Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 11, No. 10, October, 2025

A Comparative Study of a Series of Supervised Learning Models for Motorcycle Crash Injury Severity Prediction

Sonita Sum [1], Panuwat Wisutwattanasak [2], Thanapong Champahom [3], Sajjakaj Jomnonkwao [1], Vatanavongs Ratanavaraha [1*]

[1] School of Transportation Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.

[2] Institute of Research and Development, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.

[3] Department of Management, Faculty of Business Administration, Rajamangala University of Technology Isan, Nakhon Ratchasima 30000, Thailand.

Received 27 June 2025; Revised 18 September 2025; Accepted 25 September 2025; Published 01 October 2025

Abstract

Motorcycle crashes pose a major public health challenge in Thailand, where motorcyclists account for most traffic fatalities. This study aims to evaluate and compare the predictive performance of four supervised learning models—Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forest (RF)—for motorcycle crash injury severity using data from the Highway Accident Information Management System (2020–2022). After preprocessing, 36 explanatory variables covering roadway, environmental, accident causes, crash characteristics, and vehicle involvement were analyzed. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) and cost-sensitive learning were applied, and models were validated using train–test splits with cross-validation. The Random Forest model achieved the best performance with an AUC of 0.726, balanced accuracy of 0.649, and Matthews Correlation Coefficient (MCC) of 0.308, outperforming the other algorithms. SHapley Additive exPlanations (SHAP) were used to interpret the RF model, identifying nighttime crashes, large truck involvement, and roadway features (e.g., depressed medians and two-lane roads) as key predictors of severe outcomes. These insights suggest countermeasures such as improving nighttime safety, dedicating truck lanes, and designing safer medians. The novelty of this study lies in integrating model comparison, imbalance-aware metrics, and SHAP interpretability to provide actionable, context-specific policy recommendations for motorcycle safety in Thailand.

Keywords: Motorcycle Crash; Injury Severity; Machine Learning Algorithms; Supervised Learning Models; Random Forest; SHAP Analysis.

1. Introduction

Motorcycle crashes remain a critical global road safety issue, with particularly severe implications in Southeast Asian countries where motorcycles are the dominant means of travel. In Thailand, motorcycles make up a large share of the vehicle fleet and are disproportionately involved in traffic-related injuries and fatalities. According to the World Health Organization [1], Thailand has one of the highest road traffic fatality rates in Southeast Asia, and around 75% of those affected are users of two- or three-wheel vehicles, predominantly motorcyclists. This pressing public health challenge underscores the need for effective, data-driven strategies to reduce the severity of motorcycle crashes.

The multifaceted nature of elements influencing motorcycle crash severity presents a significant challenge for traditional analytical methods. These factors encompass a wide range of variables, including rider characteristics, road conditions, environmental factors, and vehicle attributes. The intricate interplay among these variables necessitates

* Corresponding author: vatanavongs@g.sut.ac.th

http://dx.doi.org/10.28991/CEJ-2025-011-10-014

sophisticated analytical approaches capable of capturing nonlinear relationships and complex interactions. Recently, the accelerated growth of data analytics and machine learning technologies has revolutionized the field of transportation safety analysis. These cutting-edge techniques offer powerful tools for processing large volumes of crash data, identifying patterns, and generating predictive models. Supervised learning models, in particular, have shown promise in analyzing crash data and predicting injury severity outcomes.

While conventional statistical methods, including multiple regression and logit approaches, have long been the cornerstone of crash analysis, they often fall short when dealing with the complex, nonlinear relationships inherent in motorcycle crash data. These traditional approaches rely on strict assumptions—such as data normality and independence of predictors—that often do not hold in real-world crash data, limiting their effectiveness in capturing complex injury risk patterns [2]. In contrast, machine learning techniques offer several advantages in this domain. They can handle large volumes of data with numerous variables, capture complex nonlinear relationships, and often provide superior predictive performance. Machine learning models are also more adept at dealing with multicollinearity and interaction effects among predictor variables, which are common in crash data [3]. Furthermore, some machine learning algorithms, including random forests and gradient boosting machines, offer built-in feature importance measures, providing an understanding of the varying impact of different elements on crash severity.

A range of supervised learning algorithms has been utilized in the study of crash data in different contexts. Decision Trees (DT), long valued in transportation safety research, have been widely used for their interpretability and ability to identify hierarchical relationships in crash risk factors [4]. K-Nearest Neighbors (KNN) has shown effectiveness in classifying crash severity based on similarity to historical data points [5]. Naïve Bayes (NB) has demonstrated utility in handling categorical data and estimating crash probabilities given various factors [6]. Random Forests (RF) have gained popularity for their ability to capture complex interactions between variables and provide insights into factor importance while being less prone to overfitting [7-9].

In recent years, comparative studies have been carried out in different countries to evaluate machine learning models for crash severity prediction. For instance, Wahab & Jiang [10, 11] conducted studies in Ghana comparing machine learning algorithms with statistical models, finding that Random Forest consistently outperformed traditional methods. Rezapour et al. [12], Rezapour et al. [13], Rezapour et al. [14] examined various algorithms in the United States, including logistic regression, decision trees, neural networks, and recurrent networks, highlighting both strengths and limitations. Mansoor et al. [15] compared machine learning and statistical models in Pakistan and confirmed the strong performance of Random Forest, supported by SHAP analysis for interpretability. Kashifi [16] applied XGBoost with SHAP in France to study two-wheeler crash severity, while Santos et al. [17] compared multiple models in Portugal, identifying Random Forest and logistic regression as the most effective.

Despite these advances, comparative evaluations focusing specifically on motorcycle injury severity remain limited in low- and middle-income countries. Most prior studies have been conducted in developed countries, where traffic conditions, infrastructure, and enforcement differ substantially from Southeast Asian contexts. This leaves an important research gap for Thailand, where motorcycles dominate daily mobility and crash fatality rates are disproportionately high. Moreover, while machine learning models often achieve higher predictive accuracy, their "black-box" nature limits their interpretability and practical policy application [18]. To address these gaps, this study introduces a unified framework that compares four supervised learning models—Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forest (RF)—while incorporating class imbalance handling and SHapley Additive exPlanations (SHAP) for interpretability. Using recent Thai crash data (2020–2022), this study provides both robust model evaluation and transparent explanations of critical factors influencing motorcycle crash severity. This contribution not only advances methodological rigor but also delivers actionable insights tailored to the high-risk conditions of Thailand and other low- and middle-income countries.

The subsequent sections of this manuscript are organized as follows: Section 2 presents a concise overview of the literature review, while Section 3 describes the data utilized. Section 4 outlines the methodological framework, followed by the results and discussion of model outcomes in Section 5. Section 6 provides the conclusion together with policy-oriented recommendations, and Section 7 presents the limitations of this study and directions for future research.

2. Literature Review

In recent years, numerous studies have applied machine learning techniques to predict motorcycle crash severity. Table 1 summarizes key findings from 2019 to 2024, highlighting methodological evolution, geographic focus, and recurring model performance trends. As shown in Table 1, scholarly investigations predominantly draw on data from developed countries, underscoring the need for expanded exploration in developing contexts like Thailand, where motorcycle accidents impose significant economic and social burdens.

Table 1. Overview of research on motorcycle crashes injury severity within the past five years

Authors (Year)	Country	Models used	Best Results
Santos et al. [17]	Portugal	Decision Tree (DT), Logistic Regression (LR), Random Forests (RF), Gradient Boosting (GB), XGBoost, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM)	Random Forests (RF) and Logistic Regression (LR)
Kashifi [16]	France	eXtreme Gradient Boosting (XGBoost) algorithm for crash severity. SHapley Additive exPlanations (SHAP) analysis for feature ranking and interaction exploration.	
Mansoor et al. [15]	Pakistan	Multinomial logit model (MNL), Random Forest (RF), Naive Bayes, Gradient-boosted trees methods	Random Forest (RF)
Rezapour et al. [12]	US	Random Forest (RF), Support Vector Machines (SVM), Multivariate Adaptive Regression Splines (MARS), and logistic regression (LR)	Random Forest (RF)
Rezapour et al. [13]	US	Deep belief networks (DBN), Recurrent neural networks (RNN), Multilayer neural networks (MLNN), and Single-layer neural networks	Recurrent neural networks (RNN)
Rezapour et al. [14]	US	Binary logistic regression and classification tree (CT)	Binary logistic regression
Rezapour & Ksaibati [19]	US	Support Vector Machines (SVM), Decision Tree (DT), Naïve Bayes (NB), Long short-term memory (LSTM), and Deep neural networks (DNN)	Long short-term memory (LSTM)
Wahab & Jiang [10]	Ghana	Multi-layer perceptron (MLP), Rule induction (PART) and Classification and Regression trees (Simple Cart)	Simple Cart model
Wahab & Jiang [11]	Ghana	Machine learning algorithms: J48, Random Forest, Instance-Based learning; Statistical model: Multinomial logit model (MNLM), Binary logit models, Binary probit model, Ordered logit model.	Random Forest (RF)

In Ghana, Wahab & Jiang [11] conducted a comprehensive study by comparing machine learning algorithms (J48, Random Forest, Instance-Based learning) with statistical models (Multinomial logit model, Binary logit models, Binary probit model, Ordered logit model). Their results showed that machine learning algorithms were more effective than conventional methods in assessing crash impact, with the Random Forest (RF) algorithm demonstrating the best agreement with experimental data. They identified location, time, and collision type as critical determinants of crash severity. A follow-up study by Wahab & Jiang [10] compared Multi-layer perceptron (MLP), Rule induction (PART), and Classification and Regression trees (Simple Cart) models. The Simple Cart model outperformed PART and MLP with 73.81% accuracy, identifying factors such as location, settlement type, time, collision type, and crash partner as significant predictors of crash severity.

In the United States, Rezapour & Ksaibati [19] compared numerous machine-learning algorithms, such as Naïve Bayes (NB), Support Vector Machines (SVM), Long short-term memory (LSTM), Decision Tree (DT), and Deep neural networks (DNN). Interestingly, they found that deeper models did not necessarily enhance performance, with a simple LSTM model outperforming more complex alternatives. In the same year, Rezapour et al. [14] compared Binary logistic regression and Classification tree (CT) for injury severity prediction in the United States. They found that Binary logistic regression performed slightly better than Classification tree, although both models identified similar predictors and showed comparable performance in crash prediction. Furthermore, Rezapour et al. [13] explored deep learning techniques, evaluating Deep belief networks (DBN), Multilayer neural networks (MLNN), Recurrent neural networks (RNN), and Single-layer neural networks. Their findings indicated that RNN outperformed other neural network models in crash severity prediction, highlighting the potential of deep learning techniques within this area.

In another study, Rezapour et al. [12] carried out a side-by-side assessment of Random Forest (RF), Support Vector Machines (SVM), Multivariate Adaptive Regression Splines (MARS), and logistic regression for injury severity prediction. Using k-fold cross-validation to assess misclassification rates, they found that the Random Forest algorithm surpassed other methods. Key factors influencing crash severity included speed, traffic volume, and rider's age. Similarly, Mansoor et al. [15] compared Multinomial logit models (MNL), Random Forest (RF), Naïve Bayes (NB), and Gradient-boosted trees in Pakistan. The Random Forest algorithm was more effective than others with 86.7% accuracy. They used the SHAP method to identify consistent determinants across both statistical methods and machine learning approaches, highlighting the statistical models' limitations due to unobserved factors.

In France, Kashifi [16] applied the XGBoost model and SHAP (SHapley Additive exPlanations) analysis to study two-wheeler crash severity. The study identified road category, urbanization level, two-wheeler category, and rider age as significant factors. It also noted that crash severity was higher for older riders and male two-wheeler riders, with rural areas, older riders, and non-helmet use associated with increased crash severity. Most recently, Santos et al. [17] conducted a study in Portugal comparing Decision Tree (DT), XGBoost, Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), Gradient Boosting (GB), and K-Nearest Neighbor (KNN). They found that RF and LR models performed best in predicting injury severity. Key risk factors identified included alcohol consumption, motorcycle age, road type, and gender. The study also noted that accidents occurring on weekends, involving older motorcycles, and on dry roads tended to increase severity.

This review reveals a consistent trend toward the superior performance of machine learning models—especially Random Forest—in motorcycle crash severity prediction. The increasing use of SHAP analysis further highlights the need for model interpretability to support actionable policy insights. Building on this foundation, the present study contributes to the literature by implementing a unified framework that integrates model comparison, class imbalance handling (via SMOTE and cost-sensitive learning), and SHAP-based interpretability. Using a nationally representative Thai dataset (2020–2022), the study applies four supervised models and evaluates performance using both conventional and imbalance-aware metrics (Balanced Accuracy and Matthews Correlation Coefficient), offering a rare combination of methodological rigor and real-world relevance in an underexplored, high-risk context.

3. Data Description

This research utilizes data obtained from the Thailand Highway Accident Information Management System (HAIMS), focusing solely on motorcycle crashes occurring within the timeframe of 2020 to 2022. Throughout this three-year interval, a cumulative total of 12,266 motorcycle crashes were documented. Following an extensive data cleansing process, 36 variables were identified and organized into five distinct categories of explanatory factors, which encompass ***roadway characteristics*** (the number of lanes (2, 4, 6, or ≥ 8), surface type (asphalt or other), median type (no median, flush and painted, raised, depressed, or barrier), and intersection type (four-leg, T, Y, U-turn, public area connection, private connection, or bridge section)), ***environmental characteristics*** (daytime, raining), ***causative factors of the accident*** (front-path interruption, illegal passing, violating the traffic signs, alcohol consumption, drowsiness), ***characteristics of the accident*** (head-on, rear-end, side swipe in parallel lane, off carriageway to the left/right, out of control on carriageway), and ***vehicles involved in the accident*** (car, van, pick-up truck 4 wheels, truck more than 6 wheels).

The dataset covers crashes on Thailand's national highway network, spanning both urban and rural regions. These roads feature mixed traffic conditions—including a high volume of motorcycles, limited lane separation, and seasonal rainfall—that are known to affect crash dynamics and severity. Thailand's notably high motorcycle ownership and crash fatality rates make it a critical setting for severity modeling in developing countries.

Table 2 summarizes the descriptive statistics of the dataset and the injury severity distribution: Severe/Fatal (44.13%) and Minor/PDO (55.87%).

Table 2. Overview of variable descriptive statistics

Variables	Frequency (%)	Mean	SD
Injury Severity			
Severe/Fatal	5,413 (44.13%)		
Minor/PDO	6,853 (55.87%)		
Roadway Characteristics			
UNDER MAINTENANCE	421 (3.43%)	0.034	0.182
UNDER CONSTRUCTION	669 (5.45%)	0.055	0.227
LANE = 2	3,606 (29.40%)	0.294	0.456
LANE = 4	5,479 (44.67%)	0.447	0.497
LANE = 6	1,161 (9.47%)	0.095	0.293
LANE ≥ 8	1,780 (14.51%)	0.145	0.352
ASPHALT	10,135 (82.63%)	0.826	0.379
NO MEDIAN	3,886 (31.68%)	0.317	0.465
FLUSH AND PAINTED MEDIAN	1,208 (9.85%)	0.098	0.298
RAISED MEDIAN	3,611 (29.44%)	0.294	0.456
DEPRESSED MEDIAN	2,183 (17.80%)	0.178	0.383
BARRIER MEDIAN	1,378 (11.23%)	0.112	0.316
PLAIN ROAD	12,004 (97.86%)	0.979	0.145
FOUR-LEG_INT	564 (4.60%)	0.046	0.209
T_INT	714 (5.82%)	0.058	0.234
Y_INT	127 (1.04%)	0.010	0.101
U_TURN	785 (6.40%)	0.064	0.245
CONNECT_PUBLIC AREA	424 (3.46%)	0.035	0.183
CONNECT_PRIVATE	211 (1.72%)	0.017	0.130
BRIDGE SECTION	177 (1.44%)	0.014	0.119

Environmental Characteristics			
DAYTIME	7,409 (60.40%)	0.604	0.489
RAINING	574 (4.68%)	0.047	0.211
Accident Causes			
FRONT-PATH INTERRUPTION	2,801 (22.84%)	0.228	0.420
ILLEGAL PASSING	95 (0.77%)	0.008	0.088
VIOLATING THE TRAFFIC SIGNS	356 (2.90%)	0.029	0.168
ALCOHOL CONSUMPTION	238 (1.94%)	0.019	0.138
DROWSINESS	107 (0.87%)	0.009	0.093
Accident Characteristics			
HEAD-ON	564 (4.60%)	0.046	0.209
REAR-END	3,515 (28.66%)	0.287	0.452
SIDE SWIPE IN PARALLEL LANE	2,107 (17.18%)	0.172	0.377
OFF CARRIAGEWAY TO THE LEFT/RIGHT	825 (6.73%)	0.067	0.250
OUT OF CONTROL ON CARRIAGEWAY	492 (4.01%)	0.040	0.196
Accident-involved Vehicles			
CAR	3,438 (28.03%)	0.280	0.449
VAN	165 (1.35%)	0.013	0.115
PICK-UP TRUCK 4 WHEELS	3,536 (28.83%)	0.288	0.453
TRUCK MORE THAN 6 WHEELS	1,298 (10.58%)	0.106	0.308

Note: SD = Standard Deviation

Originally, HAIMS categorized crash severity into four levels: property damage only (PDO), minor injury, severe injury, and fatality. However, severe and fatal crashes constituted a small fraction of the data, introducing class imbalance. To enhance model stability and predictive reliability, these were grouped into two categories: (1) Minor/PDO and (2) Severe/Fatal. This binary grouping approach is consistent with prior machine learning research on crash severity [20-22] and enhances the interpretability of results while maintaining predictive robustness.

Ethical approval: This study received the ethical approval from the Human Research Ethics Office of Suranaree University of Technology, Thailand (Approval Code: COE No.1/2568).

4. Research Methodology

4.1. Methodological Framework

The overall methodological framework of this study is illustrated in Figure 1. The flowchart outlines the sequential stages of the research process, beginning with data collection from the Highway Accident Information Management System (HAIMS) for motorcycle crashes during 2020–2022. The next steps involve data preprocessing and feature categorization, followed by handling class imbalance using Synthetic Minority Oversampling Technique (SMOTE) for KNN and NB, and class weights for DT and RF. Model training was performed with an 80/20 stratified train–test split and 5-fold cross-validation, accompanied by hyperparameter optimization using GridSearchCV and RandomizedSearchCV. Four supervised learning models—Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forest (RF)—were developed and tuned.

The optimized models were then compared using both conventional and imbalance-sensitive metrics, including accuracy, balanced accuracy, precision, recall, F1-score, area under the curve (AUC), and Matthews Correlation Coefficient (MCC). Based on performance evaluation, the best-performing model was selected and further interpreted using SHapley Additive exPlanations (SHAP) to identify the most influential predictors. Finally, the findings were discussed, and policy recommendations were proposed to improve motorcycle crash safety outcomes.

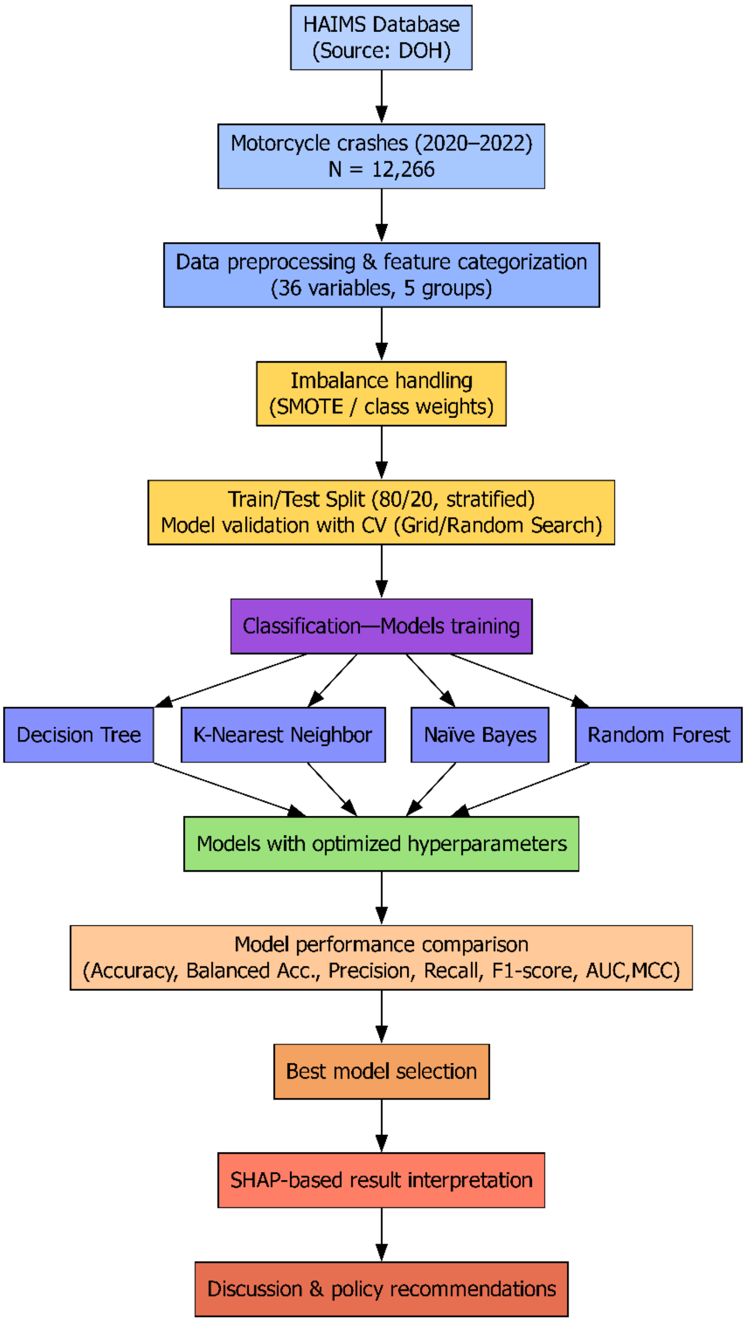

Figure 1. Methodological framework of the study

4.2. Model Descriptions

4.2.1. Decision Tree (DT)

Decision Tree (DT) represents a widely utilized methodology within the domain of machine learning, predominantly employed for classification endeavors, commonly referred to as the classification tree methodology. The DT employs a flowchart-like framework for categorization. Internal nodes indicate tests on variables, connectors denote test results, and terminal nodes indicate class categories. The routes leading from root to terminal nodes illustrate categorization guidelines. DT, along with the related influence diagram, serves as a visual and analytical decision support mechanism for classification analysis. In crash severity modeling, each node signifies a severity predictor while each branch reflects a state of the variable. A tree leaf represents the anticipated injury severity based on the training dataset's information. Upon acquiring a new crash sample from the test dataset, predictions regarding crash severity can be derived by tracing the tree from root to leaf utilizing variable values. The crux of the decision tree lies in selecting optimal attributes, aiming for branch nodes to exhibit maximal category homogeneity, thus enhancing node purity.

4.2.2. K-Nearest Neighbor (KNN)

The KNN algorithm is alternatively known as the Nearest Neighbor Classification (NNC). In the context of a predictive modeling scenario, the KNN algorithm determines a data point by examining the k closest data points. It applies the nearest neighbor rule, assigning a new sample to a particular class based on the closeness of a group of existing labeled data instances. In essence, the instance's classification is determined by the predominant class among k nearest data points [23]. The KNN methodology necessitates two critical decisions: the selection of the value of k and the selection of a proximity measure. The best k value is typically established by experimenting with various values for this parameter and identifying the one that yields the highest predictive accuracy. The Euclidean distance, which can be conceptualized as the physical distance between points in a two-dimensional space, serves as the distance metric utilized in the KNN algorithm.

4.2.3. Naïve Bayes (NB)

The Naïve Bayes algorithm represents a supervised machine learning technique often utilized as a linear model for different categorization tasks. It is rooted in Bayes' theorem and operates under the premise that the features within the dataset are independent given certain conditions. It is acknowledged that, on occasion, the assumption of independent features may be compromised; nevertheless, the Naïve Bayes algorithm demonstrates commendable performance even when operating under such unrealistic assumptions, particularly in scenarios involving limited sample sizes [24]. In the context of a classification task, let Y denote the variable subject to classification, while X signifies a collection of features represented as $X = (X_1, X_2, \dots, X_n)$. Based on Bayes' theorem, the anticipated likelihood of the class variable $Y = y_i$, contingent upon x, is articulated as follows (Equation 1):

$$P(Y = y_i|X) = \frac{P\big(X = x = (x_1, x_2, \dots, x_n)\big|Y = y_i\big)P(Y=y_i)}{P\big(X=x=(x_1,x_2,\dots,x_n)\big)} \tag{1}$$

The Naïve Bayes algorithm is characterized by its rapid computation, ease of implementation, accuracy, and robustness, making it a widely used as a classification technique applied in a diverse array of use cases [25].

4.2.4. Random Forest (RF)

The RF algorithm, as introduced by Breiman [26], constitutes a collective learning approach built on the foundation of the decision tree framework. It utilizes a resampling technique to generate k subsets from the original dataset and subsequently utilized to train k decision trees. The random forest is constructed by combining individual models. This RF method effectively reduces the problem of overfitting commonly associated with decision tree (DT) models. Each decision tree within the RF framework makes predictions using the test set, and the ultimate classification is decided by a consensus vote among the models.

To implement the RF algorithm, it is imperative to determine two critical parameters: the total count of trees (k) to be generated and the subset of features chosen at each split (m). The algorithm draws k bootstrap samples from the initial dataset, while the remaining portion, known as the out-of-sample data, is employed to evaluate the predictions' accuracy. Following the cultivation of k trees utilizing m randomly selected attributes, predictions for novel instances are established by combining the outcomes of these k trees. The initial data set is illustrated in the following manner [27]:

$$S = \big[(x_i, y_j), i = 1, 2, \dots, N, j = 1, 2, \dots, M\big] \tag{2}$$

In which x signifies an instance and y corresponds to an attribute of S. Every entry in the initial training data encompasses N samples and M features. The method for choosing resampled subsets and formulating the random forest technique using multiple trees is delineated below:

$$S_{Train} = [S_1, S_2, \dots, S_k] \tag{3}$$

In this context, S_{Train} includes a portion of k bootstrap training sets. Consequently, k decision trees are built using these k sets. These k trees are subsequently aggregated to form an RF technique, as illustrated below:

$$H(X, O_j) = \sum_{i=1}^{k} h_i(x, O_j), (j = 1, 2, \dots, m) \tag{4}$$

In this context, X denotes the input feature array derived from the training set, $h_i(x, O_j)$ represents a meta- level decision tree model, and O_j is an independently and identically distributed sequence that governs the development trajectory of the tree.

4.3. Model Training and Validation

All models were trained using an 80/20 train-test split with a fixed random seed of 42 to ensure reproducibility, a common approach in traffic safety and crash severity prediction studies [28-30]. Stratified sampling preserved the class distribution across splits. For model evaluation and tuning, 5-fold cross-validation was applied using GridSearchCV or RandomizedSearchCV, with AUC as the primary scoring metric. To avoid potential bias, SMOTE was applied only to the training set, with the test set left unchanged. Hyperparameter tuning was conducted with stratified cross-validation, and results were evaluated using both conventional and imbalance-sensitive metrics (balanced accuracy, MCC, and AUC). Preliminary comparisons with and without SMOTE confirmed that the oversampling procedure improved class balance without introducing significant synthetic noise. To address class imbalance, SMOTE was applied for KNN and NB, while DT and RF used class_weight='balanced'. All final performance metrics were computed on the independent test set. For hyperparameter optimization, DT, KNN, and NB models were tuned using GridSearchCV, while RF used RandomizedSearchCV with 100 iterations. Key parameters such as max_depth, min_samples_split, and class_weight (DT), n_neighbors and distance metrics (KNN), var_smoothing (NB), and n_estimators, max_depth, and class_weight (RF) were systematically explored. This approach ensured that each model was fairly tuned under comparable validation settings before performance comparison. All models were evaluated on the same stratified test set using consistent performance metrics to support a transparent and equitable comparison.

4.4. Model Performance Evaluation Metrics

The primary measures for judging categorization methods include accuracy, precision, recall, F1-Score and the Area Under the Curve (AUC). These metrics have been widely utilized in studies on traffic crash severity prediction [31-33]. They are derived from the confusion matrix (CM) as presented in Table 3. Columns in the CM indicate predicted class instances, while rows indicate actual class instances, with correct predictions located on the diagonal. True positives (TP) and True negatives (TN) refer to instances that are accurately identified. A false positive (FP) is an instance incorrectly labeled as positive, whereas a false negative (FN) is an instance incorrectly labeled as negative. The performance metric calculation formulas are specified in Equations 5 to 8.

Accuracy is defined as the proportion of correctly classified crashes. However, due to the moderate class imbalance in crash severity data, relying solely on accuracy may provide a misleading assessment of model performance. To address this, additional metrics were examined. ***Balanced Accuracy*** captures the average recall for both classes, providing a more equitable measure under imbalance. ***Matthews Correlation Coefficient (MCC)*** evaluates the overall quality of binary classifications by considering all elements of the confusion matrix and is particularly robust for imbalanced data. Together, these metrics offer a more comprehensive and fair evaluation of model performance.

Accuracy represents the fraction of instances that are accurately recognized:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

Precision is the ratio of accurate positive predictions:

$$Precision\ (P) = \frac{TP}{TP+FP} \tag{6}$$

Recall quantifies the correctly classified positive instances:

$$Recall\ (R) = \frac{TP}{TP+FN} \tag{7}$$

The ***F1-Score***, representing the harmonic mean of precision and recall, varies between 0 and 1, with 1 signifying optimal effectiveness and 0 signifying the least:

$$F1\_Score = \frac{2\times(R\times P)}{R+P} \tag{8}$$

The Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) curve, frequently referred to as AUROC, represents an exceptionally valuable evaluative metric. The ROC curve, alternatively termed the sensitivity/specificity curve, constitutes a probabilistic graphical representation that facilitates the assessment of classification model efficacy. A classification model exhibiting an AUC value proximate to 1 is typically proficient in accurately predicting the binary outcomes of 0 as 0 and 1 as 1.

Table 3. Confusion Matrix (CM)

	Total instances	Predicted class	
		Positive	**Negative**
Actual class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

4.5. SHapley Additive exPlanations (SHAP)

According to Lundberg & Lee [34], the SHAP methodology employs Shapley values to clarify the outcomes generated by machine learning techniques. Drawing from joint game analysis, these scores quantify the specific impact of all attributes on the overall outcome [35]. Initially, the SHapley Additive exPlanations methodology constructs a framework incorporating all input features. It then creates another version that omits the feature of interest, thereby allowing for an examination of how its exclusion influences the model's accuracy. The SHAP score associated with a feature is characterized as its incremental impact on the prediction. The subsequent equation is employed to compute the SHAP score for the feature [36]:

$$\emptyset_i = \sum_{S\subseteq X\setminus\{i\}} \frac{|S|!(|X|-|S|-1)!}{|X|!}\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right] \tag{9}$$

In this context, $\emptyset_i$ denotes the incremental impact of a feature, synonymous with its SHAP score; X signifies the entirety of features; S represents a smaller group of all features; and x_S indicates the scores corresponding to the features within S. To evaluate the effect of the specific feature in question, a framework $f_{S\cup\{i\}}$ is developed that includes this feature, while another framework f_S is developed without it. The predictions yielded by the two models are subsequently juxtaposed with the current output, represented as $f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)$. Given that the specific feature in question is also contingent upon other features within the framework, the discrepancies are computed across all conceivable the smaller group of all features [34].

5. Model Results and Discussion

5.1. Model Performance Comparison

To gauge the efficacy of multiple machine learning techniques in anticipating the seriousness of traffic injury, this study implemented and evaluated four distinct models: Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forest (RF). The effectiveness of these algorithms was quantified by utilizing a wide range of metrics to ensure thorough evaluation.

Table 4 provides a comprehensive summary of the performance metrics for each model. The models demonstrated slight variations in accuracy, with Random Forest (RF) achieving the highest accuracy at 0.666, followed by Decision Tree (DT) at 0.656, Naïve Bayes (NB) at 0.629, and K-Nearest Neighbor (KNN) at 0.623. Precision, which evaluates the proportion of true positive predictions among all predicted positives, was also highest for RF (0.663), followed by DT (0.652), NB (0.627), and KNN (0.625). Similarly, Recall, reflecting the proportion of true positive instances correctly identified, aligned with the accuracy rankings: RF at 0.666, DT at 0.656, NB at 0.629, and KNN at 0.623. The F1-Score, a harmonic mean of precision and recall, ranged from RF at 0.660 to DT at 0.648, NB at 0.628, and KNN at 0.623.

Table 4. Overview of performance metrics for each model

Models	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score	AUC	MCC
Decision Tree (DT)	0.656	0.636	0.652	0.656	0.648	0.700	0.285
K-Nearest Neighbor (KNN)	0.623	0.618	0.625	0.623	0.623	0.676	0.235
Naïve Bayes (NB)	0.629	0.620	0.627	0.629	0.628	0.673	0.241
Random Forest (RF)	0.666	0.649	0.663	0.666	0.660	0.726	0.308

Among the performance metrics, the Area Under the Curve (AUC) provided the clearest differentiation of the models' discriminative abilities. As illustrated in Figure 2 and Table 4, the AUC values emphasize each model's ability to distinguish between classes. RF achieved the highest AUC at 0.726, indicating superior discriminative performance. This was followed by DT at 0.700, KNN at 0.676, and NB at 0.673. Figures 2-a to 2-d display the individual ROC curves for each model, providing a detailed visual comparison of their performance across varying thresholds.

In addition to conventional metrics, Table 4 also reports Balanced Accuracy and Matthews Correlation Coefficient (MCC) to account for class imbalance. RF again achieved the highest values in both (Balanced Accuracy = 0.649, MCC = 0.308), reinforcing its superior generalization performance and robustness against class bias.

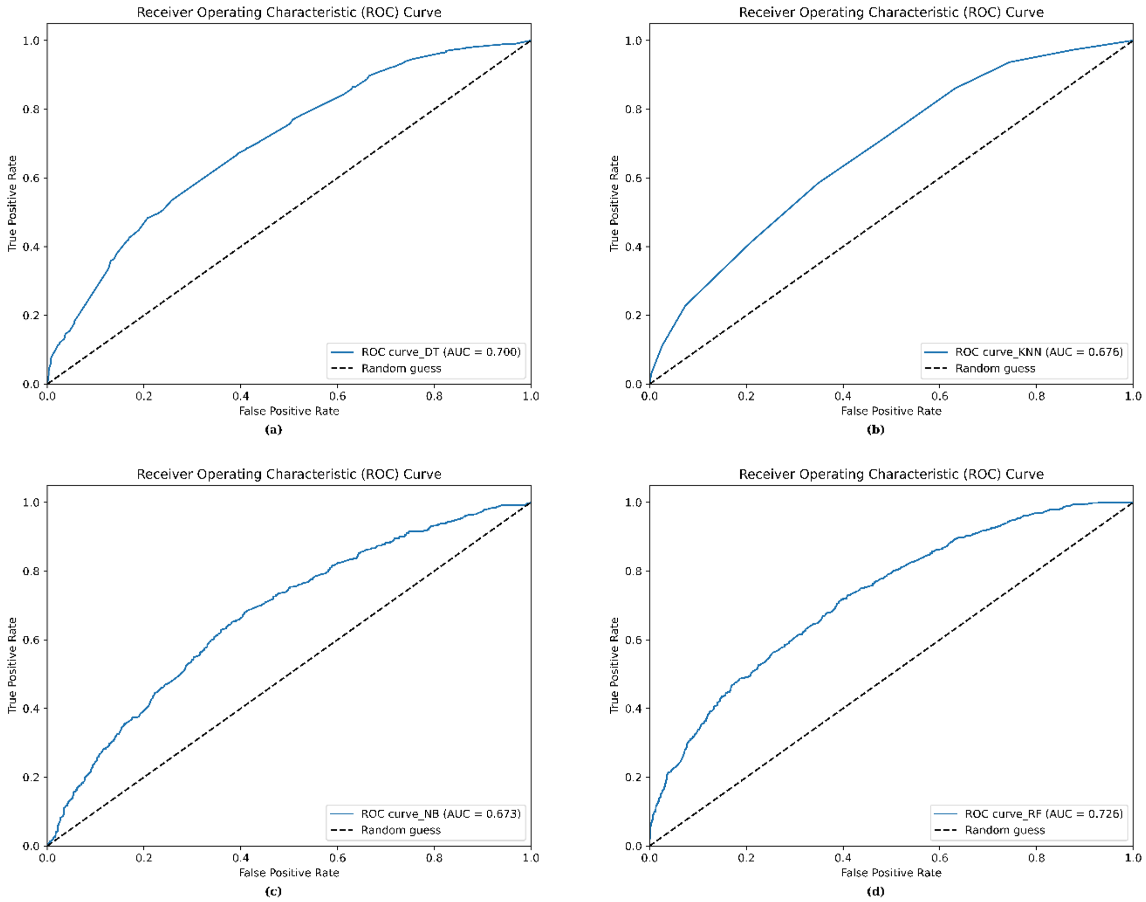

Figure 2. Comparison of Area Under the Curve (AUC): ***(a) AUC_DT; (b) AUC_KNN; (c) AUC_NB; (d) AUC_RF***

In summary, the Random Forest model consistently outperformed the other algorithms across all evaluation metrics, making it the most effective approach for predicting motorcycle crash severity in this study. This superior performance is likely due to its ensemble structure, which enhances stability, reduces overfitting, and captures complex nonlinear interactions among predictors—advantages especially relevant in heterogeneous crash data. These results align with previous studies that have demonstrated RF's consistent edge in traffic safety prediction tasks [11, 12, 15, 17]. In contrast, the lower performance of NB and KNN may stem from their respective assumptions about feature independence and local similarity, which can limit their accuracy in high-dimensional, correlated crash datasets. From a practical standpoint, RF's strong showing across conventional and imbalance-sensitive metrics (Balanced Accuracy, MCC) reinforces its suitability for real-world implementation, offering both predictive strength and reliability in imbalanced datasets common in traffic injury research.

Taken together, these comparisons confirm that ensemble methods such as Random Forest are particularly well suited for heterogeneous and imbalanced crash data. While simpler methods like KNN and NB may offer interpretability or computational efficiency, their limitations reduce their practical utility for policy applications where accuracy and robustness are essential.

Compared to previous studies, our results confirm the strong performance of Random Forest reported in other contexts such as Ghana, Pakistan, and Portugal, but also demonstrate its robustness in Thailand where motorcycle crashes dominate the road traffic landscape. This extension to a low- and middle-income country context adds new evidence that ensemble models remain effective even when data conditions and traffic patterns differ substantially from those in high-income countries.

5.2. Model Interpretation

To interpret the predictions of the most effective model, SHapley Additive exPlanations (SHAP) were used to analyze the Random Forest output. This interpretability method allows us to unpack the 'black box' nature of machine learning model, revealing the nuanced interplay of various features in predicting crash outcomes.

The SHAP bee swarm plot in Figure 3 provides valuable insights into the variables affecting the assessment of the seriousness of injuries in motorcycle crashes as predicted by the Random Forest model, particularly for predicting whether crash results in severe or fatal injuries (class 1). The SHAP scores plotted on the X-axis illustrate the degree of impact every variable has on the model's outcome: positive SHAP scores increase the likelihood of a severe outcome, while negative values decrease it. The features are ranked by their importance on the Y-axis, with the most critical factors at the top.

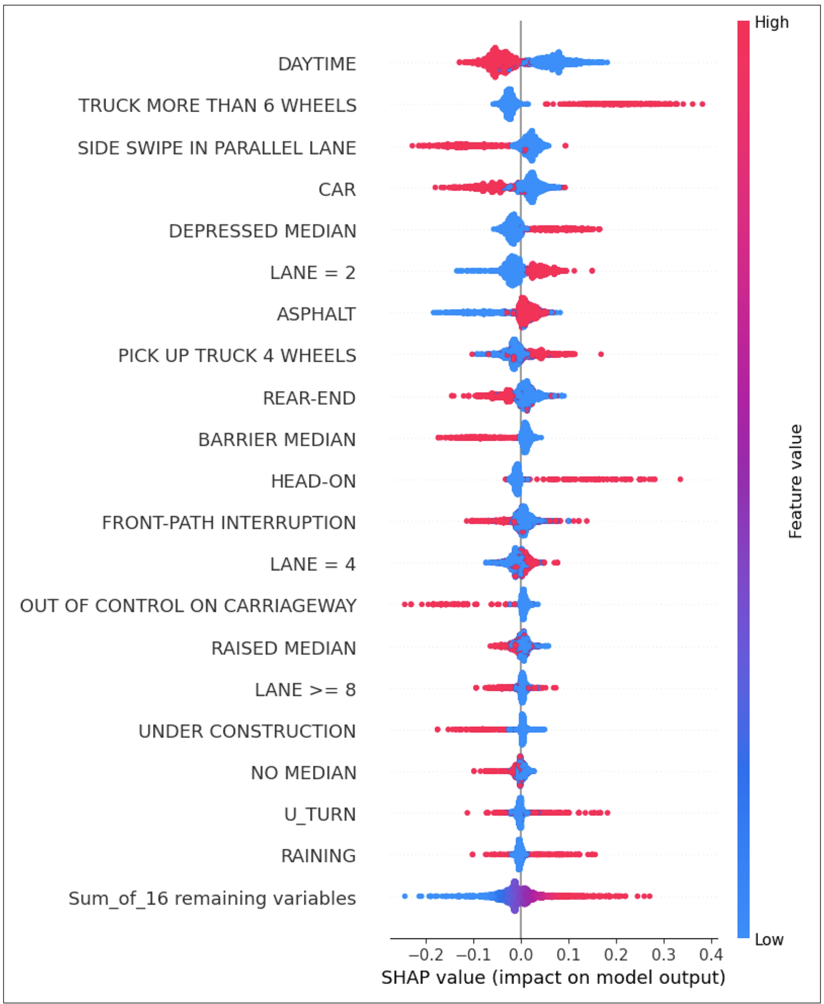

Figure 3. The impact of features on motorcycle crash injury severity

Key predictors associated with increased injury severity include nighttime crashes, involvement of large trucks, the presence of depressed medians, two-lane roads, pick-up trucks, and head-on collisions. These conditions contribute to a higher predicted probability of severe or fatal injuries (class 1). In contrast, crashes occurring during the daytime, side-swipe and rear-end collisions, and those involving passenger cars are linked to a reduced likelihood of severe outcomes. These patterns are consistent with prior research and highlight practical roadway and vehicle-related conditions that elevate or mitigate injury severity.

The following discussion expands on these findings by explaining how each key feature influences the model's predictions, supported by prior research.

At the forefront of influential factors, the result shows that the time of day plays a pivotal role in injury severity prediction. Daytime crashes are associated with lower injury severity compared to nighttime incidents. This temporal effect aligns with previous research by Chang et al. [37], who found that unlit darkness was linked to higher injury severity, which can be attributed to decreased visibility and longer reaction times during nighttime conditions [38]. The stark contrast between daytime and nighttime crash severities underscores the need for enhanced safety measures during darker hours, such as improved road lighting and increased awareness campaigns targeting night riders.

The involvement of large trucks (more than six wheels) emerges as the second most critical factor, showing a strong correlation with increased injury severity. This finding resonates with recent research by Kanitpong et al. [39], Laphrom et al. [40], who attributed this heightened risk to the substantial size and mass differentials between trucks and motorcycles. The vulnerability of motorcyclists in such collisions highlights the urgent need for strategies to mitigate these high-risk interactions, possibly through dedicated lane policies or advanced warning systems for both truck drivers and motorcyclists. This highlights the urgent need for truck lane management or separation strategies to reduce motorcycle–truck conflicts.

Furthermore, the analysis reveals that side swipes in parallel lanes lead to less serious injuries. This counterintuitive finding might be attributed to the glancing nature of these impacts, as suggested by Agyemang et al. [41] in their comprehensive study of factors influencing motorcycle collision severity. Additionally, the presence of cars in crash scenarios shows a mixed impact, with a slight tendency towards decreased severity, possibly due to the comparative protection offered by car structures in collisions with motorcycles. This trend was similarly observed in an investigation conducted by Se et al. [42].

The results point out that road infrastructure elements demonstrate significant influence on crash severity outcomes. Depressed medians are connected to more severe injuries, potentially owing to the higher risk of more severe impacts or rollovers. This aligns with the findings by Se et al. [43], who observed that flush and depressed median factors produce a favorable marginal impact on fatal injuries, consequently raising the chances of fatalities during a collision. Conversely, barrier medians slightly decrease injury severity, likely due to their protective function in preventing cross-median collisions, agreeing with the research from Champahom et al. [44]. A plausible explanation is that a barrier median limits the possibility of turning, guiding such actions to safer locations, which helps to decrease the risk of head-on accidents and dangerous overtaking scenarios.

Moreover, the number of lanes shows varied effects, with two-lane roads slightly increasing severity risk. This nuanced impact of road configuration on motorcycle safety supported by Se et al. [42], which demonstrated that collisions on two-lane roadways exhibit a significantly elevated the likelihood of severe injuries and fatalities compared to collisions on four-lane roadways. This phenomenon might arise because two-lane roadways are often non-separated and situated in rural regions characterized by elevated speed limits; consequently, incidents on these roadways are susceptible to severe impacts, including head-on crashes and accidents related to excessive speed. This suggests that targeted investment in safer rural two-lane roadways could substantially reduce motorcycle crash severity.

With regard to road surface, crashes on asphalt road surface displayed a relatively neutral impact, with a slight tendency towards reduced severity compared to other surface types, corresponding to observations by Champahom et al. [44], who uncovered that the severity of crashes is typically higher on concrete surfaces than those on asphalt pavements. One possible explanation may be the material properties of asphalt, providing better traction and smoother driving surfaces, are likely to contribute to safer driving conditions and reduced injury severity.

The analysis additionally highlights that the feature related to four-wheeled pick-up trucks shows high SHAP scores, indicating a higher likelihood of severe injuries in collisions involving such vehicles. Pick-up trucks often have higher centers of gravity, which can contribute to rollover crashes, leading to more severe injuries. Past research by Chang et al. [45] reported that collisions involving substantial vehicles, including pick-up trucks and tractor-trailers, have been demonstrated to significantly elevate the probability of death by 77% in non-intersection accidents and by 102% in intersection, correspondingly.

With respect to the collision types, the model result demonstrates distinct impacts: rear-end collisions lead to less serious injuries than those associated with alternative categories, while head-on crash, when they occur, are strongly associated with increased severity. In head-on collisions, the substantial force generated by vehicles moving directly towards each other is likely to escalate the level of injury severity [37, 46].

The findings further delineate that crashes involving a front-path interruption are linked to a higher probability of serious injuries. Positive SHAP values suggest that these interruptions, which may involve sudden obstacles or loss of control, significantly increase the likelihood of serious consequences. The finding is reasonable and supported by prior studies, owing to the significant deceleration triggered by sudden stopping [47, 48].

The model also captures the heightened risk associated with out-of-control incidents on carriageways and the varied impacts of road construction zones. Areas under construction demonstrated a minimal impact on injury severity, suggesting that reduced speeds in these zones may offset potential hazards. In construction zones equipped with amber or warning signals, motorists exhibited a decreased chance of being involved in serious accidents [49].

Notably, crashes occurring on roads with four or more than eight lanes, roads with raised medians, and roads without medians show minimal SHAP influence, suggesting negligible contribution to injury severity. Additionally, crashes involving U-turns or occurring during rainy conditions show a small but positive association with severe injuries.

While many of these results are consistent with international studies, some differences also emerge. For instance, the prominence of rural two-lane roads and depressed medians as critical predictors appears more specific to Thailand's infrastructure context, where motorcycles frequently share non-divided highways with heavy vehicles. This contrasts with findings from high-income countries, where factors such as alcohol or speeding often dominate severity models. These differences highlight the importance of tailoring countermeasures to local traffic and roadway conditions rather than directly transferring strategies across regions.

By combining these interpretations with evidence from prior studies, the analysis not only validates existing knowledge but also identifies context-specific risks unique to Thailand, strengthening the case for tailored countermeasures.

These findings highlight not only statistical associations but also practical risk factors that can guide targeted safety measures. While Section 6 presents the overall policy recommendations, the discussion here emphasizes how specific predictors—such as nighttime conditions, large trucks, and rural roadway features—directly influence crash severity and therefore warrant particular attention from policymakers.

6. Conclusion

This study evaluated the predictive performance of four supervised learning algorithms—Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forest (RF)—to model motorcycle crash injury severity using nationally representative crash data from Thailand (2020–2022). Thirty-six explanatory variables were analyzed across roadway, environmental, vehicle, crash types, and causative factors. After preprocessing and addressing class imbalance, the models were trained, validated, and compared across multiple conventional and imbalance-sensitive metrics. The Random Forest model consistently outperformed the other approaches, achieving the highest accuracy, balanced accuracy, AUC, and MCC, confirming the advantage of ensemble-based methods for heterogeneous and imbalanced traffic data. The SHAP analysis provided transparency into model predictions by identifying critical determinants of severity, including nighttime crashes, large truck involvement, depressed medians, two-lane roads, and head-on collisions. These findings highlight both behavioral and infrastructure-related conditions that exacerbate motorcyclist vulnerability in Thailand.

The contribution of this study lies in demonstrating the utility of ensemble methods for crash severity analysis in a motorcycle-dominated, low- and middle-income country context, while also addressing the interpretability challenge through SHAP analysis. Based on the insights gained, several countermeasures are recommended. First, nighttime safety improvements are needed, including better road lighting, reflective pavement markings, and stricter enforcement during hours of darkness. Second, large truck management should be prioritized by introducing truck lane separation or time-based restrictions in areas with high motorcycle traffic. Third, roadway infrastructure enhancements are critical, particularly on two-lane rural highways, where interventions such as centerline barriers, median treatments, and shoulder widening could reduce the risk of head-on and high-impact crashes. Additionally, improving the design of depressed medians and strengthening traffic control in construction zones can help mitigate severe crash outcomes. While some measures such as truck lane separation may face practical challenges in Thailand's current infrastructure, other interventions—including lighting improvements, reflective markings, enforcement, and low-cost roadway treatments—are more immediately feasible and can still provide substantial safety benefits. Collectively, these measures can substantially reduce injury severity among motorcyclists. Beyond methodological contributions, this study demonstrates how machine learning combined with SHAP interpretability can support context-specific, evidence-based policymaking to reduce the burden of motorcycle crashes and improve overall road safety.

6.1. Limitations and Future Research

Undeniably, this research, like any other, is not without its limitations. Future investigations may extend the analysis by incorporating additional machine learning methodologies, including Neural Networks, Bayesian Networks, Deep Learning, and advanced ensemble or boosting approaches such as Gradient Boosting and XGBoost. These techniques may achieve higher predictive accuracy and help uncover additional latent patterns in crash dynamics, offering further insight into the trade-offs between predictive strength, interpretability, and practical applicability in traffic safety analysis.

Another limitation is that this study did not explicitly apply dimensionality reduction or feature selection methods. Although the moderate number of predictors (36 variables) and the use of tree-based models with embedded feature selection helped mitigate overfitting risk, future work could systematically evaluate feature selection or dimensionality reduction approaches such as PCA, LASSO, or recursive feature elimination to further validate and refine predictor sets.

This study was also limited by the absence of certain contextual factors, such as road lighting conditions, traffic density, and enforcement data, which may act as unobserved confounders. Although proxy variables such as time of day and roadway type capture some of these effects, future research should integrate richer datasets, including road inventory databases, traffic monitoring systems, and enforcement records, to more directly account for these influences and improve explanatory power.

Finally, this study focused on retrospective crash severity prediction using historical crash records, which limits direct application in real-time crash risk warning systems. Future research should explore integration with real-time data streams (e.g., traffic sensors, weather stations, and GPS devices) and develop methods to adapt SHAP interpretability to streaming contexts, enabling dynamic and actionable risk warnings.

7. Declarations

7.1. Author Contributions

Conceptualization, S.S. and V.R.; methodology, S.S.; software, S.S.; validation, T.C. and S.J.; formal analysis, S.S. and P.W.; investigation, P.W. and T.C.; data curation, S.S., P.W., and T.C.; writing—original draft preparation, S.S.; writing—review and editing, P.W. and T.C.; visualization, S.S.; supervision, S.J. and V.R. All authors have read and agreed to the published version of the manuscript.

7.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.3. Funding and Acknowledgements

This research was supported by Suranaree University of Technology (SUT). The authors would like to thank the Department of Highway for supporting the road traffic crash data.

7.4. Institutional Review Board Statement

This research was approved by the Ethics Committee for Research Involving Human Subjects, Suranaree University of Technology (COE No.1/2568).

7.5. Conflicts of Interest

The authors declare no conflict of interest.

8. References

[1] WHO. (2023). Road traffic injuries. World Health Organization (WHO), Geneva, Switzerland. Available online: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (accessed on September 2025).

[2] Santos, K., Dias, J. P., & Amado, C. (2022). A literature review of machine learning algorithms for crash injury severity prediction. Journal of Safety Research, 80, 254–269. doi:10.1016/j.jsr.2021.12.007.

[3] Chan, J. Y. Le, Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. Mathematics, 10(8), 1283. doi:10.3390/math10081283.

[4] Mohamad, I., Jomnonkwao, S., & Ratanavaraha, V. (2022). Using a decision tree to compare rural versus highway motorcycle fatalities in Thailand. Case Studies on Transport Policy, 10(4), 2165–2174. doi:10.1016/j.cstp.2022.09.016.

[5] Sahu, S., Maram, B., Gampala, V., & Daniya, T. (2023). Analysis of Road Accidents Prediction and Interpretation Using KNN Classification Model. Emerging Technologies in Data Mining and Information Security. Lecture Notes in Networks and Systems, vol 490. Springer, Singapore. doi:10.1007/978-981-19-4052-1_18.

[6] Yahaya, M., Jiang, X., Fu, C., Bashir, K., & Fan, W. (2019). Enhancing Crash Injury Severity Prediction on Imbalanced Crash Data by Sampling Technique with Variable Selection. 2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019, 363–368. doi:10.1109/ITSC.2019.8917223.

[7] Scarano, A., Rella Riccardi, M., Mauriello, F., D'Agostino, C., Pasquino, N., & Montella, A. (2023). Injury severity prediction of cyclist crashes using random forests and random parameters logit models. Accident Analysis & Prevention, 192, 107275. doi:10.1016/j.aap.2023.107275.

[8] Yan, M., & Shen, Y. (2022). Traffic Accident Severity Prediction Based on Random Forest. Sustainability (Switzerland), 14(3), 1729. doi:10.3390/su14031729.

[9] Yang, J., Han, S., & Chen, Y. (2023). Prediction of Traffic Accident Severity Based on Random Forest. Journal of Advanced Transportation, 2023. doi:10.1155/2023/7641472.

[10] Wahab, L., & Jiang, H. (2020). Severity prediction of motorcycle crashes with machine learning methods. International Journal of Crashworthiness, 25(5), 485–492. doi:10.1080/13588265.2019.1616885.

[11] Wahab, L., & Jiang, H. (2019). A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. PLoS ONE, 14(4), 214966. doi:10.1371/journal.pone.0214966.

[12] Rezapour, M., Farid, A., Nazneen, S., & Ksaibati, K. (2021). Using machine leaning techniques for evaluation of motorcycle injury severity. IATSS Research, 45(3), 277–285. doi:10.1016/j.iatssr.2020.07.004.

[13] Rezapour, M., Nazneen, S., & Ksaibati, K. (2020). Application of deep learning techniques in predicting motorcycle crash severity. Engineering Reports, 2(7), 12175. doi:10.1002/eng2.12175.

[14] Rezapour, M., Mehrara Molan, A., & Ksaibati, K. (2020). Analyzing injury severity of motorcycle at-fault crashes using machine learning techniques, decision tree and logistic regression models. International Journal of Transportation Science and Technology, 9(2), 89–99. doi:10.1016/j.ijtst.2019.10.002.

[15] Mansoor, U., Jamal, A., Su, J., Sze, N. N., & Chen, A. (2023). Investigating the risk factors of motorcycle crash injury severity in Pakistan: Insights and policy recommendations. Transport Policy, 139, 21–38. doi:10.1016/j.tranpol.2023.05.013.

[16] Kashifi, M. T. (2023). Investigating two-wheelers risk factors for severe crashes using an interpretable machine learning approach and SHAP analysis. IATSS Research, 47(3), 357–371. doi:10.1016/j.iatssr.2023.07.005.

[17] Santos, K., Firme, B., Dias, J. P., & Amado, C. (2024). Analysis of Motorcycle Accident Injury Severity and Performance Comparison of Machine Learning Algorithms. Transportation Research Record, 2678(1), 736–748. doi:10.1177/03611981231172507.

[18] Ali, Y., Hussain, F., & Haque, M. M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. Accident Analysis & Prevention, 194, 107378. doi:10.1016/j.aap.2023.107378.

[19] Rezapour, M., & Ksaibati, K. (2020). Application of various machine learning architectures for crash prediction, considering different depths and processing layers. Engineering Reports, 2(8), 12215. doi:10.1002/eng2.12215.

[20] Sum, S., Se, C., Champahom, T., Jomnonkwao, S., Sinha, S., & Ratanavaraha, V. (2025). A random forest and SHAP-based analysis of motorcycle crash severity in Thailand: Urban-rural and day-night perspectives. Transportation Engineering, 21, 100369. doi:10.1016/j.treng.2025.100369.

[21] Agheli, A., & Aghabayk, K. (2025). How does distraction affect cyclists' severe crashes? A hybrid CatBoost-SHAP and random parameters binary logit approach. Accident Analysis & Prevention, 211, 107896. doi:10.1016/j.aap.2024.107896.

[22] Sadeghi, M., Aghabayk, K., & Quddus, M. (2024). A hybrid Machine learning and statistical modeling approach for analyzing the crash severity of mobility scooter users considering temporal instability. Accident Analysis & Prevention, 206, 107696. doi:10.1016/j.aap.2024.107696.

[23] Sambasivam, G., Amudhavel, J., & Sathya, G. (2020). A Predictive Performance Analysis of Vitamin D Deficiency Severity Using Machine Learning Methods. IEEE Access, 8, 109492–109507. doi:10.1109/ACCESS.2020.3002191.

[24] Domingos, P., & Pazzani, M. (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Machine Learning, 29(2–3), 103–130. doi:10.1023/a:1007413511361.

[25] Kazmierska, J., & Malicki, J. (2008). Application of the Naïve Bayesian Classifier to optimize treatment decisions. Radiotherapy and Oncology, 86(2), 211–216. doi:10.1016/j.radonc.2007.10.019.

[26] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. doi:10.1023/a:1010933404324.

[27] Chen, J., Li, K., Tang, Z., Bilal, K., Yu, S., Weng, C., & Li, K. (2017). A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment. IEEE Transactions on Parallel and Distributed Systems, 28(4), 919–933. doi:10.1109/TPDS.2016.2603511.

[28] Sonnatthanon, N., & Choocharukul, K. (2025). Crash severity prediction using a virtual geometry-group-based deep learning approach with images-based feature representation. Results in Engineering, 27, 106155. doi:10.1016/j.rineng.2025.106155.

[29] Mohsin, A. S. M., Choudhury, S. H., & Muyeed, M. A. (2025). Automatic priority analysis of emergency response systems using internet of things (IoT) and machine learning (ML). Transportation Engineering, 19, 100304. doi:10.1016/j.treng.2025.100304.

[30] Acı, Ç. İ., Mutlu, G., Ozen, M., & Acı, M. (2025). Enhanced Multi-Class Driver Injury Severity Prediction Using a Hybrid Deep Learning and Random Forest Approach. Applied Sciences (Switzerland), 15(3), 1586. doi:10.3390/app15031586.

[31] Dia, Y., Faty, L., Sarr, M. D., Sall, O., Bousso, M., & Landu, T. T. (2022). Study of Supervised Learning Algorithms for the Prediction of Road Accident Severity in Senegal. 2022 7th International Conference on Computational Intelligence and Applications (ICCIA), 123–127. doi:10.1109/iccia55271.2022.9828434.

[32] Dong, S., Khattak, A., Ullah, I., Zhou, J., & Hussain, A. (2022). Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. International Journal of Environmental Research and Public Health, 19(5), 2925. doi:10.3390/ijerph19052925.

[33] Ijaz, M., lan, L., Zahid, M., & Jamal, A. (2021). A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw. Accident Analysis & Prevention, 154, 106094. doi:10.1016/j.aap.2021.106094.

[34] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December, 2017, Long Beach, United States.

[35] Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems, 41(3), 647–665. doi:10.1007/s10115-013-0679-x.

[36] Tahfim, S. A. S., & Yan, C. (2021). Analysis of severe injuries in crashes involving large trucks using K-prototypes clustering-based GBDT model. Safety, 7(2), 32. doi:10.3390/safety7020032.

[37] Chang, F., Yasmin, S., Huang, H., Chan, A. H. S., & Haque, M. M. (2021). Injury severity analysis of motorcycle crashes: A comparison of latent class clustering and latent segmentation based models with unobserved heterogeneity. Analytic Methods in Accident Research, 32, 100188. doi:10.1016/j.amar.2021.100188.

[38] Marcoux, R., Yasmin, S., Eluru, N., & Rahman, M. (2018). Evaluating temporal variability of exogenous variable impacts over 25 years: An application of scaled generalized ordered logit model for driver injury severity. Analytic Methods in Accident Research, 20, 15–29. doi:10.1016/j.amar.2018.09.001.

[39] Kanitpong, K., Jensupakarn, A., Dabsomsri, P., & Issalakul, K. (2024). Characteristics of motorcycle crashes in Thailand and factors affecting crash severity: Evidence from in-depth crash investigation. Transportation Engineering, 16, 100227. doi:10.1016/j.treng.2024.100227.

[40] Laphrom, W., Se, C., Champahom, T., Jomnonkwao, S., Wipulanusat, W., Satiennam, T., & Ratanavaraha, V. (2024). XGBoost-SHAP and Unobserved Heterogeneity Modelling of Temporal Multivehicle Truck-Involved Crash Severity Patterns. Civil Engineering Journal (Iran), 10(6), 1890–1908. doi:10.28991/CEJ-2024-010-06-011.

[41] Agyemang, W., Adanu, E. K., & Jones, S. (2021). Understanding the Factors That Are Associated with Motorcycle Crash Severity in Rural and Urban Areas of Ghana. Journal of Advanced Transportation, 2021, 6336517. doi:10.1155/2021/6336517.

[42] Se, C., Champahom, T., Jomnonkwao, S., Wisutwattanasak, P., Laphrom, W., & Ratanavaraha, V. (2023). Temporal Instability and Transferability Analysis of Daytime and Nighttime Motorcyclist-Injury Severities Considering Unobserved Heterogeneity of Data. Sustainability (Switzerland), 15(5), 4486. doi:10.3390/su15054486.

[43] Se, C., Champahom, T., Jomnonkwao, S., Chaimuang, P., & Ratanavaraha, V. (2021). Empirical comparison of the effects of urban and rural crashes on motorcyclist injury severities: A correlated random parameters ordered probit approach with heterogeneity in means. Accident Analysis & Prevention, 161, 106352. doi:10.1016/j.aap.2021.106352.

[44] Champahom, T., Wisutwattanasak, P., Chanpariyavatevong, K., Laddawan, N., Jomnonkwao, S., & Ratanavaraha, V. (2022). Factors affecting severity of motorcycle accidents on Thailand's arterial roads: Multiple correspondence analysis and ordered logistics regression approaches. IATSS Research, 46(1), 101–111. doi:10.1016/j.iatssr.2021.10.006.

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

[46] Prentkovskis, O., Sokolovskij, E., & Bartulis, V. (2010). Investigating Traffic Accidents: A Collision of Two Motor Vehicles. Transport, 25(2), 105–115. doi:10.3846/transport.2010.14.

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

[48] Zhou, M., & Chin, H. C. (2019). Factors affecting the injury severity of out-of-control single-vehicle crashes in Singapore. Accident Analysis & Prevention, 124, 104–112. doi:10.1016/j.aap.2019.01.009.

[49] Ghasemzadeh, A., & Ahmed, M. M. (2019). Exploring factors contributing to injury severity at work zones considering adverse weather conditions. IATSS Research, 43(3), 131–138. doi:10.1016/j.iatssr.2018.11.002.