

## Article

# Predicting Crash-Related Incident Clearance Time on Louisiana's Rural Interstate Using Ensemble Tree-Based Learning Methods

Waseem Akhtar Khan <sup>1</sup>, Milhan Moomen <sup>1,2</sup>, M. Ashifur Rahman <sup>1,2,\*</sup>, Kelvin Asamoah Terkper <sup>1</sup>, Julius Codjoe <sup>1,2</sup> and Vijaya Gopu <sup>1,2</sup>

<sup>1</sup> Department of Civil Engineering, University of Louisiana at Lafayette, Lafayette, LA 70504, USA

<sup>2</sup> Louisiana Transportation Research Center, 4101 Gourrier Ave., Baton Rouge, LA 70808, USA

\* Correspondence: ashifur@louisiana.edu

**Abstract:** Traffic crashes contribute significantly to non-recurrent congestion, thereby increasing delays, congestion pollution, and other challenges. It is important to have tools that enable accurate prediction of incident duration to reduce delays. It is also necessary to understand factors that affect the duration of traffic crashes. This study developed three machine learning models, namely extreme gradient boosting (XGBoost), categorical boosting (CatBoost), and a light gradient-boosting machine (LightGBM), to predict crash-related incident clearance time in Louisiana rural interstates and utilized Shapley additive explanations (SHAP) analysis to determine the influence of factors impacting it. Four ICT levels were defined based on 30 min intervals: short (0–30), medium (31–60), intermediate (61–90), and long (greater than 90). The results suggest that XGBoost outperforms CatBoost and LightGBM in the collective model's predictive performance. It was found that different features significantly affect different ICT levels. The results indicate that crashes involving injuries, fatalities, heavy trucks, head-on collisions, roadway departure, and older drivers are the significant factors that influence ICT. The results of this study may be used to develop and implement strategies that lead to reduced incident duration and related challenges with long clearance times, providing actionable insights for traffic managers, transportation planners, and incident response agencies to enhance decision-making and mitigate the associated increases in congestion and secondary crashes.

**Keywords:** rural interstate; incident clearance time; machine learning; SHAP analysis



**Citation:** Khan, W.A.; Moomen, M.; Rahman, M.A.; Terkper, K.A.; Codjoe, J.; Gopu, V. Predicting Crash-Related Incident Clearance Time on Louisiana's Rural Interstate Using Ensemble Tree-Based Learning Methods. *Appl. Sci.* **2024**, *14*, 10964. <https://doi.org/10.3390/app142310964>

Academic Editor: Luis Javier Garcia Villalba

Received: 4 November 2024  
Revised: 20 November 2024  
Accepted: 22 November 2024  
Published: 26 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Traffic incidents are considered a high-priority problem for traffic safety. According to a recent report published by the World Health Organization (WHO), traffic incidents cause over 1.35 million fatalities and 50 million injuries annually [1]. If not cleared on time, traffic incidents have the potential to cause negative consequences, including increased non-recurring congestion, travel delays, secondary crashes, a reduction in roadway capacity, and risks to response personnel [2]. According to National Traffic Incident Management Coalition (NTIMC) estimates, traffic incidents are responsible for approximately 25% of all congestion on U.S. roadways, leading to an annual loss of about 2.8 billion gallons of gasoline [3]. Additionally, secondary crashes account for 20% of all the incidents in the U.S. [4]. Furthermore, in the year 2014, each individual in the United States was estimated to have experienced a total of 42 h of delay during peak hours, resulting in a collective loss of around USD 160 billion due to congestion [5].

Incident clearance time (ICT) can be defined as the time from when the incident is detected until clearance of all travel lanes [6]. Incidents that are not cleared on time may lead to an increase in incident duration by twice or thrice [7]. To mitigate these negative consequences resulting from traffic incidents, it is important to improve traffic incident management (TIM) efficiency. One TIM strategy involves providing travelers with accurate ICT estimates, allowing them to make an informed decision [8,9]. Another strategy would

be managing traffic flow by actively rerouting around the incident scene to avoid congestion. Both cases required accurate prediction and understanding of the factors affecting ICT. The *Highway Capacity Manual* categorizes incident duration into four phases: detection time (the time between the occurrence of the incident to the time it is reported), response time (the time from when the incident is reported to the arrival of the first responder at the scene), clearance time (the time from the first responder's arrival to the time when the incident is cleared), and recovery time (the time taken for the traffic flow to normalize after the incident is cleared) [10]. ICT is of great importance in reducing the impact of traffic incidents among the different phases of incident duration, as it directly affects travel delays and congestion. ICT is one of the key performance measures of Louisiana's TIM.

Rural interstates face the challenge of longer incident durations because of their distance from urban centers, accessibility issues, and the limited availability of emergency responders [11]. Due to their remoteness, ICT can be significantly increased, leading to increased travel delays, congestion, and secondary crashes. To overcome these challenges, it is important to reduce the duration of these incidents. Achieving this goal will require the accurate prediction and dissemination of ICT to both motorists, particularly those approaching the incident scene, and various agencies involved in incident management, such as the traffic management center (TMC), fire, police, and emergency medical services (EMS). This would allow stakeholders to make informed decisions regarding their travel.

This study aims to develop a machine learning (ML)-based ICT prediction model for crash-related incidents in Louisiana's rural interstates and uses SHAP analysis to investigate influencing factors for these incidents. The two separate approaches were utilized in a complementary framework by taking advantage of the strengths of machine learning models and SHAP analysis. While machine learning models provide better predictions, their methodologies are considered black boxes and do not excel at establishing relationships between response and independent variables. SHAP analysis is flexible and useful for analyzing the impact and direction of the influence of variables on a phenomenon. Machine learning and SHAP analysis were applied to crash datasets to analyze ICT in relation to traffic crashes.

The main objectives of this research are summarized as follows:

1. Developing ML models, namely CatBoost, XGBoost, and LightGBM, to predict crash-related ICT.
2. Comparing the performance of these developed ML models in terms of prediction accuracy.
3. Analyzing the influence of significant factors impacting ICT using SHAP analysis.

The rest of the paper is organized as follows. Section 2 presents a review of previous studies on incident duration prediction using different machine learning techniques and the significant factors that impact it. Data description and preprocessing were discussed in Section 3. Section 4 describes the methodology followed in this research. Firstly, an overview of the machine learning models employed and the SHAP analysis is presented. Secondly, the steps involved in developing these models are discussed. Lastly, the criteria for the evaluation of the developed models are given. Section 5 discusses the model results and presents a detailed evaluation of the machine learning models employed. The analysis identifies the best tools for predicting ICT while identifying influential factors that impact incidence clearance duration resulting from the SHAP analysis. The final section presents our conclusions and recommendations

## 2. Literature Review

### 2.1. Prediction of Incident Duration with ML Models

Several studies have been conducted in recent decades to model and analyze incident duration [12–15]. The first objective of this study is to predict incident duration, which is typically the primary focus of ML methods. Due to the flexible framework of these ML algorithms, they can handle the complex and highly nonlinear relationship between input and output features. The most commonly used ML methods are the tree-based

method [16], neural networks method [17], distance metric learning method [18], and hybrid methods [19].

Ensemble learning methods have also been used to predict incident duration in the previous literature [20–22]. The majority of ensemble learning models use tree-based models. Examples of tree-based models include decision trees (DT), random forests (RF), and gradient-boosting decision trees (GBDT). In contrast to other ML techniques, tree-based models are able to rank explanatory variables according to importance and predict duration with a high degree of accuracy [16,23]. Ma et al. developed an incident duration prediction model based on gradient-boosting decision trees (GBDT) using incident and weather data [24]. The results showed that the GBDT model performed better than the random forest (RF), backpropagation neural network (BPNN), and support vector machine (SVM) methods in terms of prediction accuracy and model interpretation power. Another study by Grigorev et al. for incident duration prediction utilized tree-based modeling, such as RF, XGBoost, LightGBM, and GBDT, along with K-nearest neighbors (KNN) and linear regression (LR) [25]. They found that XGBoost outperformed the other models. Zhan et al. developed an M5P tree algorithm for the prediction of lane clearance time [16]. The results showed that the M5P model outperformed both the decision tree model and a traditional regression model created with the same data. Zhao et al. applied XGBoost along with LightGBM and CatBoost for the prediction of incidence clearance time [12]. The comparative analysis showed that XGBoost outperformed the other two models.

Neural networks have been used in the past for incident duration prediction [14,26–28]. A study conducted by Wei and Lee developed an ANN model to predict ICT and used the sequential model to update the clearance time phase. The mean absolute percentage error (MAPE) for the predicted models was almost below 40%, indicating that the suggested models have a satisfactory level of prediction ability [17]. Yu et al. developed an SVM and ANN model using 235 incidents. They found that for predictions of a longer duration, ANN performed better than the SVM model [29]. Additionally, Lee et al. developed an incident prediction model employing ANN and KNN [30]. They found that the ANN model accurately predicted incident duration with an MAE of less than 30%. Another study by Kidando et al. in Florida developed a mixture density network (MDN)-based model by utilizing 4 years of data (2014–2017) from 58,167 incidents. ANN and XGBoost models were employed to assess the performance of the developed MDN model in predicting incident duration. The comparison indicated that the MDN model had superior performance when compared to the other two models because it achieved the lowest errors [31].

Distance metric learning methods work by finding the most efficient classifier by using a selected distance function [32]. Common distance metric learning methods include the support vector machine (SVM), K-nearest neighbor (KNN), and principal component analysis (PCA).

Hamad et al. developed incident prediction models employing various machine learning techniques, including DT, RF, logistic regression (LR), naive Bayes (NB), SVM, discriminant analysis (DA), Gaussian process regression (GPR), MLP, stochastic gradient descent (SGD), and extra trees (ET). The model was trained on 110,000 incidents collected from the Houston TranStar incidents archive. Results suggest that SVM accuracy outperformed the other models [33]. Obaid et al. developed models based on SVM, KNN, DT, ANN, gradient boosting, and multivariate linear regression (MLR) prediction of incidents [34]. Among these, gradient boosting, followed by ANN and MLR, was found to be the best model for better prediction accuracy. Tang et al. employed four machine learning models: SVM, K-nearest neighbor (KNN), RF, and BPNN [18]. They found that the RF achieved better results compared to other models. In another study, Zhu developed eight statistical and machine learning-based incident prediction models, including LR, Bayesian logit regression, DT, KNN, ANN, and gradient boosting, using 11,418 incidents. The results revealed that the RF performs better than the other models [35]. Additionally, Jia et al. developed four machine learning models, namely BPNN, SVM, long short-term mem-

ory (LSTM), and RF, using 550 incidents. They found that LSTM demonstrated superior performance [36].

The above existing studies are primarily focused on urban interstates. Despite their significant effectiveness in urban interstates, a significant research gap exists concerning rural interstate scenarios. Rural interstates are characterized by longer incident durations and slower emergency responses, presenting unique challenges that are not typically encountered in the urban datasets used in prior studies. This gap highlights the need to adapt and refine ML models to suit the specific conditions of rural interstates.

Neural networks have been found to be an effective methodology for classification and regression tasks. However, they are prone to overfitting when using a limited dataset. Additionally, neural networks have been observed to have limited external validity when employed in analyzing diverse datasets. Also, it is difficult to determine the optimal architecture of neural networks and the number of hidden layers in the structure. Distant learning methods are difficult to interpret when used for complex data classification tasks. Overfitting, underfitting, model interpretation, and model training time are computationally expensive and time-consuming. To address these issues inherent to neural networks and distant learning methods, it is essential to provide adequate attention to tree-based ensemble-boosting approaches. For instance, XGBoost is scalable and precise and extends the computational capacity of tree algorithms. LightGBM offers several benefits, such as efficient model training and low memory consumption. Furthermore, the CatBoost model is employed as a series of decision trees that iteratively minimize loss. These advantages of the tree-based ensemble boosting algorithms led this study to consider these models for predicting and classifying incident duration. In addition, a comparative analysis of the accuracies of these models will be beneficial to agencies in accurately predicting incident duration to put in place measures to adequately respond to incidents. Several statistical parameters were used to evaluate the performance of the developed models, and the results showed that XGBoost outperformed the other models in the prediction of ICT.

## 2.2. Factors Influencing Incident Duration

Apart from studies that have sought to predict incident duration, several others have investigated the impact of factors influencing crash-related incident duration on U.S. interstates. Several factors related to the time of the incident, crash type, number of agencies, environmental characteristics, and other crash factors impact incident duration [2,18,37]. For example, incidents that occur during nighttime, peak hours, and weekends have been found to be associated with longer incident durations [24,37,38]. Nighttime, in comparison to morning and evening time, results in longer ICT due to the darker environment and limited resource availability to response agencies [39]. Also, due to heavy traffic conditions during peak hours, it takes more time for a response team to arrive at the incident scene, which prolongs ICT [15]. During weekends, delays are due to responders often operating with fewer staff. Regarding environmental characteristics, weather conditions, like rain, wind, fog, and snow, resulted in higher incident duration [2,40,41]. This can be due to poor visibility for responders, and crashes during these adverse conditions are probably run-off-road collisions [42].

In terms of crash characteristics, such as crash severity, number of vehicles involved, manner of collisions, heavy truck involvement, heavy occupancy vehicle (HOVs), and number of lanes blocked, have been concluded as significant factors affecting ICT [42–44]. For instance, severe incidents lead to longer ICT due to the involvement of injuries and fatalities [5]. This is because these types of incidents require a response from different response agencies, like the EMS and the fire department, which may increase ICT.

Incidents involving HOVs have been found to increase incident duration compared to single-occupancy vehicles (driver only) as more people stuck in a car will require more time to rescue [45]. Additionally, ICT is longer when multiple or all travel lanes are blocked than when a single lane is blocked [46]. This is expected because when multiple lanes are blocked, additional time is required to clear the lane, as single lanes can be cleared quickly.

It was also observed that crash type significantly impacts ICT [26,47]. In terms of collision manner, rear-end and sideswipe crashes were found to be associated with decreased ICT in comparison to head-on crashes. This is because rear-end and sideswipe crashes result in fewer injuries in comparison to head-on crashes [48,49]. Furthermore, incidents involving heavy trucks have been found to prolong ICT in comparison to other vehicles [40,50]. This is because the incident involving trucks resulted in an increased number of injuries, fatalities, and property damage [51,52]. Additionally, incidents involving rescue and fire units result in a longer duration, as rescue units may be required in larger incidents [13]. Table 1 summarizes the significant factors that impact ICT.

**Table 1.** Summary of significant factors impacting ICT.

Reference	Data Used	Significant Features	Modeling Technique
Haule et al., 2019 [5]	2014–2016	Percentage of lane closures, nighttime, weekends, off-peak hours, and an increasing number of responders	Hazard-based duration models
Nam and Mannering 2000 [37]	1994 and 1995	Peak hours, nighttime, Friday, Sunday, rain, fog, fatality, single-occupancy vehicles, and pickup trucks	Hazard-based duration models
Park et al., 2016 [53]	2010 to 2011	Higher-occupancy vehicles, snow, rain, heavy vehicles, shoulder blockage, fatality, injury, property damage, vehicle fire, and disabled vehicles	Bayesian neural networks
Ding et al., 2015 [38]	2009	Travel lanes blocked, total closure, injury involved, fire involved, heavy truck involved, and traffic control	Binary probit model and switching regression model
Khattak et al., 2016 [13]	2013–2015	Injury, day of week, time, roadway geometry	Quantile regression
Cong et al., 2018 [15]	2008 to 2010	Incident type, number of vehicles, trucks, injuries, fatalities	Bayesian network

### 3. Data Description and Preprocessing

#### 3.1. Data Source

The crash data used in this research were retrieved from the Louisiana DOTD’s crash database. In addition to crash, temporal, environmental, and other factors, the database also contains incident-related information. These include crash notification time, time ambulance called, time ambulance departed, time police arrived, and time lanes opened, among other incident response information. Only crashes that occurred on highways classified as rural interstates were considered for this study. A total of 13,196 crashes that occurred between 2016 to 2019 were extracted for the analysis. A brief description of each variable is described in Table 2.

**Table 2.** Descriptive statistics of variables in the crash dataset.

Variable Category	Category (Binary)	Description	Code	Frequency	Mean	Std. Dev
Response characteristics	Response to crash	Short	-	7323	0.65	0.48
		Medium	-	1508	0.13	0.34
		Intermediate	-	1061	0.09	0.29
		Long	-	1437	0.13	0.33

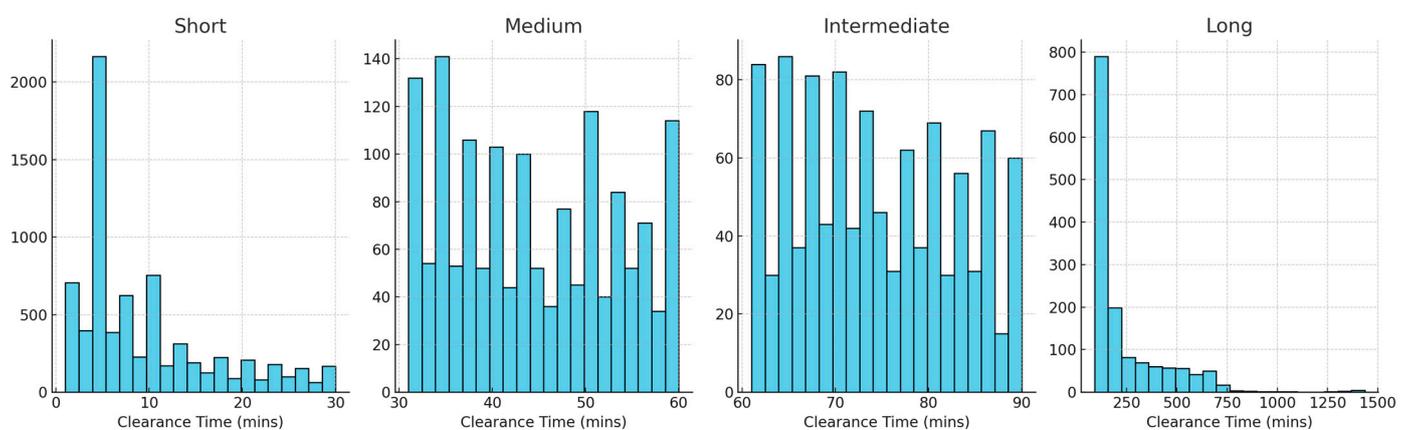
Table 2. Cont.

Variable Category	Category (Binary)	Description	Code	Frequency	Mean	Std. Dev
Temporal factors	Time of day crash occurred	AM peak	AM_Peak	914	0.08	0.27
		PM peak	PM_Peak	1519	0.13	0.34
		Night	Ngt	3444	0.3	0.46
	Day of week	Monday to Thursday	Mon_Thu	6079	0.54	0.5
		Friday to Sunday	Fri_Sun	5250	0.46	0.5
	Season	Spring (March, April, May)	Spr	2745	0.24	0.43
		Summer (June, July, August)	Sum	3179	0.28	0.45
		Fall (September, October, November)	Fall	2837	0.25	0.44
Winter (December, January, February)		Wnt	2533	0.22	0.42	
Crash characteristics	Driver gender	Male	Male	7412	0.65	0.48
	Driver age	Young age ( $\leq 25$ years)	Y_AGE	2516	0.22	0.42
		Middle age (between 25 and 65 years)	M_AGE	7379	0.65	0.48
		Old age ( $> 65$ years)	O_AGE	1434	0.13	0.33
	Crash severity	Injury crash	Inj	2638	0.23	0.42
		Fatality	Fatality	144	0.01	0.11
	Alcohol or drugs involved (binary)	Alcohol/drugs	Alc_Drugs	105	0.01	0.1
		Alcohol	Alc	399	0.04	0.18
		Drugs	Drugs	86	0.01	0.09
	Crash location	Residential	Res	168	0.01	0.12
		Business	Business	339	0.03	0.17
	Crash event	Road departure	RdwyDprt	5290	0.47	0.5
	Manner of collision (categorical)	Head-on	HeadOn	64	0.01	0.07
		Rear-ended	RearEnd	3591	0.32	0.47
		Sideswipe	SdSwp	2093	0.18	0.39
	Roadway condition	Water on roadway	Rd_WtRdwy	329	0.03	0.17
		Animal on roadway	Rd_AnRdwy	264	0.02	0.15
		Object on roadway	Rd_objRdwy_Others	784	0.07	0.25
	Driver's distraction	Cellphone	CellPh	123	0.01	0.1
		Inside the vehicle	InVeh	331	0.03	0.17
Outside the vehicle		OutVeh	293	0.03	0.16	
Vehicle type	Car	Car	4067	0.41	0.49	
	Large truck	Large_Truck	1340	0.12	0.32	
Environmental factors	Lighting conditions	Dark	Dark	3843	0.34	0.47
		Dusk/dawn	Dusk_Dawn	347	0.03	0.17
	Weather condition	Rain	Rain	2578	0.23	0.42
		Fog/snow	Fog_Snow	164	0.01	0.12
Geometric/traffic factors	Speed limit	Speed limit greater than 60 mph	PstSpd_greater_than_60 mph	9339	0.82	0.38

### 3.2. Data Preprocessing

Data preprocessing plays an important role in improving the quality of data, which includes the elimination of irrelevant data, standardization, and addressing multicollinearity [54]. Multicollinearity was checked with the Pearson correlation coefficient [55]. A correlation threshold of 0.7 was considered to identify highly correlated features. After removing highly correlated features, 45 features remained. Then, entries with null values were removed, resulting in the removal of 1867 crashes. The final number of crashes used for the analysis was 11,329.

ICT is classified into four classes based on 30 min equal intervals to improve modeling efficiency [56], ranging from short (0–30), medium (31–60), intermediate (61–90), and long (greater than 90). Among the crashes, 7322, 1507, 1061, and 1436 were classified as short, medium, intermediate, and long, respectively. The frequency distribution histograms of incident duration are shown in Figure 1.



**Figure 1.** Frequency distribution histograms of incident clearance time categories.

Categorical data are usually presented in words and cannot be directly loaded into the model [57]. Therefore, this categorical data must be converted to numerical data in advance. Label encoding, one-hot encoding, and target statistics are among the most frequently employed processing methods. This study used one-hot encoding and transformed data into a 0–1 variable.

The data set used in the analysis includes 11,329 incidents. The data set was categorized into 4 levels: 7322 short, 1507 medium, 1061 intermediate, and 1436 long. There is considerable variation in the number of categories, which has resulted in category imbalance. Such imbalances lead to bias towards the majority class since modeling classifiers prioritize the class with the most observations, generating overprediction of this class [58]. This overprediction can significantly limit the model’s generalization capacity, resulting in poor classification results. In this study, the synthetic minority oversampling technique (SMOTE) was adopted to address this issue by balancing class distribution across the data set, which was introduced by Chawla et al. in 2002 [59]. Instead of repeating existing minority cases, the SMOTE approach generates synthetic minority instances using random intervals between them.

### 3.3. Feature Selection

The feature selection process involves selecting features that significantly impact the output parameter. It is an essential step in model development, as irrelevant and noisy dataset features can negatively affect the model’s performance [60]. Feature selection was undertaken using the XGBoost algorithm, with K-fold cross-validation ( $K = 5$ ) applied to the 45 features from the dataset. During each fold, 80% of the data were allocated for training, while the remaining 20% were used for testing. Five-fold cross-validation ensured that every observation was validated exactly once, balancing the avoidance of overfitting

with reduced computational demands. The top 37 features were extracted from the best fold using the information gain (IG) method [61]. The IG estimator assesses the value of an attribute by quantifying the information gain entropy for the class in a decreasing sequence [62]. High-dimensional data are frequently processed using this methodology [63]. In the IG estimator, each attribute is assigned a score between 1 and 0, signifying the degree of relevance from most to least relevant. Subsequently, the features with the highest scores are used as inputs in the subsequent dimensionality reduction phase. The selected features are seen in Figure 2 below.

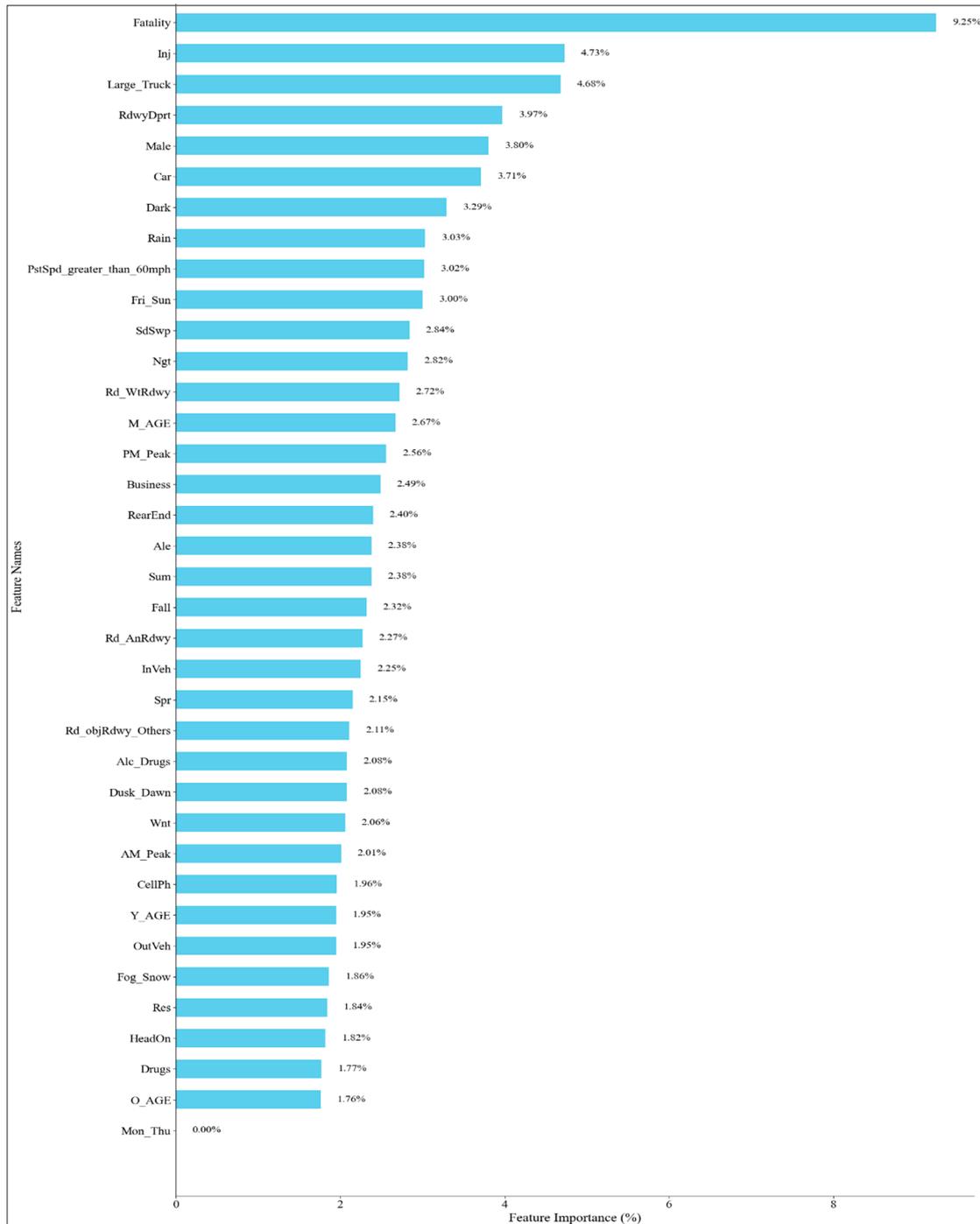


Figure 2. Features importance.

## 4. Methodology

### 4.1. Machine Learning Algorithms

This study employs three ML models, namely XGBoost, CatBoost, and LightGBM, to predict ICT. This section provides an overview of the algorithms.

#### 4.1.1. XGBoost

Extreme gradient boosting (XGBoost), proposed by Chen and Guestrin, is a decision tree-based ensemble algorithm that utilizes a gradient-boosting (GB) framework technique for classification and regression problems [64]. XGBoost is a package for boosting trees that uses  $k$  essential tree functions. GB develops a prediction model by integrating weak learners (such as decision trees) to form strong learners [65]. The model sequentially updates the weights based on the information derived from the errors of previous iterations. To enhance accuracy, the gradient descent approach is used during each iteration to minimize the arbitrary loss function, which serves as a measure of predictive ability. In contrast to GB, XGBoost employs a more accurate estimation approach [66]. Unlike GB, XGBoost calculates the second-order gradient of the loss function, L1 and L2 regularization, which offers additional details for achieving the minimum loss function and direction of the gradient [67]. Furthermore, XGBoost applies an additional regularized approach ( $\lambda$ ) that allows for better control over the complexity of the model and mitigates the probability of overfitting, which results in improved accuracy and model performance [68,69]. The detailed mathematical procedure of XGBoost is outlined as follows.

Equation (1) illustrates the boosting procedure using a dataset with  $n$  observations. Each observation is made up of several features designated as  $a_i$ , along with a corresponding response variable,  $b_i$ . The predicted response value after  $j$ th iterations, denoted as  $\hat{b}_i^{(j)}$ , is obtained by incorporating the predicted value from the  $(j - 1)$ th iteration for the  $i$ th observation with the addition of a tree function  $f(a_i)$ .

$$\hat{b}_i^{(j)} = \sum_{k=1}^j f_k(a_i) = \hat{b}_i^{(j-1)} + f_j(a_i) \tag{1}$$

The goal of this process is to minimize Equation (2), where  $l(a_i, \hat{b}_i)$  and  $\Omega(f_j) = YG + \frac{1}{2}\lambda\sum_{m=1}^G w_m^2$  represent the loss function and the model's complex penalty, respectively.  $Y$  represents a minimum number of instance weights required in a child node.  $G$  refers to the number of leaves, and  $w_m^2$  shows the L2 norm of the scores associated with the  $m$ th leaf. In order to avoid overfitting, this term is used.

$$O = \sum_{i=1}^n l(b_i, \hat{b}_i) + \sum_{k=1}^j \Omega(f_k) \tag{2}$$

After solving Equations (1) and (2), the ideal value of  $W_m$  is as follows:

$$W_m^* = -\frac{\sum_i \partial_{\hat{b}_i^{(j-1)}} l(b_i, \hat{b}_i^{(j-1)})}{\sum_i \partial_{\hat{b}_i^{(j-1)}}^2 l(b_i, \hat{b}_i^{(j-1)}) + \lambda} \tag{3}$$

The associated minimum objective value is as follows:

$$O_{min} = -\frac{1}{2} \sum_{m=1}^G \frac{\left(\sum_i \partial_{\hat{b}_i^{(j-1)}} l(b_i, \hat{b}_i^{(j-1)})\right)^2}{\sum_i \partial_{\hat{b}_i^{(j-1)}}^2 l(b_i, \hat{b}_i^{(j-1)}) + \lambda} + YG \tag{4}$$

#### 4.1.2. CatBoost

CatBoost, introduced by Dorogush et al., is a novel gradient-boosting algorithm that effectively operates with categorical features while minimizing information loss [70]. This technique is distinct from other GB algorithms. Initially, it employs ordered boosting, a

highly efficient modification of gradient boosting algorithms, to resolve the issue of target leakage [71]. Secondly, this algorithm is advantageous for relatively small datasets. Third, CatBoost is capable of managing categorical features, which is typically carried out during the preprocessing phase and involves substituting the original categorical variables with one or more numerical values.

The algorithm also has the advantage of performing random permutations to estimate leaf values when selecting the tree structure, which helps mitigate the overfitting common in traditional GB algorithms. Binary decision trees work as the main predictor in CatBoost. The estimated output of CatBoost is described in Equation (5), as follows:

$$G = K(x_i) = \sum_{l=1}^L b_l * 1(x \in P_l) \quad (5)$$

where  $K(x_i)$  is a decision tree function generated by the explanatory variables  $x_i$ , and  $P_l$  is the disjoint region that corresponds to the tree's leaves.

#### 4.1.3. LightGBM

LightGBM is a high-performance implementation of the gradient-boosting technique, developed in 2017 as an effective tool for many ML tasks [72]. The system utilizes two cutting-edge methods, namely gradient-based one-sided sampling (GOSS) and exclusive feature bundling (EFB), to speed up training and achieve exceptional accuracy. The GOSS technique is a gradient-boosting variant that prioritizes cases with greater gradients. This results in faster learning and reduced model complexity [73]. The EFB technique employs a feature selection method that involves merging sparse and mutually exclusive attributes and subsequently grouping these features to decrease the dimensionality of the feature matrix. Although LightGBM has its benefits, it is susceptible to overfitting when dealing with limited training datasets. This is primarily due to the creation of complex trees through leaf-wise tree splitting [74].

#### 4.2. Model Development

The current study used the Python Jupyter Notebook to develop all three machine-learning models, namely CatBoost, XGBoost, and LightGBM. The performance of any ML model is directly and significantly impacted by choosing the optimal hyperparameters [75]. The current research utilizes the grid search and five Kfold cross-validation ( $K = 5$ ) techniques to identify the optimal hyperparameters for all three models. Grid search is a conventional approach for optimizing hyperparameters that involves a thorough search within predefined boundaries [76]. The training and test folds consist of 60% (6797) and 40% (4532) of the total data set, respectively. Models were trained using the training data after the K-fold and tested on the testing data. All the data preprocessing, standardization, K-fold, grid search cross-validation for hyperparameters, and model training were performed using a Python Jupyter notebook. The hyperparameter settings of the models are outlined as follows:

- XGBoost: subsample = 0.8, reg\_lambda = 2.5, reg\_alpha = 0.1, n\_estimators = 300, max\_depth = 9, learning\_rate = 0.1, gamma = 0, colsample\_bytree = 1.0.
- CatBoost: iterations = 100, learning\_rate = 0.1, depth = 6, l2\_leaf\_reg = 3, border\_count = 32, ctr\_border\_count = 50.
- LightGBM: num\_leaves = 31, n\_estimators = 200, max\_depth = 20, learning\_rate = 0.1, lambda\_l2 = 0, lambda\_l1 = 0.5, feature\_fraction = 0.8, boosting\_type = gbdt.

#### 4.3. SHAP Analysis

SHAP (Shapley additive explanations) analysis, introduced by Lundberg and Lee, was employed in this study to improve the interpretability of ML models by providing a more accurate influence of each input feature on output [77]. SHAP analysis uses game theory and local explanations to quantify the contribution of each input feature [78]. The

following Equation (10) is used to calculate the Shapley coefficient  $\phi_i$ , which is the basis for this analysis:

$$\phi_i = \sum_{G \subseteq H \setminus \{i\}} \frac{|G|!(|G| - |G| - 1)!}{|G|} \left[ h_{G \cup \{i\}}(x_{G \cup \{i\}}) - h_G(x_G) \right] \quad (6)$$

where H represents all the inputs, while G indicates a subset of H that does not contain the input feature with index i. Equation (6) evaluates a feature’s influence by evaluating the variation in the model’s results when the attribute is added or removed from the input variables.

#### 4.4. Model Evaluation

Several metrics were used to evaluate the performance of the developed machine-learning models. All of these metrics were derived from the confusion matrix, a two-dimensional contingency table that presents the relation between the actual and predicted classes [79]. A confusion matrix is comprised of four possible scenarios: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP represents the number of incidents correctly classified within their respective duration categories (short, medium, intermediate, and long). TN represents the number of incidents that do not lie within the specific duration range and are correctly classified by the models as not being present within that range. FP represents the number of incidents that do not lie within the particular duration range but which are incorrectly classified by the models as present within that range, and FN represents the number of incidents that do lie within the specified duration range but which are incorrectly classified as not present within that range.

In this study, four performance evaluation metrics—accuracy, recall/sensitivity, precision, and F1 score—were used. The mathematical formulas for each metric are shown in Equations (7)–(10). In addition, the macro average is introduced as a comprehensive evaluation metric for model performance. The macro average is calculated by averaging the metrics in each category.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

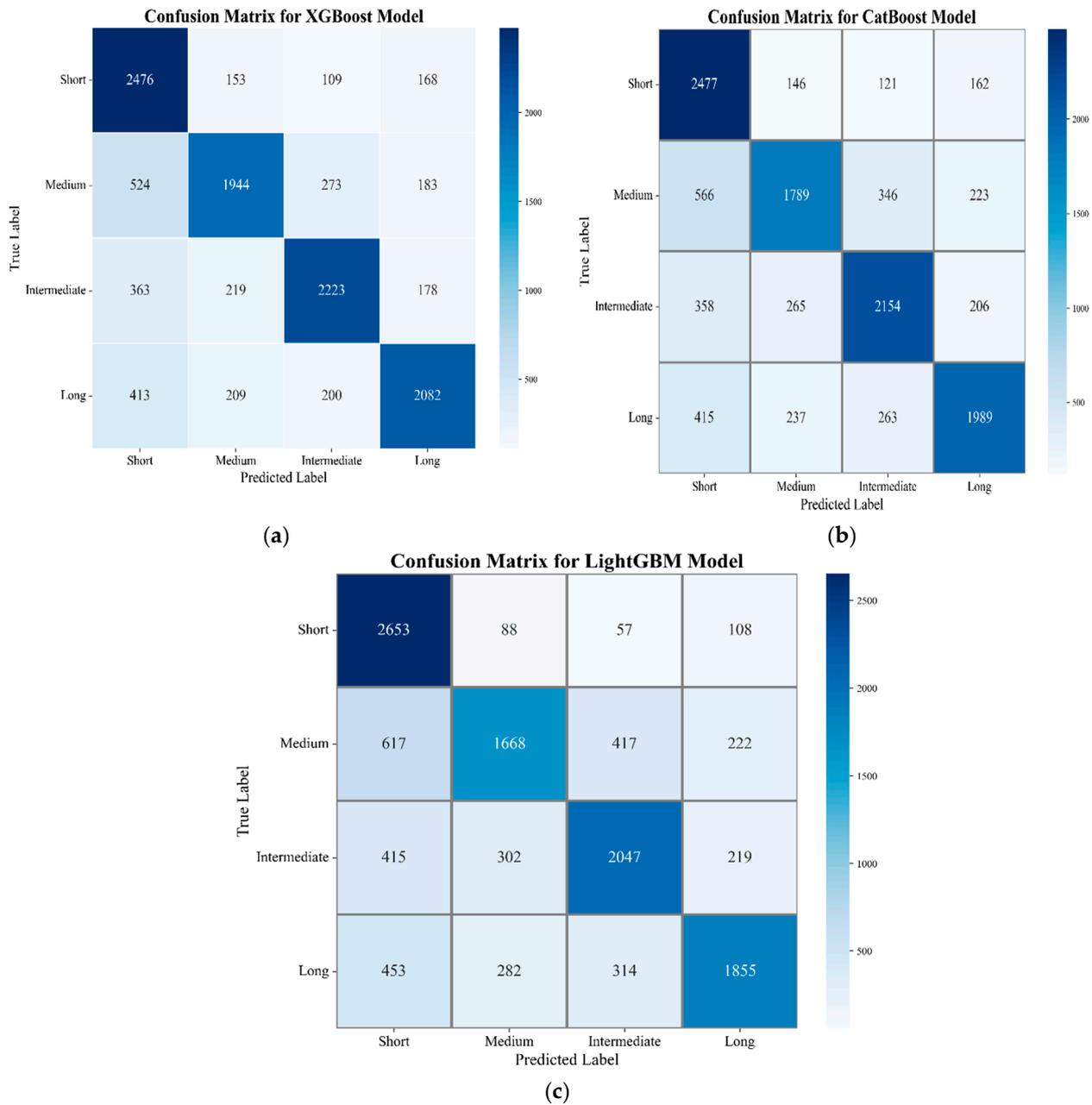
$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

## 5. Results and Discussion

The results from the ML and SHAP analyses are discussed in the following sections.

### 5.1. Machine Learning Model Results

Figure 3 presents the model’s predictive performance of each ML method in terms of their respective confusion matrix. As shown in Figure 3, the XGBoost model shows better balance performance results than the other ML methods. For the XGBoost model, 2476 (85.20%) crashes were correctly classified as short durations, whereas only 430 (4.08%) were not classified as short-duration incidents, of which 153 (5.26%) were misclassified as medium, 109 (3.75%) crashes as intermediate, and 168 (5.78%) as long. For the medium duration, the model accurately classified 1944 (66.48%) crashes as medium duration, whereas 980 (33.51%) were misclassified, with an overall accuracy of 85.08%. Similarly, for the intermediate duration, around 74.52% of crashes were correctly classified. Considering the long durations, 2082 (71.69%) crashes were correctly classified, with an overall accuracy of 87.13%.

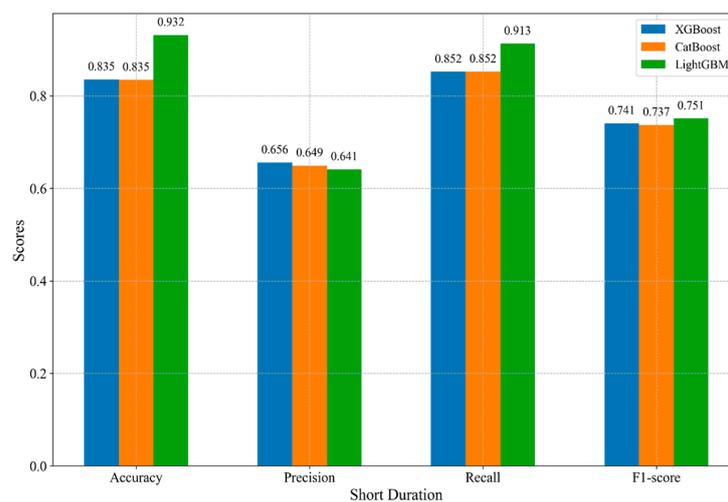


**Figure 3.** Confusion matrixes for ML models: (a) XGBoost; (b) CatBoost; (c) LightGBM.

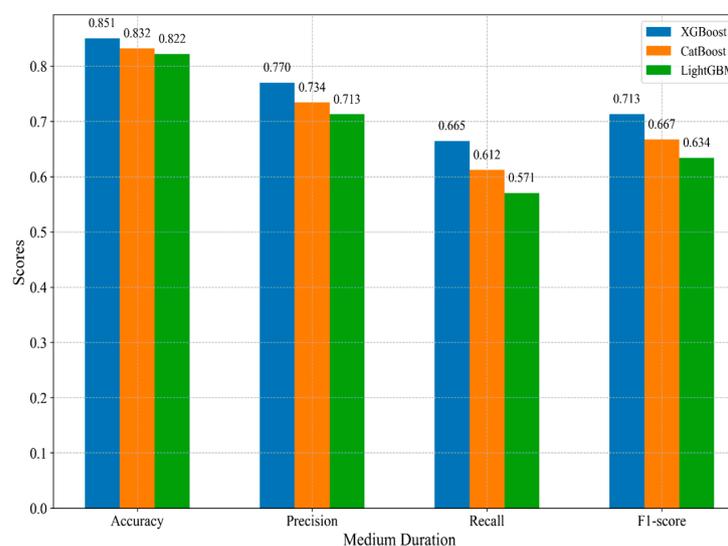
The confusion matrixes of CatBoost and LightGBM can be interpreted in the same manner. For a short duration, the LightGBM achieved the highest accuracy of 93.17% and recall of 91.30%, followed by XGBoost and CatBoost, which indicates that LightGBM is the best model for predicting short durations. XGBoost outperforms other models in predicting medium, intermediate, and long durations, with accuracies of 85.08%, 87.15%, and 87.13%, respectively. Table 3 presents the performance metrics by individual ICT levels for the three developed models. Figures 4–7 is a graphical representation of the model comparisons based on the selected performance metrics.

**Table 3.** Performance metrics by individual ICT levels for different models.

Model	Classes	TP	FP	FN	TN	Accuracy	Precision	Recall	F1-Score
XGBoost	Short	2476	1300	430	6331	0.8354	0.6557	0.8520	0.7407
	Medium	1944	581	980	7032	0.8508	0.7702	0.6648	0.7134
	Intermediate	2223	582	760	6972	0.8715	0.7926	0.7453	0.7683
	Long	2082	529	822	7104	0.8713	0.7970	0.7172	0.7551
CatBoost	Short	2477	1339	429	6472	0.8347	0.6492	0.8524	0.7372
	Medium	1789	648	1135	7145	0.8323	0.7342	0.6120	0.6674
	Intermediate	2154	730	829	7004	0.8537	0.7467	0.7221	0.7342
	Long	1989	591	915	7222	0.8575	0.7709	0.6848	0.7255
LightGBM	Short	2653	1485	253	7374	0.9317	0.6412	0.9130	0.7513
	Medium	1668	672	1256	7169	0.8221	0.7128	0.5707	0.6342
	Intermediate	2047	788	936	6994	0.8396	0.7221	0.6865	0.7038
	Long	1855	549	1049	8312	0.9461	0.7718	0.6389	0.6993



**Figure 4.** Comparison of classification ML models for short durations.



**Figure 5.** Comparison of classification ML models for medium durations.

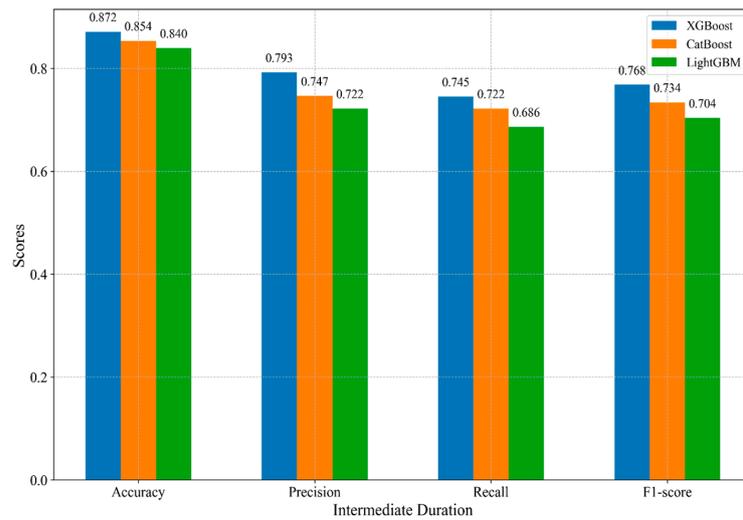


Figure 6. Comparison of classification ML models for intermediate durations.

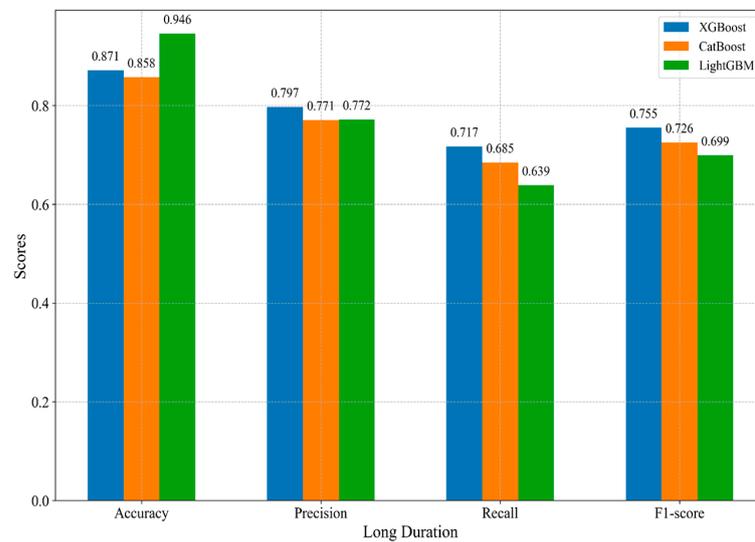


Figure 7. Comparison of classification ML models for longer durations.

The average performance evaluation metrics, such as average accuracy, macro-average precision, macro-average recall, and macro-average F1-score of all the developed models, are summarized in Table 4. LightGBM, with an average accuracy of (88.49%) outperformed the other models. XGboost, on the other hand, has the best predictive capabilities in terms of its high macro-average recall, macro-average precision, and macro-average F1 score. XGBoost outperforms CatBoost and LightGBM in the collective model’s predictive performance.

Table 4. Overall performance of XGBoost, CatBoost, and LightGBM.

Models	Average Accuracy	Macro Average Recall	Macro Average Precision	Macro Average F1-Score
XGBoost	0.8573	0.7448	0.7539	0.7444
CatBoost	0.8446	0.7178	0.7252	0.7161
LightGBM	0.8849	0.7023	0.7120	0.6972

### 5.2. SHAP Results

Figure 8 shows the SHAP summary plot, which orders features based on their importance in classifying short, medium, intermediate, and long ICT. Crashes involving

injuries and fatalities significantly increase the probability of intermediate and longer ICT, as seen in Figure 8c,d. The reason for the increased ICT with crashes involving injuries and fatalities is the need for response agencies and emergency medical services (EMS). For short ICT, as evident from Figure 8a, injuries and fatalities have a very low impact as incidents cleared during this duration are mainly less severe. This finding is consistent with prior research [2,40]. Additionally, it was found that crashes that involve trucks significantly increase the probability of long ICT. This is because the incident involving trucks resulted in an increased number of injuries, fatalities, and property damage. This finding is consistent with other studies that have found that heavy truck crashes increase incident durations significantly in comparison to other vehicles [51,52]. For the manner of collision, the result shows various impacts on different duration levels. For example, sideswipe and rear-end crashes slightly increase the short, intermediate, and long ICT probability. However, head-on crashes tend to significantly increase the probability of long ICT. This result is expected, as head-on crashes typically result in a higher number of injuries in comparison to sideswipes and rear-end [48,49].

In terms of driver age, the variables found to be significant are young age and old age. Young age increases the probability of intermediate ICT, whereas old age significantly increases long ICT. This is due to older drivers having health complications and slower reaction times, which may lead to severe crashes [80]. The results also show that road departure crashes increase the probability of long ICT. This finding is intuitive as road departure crashes are associated with severe injury crashes and are more devastating when such crashes involve encroachment into roadside obstacles [81,82]. For driver distraction features, the result in Figure 8d indicates that the use of cell phones and distractions inside the vehicles and outside both tend to increase the probability of long ICT. This happens because distraction diverts the driver's attention from the road and surrounding traffic, which may lead to severe crashes [83]. In addition, incidents during peak hours tend to have an increased probability of long ICT. This is due to peak-hour traffic and increased congestion [45,84].

Regarding environmental and temporal characteristics, dark, dusk, and dawn are associated with increased probabilities of intermediate and long ICT. This can be due to the lesser availability of incident response resources and poor visibility during dark, which can make it harder for responders to assess the situation quickly and can lead to increased ICT [47]. Also, incidents that occurred during rain, fog, and snow increased the probability of longer ICT. It was reasonable that adverse weather conditions might cause a delay in incident response. This effect of rainy and snowy conditions on ICT is supported by previous research [37].

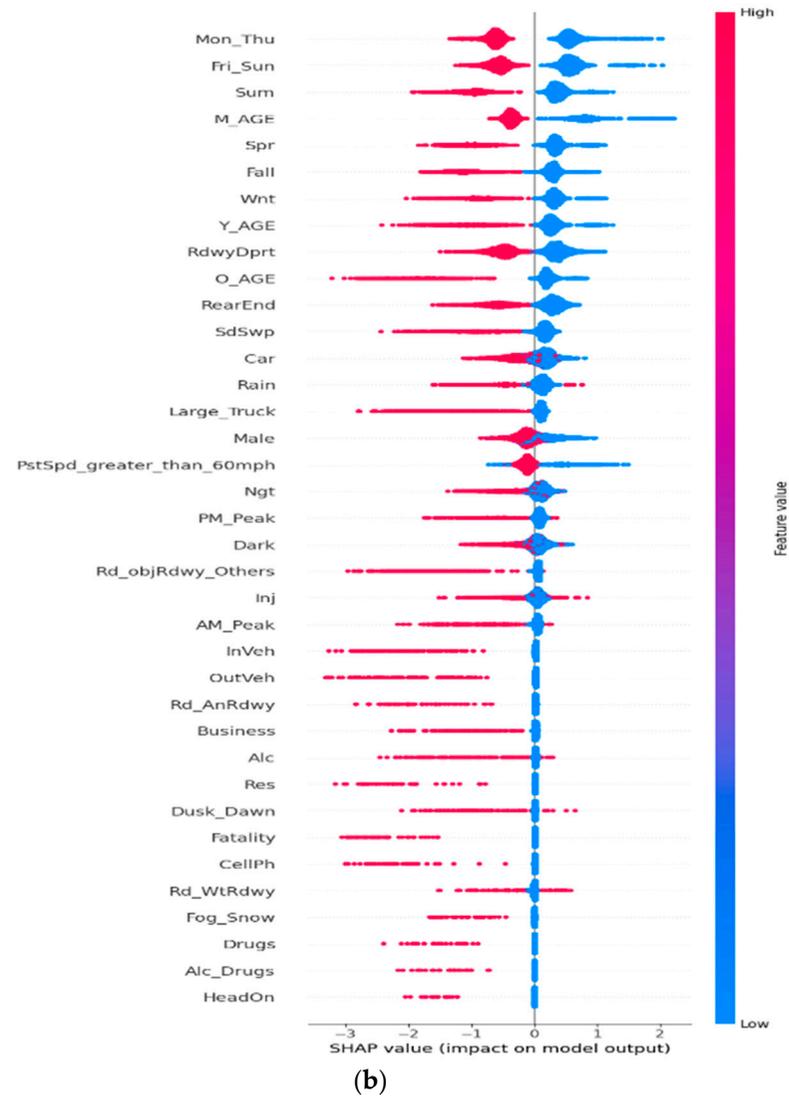
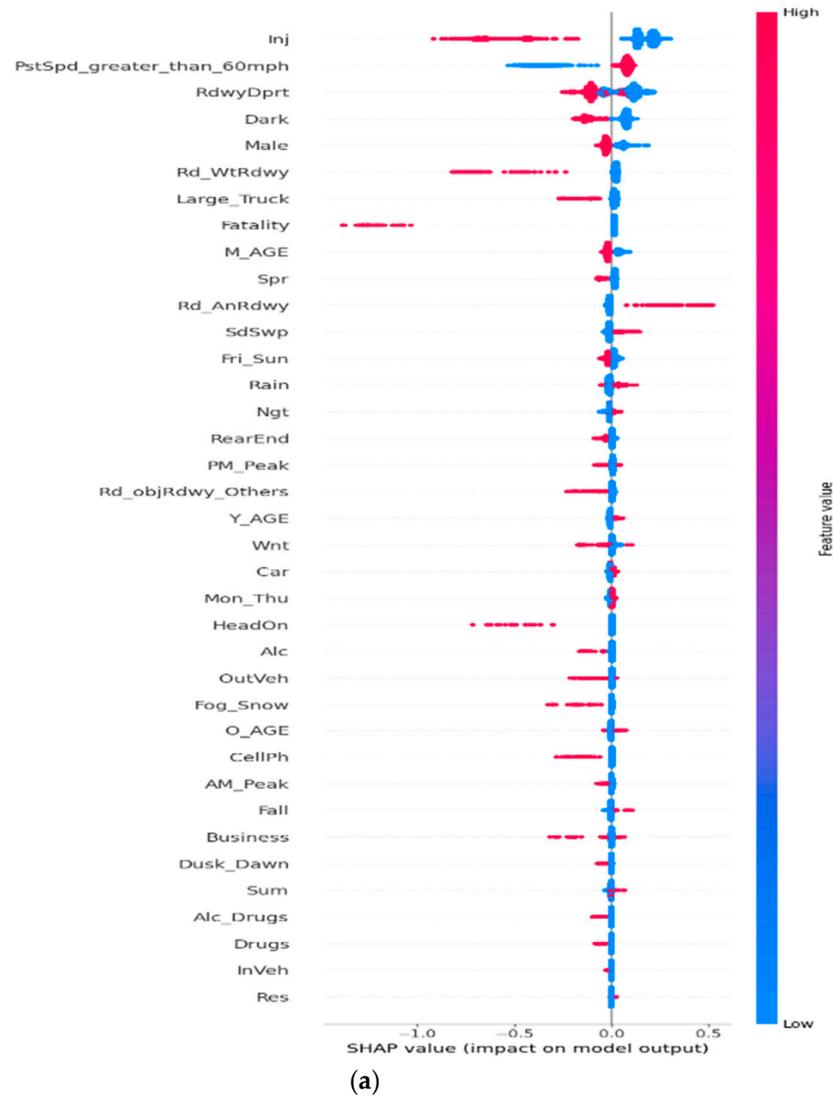


Figure 8. Cont.

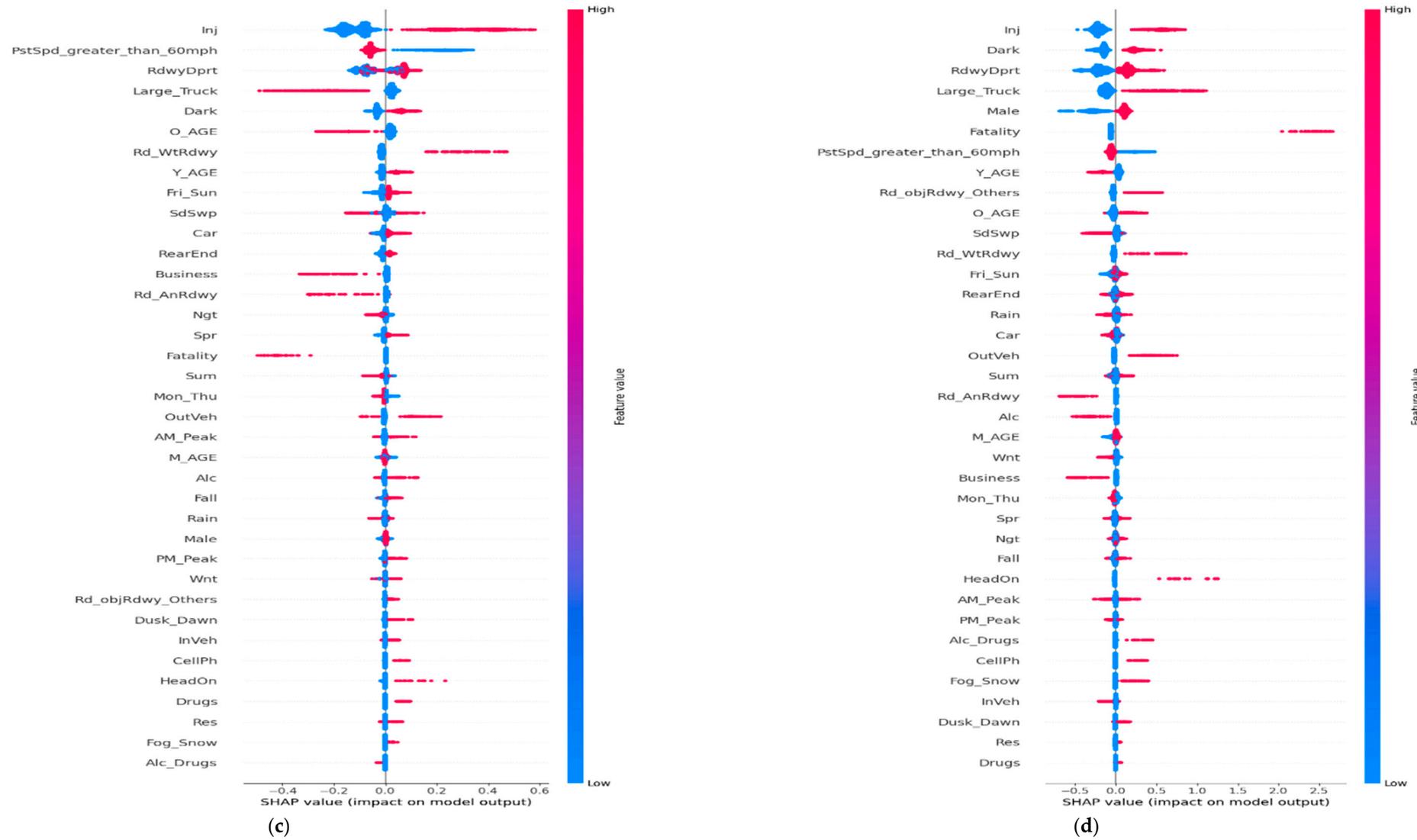


Figure 8. SHAP summary plot: (a) short duration; (b) medium duration; (c) intermediate duration; (d) long duration.

## 6. Conclusions

This study developed machine learning models to predict incident clearance time and used SHAP analysis to identify significant factors impacting crash-related ICT on Louisiana's rural interstates. The two separate approaches were utilized in a complementary framework by taking advantage of the strengths of machine learning models and SHAP analysis. While machine learning models provide better predictions, their methodologies are considered black boxes and do not excel at establishing relationships between response and independent variables. SHAP analysis is flexible and useful for analyzing the impact and direction of the influence of variables on a phenomenon. Machine learning and SHAP analysis were applied to crash datasets to analyze ICT in relation to traffic crashes.

This study developed three ML models, namely XGBoost, CatBoost, and LightGBM, to predict incident clearance time in rural interstates in Louisiana, utilizing data collected from the Louisiana DOTD's crash database. Four ICT levels were defined based on 30 min intervals: short (0–30), medium (31–60), intermediate (61–90), and long (greater than 90). Based on different performance evaluation metrics, the results revealed that LightGBM performs better when predicting short-duration incidents, followed by XGBoost and CatBoost. XGBoost outperforms other models in predicting medium, intermediate, and long durations with the highest accuracy. In terms of average performance measures, XGBoost outperforms CatBoost and LightGBM in the collective model's predictive performance. By implementing these developed models, operators at the traffic management centers (TMC) can predict and disseminate accurate information about the duration of the incident to motorists and other agencies involved in managing road incidents.

SHAP analysis was performed to identify significant features impacting ICT. It was found that crashes involving injuries, fatalities, heavy trucks, head-on collisions, roadway departure, and older drivers tend to increase the probability of longer ICT. In addition, other factors, like cellphone distraction and dark, rainy, snowy, and foggy conditions, also prolong ICT. The identification of these significant features impacting ICT can help incident management agencies, transportation planners, and policymakers take countermeasures to reduce ICT on rural interstates. For instance, strict implementation of distracted driving laws should be implemented to reduce the use of cell phones while driving. Also, lighting conditions on rural interstates in Louisiana should be improved. Taking seasonal countermeasures like roadside warning signs during foggy, snowy, or rainfall conditions is also recommended.

The findings of this study are beneficial for traffic managers, transportation planners, and incident response agencies because they provide good prediction and analytical results for crash-related ICT by using machine learning models and SHAP analysis. The following are the recommendations for future research work:

- It is recommended to predict different phases of incident duration, such as detection and response time, to better understand the factors responsible for prolonged ICT.
- This research study is limited to Louisiana's rural interstates. Future studies are recommended to include urban interstates and other road types across Louisiana to improve the model's applicability.
- To improve the model's accuracy and predictive performance, it is recommended that the model be trained on comprehensive data sets spanning more than 10 years
- It is also recommended that real-time traffic and weather data be used during modeling through the use of installed sensors to improve the model's prediction accuracy.

**Author Contributions:** The authors confirm their contributions to the paper as follows: study conception and design: M.M. and M.A.R.; data collection and preparation: M.A.R. and W.A.K.; analysis and interpretation of results: W.A.K. and M.M.; draft manuscript preparation: W.A.K., M.M., K.A.T., J.C. and V.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is a complementary work of a research project (SIO No. DOTDLT 1000468) funded by the Louisiana Department of Transportation and Development (DOTD) and was conducted by the Louisiana Transportation Research Center (LTRC).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data will be available only upon request and with permission from the Louisiana Department of Transportation and Development (DOTD).

**Conflicts of Interest:** The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

- World Health Organization. *Global Status Report on Road Safety 2018*; World Health Organization: Geneva, Switzerland, 2019.
- Alkaabi, A.M.S.; Dissanayake, D.; Bird, R. Analyzing Clearance Time of Urban Traffic Accidents in Abu Dhabi, United Arab Emirates, with Hazard-Based Duration Modeling Method. *Transp. Res. Rec.* **2011**, *2229*, 46–54. [\[CrossRef\]](#)
- National Traffic Incident Management Coalition (NTIMC). *Benefits of Traffic Incident Management*; NTIMC: Washington, DC, USA, 2006.
- Wang, S.; Li, R.; Guo, M. Application of Nonparametric Regression in Predicting Traffic Incident Duration. *Transport* **2018**, *33*, 22–31. [\[CrossRef\]](#)
- Haule, H.J.; Sando, T.; Lentz, R.; Chuan, C.-H.; Alluri, P. Evaluating the Impact and Clearance Duration of Freeway Incidents. *Int. J. Transp. Sci. Technol.* **2019**, *8*, 13–24. [\[CrossRef\]](#)
- Lee, Y.; Wei, C. A Computerized Feature Selection Method Using Genetic Algorithms to Forecast Freeway Accident Duration Times. *Comput.-Aided Civil. Infrastruct. Eng.* **2010**, *25*, 132–148. [\[CrossRef\]](#)
- Madanat, S.; Feroze, A. *Prediction Models for Incident Clearance Time for Borman Expressway (Vol. 1; Vol. 2: 96/11)*; Joint Highway Research Project, Indiana Department of Transportation and Purdue University: West Lafayette, India, 1997.
- Zhang, H.; Khattak, A.J. Analysis of Cascading Incident Event Durations on Urban Freeways. *Transp. Res. Rec.* **2010**, *2178*, 30–39. [\[CrossRef\]](#)
- Khattak, A.J.; Wang, X.; Zhang, H. Spatial Analysis and Modeling of Traffic Incidents for Proactive Incident Management and Strategic Planning. *Transp. Res. Rec.* **2010**, *2178*, 128–137. [\[CrossRef\]](#)
- Special Report 209: Highway Capacity Manual*, 3rd ed.; TRB, National Research Council: Washington, DC, USA, 1994.
- Lee, J.; Abdel-Aty, M.; Cai, Q.; Wang, L. Analysis of Fatal Traffic Crash-Reporting and Reporting-Arrival Time Intervals of Emergency Medical Services. *Transp. Res. Rec.* **2018**, *2672*, 61–71. [\[CrossRef\]](#)
- Zhao, Y.; Deng, W. Prediction in Traffic Accident Duration Based on Heterogeneous Ensemble Learning. *Appl. Artif. Intell.* **2022**, *36*, 2018643. [\[CrossRef\]](#)
- Khattak, A.J.; Liu, J.; Wali, B.; Li, X.; Ng, M. Modeling Traffic Incident Duration Using Quantile Regression. *Transp. Res. Rec.* **2016**, *2554*, 139–148. [\[CrossRef\]](#)
- Valenti, G.; Lelli, M.; Cucina, D. A Comparative Study of Models for the Incident Duration Prediction. *Eur. Transp. Res. Rev.* **2010**, *2*, 103–111. [\[CrossRef\]](#)
- Cong, H.; Chen, C.; Lin, P.-S.; Zhang, G.; Milton, J.; Zhi, Y. Traffic Incident Duration Estimation Based on a Dual-Learning Bayesian Network Model. *Transp. Res. Rec.* **2018**, *2672*, 196–209. [\[CrossRef\]](#)
- Zhan, C.; Gan, A.; Hadi, M. Prediction of Lane Clearance Time of Freeway Incidents Using the M5P Tree Algorithm. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 1549–1557. [\[CrossRef\]](#)
- Wei, C.-H.; Lee, Y. Sequential Forecast of Incident Duration Using Artificial Neural Network Models. *Accid. Anal. Prev.* **2007**, *39*, 944–954. [\[CrossRef\]](#)
- Tang, J.; Zheng, L.; Han, C.; Yin, W.; Zhang, Y.; Zou, Y.; Huang, H. Statistical and Machine-Learning Methods for Clearance Time Prediction of Road Incidents: A Methodology Review. *Anal. Methods Accid. Res.* **2020**, *27*, 100123. [\[CrossRef\]](#)
- Shang, Q.; Tan, D.; Gao, S.; Feng, L. A Hybrid Method for Traffic Incident Duration Prediction Using BOA-Optimized Random Forest Combined with Neighborhood Components Analysis. *J. Adv. Transp.* **2019**, *2019*, 4202735. [\[CrossRef\]](#)
- Zhu, W.; Wu, J.; Fu, T.; Wang, J.; Zhang, J.; Shanguan, Q. Dynamic Prediction of Traffic Incident Duration on Urban Expressways: A Deep Learning Approach Based on LSTM and MLP. *J. Intell. Connect. Veh.* **2021**, *4*, 80–91. [\[CrossRef\]](#)
- Shang, Q.; Xie, T.; Yu, Y. Prediction of Duration of Traffic Incidents by Hybrid Deep Learning Based on Multi-Source Incomplete Data. *Int. J. Environ. Res. Public Health* **2022**, *19*, 10903. [\[CrossRef\]](#)
- Chen, J.; Tao, W. Traffic Accident Duration Prediction Using Text Mining and Ensemble Learning on Expressways. *Sci. Rep.* **2022**, *12*, 21478. [\[CrossRef\]](#)
- He, Q.; Kamarianakis, Y.; Jintanakul, K.; Wynter, L. Incident Duration Prediction with Hybrid Tree-Based Quantile Regression. In *Advances in Dynamic Network Modeling in Complex Transportation Systems*; Springer: New York, NY, USA, 2013; pp. 287–305.
- Ma, X.; Ding, C.; Luan, S.; Wang, Y.; Wang, Y. Prioritizing Influential Factors for Freeway Incident Clearance Time Prediction Using the Gradient Boosting Decision Trees Method. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2303–2310. [\[CrossRef\]](#)
- Grigorev, A.; Mihaita, A.-S.; Lee, S.; Chen, F. Incident Duration Prediction Using a Bi-Level Machine Learning Framework with Outlier Removal and Intra-Extra Joint Optimisation. *Transp. Res. Part C Emerg. Technol.* **2022**, *141*, 103721. [\[CrossRef\]](#)

26. Li, R.; Pereira, F.C.; Ben-Akiva, M.E. Overview of Traffic Incident Duration Analysis and Prediction. *Eur. Transp. Res. Rev.* **2018**, *10*, 22. [[CrossRef](#)]
27. Rahmat-Ullah, Z.; Alsmadi, S.; Hamad, K. Classifying and Forecasting Traffic Incident Duration Using Various Machine Learning Techniques. In Proceedings of the 2021 14th International Conference on Developments in eSystems Engineering (DeSE), Sharjah, United Arab Emirates, 7–10 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 388–393.
28. Hamad, K.; Khalil, M.A.; Alozi, A.R. Predicting Freeway Incident Duration Using Machine Learning. *Int. J. Intell. Transp. Syst. Res.* **2020**, *18*, 367–380. [[CrossRef](#)]
29. Yu, B.; Wang, Y.T.; Yaoz, J.B.; Wang, J.Y. A Comparison of the Performance of Ann and Svm for the Prediction of Traffic Accident Duration. *Neural Netw. World* **2016**, *26*, 271–287. [[CrossRef](#)]
30. Lee, Y.; Wei, C.-H.; Chao, K.-C. Non-Parametric Machine Learning Methods for Evaluating the Effects of Traffic Accident Duration on Freeways. *Arch. Transp.* **2017**, *43*, 91–104. [[CrossRef](#)]
31. Kidando, E.; Mihayo, M.; Salum, J.H.; Kutela, B.; Kitali, A.E.; Alluri, P.; Sando, T. Prediction of Traffic Incident Clearance Duration Using Neural Network for Multimodal Data Distribution. *J. Transp. Eng. A Syst.* **2024**, *150*, 04024052. [[CrossRef](#)]
32. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*; O'Reilly Media, Inc.: Sebastopol, MA, USA, 2022; ISBN 1098122461.
33. Hamad, K.; Obaid, L.; Nassif, A.B.; Abu Dabous, S.; Al-Ruzouq, R.; Zeiada, W. Comprehensive Evaluation of Multiple Machine Learning Classifiers for Predicting Freeway Incident Duration. *Innov. Infrastruct. Solut.* **2023**, *8*, 177. [[CrossRef](#)]
34. Obaid, L.; Hamad, K.; Khalil, M.A.; Nassif, A.B. Effect of Feature Optimization on Performance of Machine Learning Models for Predicting Traffic Incident Duration. *Eng. Appl. Artif. Intell.* **2024**, *131*, 107845. [[CrossRef](#)]
35. Zhu, S. Comparative Study of Statistical and Machine Learning Methods for Streetcar Incident Duration Analysis. *Int. J. Crashworthiness* **2024**, *29*, 16–21. [[CrossRef](#)]
36. Jia, X.L.; Li, S.Q.; Yang, H.Z.; Chen, X.P. Prediction of the Duration of Freeway Traffic Incidents Based on an ATT-LSTM Model. *J. Transp. Informat. Safet.* **2022**, *40*, 61–69.
37. Nam, D.; Mannering, F. An Exploratory Hazard-Based Analysis of Highway Incident Duration. *Transp. Res. Part A Policy Pract.* **2000**, *34*, 85–102. [[CrossRef](#)]
38. Ding, C.; Ma, X.; Wang, Y.; Wang, Y. Exploring the Influential Factors in Incident Clearance Time: Disentangling Causation from Self-Selection Bias. *Accid. Anal. Prev.* **2015**, *85*, 58–65. [[CrossRef](#)] [[PubMed](#)]
39. Zeng, Q.; Wang, F.; Chen, T.; Sze, N.N. Incorporating Real-Time Weather Conditions into Analyzing Clearance Time of Freeway Accidents: A Grouped Random Parameters Hazard-Based Duration Model with Time-Varying Covariates. *Anal. Methods Accid. Res.* **2023**, *38*, 100267. [[CrossRef](#)]
40. Hou, L.; Lao, Y.; Wang, Y.; Zhang, Z.; Zhang, Y.; Li, Z. Time-Varying Effects of Influential Factors on Incident Clearance Time Using a Non-Proportional Hazard-Based Model. *Transp. Res. Part A Policy Pract.* **2014**, *63*, 12–24. [[CrossRef](#)]
41. Adeel, M.; Khattak, A.J.; Mishra, S.; Thapa, D. Enhancing Work Zone Crash Severity Analysis: The Role of Synthetic Minority Oversampling Technique in Balancing Minority Categories. *Accid. Anal. Prev.* **2024**, *208*, 107794. [[CrossRef](#)]
42. Islam, N.; Adanu, E.K.; Hainen, A.M.; Burdette, S.; Smith, R.; Jones, S. A Comparative Analysis of Freeway Crash Incident Clearance Time Using Random Parameter and Latent Class Hazard-Based Duration Model. *Accid. Anal. Prev.* **2021**, *160*, 106303. [[CrossRef](#)]
43. Tirtha, S.D.; Yasmin, S.; Eluru, N. Modeling of Incident Type and Incident Duration Using Data from Multiple Years. *Anal. Methods Accid. Res.* **2020**, *28*, 100132. [[CrossRef](#)]
44. Garib, A.; Radwan, A.E.; Al-Deek, H. Estimating Magnitude and Duration of Incident Delays. *J. Transp. Eng.* **1997**, *123*, 459–466. [[CrossRef](#)]
45. Junhua, W.; Haozhe, C.; Shi, Q. Estimating Freeway Incident Duration Using Accelerated Failure Time Modeling. *Saf. Sci.* **2013**, *54*, 43–50. [[CrossRef](#)]
46. Zou, Y.; Tang, J.; Wu, L.; Henrickson, K.; Wang, Y. Quantile Analysis of Factors Influencing the Time Taken to Clear Road Traffic Incidents. In Proceedings of the Institution of Civil Engineers-Transport; Thomas Telford Ltd.: London, UK, 2017; Volume 170, pp. 296–304.
47. Chung, Y. Development of an Accident Duration Prediction Model on the Korean Freeway Systems. *Accid. Anal. Prev.* **2010**, *42*, 282–289. [[CrossRef](#)]
48. Huang, H.; Li, C.; Zeng, Q. Crash Protectiveness to Occupant Injury and Vehicle Damage: An Investigation on Major Car Brands. *Accid. Anal. Prev.* **2016**, *86*, 129–136. [[CrossRef](#)]
49. Zeng, Q.; Wen, H.; Huang, H. The Interactive Effect on Injury Severity of Driver-Vehicle Units in Two-Vehicle Crashes. *J. Saf. Res.* **2016**, *59*, 105–111. [[CrossRef](#)] [[PubMed](#)]
50. Lee, J.-T.; Fazio, J. Influential Factors in Freeway Crash Response and Clearance Times by Emergency Management Services in Peak Periods. *Traffic Inj. Prev.* **2005**, *6*, 331–339. [[CrossRef](#)] [[PubMed](#)]
51. Chimba, D.; Kutela, B.; Ogletree, G.; Horne, F.; Tugwell, M. Impact of Abandoned and Disabled Vehicles on Freeway Incident Duration. *J. Transp. Eng.* **2014**, *140*, 04013013. [[CrossRef](#)]
52. Golob, T.F.; Recker, W.W.; Leonard, J.D. An Analysis of the Severity and Incident Duration of Truck-Involved Freeway Accidents. *Accid. Anal. Prev.* **1987**, *19*, 375–395. [[CrossRef](#)] [[PubMed](#)]

53. Park, H.; Haghani, A.; Zhang, X. Interpretation of Bayesian Neural Networks for Predicting the Duration of Detected Incidents. *J. Intell. Transp. Syst.* **2016**, *20*, 385–400. [[CrossRef](#)]
54. Teng, C.-M. Correcting Noisy Data. *ICML* **1999**, *99*, 239–248.
55. Cohen, I.; Huang, Y.; Chen, J.; Benesty, J.; Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson Correlation Coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
56. Ozbay, K.; Noyan, N. Estimation of Incident Clearance Times Using Bayesian Networks Approach. *Accid. Anal. Prev.* **2006**, *38*, 542–555. [[CrossRef](#)]
57. Rodríguez, P.; Bautista, M.A.; Gonzalez, J.; Escalera, S. Beyond One-Hot Encoding: Lower Dimensional Target Embedding. *Image Vis. Comput.* **2018**, *75*, 21–31. [[CrossRef](#)]
58. Thabtah, F.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data Imbalance in Classification: Experimental Evaluation. *Inf. Sci.* **2020**, *513*, 429–441. [[CrossRef](#)]
59. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
60. Sun, L.; Lin, Z.; Li, W.; Xiang, Y. Freeway Incident Detection Based on Set Theory and Short-Range Communication. *Transp. Lett.* **2019**, *11*, 558–569. [[CrossRef](#)]
61. Frénay, B.; Doquire, G.; Verleysen, M. Is Mutual Information Adequate for Feature Selection in Regression? *Neural Netw.* **2013**, *48*, 1–7. [[CrossRef](#)] [[PubMed](#)]
62. Holmes, G.; Donkin, A.; Witten, I.H. Weka: A Machine Learning Workbench. In Proceedings of the ANZIIS'94-Australian New Zealand Intelligent Information Systems Conference, Brisbane, QLD, Australia, 29 November–2 December 1994; IEEE: Piscataway, NJ, USA, 1994; pp. 357–361.
63. Koprinska, I. Feature Selection for Brain-Computer Interfaces. In Proceedings of the New Frontiers in Applied Data Mining: PAKDD 2009 International Workshops, Bangkok, Thailand, 27–30 April 2009; Revised Selected Papers 13. Springer: Berlin/Heidelberg, Germany, 2010; pp. 106–117.
64. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
65. Subasi, A. *Practical Machine Learning for Data Analysis Using Python*; Academic Press: Cambridge, MA, USA, 2020; ISBN 0128213809.
66. Raschka, S.; Liu, Y.H.; Mirjalili, V.; Dzhulgakov, D. *Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python*; Packt Publishing Ltd.: Birmingham, UK, 2022; ISBN 1801816387.
67. Bonaccorso, G. *Machine Learning Algorithms*; Packt Publishing Ltd.: Birmingham, UK, 2017; ISBN 1785884514.
68. Das, A.; Khan, M.N.; Ahmed, M.M. Detecting Lane Change Maneuvers Using SHRP2 Naturalistic Driving Data: A Comparative Study Machine Learning Techniques. *Accid. Anal. Prev.* **2020**, *142*, 105578. [[CrossRef](#)] [[PubMed](#)]
69. Mousa, S.R.; Bakhit, P.R.; Osman, O.A.; Ishak, S. A Comparative Analysis of Tree-Based Ensemble Methods for Detecting Imminent Lane Change Maneuvers in Connected Vehicle Environments. *Transp. Res. Rec.* **2018**, *2672*, 268–279. [[CrossRef](#)]
70. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient Boosting with Categorical Features Support. *arXiv* **2018**, arXiv:1810.11363.
71. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *Adv. Neural Inf. Process Syst.* **2018**, *31*, 6639–6649.
72. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process Syst.* **2017**, *30*, 3149–3157.
73. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [[CrossRef](#)]
74. Mahesh, B. Machine Learning Algorithms-a Review. *Int. J. Sci. Res. (IJSR)* **2020**, *9*, 381–386. [[CrossRef](#)]
75. Vincent, A.M.; Jidesh, P. An Improved Hyperparameter Optimization Framework for AutoML Systems Using Evolutionary Algorithms. *Sci. Rep.* **2023**, *13*, 4737. [[CrossRef](#)]
76. Krstajic, D.; Buturovic, L.J.; Leahy, D.E.; Thomas, S. Cross-Validation Pitfalls When Selecting and Assessing Regression and Classification Models. *J. Cheminform.* **2014**, *6*, 10. [[CrossRef](#)] [[PubMed](#)]
77. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process Syst.* **2017**, *30*, 4765–4774.
78. Štrumbelj, E.; Kononenko, I. Explaining Prediction Models and Individual Predictions with Feature Contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665. [[CrossRef](#)]
79. Deng, X.; Liu, Q.; Deng, Y.; Mahadevan, S. An Improved Method to Construct Basic Probability Assignment Based on the Confusion Matrix for Classification Problem. *Inf. Sci.* **2016**, *340*, 250–261. [[CrossRef](#)]
80. Silva, V.C.; Dias, A.S.; Greve, J.M.D.; Davis, C.L.; Soares, A.L.d.S.; Brech, G.C.; Ayama, S.; Jacob-Filho, W.; Busse, A.L.; de Biase, M.E.M. Crash Risk Predictors in Older Drivers: A Cross-Sectional Study Based on a Driving Simulator and Machine Learning Algorithms. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4212. [[CrossRef](#)]
81. Hosseinpour, M.; Smith, J.; Williams, B.; Clouser, J.; Anastasio, I.; Haleem, K. Comparative Analysis of Aggressive-Driving and Distracted-Driving Crashes Involving Commercial Motor Vehicles in Kentucky. In Proceedings of the International Conference on Transportation and Development 2021, Virtual, 8–10 June 2021; pp. 272–284.
82. Kusano, K.D.; Gabler, H.C. Characterization of Opposite-Direction Road Departure Crashes in the United States. *Transp. Res. Rec.* **2013**, *2377*, 14–20. [[CrossRef](#)]

- 
83. Karl, J.B.; Nyce, C.M.; Powell, L.; Zhuang, B. How Risky Is Distracted Driving? *J. Risk Uncertain.* **2023**, *66*, 279–312. [[CrossRef](#)]
  84. Rahman, M.A.; Moomen, M.; Khan, W.A.; Codjoe, J. An Analysis of the Impact of Injury Severity on Incident Clearance Time on Urban Interstates Using a Bivariate Random-Parameter Probit Model. *Stats* **2024**, *7*, 863–874. [[CrossRef](#)]

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