–54.
- [10] Dhananjaya, S., & Alibuhito, M. (2016). Factors influencing road accidents in Sri Lanka: a logistic regression approach. *5th Annual Science Research Sessions*, (pp. 157–173).
- [11] A. Iqbal, Z.U. Rehman, S. Ali, K. Ullah, U. Ghani, Road traffic accident analysis and identification of black spot locations on highway, *Civ. Eng. J.* (2020) 2448–2456.
- [12] L. Zheng, T. Sayed, F. Mannering, Modeling traffic conflicts for use in road safety analysis: a review of analytic methods and future directions, *Anal. Methods Accid. Res.* (2020).
- [13] M.M. Chong, A. Abraham, M. Paprzycki, Traffic accident analysis using machine learning paradigms, *Informatica* (2005) 89–98.
- [14] G. Guido, S. Haghshenas, S. Haghshenas, A. Vitale, V. Astarita, A. Haghshenas, Feasibility of stochastic models for evaluation of potential factors for safety: a case study in Southern Italy, *Sustainability* (2020).
- [15] M. Kushan, N. Chandrasekara, Identification of factors and classifying the accident severity in Colombo—Katunayake expressway, Sri Lanka. *Smart Computing and Systems Engineering*, IEEE, Kelaniya, 2020, pp. 166–173.
- [16] J. Zhang, Z. Li, Z. Pu, C. Xu, Comparing prediction performance for crash injury severity among various machine learning and statistical methods, *IEEE Access* 6 (2018) 60079–60087.
- [17] F. Wang, J. Wang, X. Zhang, D. Gu, Y. Yang, H. Zhu, Analysis of the causes of traffic accidents and identification of accident-prone points in long downhill tunnel of mountain expressways based on data mining, *Sustainability* (2022).
- [18] M. Pljakic, D. Jovanovic, B. Matovic, S. Micic, Macro-level accident modeling in Novi Sad: a spatial regression approach, *Accid. Anal. Prev.* (2019).
- [19] L. Wang, M. Abdel-Aty, Microscopic safety evaluation and prediction for freeway-to-freeway interchange ramps, *Transp. Res. Rec.* (2016) 56–64.
- [20] N. Casado-Sanz, B. Guirao, M. Attard, Analysis of the risk factors affecting the severity of traffic accidents on Spanish crosstown roads: sustainability and sustainable development in China, *Sustainability* (2020).
- [21] Z. Ma, C. Shao, H. Yue, S. Ma, Analysis of the logistic model for accident severity on urban road environment, in: *Intelligent Vehicles Symposium, IEEE*, 2009, pp. 983–987.
- [22] V. Ratanavaraha, S. Suangka, Impacts of accident severity factors and loss values of crashes on expressways in Thailand, *IATSS Res.* (2014) 130–136.
- [23] B. Claros, P. Edara, C. Sun, When driving on the left side is safe: safety of diverging diamond interchange ramp terminals, *Transp. Res. C* (2020) 133–142.

- [24] E. Edirisinghe, A. Edirisinghe, Analysis of accidents on the Southern Expressway, in: R4TLI Conference Proceedings, 2017, pp. 195–200.
- [25] T. Shantajit, C.R. Kumar, Z.Q. Zyed, Road traffic accidents in India: an overview, *Int. J. Clin. Biomed. Res.* (2018) 36–38.
- [26] B.K. Brimley, M. Saito, G. Schultz, Calibration of highway safety manual safety performance function, *Transp. Res. Rec.* (2012) 82–89.
- [27] G. Mehta, Y. Lou, Safety performance function calibration and development for the state of Alabama: two-lane two-way rural roads and four-lane divided highways, in: 92nd Annual Meeting of the Transportation Research Board, Washington, DC, USA, 2013.
- [28] A. Montella, Safety reviews of existing roads: quantitative safety assessment methodology, *Transp. Res. Rec.* (2005) 62–72.
- [29] I. Paolo, B. Nicola, C. Pasquale, R. Vittorio, R. Eirin, Exploring the relationships between drivers' familiarity and two-lane rural road accidents. A multi-level study, *Accid. Anal. Prev.* (2018) 280–296.
- [30] R. Elvik, Regression analysis and road safety research, *Accid. Anal. Prev.* (2017) 23–34.
- [31] Chakraborty, M., Gates, T., & Sinha, S. (2021). *Causal Analysis and Classification of Traffic Crash Injury*.
- [32] A. Cigadem, O. Cevher, Predicting the severity of motor vehicle accident injuries in adana-turkey using machine learning methods and detailed meteorological data, *Int. J. Intell. Syst. Appl. Eng.* (2018) 72–79.
- [33] D. Yang, Y. Wu, F. Sun, J. Chen, D. Zhai, C. Fu, Freeway accident detection and classification based on the multi-vehicle trajectory data and deep learning model, *Transp. Res. C* (2021) 130.
- [34] S. Das, M. Le, B. Dai, Application of machine learning tools in classifying pedestrian crash types: a case study, *Transp. Saf. Environ.* (2020) 106–119.
- [35] P.B. Silva, M. Andrade, S. Ferreira, Machine learning applied to road safety modeling: a systematic literature review, *J. Traffic Transp. Eng.* 7 (6) (2020) 775–790.
- [36] Y. Wang, J. Xu, X. Liu, Z. Zheng, H. Zhang, C. Wang, Analysis on risk characteristics of traffic accidents in small-spacing expressway interchange, *Int. J. Environ. Res. Public Health* (2022).
- [37] J. Wang, B. Liu, T. Fu, S. Liu, J. Stipanovic, Modeling when and where a secondary accident occurs, *Accid. Anal. Prev.* (2019).
- [38] K. Yau, H. Lo, S. Fung, Multiple-vehicle traffic accidents in Hong Kong, *Accid. Anal. Prev.* 38 (2006) 1157–1161.
- [39] S.Y. Sohn, S.H. Lee, Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea, *Saf. Sci.* 41 (1) (2003) 1–14.
- [40] L. Wahab, H. Jiang, A comparative study on machine learning based algorithms for prediction of motorcycle crash severity, *PLoS One* 14 (4) (2019).
- [41] M. Ijaz, L. Liu, M.Z. Khattak, A. Jamal, A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw, *Accid. Anal. Prev.* (2021) 154.
- [42] J. de Ona, G. Lopez, J. Abellan, Extracting decision rules from police accident reports through decision trees, *Accid. Anal. Prev.* (2012) 1151–1160.
- [43] O.H. Kwon, W. Rhee, Y. Yoon, Application of classification algorithms for analysis of road safety risk factor dependencies, *Accid. Anal. Prev.* 75 (2015) 1–15.
- [44] A. Iranitalab, A. Khattak, Comparison of four statistical and machine learning methods for crash severity prediction, *Accid. Anal. Prev.* 108 (2017) 27–36.
- [45] W. Lu, J. Liu, X. Fu, J. Yang, S. Jones, Integrating machine learning into path analysis for quantifying behavioral pathways in bicycle-motor vehicle crashes, *Accid. Anal. Prev.* (2022) 168.
- [46] C. Chen, G. Zhang, Z. Qian, R.A. Tarefder, Z. Tian, Investigating driver injury severity patterns in rollover crashes using support vector machine models, *Accid. Anal. Prev.* (2016) 128–139.
- [47] A.T. Kashani, A.S. Mohaymany, Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models, *Saf. Sci.* 49 (10) (2011) 1314–1320.
- [48] M.A. Abdel-Aty, H.T. Abdelwahab, Predicting injury severity levels in traffic crashes: a modeling comparison, *J. Transp. Eng.* (2004) 204–210.
- [49] Q. Zeng, H. Huang, A stable and optimized neural network model for crash injury severity prediction, *Accid. Anal. Prev.* 73 (2014) 351–358.
- [50] A.M. Amiri, A. Sadri, N. Nadimi, M. Shams, A comparison between Artificial Neural Network and Hybrid Intelligent Genetic Algorithm in predicting the severity of fixed object crashes among elderly drivers, *Accid. Anal. Prev.* (2020) 138.
- [51] K. Alzaffin, S.A. Kaye, A. Watson, M.M. Haque, A data fusion approach of police-hospital linked data to examine injury severity of motor vehicle crashes, *Accid. Anal. Prev.* (2023) 179.
- [52] Road Development Authority, Retrieved from Road Development Authority, 2022. <http://www.rda.gov.lk/>.
- [53] A. Suresh, Infrastructure, Investment and Development Evaluation: a Complex Stakeholder Perception Mapping Approach For Improving Local and Regional Economic Development Outcomes, University of Moratuwa, Colombo, 2015.
- [54] Road Development Authority, E01 - Southern Expressway (SE), Expressway Operation Maintenance and Management Division, 2023. Retrieved from, http://www.exway.rda.gov.lk/index.php?page=expressway_network/e01.
- [55] Q. Wu, F. Chen, G. Zhang, X.C. Liu, H. Wang, S.M. Bogus, Mixed logit model-based driver injury severity investigations in single and multi-vehicle crashes on rural two-lane highways, *Accid. Anal. Prev.* (2014) 105–115.
- [56] I.U. Ekanayake, D.P.P. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP), *Case Stud. Constr. Mater.* 16 (2022) e01059.
- [57] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, *Classification and Regression Trees*, Taylor & Francis, 1984.
- [58] J. Han, M. Kamber, J. Pei, *Data Mining Third Edition*, Morgan Kaufmann, Waltham, 2012.
- [59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, M.B. Olivier Grisel, Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [60] D.P.P. Meddage, I.U. Ekanayake, A.U. Weerasuriya, C.S. Lewangamage, Tree-based regression models for predicting external wind pressure of a building with an unconventional configuration, in: 2021 Moratuwa Engineering Research Conference (MERCon), IEEE, 2021, pp. 257–262.
- [61] L. Chen, S. Huang, C. Yang, Q. Chen, Analyzing factors that influence expressway traffic crashes based on association rules: using the Shaoyang–Xinhuang section of the Shanghai–Kunming Expressway as an example, *J. Transp. Eng. A Syst.* (2020).
- [62] P. Meddage, I. Ekanayake, U.S. Perera, H.M. Azamathulla, M.A. Md Said, U. Rathnayake, Interpretation of machine-learning-based (black-box) wind pressure predictions for low-rise gable-roofed buildings using Shapley additive explanations (SHAP), *Buildings* 12 (6) (2022) 734.
- [63] R.C. Fong, A. Vedaldi, Interpretable explanations of black boxes by meaningful perturbation, in: IEEE International Conference on Computer Vision (ICCV), 2017, pp. 3449–3457.
- [64] Y. Liang, S. Li, C.Y. Yan, M. Li, C. Jiang, Explaining the black-box model: a survey of local interpretation methods for deep neural networks, *Neurocomputing* 419 (2021) 168–182.
- [65] D.P.P. Meddage, I.U. Ekanayake, S. Herath, R. Gobirahavan, N. Muttill, U. Rathnayake, Predicting bulk average velocity with rigid vegetation in open channels using tree-based machine learning: a novel approach using explainable artificial intelligence, *Sensors* 22 (12) (2022) 4398.
- [66] M.T. Ribeiro, S. Singh, C. Guestrin, Why Should I Trust You?": Explaining the Predictions of Any Classifier, *HLT-NAACL Demos*, 2016.
- [67] Petsiuk, V., Das, A., & Saenko, K. (2018). RISE: Randomized Input Sampling for Explanation of Black-box Models. [arXiv:1806.07421](https://arxiv.org/abs/1806.07421) (cs).
- [68] S.M. Lundburg, S.I. Lee, A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 4768–4777.
- [69] I.U. Ekanayake, Sandini Palitha, Sajani Gamage, D.P.P. Meddage, Kasun Wijesooriya, Damith Mohotti, Predicting adhesion strength of micropatterned surfaces using gradient boosting models and explainable artificial intelligence visualizations, *Mater. Today Commun.* (2023), <https://doi.org/10.1016/j.mtcomm.2023.106545>.
- [70] D.P.P. Meddage, I.U. Ekanayake, A.U. Weerasuriya, C.S. Lewangamage, K.T. Tse, T. P. Miyanawala, C.D.E. Ramanayaka, Explainable machine learning (XML) to predict external wind pressure of a low-rise building in urban-like settings, *J. Wind Eng. Ind. Aerodyn.* 226 (2022), 105027.