Journal of Current Science and Technology, April - June 2025 Vol. 15 No. 2, Article 99

ISSN 2630-0656 (Online)

Cite this article: Simmachan, T., & Boonkrong, P. (2025). Effect of resampling techniques on machine learning models for classifying road accident severity in Thailand. *Journal of Current Science and Technology*, *15*(2), Article 99. https://doi.org/10.59796/jcst.V15N2.2025.99

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Effect of Resampling Techniques on Machine Learning Models for Classifying Road Accident Severity in Thailand

Teerawat Simmachan[1,2] and Pichit Boonkrong[3,*]

[1]Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand
[2]Thammasat University Research Unit in Statistical Theory and Applications, Thammasat University, Pathum Thani 12120, Thailand
[3]College of Biomedical Engineering, Rangsit University, Pathum Thani 12000, Thailand

*Corresponding author; E-mail: pichit.bk@rsu.ac.th

Received 1 September 2024; Revised 17 October 2024; Accepted 28 November 2024; Published online 25 March 2025

Abstract

Road traffic accidents (RTAs) pose a significant global challenge, particularly in Thailand. This study investigates the impact of resampling techniques on machine learning (ML) models for classifying road accident severity in Thailand, utilizing data from 31,817 road traffic accidents collected between January 1, 2021, and December 31, 2022. The primary challenge addressed is class imbalance, where fatal accidents represent a small fraction of the dataset. Three popular ML models, including Random Forest (RF), K-Nearest Neighbors (KNN), and Extreme Gradient Boosting (XGB), were evaluated with four resampling techniques: Imbalanced (IB), Under-sampling (US), Over-sampling (OS), and Combined Sampling (CS). These resampling approaches generated 12 ML models, whose performance was evaluated under three different train/test split ratios: 70/30, 80/20, and 90/10. Compared to the IB approach, the results demonstrate that all US, OS and CS techniques significantly improved model performance, particularly in terms of F1 score, G-mean, and balanced accuracy. Among the models, RF-CS, KNN-OS, and XGB-CS exhibited the best classification performance. Although these evaluation metrics improved over the imbalanced scheme, KNN's overall performance in detecting fatal accidents was weaker compared to RF and XGB. Specifically, KNN struggled more with the imbalanced dataset, even after applying resampling techniques. These findings suggest that choosing the appropriate resampling techniques is crucial for enhancing model performance in classifying accident severity.

Keywords**:** *gradient boosting; imbalanced data; KNN; over-sampling; random forest; road safety; SDGs 3*

1. Introduction

Thailand has the highest Road Traffic Fatality (RTF) rate among ASEAN countries and ranks 9th among 175 countries globally, with approximately 36.2 deaths per 100,000 population (WHO, 2018). Since the ratio exceeds the global average, Thai people are at quite high risk of dying from road accidents, and road accident statistics continue to rise. Notably, the incidence of road accidents is even higher during festival periods (Lerdsuwansri et al., 2022). One of the major challenges of road accidents in Thailand is their severity. In other words, road users must exercise caution to reduce the risk of fatalities, injuries, and property damage.

Road Traffic Accidents (RTAs) can occur due to various factors, including human errors (such as drowsiness, intoxication, traffic violations, and lack of road familiarity) and vehicle-related issues (such as

operating an unroadworthy vehicle). Animals, particularly those that wander too close to move vehicles, can also contribute to accidents. Furthermore, a multitude of factors, including road conditions, weather conditions, driving duration, and geographical location, might impact the likelihood of road accidents (Simmachan et al., 2022; Taveekal et al., 2023). It is critical to implement policies that aim to reduce RTA and RTF rates from such accidents. Building a model to classify RTA severity in Thailand is crucial. Accident severity classification (i.e., non-fatal vs. fatal accidents) often results in imbalanced data, where one class significantly outnumbers the other. This issue has an impact on the classification model's performance. More specifically, it reduces the predictive model's accuracy and effectiveness (Kotb, & Ming, 2021; Simmachan et al., 2023). To improve model efficiency, class-balancing techniques are essential before developing the predictive model. Numerous techniques exist to enhance model performance. To address Thailand's urgent road safety concerns in alignment with Sustainable Development Goal 3, this study proposes a framework for predicting RTA severity using multiple resampling techniques, including under-sampling, over-sampling, and combined sampling, across various Machine learning (ML) models such as Random Forest (RF), K-Nearest Neighbors (KNN), and Extreme Gradient Boosting (XGBoost). This paper will provide deeper insights into how these techniques can enhance the performance of ML models, offering valuable recommendations for future studies and applications in road safety.

1.1 Related Works

Over the past two decades, Thai road safety research has primarily focused demography and human behavior rather than road and environmental factors (Chantith et al., 2021; Wisutwattanasak et al., 2022; Phaphan et al., 2023). Motorcycles, young individuals, intoxicated driving, and lack of helmet usage have been the primary factors contributing to road traffic accidents, fatalities, and injuries (Tanaboriboon, & Satiennam, 2005; Siviroj et al., 2012a, 2012b; Riyapan et al., 2018). This trend highlights the limited effectiveness of road safety programs and law enforcement efforts in Thailand. Research focusing on road and environmental factors contributing to RTAs remains limited. A significant increase in RTA incidence was observed in Thailand's southern and northern provinces between 2012 and 2018, which can be attributed to increased precipitation levels. The correlation between rainfall and RTA frequency underscores the need for further investigation into meteorological influences on road safety in these regions (Boonserm, & Wiwatwattana, 2021; Sangkharat et al., 2021; Worachairungreung et al., 2021). Thai highway data from 2011 to 2017 indicated that segment length and average annual traffic volume influenced accident rates. (Champahom et al., 2021).

Table 1 Recent studies on road safety in Thailand using ML approaches

	Model / Algorithm											**Balancing**		
Author (Year)	Logistic regression	Decision Tree	Random Forest	Gradient Boosting	K-Nearest Neighbor	Naïve Bayes	Neural Network	Multinomial Logit Model	Support Vector Machines	Apriori Algorithm	Extreme Gradient Boosting	Under-Sampling	Over-Sampling	Combined Sampling
Chaiwuttisak (2019)	★	★				★	★							
Worachairungreung et al., (2021)			★			★			★				★	★
Boonserm, & Wiwatwattana (2021)			★									★	★	★
Vanishkorn, & Supanich (2022)	★	★	★	★	★								★	
Mahikul et al., (2022)								★						
Chaiyapet et al., (2022)								★						
Almannaa et al., (2023)										★				
Champahom et al., (2023a)	★	★												
Mahikul et al., (2024)			★						★		★	★	★	

Many efforts have been made to reduce road traffic fatality and injury rates using various approaches. By analyzing the relationship between motorization and RTFs per 100,000 people, the safety status of Thailand and other Asian countries has been studied (Klungboonkrong et al., 2019). Count regressions models have been implemented to predict road traffic injuries, fatalities, and their combined impact. Thai RTAs from 2015 were used in the investigations (Simmachan et al., 2022; Lerdsuwansri et al., 2022; Taveekal et al., 2023). Several studies examined ML framework for RTA severity prediction in Thailand. Table 1 reviews recent ML-based works. Statistical models for binary and multi-class classifications were frequently applied (Champahom et al., 2023a, 2023b; Mahikul et al., 2022; Mahikul et al., 2024). However, a notable gap in these studies is the limited exploration of resampling techniques and their effects on various models. Most studies focus on traditional models like logistic regression, decision trees, and random forests, with gradient boosting being used more frequently in recent years. However, advanced techniques such as XGB and KNN are underrepresented. While over-sampling techniques have been applied in several studies under-sampling and combined sampling are less frequently applied across studies. Interestingly, there is a lack of detailed analysis on how these resampling strategies influence model performance, particularly with imbalanced datasets such as those involving road accident severity in Thailand.

1.2 Problem Formation

To evaluate the effect of resampling techniques on ML model performance in classifying road accident severity in Thailand, two main tasks were undertaken: (1) balancing the dataset by ensuring an equal number of class instances and (2) optimizing ML model performance using Thailand's RTA data. Let *D* represent the dataset, consisting of *N* instances (road traffic accidents) and *k* features or predictors related to each accident, such as road section, weather condition, crash type, etc., along with a target variable *Y*. The dataset can be written as $D=\{(X_1,X_2,\ldots,X_k),Y\}$ where their descriptions are given in Table 2. Given the imbalance in the dataset, where the number of non-fatal accidents ($Y=0$) is practically much greater than fatal accidents ($Y=1$), we will firstly apply resampling techniques to balance the dataset. Employing the resulting data in the training set, the classification performance was evaluated in the testing set. Subsequently, the optimizations were conducted as follows:

1.2.1 Resampling

To minimize the class imbalance between fatal and non-fatal accidents by selecting the appropriate resampling technique, the objective is to achieve class balance by minimizing the difference between the number of samples in the two classes after resampling, i.e., $\min_{r}\left|N_0^{(r)}-N_1^{(r)}\right|$ where *r* denotes the replication. The constraint for each resampling technique is described in section 3.4.

1.2.2 ML Performance

To minimize the classification error while ensuring that the model achieves acceptable performance metrics, the least classification error of the model *M* was considered via $\min_{M}\text{Error(M)}$. By training the ML model on the resampled data and ensuring it generalizes well, the better classification performance of the model *M* can be determined by evaluation metrices such as F1-score or G-mean. Thus, maximizing the evaluation metrics, e.g., $\max_{M}\text{F1-score(M)}$ or $\max_{M}\text{G-mean(M)}$, is more convenient and utilized instead of $\min_{M}\text{Error}(M)$ in this study. Each evaluation metric is explained in section 3.7.

2. Objectives

The goal of this research is to compare resampling techniques for class-balancing. Under-sampling, over-sampling, and both are used to resample imbalanced data. The work combines three popular and efficient ML algorithms: RF, XGB, and KNN, a basic nonparametric distance function-based method. ML models and resampling are used to classify RTA severity in Thailand. Thai Ministry of Transport open data website includes 2021–2022 road accident data (Open Government Data of Thailand, 2023). The research aims to assist government agencies in establishing guidelines for managing traffic accident injuries and fatality guidelines. The nation's economy and society will lose less.

3. Methodology

Classifying RTA severity is essential for preventing RTAs. Figure 1 presents the RTA severity prediction framework. The methodology encompasses a comprehensive range of processes, spanning from initial dataset acquisition and description through rigorous data pre-processing, strategic data splitting, the implementation of three distinct techniques to address imbalanced data, the application of diverse machine learning algorithms, and thorough evaluations of their respective performance metrics. The predictive

models were constructed using various resampling techniques to achieve the study's goal. ML algorithms and a data-driven approach were used. Data on RTA factors, including accident causes, road sections, and incident regions, was presented first. Then, data quality and analytical readiness were ensured by data pre-processing. Train-test splits were utilized to split data. ML algorithms were implemented to develop and operate classification models in training sets to discover accident severity patterns that predict accident severity. Finally, test sets generated evaluation metrics for forecasting models.

3.1 Dataset and Description

This study utilizes a dataset including information on RTAs that occurred in Thailand between January 1, 2021, and December 31, 2022. The dataset includes a total of 31,817 accidents. Table 2 demonstrates relevant variable descriptions and their descriptive statistics. The accident severity is the binary dependent variable, i.e., 0 indicates non-fatal accident while 1 represents fatal accident. Accidents resulted in 28,634 injuries (89.99%), and 3,183 deaths (10.01%). It is evident that the number of fatal accidents was significantly lower than the number of non-fatal accidents, indicating imbalanced data. In this case, a fatal accident represents a positive class whereas a non-fatal accident denotes a negative class. The other nine variables are categorical features used to establish predictive models. The bold face represents the top value.

3.2 Data Pre-Processing

Thoroughly, data pre-processing is essential before integrating data into prediction models. This essential stage involves ensuring data meets ML models, processing parameters and methodically correcting missing values. Firstly, the dataset was checked to see if there was any missing data. If there is missing data, it will be removed before further analysis. Moreover, data transformation or data encoding is necessary. This procedure transforms raw data into an appropriate format for analysis, improving the capacity to understand models, optimizing computational speed, and reducing the impact of outliers (Aksoy, & Haralick, 2001; Ioffe, & Szegedy, 2015). Dummy variable encoding was used for categorical features. This method, which represents each category using binary vectors of 0s and 1s, overcomes the constraints of numerical input models.

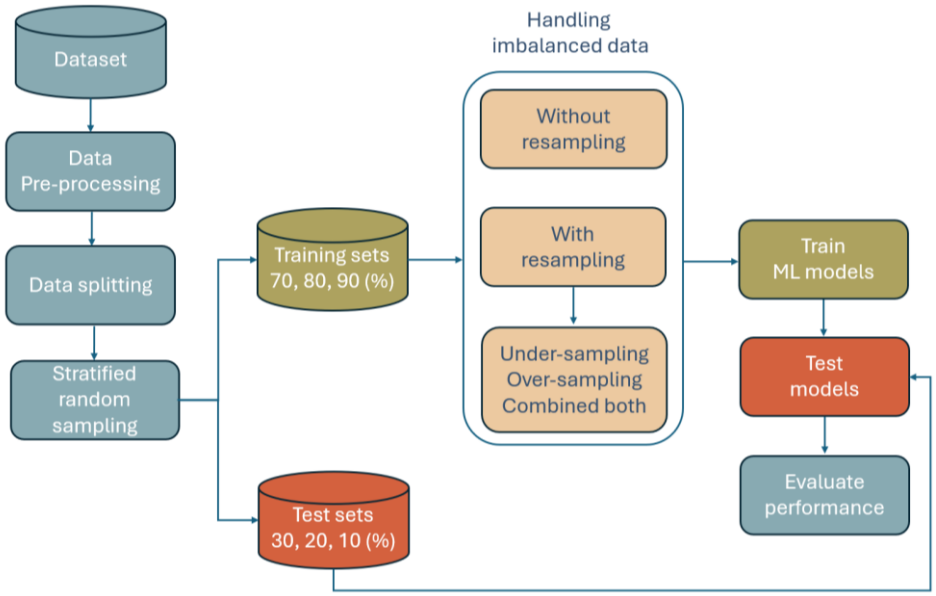

Figure 1 Framework for classifying RTA severity

3.3 Data Splitting

To evaluate the efficacy of ML models in predicting road accident severity in Thailand, rigorous data validation techniques are imperative. Given the substantial sample size (n = 31,817), a train-test split was deemed appropriate. To mitigate potential bias and ensure equitable representation of both positive and negative classes in the training and test sets, stratified random sampling was initially employed. Subsequently, three widely accepted train-test split ratios were implemented, i.e., 70/30, 80/20, and 90/10 percent. This approach enables a comprehensive assessment of model performance across different data distributions, thereby enhancing the robustness and generalizability of predictive models. The stratification process is crucial for maintaining the original class distribution in both subsets, thus preserving the dataset's inherent characteristics (Moon et al., 2019; Na Bangchang et al., 2023; Simmachan et al., 2023). Varying the train-test split ratios in ML classification tasks, such as classifying road accident severity in Thailand, serves several important purposes. First, it allows for the exploration of model performance stability under different training data sizes. A smaller training set (e.g., 70%) might lead to a model that has not learned the full complexity of the data, increasing the likelihood of underfitting, while a larger test set ensures a robust evaluation of model performance. Conversely, a larger training set (e.g., 90%) helps the model capture more patterns in the data but may leave insufficient data for a rigorous test set evaluation. Thus, data splitting provides insights into the amount of data required for the model to achieve optimal performance.

Table 2 Variable descriptions and their descriptive statistics of Thai RTAs in 2021-2022

Variable role	Variable name	Description	RTAs (Percentage)
Dependent variable/ Target variable	Accident severity (Y)	1: Fatal accident	3,183 (10.01)
		0: Non-fatal accident	**28,634 (89.99)**
Independent variables/ Predictor variables/ Features/ Attributes	Cause of accident (X_1)	1: Caused by a person	**27,876 (87.61)**
		2: Caused by a vehicle	1,323 (4.16)
		3: Others	2,618 (8.23)
	Road section (X_2)	1: Straight	**25,654 (80.63)**
		2: Curve	5,482 (17.23)
		3: Others	681 (2.14)
	Region of incidence (X_3)	1: North	7,594 (23.86)
		2: Central	**9,333 (29.33)**
		3: East	4,176 (13.13)
		4: Northeast	6,432 (20.21)
		5: South	4,282 (13.45)
	Crash type (X_4)	1: Overturned	**18,869 (59.30)**
		2: Collision	9,703 (30.50)
		3: Others	3,245 (10.20)
	Road type (X_5)	1: National highway	**31,762 (99.83)**
		2: Rural road	55 (0.17)
	Weather condition (X_6)	1: Clear	**26,909 (84.57)**
		2: Rain	4,679 (14.70)
		3: Others	235 (0.73)
	Time of incidence (X_7)	1: Day	**18,448 (57.98)**
		2: Night	13,369 (42.02)
	Day of incidence (X_8)	1: Weekdays	**22,574 (70.95)**
		2: Weekends	9,243 (29.05)
	Month of incidence (X_9)	1: Festive month	10,582 (33.36)
		2: Others	**21,235 (66.74)**

3.4 Handling Imbalanced Data

Imbalanced datasets, where there is a disproportionate ratio of training samples in each class, pose a fundamental challenge in machine learning (Polvimoltham, & Sinapiromsaran, 2021; Zha et al., 2022, Arockia Panimalar, & Krishnakumar, 2023). Practical applications often encounter this scenario, such as the RTA dataset used in this study. Since non-fatal accidents outnumber fatal accidents, ML models favor the majority class, making them untrustworthy (Kotb, & Ming, 2021; Simmachan et al., 2023). Consequently, class-balancing techniques are necessary for handling the imbalanced dataset. These techniques modify the distribution of classes to boost model performance. Classes can be balanced in many ways. The basic class balancing approaches are data-based and algorithm-based (Mathew, 2022). Both approaches aim to reduce the effects of class imbalance, but differently. The main distinction lies in whether the emphasis is on modifying the data or the learning algorithm. After data splitting, resampling techniques under data-based approaches were applied to the training sets. There were four different resampling techniques employed in this study: one of which did not use any resampling techniques, while the other three did. Focusing on the binary target variable *Y*, where Y = 0 represents a non-fatal accident (majority class) and Y = 1 represents a fatal accident (minority class), the resampling techniques aim to balance the number of instances in these classes without considering the features X_i. The details of the four techniques were described as follows:

3.4.1 Imbalanced Scheme (IB)

The original dataset remains imbalanced, with the number of instances of Y = 0 is much larger than Y = 1 leading to biased model predictions in its favor. The class distribution can be written as $|Y=0| \gg |Y=1|$. In other words, the original dataset is directly assessed by classification model without any modification.

3.4.2 Under-sampling (US)

Under-sampling is a technique that reduces the number of observations in the majority class to achieve a balanced dataset (Polvimoltham, & Sinapiromsaran, 2021). This technique can mitigate bias toward the majority class but may lead to the loss of crucial information. The number of instances in the majority class Y = 0 is decreased to balance the dataset with the minority class Y = 1. Thus, the new size of both classes after under-sampling is denoted by $n_{US} = \min\{|Y=0|,|Y=1|\}$. Under-sampling selects a random subset of $Y = 0$ instance such that

$$Y_{US}=\{Y_i=0|i\in\text{random subset}, |Y=0|=n_{US}\}\cup|Y=1|.$$

As a result, the dataset becomes balanced, but crucial information from the majority class $Y = 0$ may be lost.

3.4.3 Over-sampling (OS)

Over-sampling is a technique that involves augmenting the number of observations in the minority class to achieve a balanced dataset (Zha et al., 2022, Pasangthien & Yimwadsana, 2022). To balance the dataset by over-sampling, the number of instances in the minority class $Y = 1$ is increased to be equal with $Y = 0$. Then, the new size of both classes after over-sampling is $n_{OS} = \max\{|Y=0|,|Y=1|\}$. Simple duplication is used to balance the class sizes, ensuring equal representation:

$$Y_{OS}=\{Y_i=0|i\in D_{majority}\}\cup\{Y_j=1|j\in D_{minority}, \text{randomely duplicate}|Y=1|=n_{OS}\}$$

.

This strategy can augment the training data for the minority class, but it may lead to overfitting due to repeated instances.

3.4.4 Combined sampling (CS)

Combined sampling combines under-sampling of the majority class and over-sampling of the minority class to create a balanced dataset. It addresses the limitations of both methods: under-sampling, which can result in loss of important data, and over-sampling, which may lead to overfitting (He et al., 2005; Polvimoltham, & Sinapiromsaran, 2021). The technique uses a balance factor, $\alpha\in[0,1]$, to adjust the proportions of under-sampled and over-sampled data, forming the new dataset. CS combines simplified versions of US and OS, providing $n_{CS}=\alpha\cdot n_{US}+(1-\alpha)\cdot n_{OS}$. The new dataset is formed as $Y_{CS}=Y_{US}\cup Y_{OS}$, which $\alpha=0.5$ is used in this study. This approach balances the binary target variable, *Y*, ensuring that the classifier generalizes better and reduces bias toward the majority class. Regarding four data handling techniques, Table 3 presents train and test distributions for road accident severity classification from four sampling schemes.

.

Table 3 Train/test ratios for different data sampling techniques in road accident severity classification

Train/Test	Class of *Y*	IB		US		OS		CS	
		Train	Test	Train	Test	Train	Test	Train	Test
70/30	1: Fatal accident	2,367	994	2,367	994	19,904	994	11,206	994
	0: Non-fatal accident	19,904	8,552	2,367	8,552	19,904	8,552	11,065	8,552
80/20	1: Fatal accident	2,707	654	2,707	654	22,746	654	12,621	654
	0: Non-fatal accident	22,746	5,710	2,707	5,710	22,746	5,710	12,832	5,710
90/10	1: Fatal accident	3,032	329	3,032	329	25,603	329	14,149	329
	0: Non-fatal accident	25,603	2,853	3,032	2,853	25,603	2,853	14,486	2,853

Regarding four data handling techniques, Table 3 presents train and test distributions for road accident severity classification from four sampling schemes. For IB, the original dataset is used with the imbalance preserved, showing a significant difference between non-fatal and fatal accident cases. In US, non-fatal cases are under-sampled to match the number of fatal accidents, resulting in a balanced train set, but the test set remains imbalanced. OS applies oversampling to fatal cases, balancing them with non-fatal accidents for training while keeping the test set unchanged. Lastly, CS combines both under- and over-sampling, balancing fatal and non-fatal cases within the train set, while the test set retains its original distribution. Each approach addresses data imbalance uniquely, ensuring more reliable model training and improving the accuracy of predicting fatal versus non-fatal road accidents.

3.5 Machine Learning Algorithms

To address the effect of resampling techniques on ML models for classifying road accident severity, several studies have explored various methods to handle imbalanced datasets, which is a common issue in this domain. Resampling techniques play a crucial role in improving the classification performance of machine learning models. Thus, this study utilizes three efficient and widely used ML algorithms, including RF, XGB, and KNN. These nonparametric algorithms do not rely on explicit assumptions and are user-friendly. The key details of each algorithm are outlined below:

3.5.1 Random Forest

Random Forest (RF) was first presented by Breiman (2001). It is a type of ensemble learning algorithm called bagging or bootstrap aggregation (Breiman, 2001). This algorithm comprises multiple independent decision trees functioning cooperatively. Initially, in bootstrap samples, *n* instances are randomly selected from the training data to create *n* decision trees. The prediction for each tree is based on feature importance scores, such as Mean Decrease in Accuracy (MDA) or Mean Decrease in Impurity (MDI) (James et al., 2013). For each tree, a set of features $X_1,X_2,\ldots,X_k$is randomly selected, and a majority vote among the predicted values from all trees determines the final prediction for a new data point:

$$\hat{y}=\text{majority vote}(\hat{y}_1,\hat{y}_2,...,\hat{y}_n)\ ,$$

where $\hat{y}_i$ is the prediction from the i^{th} tree, and i = 1, 2, …, n. This method enhances the overall accuracy of the model through aggregation (Kowshalya, & Nandhini, 2018; Simmachan et al., 2023; Na Bangchang et al., 2023).

3.5.2 Extreme Gradient Boosting (XGB)

XGB is a robust ML algorithm devised by Friedman (2001) and has been implemented in various fields (Friedman, 2001; Bentéjac et al., 2021). XGB is another type of ensemble algorithm, specifically boosting. Boosting is a technique that combines numerous weak learners, such as decision trees, into a single strong learner. Mathematically, XGB minimizes the loss function L(θ) over multiple trees, each tree t making prediction $\hat{y}^{(t)}$:

$$L(\theta)=\sum_{i=1}^{n} l\left(y_i,\hat{y}_i^{(t)}\right)+\sum_{t}\Omega(f_t)$$

where $l(\cdot)$is the loss function (e.g., squared error for regression, log loss for classification), and $\Omega(f_t)$ is a regularization term for controlling model complexity. The goal is to iteratively correct the prediction errors from previous trees by adding new trees that minimize the current error, enhancing the model's accuracy (Bentéjac et al., 2021; Na Bangechang et al., 2023).

3.5.3 K-Nearest Neighbor

The K-Nearest Neighbor (KNN) algorithm is a nonparametric, instance-based learning method that

classifies observations based on proximity to other data points in the feature space (Moulaei et al., 2022; Pechprasarn et al., 2025). It operates by identifying the k closest instances to a query point using a distance metric, typically Euclidean distance. The Euclidean distance $D(p_i,q_j)$ between two data points p and q with k features is computed as:

$$D(p_i,q_j)=\sqrt{\sum_{l=1}^{k}(p_{il},q_{jl})^2}$$

Classification is then performed by majority voting among the k-nearest neighbors (Prasasti et al., 2020):

$$\hat{y}=\text{majority vote}(y_1,y_2,...,y_k)$$

where $y_1,y_2,...,y_k$ are the labels of the k nearest neighbors. This approach leverages local information and assumes that nearby data points likely share the same class label, making it effective for many pattern recognition tasks (Wang et al., 2007; Boonkrong, & Simmachan, 2016; Mamdouh Farghaly et al., 2023).

This study explores the combined application of three ML algorithms and four resampling strategies to explore and enhance predictive performance. Subsequently, 12 ML possible algorithms and their descriptions are listed in Table 4.

3.6 Hyperparameter Tuning

Hyperparameter tuning is crucial in developing ML models for several reasons. Proper tuning hyperparameters can significantly enhance a model's performance by optimizing complexity and learning rate (Geron, 2019). It helps prevent overfitting by limiting model complexity and underfitting by allowing more flexibility (Goodfellow et al., 2016). Hyperparameters, such as regularization strength, influence the bias-variance trade-off, enabling models to generalize well (Hastie et al., 2009). Moreover, tuning ensures models perform well on unseen data, minimizing overfitting and maximizing generalization performance (Kuhn, & Johnson, 2013). The grid search technique, a popular tuning tool, was used to optimize the parameters of the ML models. This technique explores different combinations of hyperparameters to find the set that yields the best classification performance. Therefore, 12 ML models across all data splitting ratios were trained using different hyperparameter combinations. Table 5 shows parameter settings and the best values for the ML model in classifying the RTA severity. The best values, such as 800 trees for RF, 9 neighbors for KNN, and 150 boosting rounds in XGB, were obtained through the hyperparameter tuning process.

Table 4 Description of algorithms used for predicting RTA severity

Algorithm	Description
RF-IB	Random Forest without resampling techniques or under imbalanced data.
RF-US	Random Forest with under-resampling technique.
RF-OS	Random Forest with over-resampling technique.
RF-CS	Random Forest with combined resampling technique.
KNN-IB	K-Nearest Neighbor without resampling techniques or under imbalanced data.
KNN-US	K-Nearest Neighbor with under-resampling technique.
KNN-OS	K-Nearest Neighbor with over-resampling technique.
KNN-CS	K-Nearest Neighbor with combined resampling technique.
XGB-IB	Extreme Gradient Boosting without resampling techniques or under imbalanced data.
XGB-US	Extreme Gradient Boosting with under-resampling technique.
XGB-OS	Extreme Gradient Boosting with over-resampling technique.
XGB-CS	Extreme Gradient Boosting with combined resampling technique.

Table 5 Machine learning models with their parameter settings

Model	Hyperparameter	Best Value
RF	The number of trees $\in$ {100, 200, …, 1000}	800
	The maximum depth of each tree $\in$ {1, 2, …, 5}	5
	The number of features to consider when looking for the best split $\in$ {1, 2, 3}	3
	The minimum number of samples required to split a node $\in$ {1, 2, …, 10}	3
	The minimum number of samples required to be at a leaf node $\in$ {1, 2, …, 5}	3
KNN	The number of neighbors (k) $\in$ {3, 5, …, 31}	9
XGB	Number of boosting rounds $\in$ {50,100, 150, 200}	150
	Learning rate $\in$ {0.1-0.5}	0.1
	The maximum depth of each tree $\in$ {1, 2, …, 10}	3

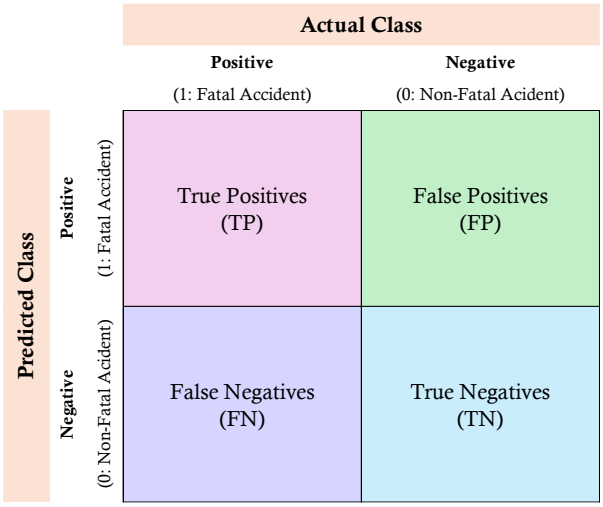

Figure 2 Confusion matrix for RTA Severity Classification

3.7 Model Evaluation

Assessing the performance of an ML model is essential. A confusion matrix is a useful tool for computing evaluation metrics. When selecting evaluation metrics for RTA severity classification model, it is critical to prioritize metrics that provide the most insight into the model's performance, especially given the class imbalance.

3.7.1 Confusion Matrix

A confusion matrix is a commonly utilized tool in classification tasks (Akarajarasroj et al., 2023; Yilmaz, & Demirhan, 2023). In RTA severity binary classification problems, we treat a fatal accident as a positive class and a non-fatal accident as a negative class. Four possible outcomes are used in the matrix, which depicts predicted and actual counts as shown in Figure 2. TP represents the model's fatal accident accuracy. TN denotes the model's correct non-fatal accident identification. FP indicates the number of non-fatal accidents misclassified as fatal. FN shows fatal accidents misclassified as non-fatal. To measure how good the models have performed, all TP, TN, FP and FN are necessary for computing evaluation metrics.

3.7.2 Evaluation Metrics

Typically, road safety practitioners prioritize raising awareness of fatal accidents (positive class) over non-fatal accidents (negative class). In RTA severity classification, missing a fatal accident (false negative) could have severe consequences, making it crucial to maximize the detection of fatal accidents. However, this research considers both fatal and non-fatal accident classes. Given this goal and class imbalance characteristics of RTA dataset, therefore, priority metrics include G-mean, balanced accuracy, F1-score, and accuracy. These metrics assess the model's ability to detect fatal accidents while minimizing false alarms, optimizing safety and resource allocation. The corresponding characteristics and formulas of the metrics are provided as follows:

- Accuracy measures the proportion of correct predictions (both true positives and true negatives) out of the total predictions. It is often used as the default evaluation metric but can be misleading in imbalanced datasets. For this reason, accuracy is often less prioritized in this study. Generally, the accuracy is defined as
$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

- F1-score is the harmonic means of both recall and precision. The F1-score achieves a balance between precision and recall, offering a single statistic considering both false positives and false negatives, which is useful in situations when both are costly. F1-score is given as
$$\text{F1-score} = \frac{2TP}{2TP + FP + FN}$$
where $\text{precision} = \frac{TP}{TP + FP}$ and $\text{recall} = \frac{TP}{TP + FN}$.

- G-mean (Geometric mean) is the geometric mean of sensitivity (recall) and specificity. G-mean

provides an overall balance of the classifier's performance across both classes. G-mean is particularly useful in imbalanced data because it considers the balance between sensitivity (true positive rate) and specificity (true negative rate). G-mean is derived as

$$\text{G-mean} = \sqrt{\frac{\text{TP}}{\text{TP} + \text{FN}} \times \frac{\text{TN}}{\text{TN} + \text{FP}}}$$

where specificity $= \frac{\text{TN}}{\text{TN} + \text{FP}}$.

- Balanced Accuracy adjusts the traditional accuracy metric to account for class imbalance. It is the arithmetic means of sensitivity and specificity. It helps capture the overall performance while accounting for both false negative and false positives. Balanced Accuracy is computed by

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}\right).$$

as the value of each evaluation metric rises, the model's performance improves.

4. Results and Discussion

This section presents a detailed analysis of numerical outcomes, primarily focusing on the performance of various machine learning models under different resampling techniques. Additionally, the discussion interprets these findings, emphasizing key trends, potential limitations, and practical implications for optimizing model selection in imbalanced datasets.

4.1 Numerical Results

This work has effectively highlighted the integration of ML methodologies and resampling techniques to accurately predict the severity of RTAs in Thailand. Three classifiers, including RF, KNN, and XGB, were used in conjunction with four resampling strategies, generating 12 predictive models. Table 6 shows the overall performance of all predictive models. Bold text indicates the best models in each evaluation metric for each train-test split. Model performance across the three train-test splitting options was consistent. For easier interpretation, the graphical results are shown in Figure 3. The evaluation metrics were averaged over all train-test splits. Based on the main objective of the study, the numerical findings were given as follows:

4.1.1 Classification without Resampling

Using different resampling techniques, Table 6 provides model performance metrics for RF, KNN and XGB classifiers across three train/test splits (70/30, 80/20, 90/10). The IB results demonstrate how each model performs with the original data, highlighting the challenges of predicting the minority class (fatal accidents). For RF, IB achieves the highest accuracy at 89.59% for the 70/30 split, 89.49% for 80/20, and 89.53% for 90/10, but the F1-score and G-mean are significantly low, indicating poor performance in classifying the minority class. For KNN, IB produces consistently high accuracy, but the F1 score remains low (ranging from 17.63 to 17.64), reflecting poor minority class performance. XGB's IB accuracy is similarly high, reaching 89.60% to 89.85%, but the F1 scores are low, particularly for the fatal accidents class. While IB yields high accuracy, this metric is misleading due to the class imbalance, with non-fatal accidents dominating the dataset. Accuracy alone fails to account for the performance disparity between majority and minority classes. The very low F1 scores, particularly for RF (0.20–0.60) and KNN (17.63–17.64), demonstrate poor precision and recall for fatal accidents. G-mean and balanced accuracy, both of which consider performance on both classes, are also much lower for IB compared to the other resampling techniques. This highlights the need for improvement in IB, as models trained on imbalanced data are biased toward the majority class, leading to poor generalization in predicting fatal accidents, which is critical in this scenario. To avoid misleading conclusions in an imbalanced dataset, G-mean and balanced accuracy were more suitable than accuracy when considering both classes of RTA severity. Meanwhile, the F1 score measured the model's capability in predicting the positive class (fatal accidents). Clearly, the imbalanced scheme resulted in an extremely low F1 score.

Table 6 Model performance (%) along with different resampling techniques

Train/ Test	**RF**				**KNN**				**XGB**			
70/30	IB	US	OS	CS	IB	US	OS	CS	IB	US	OS	CS
Accuracy	89.59	73.38	73.88	73.79	**89.72**	62.95	64.39	67.69	89.60	71.36	72.35	71.85
F1-score	0.20	32.26	32.53	**32.67**	17.63	26.63	26.57	25.94	8.31	30.93	31.72	31.92
G-mean	3.17	67.49	67.54	67.80	32.43	63.87	63.46	61.53	21.22	66.81	67.37	**67.95**
Balanced Acc.	50.04	67.85	67.95	**68.17**	54.74	63.88	63.47	61.95	52.01	67.03	67.63	68.11
Train/ Test	**RF**				**KNN**				**XGB**			
80/20	IB	US	OS	CS	IB	US	OS	CS	IB	US	OS	CS
Accuracy	89.49	72.69	75.16	74.83	89.73	62.96	64.41	67.70	**89.85**	71.59	73.21	72.41
F1-score	0.30	33.58	32.58	**33.82**	17.63	26.64	26.58	25.96	8.31	30.93	31.72	31.92
G-mean	3.87	68.74	68.77	**68.83**	32.45	63.87	63.47	61.54	21.22	66.81	67.37	67.95
Balanced Acc.	50.06	68.94	**69.23**	69.12	54.75	63.89	63.47	61.96	52.01	67.03	67.63	68.11
Train/ Test	**RF**				**KNN**				**XGB**			
90/10	IB	US	OS	CS	IB	US	OS	CS	IB	US	OS	CS
Accuracy	89.53	74.54	74.39	75.20	**89.73**	62.99	64.48	67.72	89.57	72.34	74.10	73.82
F1-score	0.60	32.27	32.25	**33.08**	17.71	26.65	26.62	25.96	4.60	32.13	32.00	31.36
G-mean	5.48	66.94	67.02	**67.58**	32.45	63.90	63.51	61.60	15.56	67.03	67.05	67.08
Balanced Acc.	50.13	67.52	67.57	**68.16**	54.80	63.91	63.50	61.99	51.02	67.37	67.54	67.52

4.1.2 Effect of Resampling Techniques

Investigating the effect of different resampling techniques across different train/test ratios and classification models, Table 6 and Figure 3 compare their performance in classifying the accident severity. It is observed that resampling techniques, including US, OS and CS significantly improve performance over IB in classifying road accident severity. Each resampling technique affects the model performance as follows:

- **US**: It is seen that non-fatal accident cases are reduced to balance the dataset. While this method lowers accuracy across models compared to IB, it improves the F1-score, G-mean, and balanced accuracy. For example, KNN sees F1-score rise from 17.63% to 26.63% in the 70/30 split, and RF achieves a G-mean of 68.74% for 80/20. This indicates better performance in handling minority classes (fatal accidents). However, under-sampling risks discarding valuable information from the majority class, potentially reducing overall accuracy, which is why its F1-score, and G-mean improvements must be carefully weighed against accuracy drops.

- **OS**: Increasing the number of fatal accident cases to balance the dataset. It typically shows performance improvements across F1-score, G-mean, and balanced accuracy. For example, XGB's F1-score increases from 8.31% in IB to 31.72% in OS for the 70/30 split, and RF's G-mean improves to 67.58% for 90/10. Oversampling prevents information loss from the majority class but can lead to overfitting, especially in models like KNN. F1-score, G-mean, and balanced accuracy are crucial in this context, as they reflect how well the model captures both fatal and non-fatal accidents, regardless of class imbalance.

- **CS**: Merging both US and OS techniques to balance the dataset, this method achieves high F1-score, G-mean, and balanced accuracy without sacrificing as much information as US or overfitting like OS. For instance, RF in the 70/30 split achieves the highest balanced accuracy (68.17%) and a significant F1-score improvement (32.67%). XGB also benefits, with balanced accuracy rising to 68.11% for 80/20. The combined approach provides a balanced trade-off between handling class imbalance and maintaining model generalizability. F1-score, G-mean, and balanced accuracy are crucial in evaluating CS's performance, as they ensure the model predicts both accident types effectively.

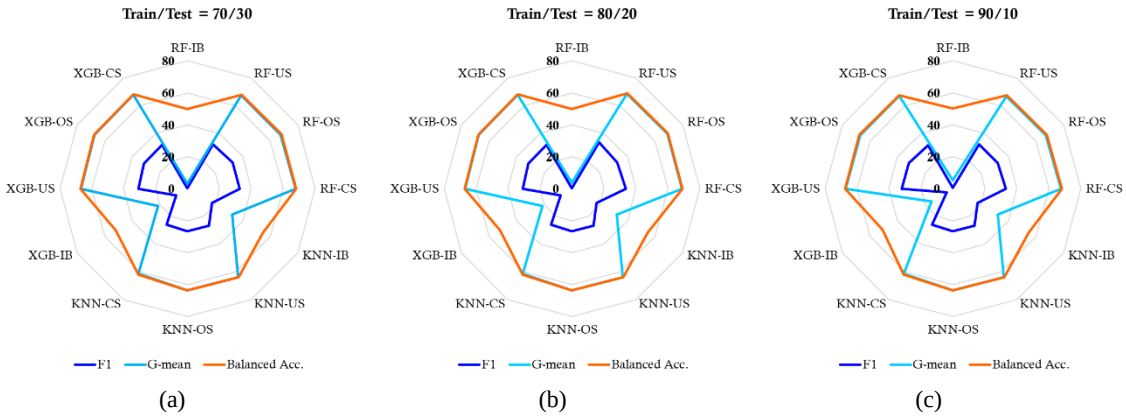

Figure 3 Model performance over three train/test ratios

Figure 4 The confusion matrices across four different samplings in three classification models with train/test ratio of 70/30: (a) – (d) RF models; (e) – (h) KNN models; (i) – (l) XGB models.

As shown in Figure 3, the radar plots illustrate model performance across three train/test splits (70/30, 80/20, and 90/10) using F1-score, G-mean, and Balanced Accuracy metrics. Each axis represents the 12 ML algorithms generated by a combination of the 3 main models (RF, KNN, XGB) and resampling techniques (IB, US, OS, CS). Balanced Accuracy (orange) consistently scores higher, whereas F1-score (blue) and G-mean (sky blue) show greater variability across models and techniques. Focusing on the 70/30 train/test split, the confusion matrices in Figure 4 display classification results for 12 ML algorithms. The 1st – 3rd rows respectively indicate the performance of RF, KNN, and XGB models. The 1st – 4th columns represent that of IB, US, OS, and CS, respectively. Most models perform better in non-fatal accidents (class 0) but struggle with fatal accidents (class 1). For instance, US and CS resampling techniques slightly yield improved performance on minority class predictions, particularly in XGB as shown in Figure 4: (j)-(l). These techniques balance the dataset and enhance F1-score, G-mean, and

balanced accuracy by addressing the imbalance between fatal and non-fatal accident predictions, especially for models like XGB and RF. In conclusion, the results indicate that resampling techniques, particularly US, OS, and CS, significantly improve model performance in classifying RTA severity compared to IB. The metrics F1-score, G-mean, and balanced accuracy consistently show improvements when these resampling methods are applied, addressing the class imbalance challenge. The CS method appears to offer a balanced trade-off between handling class imbalance and preserving overall model performance, making it a suitable technique for improving classification outcomes in road accident severity prediction.

4.2 Discussion

The numerical results obtained from our analysis shed light on the critical role of resampling techniques in addressing the challenge of imbalanced data in road accident severity classification for Thailand. This section delves into the implications of these findings, discussing how different resampling strategies influenced model performance. Furthermore, we explore the practical implementation of these insights in real-world scenarios, considering how improved prediction accuracy can enhance road safety measures, inform policy decisions, and ultimately contribute to reducing the severity and frequency of road accidents in Thailand.

4.2.1 Effect of Resampling Techniques

The comparative analysis of RF, KNN and XGB models under three resampling techniques reveals distinct performance patterns. Resampling techniques play a crucial role in addressing class imbalances in RTA severity prediction models. The empirical results demonstrate that IB often leads to models that are biased toward the majority class (non-fatal accidents), resulting in poor detection of fatal accidents (the minority class). US balances the dataset by reducing majority class instances. While this method improves minority class detection, it often sacrifices overall accuracy due to the loss of valuable data. For instance, KNN sees a substantial F1-score increase from 17.63% (IB) to 26.63% (US) in the 70/30 split, but the overall accuracy decreases. Oversampling (OS), on the other hand, balances the dataset by increasing the minority class cases. This method prevents the loss of information, enhancing both F1-score and G-mean without significantly compromising accuracy. XGB-OS, for example, delivers exceptional performance, with its F1- score rising from 8.31% (IB) to 31.72% (OS) and its G-mean increasing as well. However, oversampling may lead to overfitting, especially in simpler models like KNN. CS achieves high F1, G-mean, and balanced accuracy across all models without losing as much information as US or overfitting like OS. RF-CS in the 70/30 split, for example, achieves the highest balanced accuracy (68.17%) with significant F1-score improvements (32.67%). Thus, the choice of resampling technique significantly influences the model's ability to classify RTA severity accurately. The mechanism of resampling techniques affecting the model performances are discussed as follows:

- RF performs well with CS because it relies on aggregating multiple decision trees, each trained on different subsets of the data. Thus, using a mix of OS and US techniques can further enhance RF performance (Ran, 2023; Sainin et al., 2017; Sun et al., 2021). CS balances both the majority and minority classes, reducing the likelihood of bias in individual trees towards the majority class. This balanced data allows RF to make better splits at each tree level, improving the prediction of both classes and yielding higher overall performance. CS ensures that valuable information from both classes is preserved, enhancing RF's ensemble nature, which thrives on diversity in the data

- KNN shows high performance with US because it relies on the proximity of data points to make predictions. When the dataset is heavily imbalanced, the majority class dominates the decision boundaries, making it difficult for KNN to predict the minority class accurately. By under-sampling the majority class, the dataset becomes more balanced, allowing KNN to focus more on identifying patterns within the minority class. US eliminates the overwhelming influence of the majority class, helping KNN to form clearer decision boundaries, improving the detection of the minority class. However, KNN struggles with imbalanced datasets, which can lead to biased results (Hao et al. 2008; Nair, & Kashyap, 2019; Aryanti et al., 2023; Simmachan et al., 2025). KNN's sensitivity to class distributions makes it susceptible to majority class bias. Under-sampling and other preprocessing techniques effectively address this issue, allowing KNN to

focus more on class proximity and improve classification performance.

- XGB's strength lies in its ability to iteratively improve upon the mistakes of prior models (boosting). With CS technique, the dataset becomes more balanced, allowing XGB to correct errors more effectively for both classes (Xu et al, 2014; Aggarwal, & Jacob, 2020; Sarac, & Guvenis, 2023). In an imbalanced dataset, XGB would focus primarily on the majority class, but CS ensures that it pays equal attention to the minority class. The improved class balance enhances the model's ability to optimize the loss function, resulting in better predictions for both the minority and majority classes. Therefore, combining over-sampling and under-sampling techniques with XGB classifiers can substantially improve model performance across different domains by effectively addressing class imbalance issues.

4.2.2 Model Implementation

The research findings on classifying RTA severity in Thailand offer valuable insights into improving predictive model performance, particularly when handling imbalanced data. For RF and XGB, the OS and CS techniques significantly improve performance in handling imbalanced data. RF with CS achieves the best balanced accuracy, particularly in the 70/30 split, while maintaining a strong F1 score, making it effective for detecting both fatal and non-fatal accidents. XGB consistently outperforms RF, especially with OS, showing higher F1 scores, G-mean values and balanced accuracy, indicating better prediction of fatal accidents. XGB with CS also achieves balanced accuracy, making it the most effective approach for predicting road accident severity in Thailand. Additionally, it prevents the erroneous classification of non-fatal accidents as fatal, thereby averting unnecessary panic and resource misallocations. Therefore, XGB is treated as the most effective algorithm based on the priority of evaluation metrics, and this finding corresponds to the study of Vanishkorn, & Supanich (2022). The proposed models may guide road safety practitioners or authorities to raise road safety planning or policies to the RTAs in Thailand as well as fatality rate. Other features, such as the festive period mentioned in Lerdsuwansri et al., (2022) and the driver demographic factors used in Phaphan et al., (2023), should also be considered for more accurate model performance. Another limitation of this work is that the features used are all categorical. These features may not accurately reflect the actual patterns of RTA severity. Therefore, the quantitative features, such as the number of vehicles of various types utilized in Vanishkorn, & Supanich (2022), should be looked on. Due to the class imbalance in the dataset, where fatal accidents represent a small number of observations compared to non-fatal accidents, classification models favor the majority class, leading to low performance. Additionally, the limited feature set restricts the model's ability to capture key patterns or relationships that could improve overall classification performance in future research. Integrating additional quantitative and environmental features, such as traffic density, road quality, or driver demographics, could improve model performance.

5. Conclusion

Addressing class imbalance issues, this study highlights the critical role of resampling techniques in improving machine learning model performance for classifying road accident severity in Thailand. By applying three popular models including RF, KNN, and XGB to the imbalanced dataset using four different resampling methods including IB, US, OS, and CS, the results demonstrate that the resampling methods significantly enhance the models' ability to classify the RTA severity. While models trained on the imbalanced dataset showed high accuracy, they performed poorly in terms of F1-score, G-mean, and balanced accuracy, particularly in detecting fatal accidents. Under-sampling (US) improved the detection of minority class instances but led to a loss in overall accuracy, but KNN showed its best performance in this case. Oversampling (OS) showed superior performance across all metrics without information loss. Based on our numerical findings, combined sampling (CS) balanced the benefits of both under- and oversampling, achieving the highest performance of RF and XGB across all metrics without sacrificing data or overfitting. These findings underscore the importance of selecting appropriate resampling techniques customized to specific performance objectives. Increasing RTAs is a critical issue worldwide, especially in Thailand. Future work should explore the integration of additional features, other ML methods, and address data collection challenges to further enhance model performance for road traffic accident severity classification. Advanced machine learning techniques, such as ensemble methods and deep learning could be explored to

enhance accuracy and robustness. Addressing data collection limitations and utilizing external methods to handle class imbalance more effectively should also be prioritized. Expanding the dataset and incorporating real-time data analysis could further improve prediction capabilities and support better road safety interventions.

6. Acknowledgements

We acknowledge the referees for their informative manuscript suggestions. The authors gratefully acknowledge the financial support provided by the Faculty of Science and Technology, Thammasat University, Contract No. SciGR 4/2567.

7. References

Aggarwal, H. K., & Jacob, M. (2020). J-MoDL: Joint model-based deep learning for optimized sampling and reconstruction. *IEEE Journal of Selected Topics in Signal Processing*, *14*(6), 1151-1162. https://doi.org/10.1109/JSTSP.2020.3004094

Akarajarasroj, T., Wattanapermpool, O., Sapphaphab, P., Rinthon, O., Pechprasarn, S., & Boonkrong, P. (2023, October 28-31). *Feature selection in the classification of erythemato-squamous diseases using machine learning models and principal component analysis* [Conference presentation]. 2023 15th Biomedical Engineering International Conference (BMEiCON). IEEE, Tokyo, Japan. https://doi.org/10.1109/BMEiCON60347.2023.10322034

Aksoy, S., & Haralick, R. M. (2001). Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, *22*(5), 563-582. https://doi.org/10.1016/S0167-8655(00)00112-4

Almannaa, M., Zawad, M. N., Moshawah, M., & Alabduljabbar, H. (2023). Investigating the effect of road condition and vacation on crash severity using machine learning algorithms. *International Journal of Injury Control and Safety Promotion*, *30*(3), 392-402.

Arockia Panimalar, S., & Krishnakumar, A. (2023). A review of churn prediction models using different machine learning and deep learning approaches in cloud environment. *Journal of Current Science and Technology*, *13*(1), 136-161. https://doi.org/10.14456/jcst.2023.12

Aryanti, R., Arifin, Y. T., Khairunas, S., Misriati, T., Dalis, S., Baidawi, T., ... & Marlina, S. (2023). *The use of resampling techniques to overcome imbalance of data on the classification algorithm* [Conference presentation]. AIP Conference Proceedings. AIP Publishing, Jakarta, Indonesia. https://doi.org/10.1063/5.0128424

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, 1937-1967. https://doi.org/10.1007/s10462-020-09896-5

Boonkrong, P., & Simmachan, T. (2016). A Multigroup SEIR Epidemic Model with Vaccination on Heterogeneous Network. *Chiang Mai Journal of Science*, *43*(4), 897-903.

Boonserm, E., & Wiwatwattana, N. (2021). *Using Machine Learning to Predict Injury Severity of Road Traffic Accidents During New Year Festivals from Thailand's Open Government Data* [Conference presentation]. The 2021 9th International Electrical Engineering Congress (iEECON). IEEE, March 10-12, 2021, Pattaya, Thailand. https://doi.org/10.1109/iEECON51072.2021.9440287

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chaiwuttisak, P. (2019). *Analysis of Accidental Deaths During Songkran Festival Using Data Mining* [Conference presentation]. The International Conference on Industrial Engineering and Operations Management Pilsen, IEOM Society International, July 23-26, 2019, Czech Republic.

Chaiyapet, C., Phakdeekul, W., & Kedthongma, W. (2022). Risk factors of severity of road accident injury incidence at Kut Bak district Sakon Nakhon province, Thailand. *Res Militaris*, *12*(5), 835-45.

Champahom, T., Jomnonkwao, S., Banyong, C., Nambulee, W., Karoonsoontawong, A., & Ratanavaraha, V. (2021). Analysis of crash frequency and crash severity in Thailand: Hierarchical structure models approach. *Sustainability*, *13*(18), Article 10086. https://doi.org/10.3390/su131810086

Champahom, T., Wisutwattanasak, P., Se, C., Banyong, C., Jomnonkwao, S., & Ratanavaraha, V. (2023a). Analysis of Factors

Associated with Highway Personal Car and Truck Run-Off-Road Crashes: Decision Tree and Mixed Logit Model with Heterogeneity in Means and Variances Approaches. *Informatics*, *10*(3), Article 66. https://doi.org/10.3390/informatics10030066

Champahom, T., Se, C., Aryuyo, F., Banyong, C., Jomnonkwao, S., & Ratanavaraha, V. (2023b). Crash Severity Analysis of Young Adult Motorcyclists: A Comparison of Urban and Rural Local Roadways. *Applied Sciences*, *13*(21), Article 11723. https://doi.org/10.3390/app132111723

Chantith, C., Permpoonwiwat, C. K., & Hamaide, B. (2021). Measure of productivity loss due to road traffic accidents in Thailand. *IATSS Research*, *45*(1), 131-136. https://doi.org/10.1016/j.iatssr.2020.07.001

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, *29*(5), 1189-1232. https://www.jstor.org/stable/2699986

Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. Cambridge: MIT press.

Hao, X., Zhang, C., Xu, H., Tao, X., Wang, S., & Hu, Y. (2008). *An improved condensing algorithm* [Conference presentation]. Seventh IEEE/ACIS International Conference on Computer and Information Science (icis 2008). IEEE, May 14-16, 2008, Portland, OR, USA. https://doi.org/10.1109/ICIS.2008.67

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

He, G., Han, H., & Wang, W. (2005). *An over-sampling expert system for learing from imbalanced data sets* [Conference presentation]. the 2005 international conference on neural networks and brain. IEEE, October 13-15, 2005, Beijing. https://doi.org/10.1109/ICNNB.2005.1614671

Ioffe, S., & Szegedy, C. (2015). *Batch normalization: Accelerating deep network training by reducing internal covariate shift* [Conference presentation]. Proceedings of the 32nd International Conference on Machine Learning. July 6-11, 2015, PMLR, Lille, France. https://proceedings.mlr.press/v37/ioffe15.html

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: springer.

Klungboonkrong, P., Woolley, J., Pramualsakdikul, S., Tirapat, S., Yotmeeboon, W., Pattulee, N., & Faiboun, N. (2019). Road safety status and analysis in Thailand and other Asian countries. *Engineering & Applied Science Research*, *46*(4), 340-348. https://ph01.tci-thaijo.org/index.php/easr/index

Kotb, M. H., & Ming, R. (2021). Comparing SMOTE Family Techniques in Predicting Insurance Premium Defaulting using Machine Learning Models. *International Journal of Advanced Computer Science and Applications*, *12*(9), 621-629. https://doi.org/10.14569/IJACSA.2021.0120970

Kowshalya, G., & Nandhini, M. (2018). *Predicting fraudulent claims in automobile insurance* [Conference presentation]. the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). April 20-21, 2018, IEEE, Coimbatore, India. https://doi.org/10.1109/ICICCT.2018.8473034

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York: Springer.

Lerdsuwansri, R., Phonsrirat, C., Prawalwanna, P., Wongsai, N., Wongsai, S., & Simmachan, T. (2022). Road traffic injuries in Thailand and their associated factors using Conway-Maxwell-Poisson regression model. *Thai Journal of Mathematics*, 240-249.

Mahikul, W., Aiyasuwan, O., Thanartthanaboon, P., Chancharoen, W., Achararit, P., Sirisombat, T., & Singkham, P. (2022). Factors affecting bus accident severity in Thailand: A multinomial logit model. *PLoS One*, *17*(11), Article e0277318. https://doi.org/10.1371/journal.pone.0277318

Mahikul, W., Thongbun, T., Tungparamutsakul, A., Kitudom, P., Phun, S., ..., & Onuean, A. (2024). *Machine Learning for Predicting the Severity of Road Accident Victims at a University Hospital Emergency Center* [Conference presentation]. the 2024 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, February 18-21, 2024, Bangkok, Thailand. https://doi.org/10.1109/BigComp60711.2024.00091

Mamdouh Farghaly, H., Shams, M. Y., & Abd El-Hafeez, T. (2023). Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt. *Knowledge and Information Systems*, *65*(6), 2595-2617. https://doi.org/10.1007/s10115-023-01851-4

Mathew, T. E. (2022). Appositeness of Hoeffding tree models for breast cancer classification. *Journal of Current Science and Technology*, *12*(3), 391-407. https://ph04.tci-thaijo.org/index.php/JCST/article/view/253

Moon, H., Pu, Y., & Ceglia, C. (2019). A Predictive Modeling for Detecting Fraudulent Automobile Insurance Claims. *Theoretical Economics Letters*, *9*(6), Article 1886. https://doi.org/10.4236/tel.2019.96120

Moulaei, K., Bahaadinbeigy, K., Ghasemian, F., & Taghiabad, Z. M. (2022). Predicting the Mortality in the Patients Hospitalized in Intensive Care Units (ICU) Based on Machine Learning Techniques. *Science & Technology Asia*, *27*(2), 98–114. https://ph02.tci-thaijo.org/index.php/SciTechAsia/article/view/242886

Na Bangchang, K., Wongsai, S., & Simmachan, T. (2023). *Application of Data Mining Techniques in Automobile Insurance Fraud Detection* [Conference presentation]. Proceedings of the 2023 6th International Conference on Mathematics and Statistics. July 14-16, 2023, ACM, New York, NY, USA. https://doi.org/10.1145/3613347.3613355

Nair, P., & Kashyap, I. (2019). Optimization of kNN classifier using hybrid preprocessing model for handling imbalanced data. *International Journal of Engineering Research and Technology*, *12*(5), 697-704.

Open Government Data of Thailand. (2023). *Road accident data set*. Retrieved December 20, 2023, from https://data.go.th/en/

Pasangthien, T., & Yimwadsana, B. (2022). Rebalancing Clinical Data with Probabilistic Random Oversampling. *Journal of the Thai Medical Informatics Association*, *8*(2), 68–72. https://he03.tci-thaijo.org/index.php/jtmi/article/view/480

Pechprasarn, S., Srisaranon, N., & Yimluean, P. (2025). Optimizing diabetes prediction: an evaluation of machine learning models through strategic feature selection. *Journal of Current Science and Technology*, *15*(1), Article 75. https://doi.org/10.59796/jcst.V15N1.2025.75

Phaphan, W., Sangnuch, N., & Piladaeng, J. (2023). Comparison of the Effectiveness of Regression Models for the Number of Road Accident Injuries. *Science & Technology Asia*, *28*(4), 54–66. https://ph02.tci-thaijo.org/index.php/SciTechAsia/article/view/249723

Polvimoltham, P., & Sinapiromsaran, K. (2021). Mass Ratio Variance Majority Undersampling and Minority Oversampling Technique for Class Imbalance. In *Fuzzy Systems and Data Mining VII* (pp. 152-161). IOS Press. https://doi.org/10.3233/FAIA210186

Prasasti, I. M. N., Dhini, A., & Laoh, E. (2020). *Automobile insurance fraud detection using supervised classifiers* [Conference presentation]. The 2020 International Workshop on Big Data and Information Security (IWBIS). October 17-18, 2020, IEEE, Depok, Indonesia. https://doi.org/10.1109/IWBIS50925.2020.9255426

Ran, C. (2023). *An Imbalanced Data Classification Algorithm Based on Mixed Sampling* [Conference presentation]. 2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). December 8-10, 2023, IEEE, Chongqing, China. https://doi.org/10.1109/ITAIC58329.2023.10409074

Riyapan, S., Thitichai, P., Chaisirin, W., Nakornchai, T., & Chakorn, T. (2018). Outcomes of emergency medical service usage in severe road traffic injury during Thai holidays. *Western Journal of Emergency Medicine*, *19*(2), 266-275. https://doi.org/10.5811/westjem.2017.11.35169

Sainin, M. S., Alfred, R., Adnan, F., & Ahmad, F. (2017). *Combining sampling and ensemble classifier for multiclass imbalance data learning* [Conference presentation]. Computational Science and Technology: 4th ICCST 2017, November 29-30, 2017, Kuala Lumpur, Malaysia. Springer Singapore. https://doi.org/10.1007/978-981-10-8276-4_25

Sangkharat, K., Thornes, J. E., Wachiradilok, P., & Pope, F. D. (2021). Determination of the impact of rainfall on road accidents in

Thailand. *Heliyon*, *7*(2), Article e06061. https://doi.org/10.1016/j.heliyon.2021.e06061

Sarac, K., & Guvenis, A. (2023). *Determining HPV status in patients with oropharyngeal cancer from 3D CT images using radiomics: Effect of sampling methods* [Conference presentation]. International Work-Conference on Bioinformatics and Biomedical Engineering. Cham, July 12-14, 2023, Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34960-7_3

Simmachan, T., Manopa, W., Neamhom, P., Poothong, A., & Phaphan, W. (2023). Detecting fraudulent claims in automobile insurance policies by data mining techniques. *Thailand Statistician*, *21*(3), 552-568. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/250065

Simmachan, T., Wongsai, N., Wongsai, S., & Lerdsuwansri, R. (2022). Modeling road accident fatalities with underdispersion and zero-inflated counts. *PLoS One*, *17*(11), Article e0269022. https://doi.org/10.1371/journal.pone.0269022

Simmachan, T., Wongsai, S., Lerdsuwansri, R., & Boonkrong, P. (2025). Impact of COVID-19 Pandemic on Road Traffic Accident Severity in Thailand: An Application of K-Nearest Neighbor Algorithm with Feature Selection Techniques. *Thailand Statistician*, 23(1), 129-143.

Siviroj, P., Peltzer, K., Pengpid, S., & Morarit, S. (2012a). Helmet use and associated factors among Thai motorcyclists during Songkran festival. *International Journal of Environmental Research and Public Health*, *9*(9), 3286-3297. https://doi.org/10.3390/ijerph9093286

Siviroj, P., Peltzer, K., Pengpid, S., & Morarit, S. (2012b). Non-seatbelt use and associated factors among Thai drivers during Songkran festival. *BMC Public Health*, *12*, Article 608. https://doi.org/10.1186/1471-2458-12-608

Sun, H., Wang, A., Feng, Y., & Liu, C. (2021). *An optimized random forest classification method for processing imbalanced data sets of alzheimer's disease* [Conference presentation]. 2021 33rd Chinese Control and Decision Conference (CCDC). IEEE. https://doi.org/10.1109/CCDC52312.2021.9602177.

Tanaboriboon, Y., & Satiennam, T. (2005). Traffic accidents in Thailand. *IATSS Research*, *29*(1), 88-100. https://doi.org/10.1016/S0386-1112(14)60122-9

Taveekal, P., Rajchanuwong, P., Wongwiangjan, R., Lerdsuwansri, R., Intrakul, J., Simmachan, T., & Wongsai, S. (2023). Modelling Road Accident Injuries and Fatalities in Suratthani Province of Thailand Using Conway-Maxwell-Poisson Regression. *Thailand Statistician*, *21*(3), 569-579. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/250067

Vanishkorn, B., & Supanich, W. (2022). *Crash severity classification prediction and factors affecting analysis of highway accidents* [Conference presentation]. 2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA). September 28-29, 2022, IEEE, Tokoname, Japan. https://doi.org/10.1109/ICAICTA56449.2022.9932998

Wang, J., Neskovic, P., & Cooper, L. N. (2007). Improving nearest neighbor rule with a simple adaptive distance measure. *Pattern Recognition Letters*, *28*(2), 207-213. https://doi.org/10.1016/j.patrec.2006.07.002

Wisutwattanasak, P., Jomnonkwao, S., Se, C., & Ratanavaraha, V. (2022). Influence of psychological perspectives and demographics on drivers' valuation of road accidents: a combination of confirmatory factor analysis and preference heterogeneity model. *Behavioral Sciences*, *12*(9), Article 336. https://doi.org/10.3390/bs12090336

Worachairungreung, M., Ninsawat, S., Witayangkurn, A., & Dailey, M. N. (2021). Identification of road traffic injury risk prone area using environmental factors by machine learning classification in Nonthaburi, Thailand. *Sustainability*, *13*(7), Article 3907. https://doi.org/10.3390/su13073907

WHO. (2018). *Global status report on alcohol and health 2018*. Retrieved December 20, 2023, from https://www.who.int/publications/i/item/9789241565684

Xu, J., Yao, L., Li, L., & Chen, Y. (2014). *Sampling based multi-agent joint learning for association rule mining* [Conference presentation]. The 2014 international conference on Autonomous agents and multi-agent systems. ACM. https://dl.acm.org/doi/abs/10.5555/2615731.2617527

Yilmaz, A. E., and Demirhan, H. (2023). Weighted kappa measures for ordinal multi-class classification performance. *Applied Soft Computing, 134*, Article 110020. https://doi.org/10.1016/j.asoc.2023.110020

Zha, D., Lai, K. H., Tan, Q., Ding, S., Zou, N., & Hu, X. B. (2022). *Towards automated imbalanced learning with deep hierarchical reinforcement learning* [Conference presentation]. The 31st ACM International Conference on Information & Knowledge Management. October 17-21, 2022, ACM, Atlanta GA USA. https://doi.org/10.1145/3511808.3557474