OPTUNA-OPTIMIZED MACHINE LEARNING FOR TRAFFIC ACCIDENT SEVERITY PREDICTION IN JORDAN UNDER CLASS IMBALANCE CONDITIONS

Maen Qaseem Ghadi

Al-Zaytoonah University of Jordan, Department of Civil and Infrastructure Engineering, Amman, Jordan

*E-mail of corresponding author: m.ghadi@zuj.edu.jo

Maen Qaseem Ghadi 0000-0002-3490-9904

Resume

A machine learning framework for predicting the traffic accident severity under class imbalance conditions is predented in Amman, Jordan. The methodology began with a preprocessing pipeline consisting of IQR-based (Interquartile range) outlier removal, label encoding, feature scaling, and Synthetic Minority Over-sampling Technique (SMOTE). Sixteen model configurations were considered, comprising four ensemble learning algorithms (XGBoost, LightGBM, Random Forest, CatBoost) under oversampling, with Optuna-based Bayesian hyperparameter optimisation. The result indicated that the class balancing using SMOTE had a greater influence on the predictive performance than hyperparameter optimisation, specifically for CatBoost model. Feature importance showed that vehicles involved, collision type, and speed are the strongest predictors of accident severity.

Article info

Received 19 April 2026
Accepted 15 June 2026
Online 3 July 2026

Keywords:

hyperparameter tuning
SMOTE class imbalance
traffic accident severity
machine learning
Optuna

Available online: https://doi.org/10.26552/com.C.2026.039

ISSN 1335-4205 (print version)
ISSN 2585-7878 (online version)

1 Introduction

Road traffic accidents are among the most pressing issues for researchers and decision-makers, from both a health and economic perspective. In Jordan, this serious issue causes the death for about two people per day and burdens its economy with 2-3 percent of Gross Domestic Product (GDP) [1]. In the past 20 years alone, Jordan has had to deal with a large number of migrants, approximately 3.7 million, from neighboring countries that experienced civil unrest. This constitutes about 33 percent of its total population [2]. What burdens the road networks more in Jordan is the steady increase in the number of private car owners during the same period, rising from one car for every 58 people to one car for every six people. These factors have been linked to an increase in the number and severity of accidents, especially in urban Jordan. This demographic and cultural makeup has not only affected Jordan's economy but also the nature and rate of road accidents [3]. Therefore, it was necessary to re-examine the causes of accidents in Jordan and to predict them more accurately.

Traffic accident severity has been widely investigated using traditional and machine learning (ML) techniques. Traditional techniques have relied on simple explainable algorithms such as regression and decision analysis. In [4] a multilevel probit regression was applied to investigate changes in accident counts across different roadway types. The multilevel approach allows to detect the micro and macro-level characteristics of accident scenes within distinct roadway categories. In [5-6] a decision tree was applied to investigate the environmental characteristics of traffic accidents spatially. Moreover, several direct and indirect features related to road accidents were examined. Authors of [7-9] investigated road geometrics and environmental characteristics. The economic and demographic factors related to the accidents caught attention from [3, 10-11]. Even the effects of climate change, tourism activities, and other indirect factors have been researched by [12-14]. However, accident data is heterogeneous in nature and has complex, nonlinear structures, which are not well captured using the traditional models. Therefore, the machine learning techniques have been recently used to identify patterns in accident data and provide better predictive capabilities.

Modern machine learning (ML) techniques have become widespread across many fields. ML is a computational approach in which algorithms analyse data structures and discover patterns or decisions without being explicitly explained. ML techniques have recently been used in road safety research to identify patterns in accident data and provide better predictive capabilities. Ensemble techniques yield a better prediction of outcomes in injury severity studies, especially in cases of multi-class outcomes and large heterogeneity in the data. Recent studies have shown that ensemble methods, such as Random Forest and Gradient Boosting models, achieve better performance than traditional techniques [15-16]. In [17], a predictive random forest model was constructed using 10-fold cross-validation to predict traffic accident severity in Abu Dhabi. The author compared the random forest technique with the traditional Probit method, demonstrating the superiority of the random forest. Authors of [18] developed a two-stage gradient-boosting predictive model to analyze traffic accidents requiring firefighter intervention in Portugal. Support vector machine has also been used in safety research due to its robustness against overfitting, its ability to handle nonlinear data, and its efficiency with limited datasets [19-20].

Ensemble learning techniques have demonstrated superior performance in traffic accident severity prediction due to their ability to capture complex nonlinear relationships and interactions among explanatory variables. Among these methods, eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Random Forest (RF), and Categorical Boosting (CatBoost) are widely recognized as state-of-the-art algorithms in classification tasks [21-22]. Numerous studies have reported that XGBoost achieves high accuracy, F1-score, and balanced performance across severity categories in comparison to logistic regression and decision trees, in use cases with large highway and arterial datasets [23-25]. Furthermore, recent research has shown that combining ensemble learning algorithms with class-balancing techniques can significantly improve the prediction of minority severity classes. For instance, authors of [26] combined ensemble learning models with the Synthetic Minority Oversampling Technique (SMOTE) to predict accident severity in China. According to their findings, generated models outperform baseline algorithms, including BP Neural Network and CNN.

To ensure high accuracy and better generalization of ensemble models, hyperparameter tuning is used to find the optimal configuration of settings. Hyperparameter tuning is crucial to reveal the full potential of complex ML models. Recent accident severity studies have begun to systematically explore grid search and cross-validation to optimise parameters such as learning rate, maximum tree depth, and number of estimators [27-28]. These hyperparameter search techniques are computationally expensive, especially when combined with high-dimensional feature spaces and multiple models. Optuna is a modern optimisation library used to automate the search for optimal hyperparameters in ML. Unlike the other tuning methods (i.e., Grid Search and Random Search), it utilises Bayesian optimisation to learn from previous trials and find the best parameter combinations.

Within this emerging body of work, the proposed methodology contributes by developing a robust framework for evaluating the integration of the SMOTE oversampling technique with Optuna-based optimization to improve the accuracy of machine learning-based prediction models for traffic accident severity in Jordan. This work offers a practically-oriented methodology that can support targeted safety interventions, resource allocation, and real-time decision-support systems in Jordan.

2 Data description

In this case study, the traffic accidents for the capital of Jordan, Amman, were collected from the Jordanian Traffic Institute for 2022, [29]. The institute provided information on 14,972 fatalities, serious, moderate, and minor injury accidents. Injury severity was chosen as the target variable after combining the data on fatalities and serious injuries. Fatalities and serious injuries were merged into a single high-severity class, consistent with international practice such as the U.S. KABCO scale, the ITF/EuroStat/UNECE definitions adopted in Europe, and recent scholarly studies [30-32]. Statistically, this grouping is widely used in traffic safety literature to address the relatively low fatality rate and avoid model bias toward high-frequency classes.

The other explanatory variables presented are divided into numerical and categorical variables. The relevant numerical variables include speed, number of vehicles involved in the accident, and drivers' age, while seven categorical variables are considered: accident type (vehicle-vehicle, vehicle-pedestrian, roll-over), median separated (separated, not-separated, one-way road), surface condition (dry, wet, snowy), light conditions (day, night with light, night without light), weather conditions (clear, snow/rain, fog), vehicle type (passenger car, bus/heavy vehicle, motor cycle), and gender (male, female). Description statistics for numerical and categorical variables are presented in Table 1 and Figure 1, respectively.

As shown in Table 1 and Figure 1, vehicle-vehicle accidents, dry road surface, clear weather, passenger vehicles, and males are the predominant variables within their respective categories. However, accident severity cannot be attributed to these variables without applying a precise explanatory model, as is demonstrated in the methodology, Figure 2.

Table 1 *Descriptive statistics for numerical variables*

Variable	Count	Mean	Std	Min	Max
Speed	14972	53.50	14.00	10	120
Vehicles involved	14972	2.07	1.17	1	14
Driver Age	14972	36.56	12.27	14	80

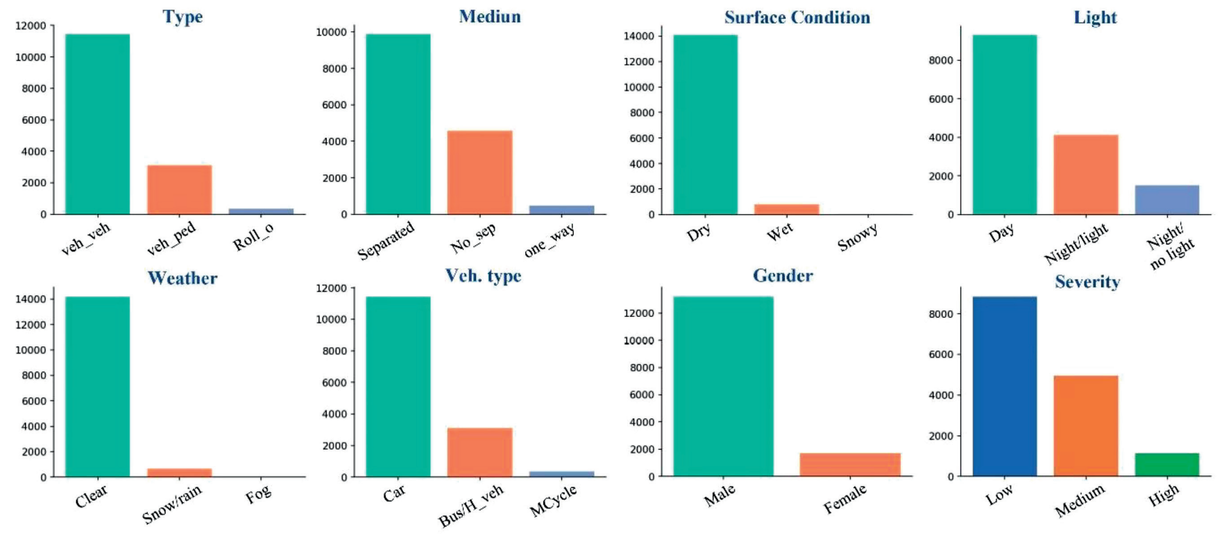


Figure 1 *Categorical features distributions*

3 Methodology

Accuracy in predicting traffic accidents is a major concern for safety research. Building a comprehensive safety prediction model is challenging due to the heterogeneity and complexity of accident-related features. An optimised ML-based model is proposed that handles class imbalance and utilises Optuna for hyperparameter tuning. In this study is utilized the benefit of eXtreme Gradient Boosting (XGBoost), LightGBM, Random Forest (RF), and CatBoost as state-of-the-art algorithms in classification tasks. Such ML models have proven their efficiency in dealing with heterogeneous and non-linear traffic accident data [26]. The effect of SMOTE, Optuna, and their combination are alternatively examined on the performance of the four ML algorithms, producing 16 predictive models. All the models are additionally optimised via Optuna Bayesian optimization with 60 trials. The adequacy of the number of trials was verified through convergence analysis by monitoring the best Macro-F1 curve for each model. All the models met this criterion between trial 35 and trial 43, and the marginal improvement in Macro-F1 after convergence was less than 0.001, which confirms the sufficiency of the selected number of trials in this study. The Python pipeline is provided in this link (https://github.com/M-ghadi/Optuna-Optimized-Ensemble-Pipeline-with-SMOTE-for-Accident-Severity-Prediction.git).

3.1 Data preprocessing

The collected data was passed through a filtering and structuring process to improve the model accuracy and ensure reliable predictions before fitting to the model. Initially, missing rows of data were removed to avoid bias and improve the prediction process, resulting in a net of 10,577 accidents. Next, Label-encoding and One-hot encoding are applied for categorical variables to meet the model requirement for numerical input. One-hot encoding creates dummy variables for each nominal attribute. However, in the cases of multiple unique and ordinal classes, label-encoding is preferred.

The Interquartile Range (IQR) is used to handle outliers. Any value outside the defined lower and upper values within the IQR is considered an outlier. The first (Q_1 = 25%) and third (Q_3 = 75%) quantiles are used to define the upper and lower limits for this study as follows:

$$Lower\ limit = Q_1 - 1.5\,(IQR),$$
$$Upper\ limit = Q_3 + 1.5\,(IQR), \tag{1}$$

where

$$IQR = Q_3 - Q_1. \tag{2}$$

Next, the normalization method is used for the dataset to overcome the scale bias of the different

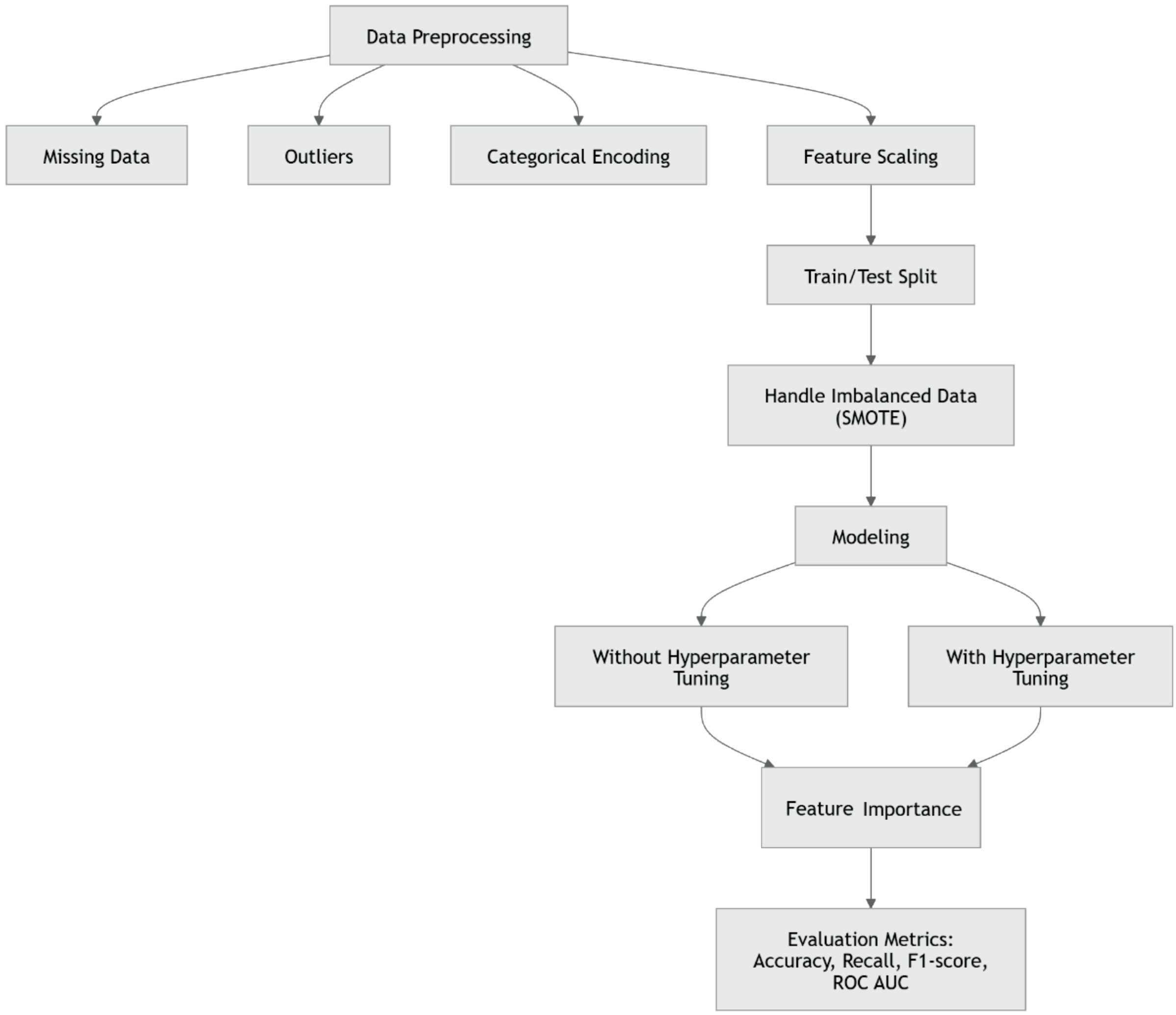


Figure 2 *Main steps of the proposed methodology*

weighted variables. The data preprocessing step ultimately resulted in 13222 rows of filtered data.

3.2 Handling the class imbalance

Class imbalance in ML models can lead to an incorrect preference for the dominant categories (majority), and a failure to detect rare and crucial cases (minority). For instance, in this case study, the low accident severity (8,861 accidents) may overshadow the importance of the serious accident severity (1,154 accidents), resulting in a biased model prediction. Synthetic Minority Over-sampling Technique (SMOTE) [33] is a common resampling technique that is adapted in this work. Instead of duplicating samples, SMOTE creates synthetic new samples by interpolating between the existing minority class samples and their nearest neighbors, improving model generalisation. Oversampling also reduces the risk of overfitting by avoiding the production of synthetic points that are too similar to original samples.

Stratified sampling was used to divide the data to improve model evaluation accuracy and ensure representation of minority classes in the test set, i.e., high-risk accidents. Moreover, the SMOTE was applied exclusively to the training set, leaving the test set unchanged for more reliable performance evaluation.

3.3 Optuna hyperparameter tuning with K-Fold Cross Validation

Machine learning and ensemble learning techniques are applied, optimized, and compared. The optimization step was performed using Optuna hyperparameter tuning library. Optuna utilizes Bayesian optimization to learn from previous trials and search for the best parameter combinations automatically. Unlike conventional methods, such as GridSearchCV, Optuna adapts search spaces in real time based on past evaluations, improving efficiency and model performance. In this study, ten main hyperparameter tuning options were selected based on the applied machine learning model (i.e., max_depth, n_estimators, learning_rate, gamma, colsample_bytree, reg_alpha (L1), reg_lambda (L2), subsample, num_leaves, min_child_samples).

To ensure the fair evaluation and prevent overfitting, cross-validation was applied for the training set. The

Table 2 *Cross-Validation Results*

Folds (k)	Mean Macro-F1	Std. Dev.
3	0.6418	0.0031
4	0.6381	0.0039
5	0.6367	0.0074
6	0.6461	0.0070
7	0.6438	0.0091
8	0.6417	0.0070
9	0.6419	0.0092
10	0.6453	0.0056

cross-validation results show nearly identical Mean Macro-F1 across all the evaluated folds (i.e., 3 to 10 folds), as shown in Table 2. The 6-fold configuration achieves the highest Mean Macro-F1 (0.6461) with moderate stability (Std. Dev. = 0.0070), indicating a good balance between accuracy and consistency. Thus, six folds were selected to improve the computational efficiency and help avoid the risk of overfitting.

3.4 Feature importance

The feature importance is typically assessed using metrics such as gain and Gini significance. Gain measures how much the model's accuracy improves when a particular feature is used in the segmentation process, while Gini significance (or segmentation significance) calculates how often a feature is used across all trees. These mechanisms enable the model to prioritise variables that contribute most to reducing prediction error. Benefits include improved model interpretability, feature selection, and reduced allocation overshoot by focusing on the most relevant inputs. In traffic accident data analysis, feature significance helps to identify critical features that significantly influence accident severity.

3.5 Evaluation metrics

In this study, multiple evaluation measures, including weighted accuracy, Macro-F1, recall, F1-score, and area under the curve (AUC), were used to ensure a comprehensive assessment of model performance. Accuracy measures the overall correctness of the model, Precision focuses on the accuracy of positive predictions (minimising false positives), Recall measures the coverage of actual positive results (minimising false negatives), F1-Score balances precision and recall, and AUC represents performance across all classification thresholds. Macro-F1 is the unweighted average of the F1-scores calculated separately for each class. It is particularly suitable for imbalanced datasets as it evaluates the performance of each class individually, regardless of its proportion. These metrics are extremely useful when using the ML algorithm with hyperparameter tuning and SMOTE, as they guide the optimisation process towards the robust and generalizable performance. Their integrated application ensures reliable prediction of traffic accident severity, even in cases of class imbalances.

4 Results and discussion

The proposed methodology (Figure 2) is applied to study the severity of traffic accidents in Jordan. After processing the original data, class imbalances are addressed using the SMOTE technique for training data with stratified sampling. Figure 3 presents the distribution of accident severity classes before and after applying the SMOTE. As shown in Figure 3, the distribution of classes before the application of SMOTE was uneven, with 60% of the samples being for the low-risk accidents and only 7.1% for the high-risk accidents. Thus, the selected oversampling technique is applied to generate synthetic samples for low and medium-risk accident classes to balance the high-risk accident class. The uniform distribution after oversampling helps to reduce the model bias and enhance learning.

4.1 Model construction and evaluation

Before training the processed data on ML models, hyperparameters were tuned using Optuna to facilitate the modelling process. Table 3 shows the selected parameters, the selected ranges, and the best selected values after tuning for each model.

The filtered data and selected, tuned parameters are fed into the predictive models' algorithms. Table 4 presents and compares the performance metrics of the resulting sixteen model configurations, including XGBoost, LightGBM, Random Forest, and CatBoost, under both oversampling and non-oversampling conditions, with and without Optuna-based optimization. The results of the sixteen model variants reveal that, while the hyperparameter tuning had a significant effect on the predictive performance of the models,

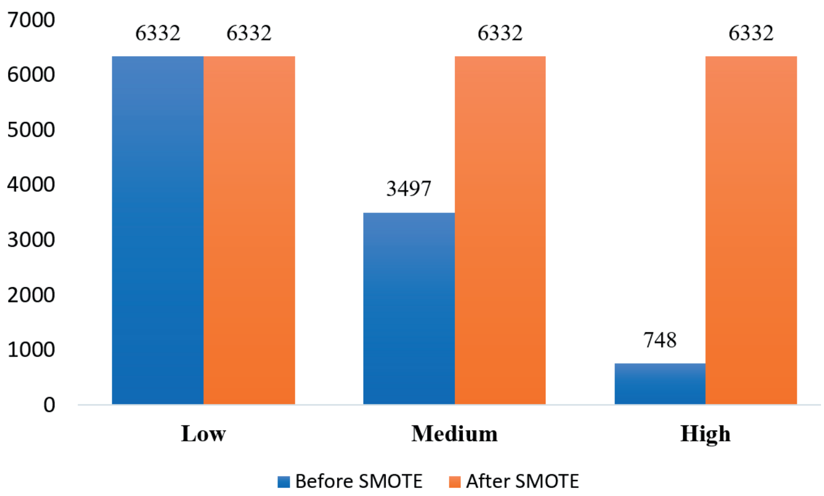


Figure 3 *Class distribution for accident severity before and after SMOTE (Train-set)*

Table 3 *Parameter ranges for Optuna hyperparameter tuning and selected best values*

Parameter	Range	Best Value			
		XGBoost	LightGBM	CatBoost	Random Forest
n_estimators	100-1000	718	698	881	665
learning_rate	0.005-0.3	0.0474	0.103	0.141	-
max_depth	3-20	12	11	8	20
gamma	0-1	0.2638	-	-	-
colsample_bytree	0.4-1	0.946	0.9193	0.9374	0.4842
reg_alpha (L1)	0-10	0.3762	0.0008	-	-
reg_lambda (Ls)	0-10	0.0016	6.6729	0.0073	-
subsample	0.5-1	0.7751	0.5027	-	-
num_leaves	20-300	-	180	-	-
min_child samples	5-100	-	56	-	-

class imbalance handling had a much larger effect. CatBoost with SMOTE and without Optuna had the highest Macro-F1 score (0.4305) among all evaluated models, followed by XGBoost with SMOTE (0.4198) and Random Forest with SMOTE (0.4130). All the models trained without SMOTE had significantly lower Macro-F1 values between 0.328 and 0.374, in comparison. This result suggests that the main issue of the accident severity dataset is the imbalance of the different severity classes and not necessarily suboptimal hyperparameters of the model.

The application of SMOTE improved the model's ability to identify minority classes, particularly high-severity accidents. For instance, the recall of the class with high severity, when non-SMOTE methods were used, ranged between 0.005-0.043, but with SMOTE application, the recall range is between 0.214-0.433. As mentioned above, synthetic minority samples are created, and severe accidents are more balanced in the training of the learning algorithm, which contributes to such improvements. The finding validates the need for class-balancing methods in traffic safety analysis, where severe collisions are, by nature, a small share of the total number of collisions, but are of greatest relevance for the analysis.

The performance of the SMOTE-based models did not increase with the inclusion of Optuna tuning in some cases and decreased slightly in the others for Macro-F1 and AUC. This behaviour can be justified by the characteristics of the dataset. The accident database consists mostly of categorical variables and a few numerical variables with only a few discrete values, so that the models that can be learned are not very complex. In such circumstances, a sophisticated hyperparameter optimisation can only provide information that was already available in the data that is predictive. On the other hand, Optuna may have gravitated to values that were close to optimising cross-validation, but not much else with regard to generalisation on the unseen test set. The differences between tuned and untuned models are quite small and are generally not considered to be practically meaningful in applied machine-learning research (typically < 0.03 Macro-F1), and they do not show any performance improvement. The wide search

***Table 4** Performance evaluation metrics for all models*

Model		Accuracy	Macro-F1	Macro-Recall	Recall	F1-Score	ROC-AUC(macro)
No SMOTE No Optuna	XGBoost	0.586	0.328	0.354	0.354	0.328	0.599
	LightGBM	0.587	0.361	0.372	0.372	0.361	0.630
	Random Forest	0.601	0.346	0.368	0.368	0.346	0.655
	CatBoost	0.602	0.366	0.379	0.379	0.366	0.652
Only Optuna	XGBoost	0.593	0.337	0.361	0.361	0.337	0.590
	LightGBM	0.555	0.374	0.374	0.374	0.374	0.592
	Random Forest	0.565	0.351	0.360	0.360	0.351	0.610
	CatBoost	0.576	0.362	0.370	0.370	0.362	0.605
Only SMOTE	XGBoost	0.563	0.420	0.424	0.424	0.420	0.622
	LightGBM	0.570	0.402	0.400	0.400	0.402	0.620
	Random Forest	0.530	0.413	0.460	0.460	0.413	0.629
	CatBoost	0.555	0.431	0.450	0.450	0.431	0.636
SMOTE + Optuna	XGBoost	0.522	0.376	0.378	0.378	0.376	0.588
	LightGBM	0.533	0.377	0.376	0.376	0.377	0.592
	Random Forest	0.512	0.386	0.394	0.394	0.386	0.593
	CatBoost	0.545	0.394	0.397	0.397	0.394	0.608

space for hyperparameters may have enabled Optuna to identify highly complex configurations that reduced training bias on SMOTE-balanced data, but increased training overfitting and degraded generalisation. However, constraining the search to narrower ranges may improve the compatibility of optimized configurations, but it may exclude truly optimal areas of the search space.

Appendix 1 shows the confusion matrices generated by the base models, the SMOTE-only models, and the SMOTE+Optuna models. The baseline models show severe majority-class biases. This is further illustrated by the number of correct predictions for the high-severity class, which ranged from only 1 to 8 out of 187 cases. The results confirm that unbalanced training data renders all four classifiers unable to detect the most serious accident category. SMOTE was the single most decisive intervention, increasing High-severity recall by an order of magnitude across all models (e.g., Random Forest: 1 to 82, CatBoost: 8 to 60). In contrast, applying Optuna surprisingly showed an adverse impact on model performance. For instance, the high-risk correct predictions declined in every ML model (e.g., CatBoost: 60 to 18, Random Forest: 82 to 37). The decline may be attributed to the optimiser overfitting tuned parameters to the synthetic training distribution rather than the true data structure. CatBoost with SMOTE achieved the highest Macro-F1 score (0.4335) and the most balanced confusion matrices across all models tested. With a balance of 1,147 Low, 269 Medium, and 60 High predictions, it is the model that is most likely to achieve the best result. The boosted mechanism and categorical nature of the algorithm show a more stable and generalised response to the resampled data compared to XGBoost, LightGBM, or Random Forest.

Although CatBoost+SMOTE achieved the highest Macro-F1 score, the performance differences among the leading SMOTE-based ensemble models were relatively small. Overall, the findings indicate that balancing the severity classes is more critical than extensive hyperparameter tuning for accident severity data in Jordan.

4.2 Feature importance results

The importance of class variables was investigated to complement the data analysis. Figure 4 illustrates the relative importance of class features on accident severity. The number of vehicles involved in accidents stands out as the most important factor for accident severity. This is logical, as increasing the number of vehicles involved increases the likelihood of the accident worsening. Collision (vehicle-vehicle) accidents come in second importance. This can be attributed to the high percentage of this type of accident (approximately 76% of all injury accidents) and its inclusion of the largest percentage of severe accidents (i.e., 60%). As is usually the case, the effect of speed on the accident severity is also significant and comes next in importance. The effects of the remaining variables can be traced from the Figure 4, noting the low impact of lighting and road surface conditions on accident severity in Jordan.

However, the applied methodology, combining engineering expertise with the data-driven validation, strengthens confidence in the model's interpretability and predictive reliability.

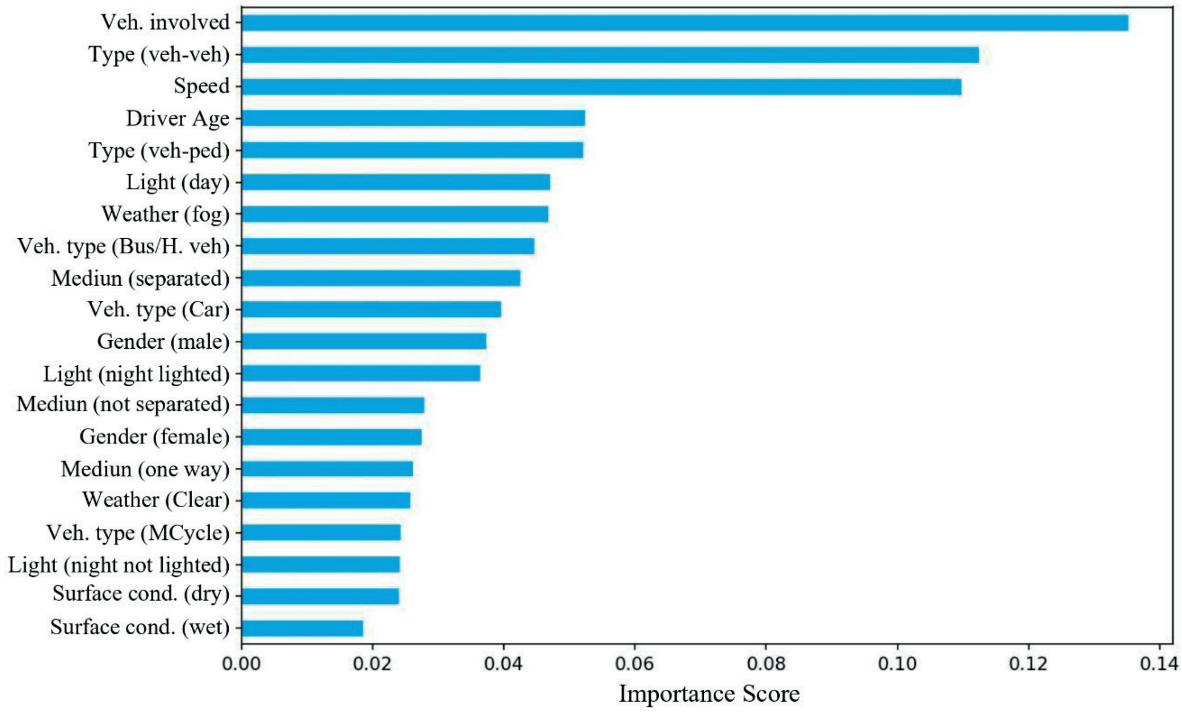


Figure 4 *Top-20 feature importances with Optuna-tuned XGBoost*

5 Conclusion

Accuracy in predicting traffic accidents continues to be a major focus for researchers due to the heterogeneous and diverse nature of traffic accident data. An optimised machine learning-based model was applied in this study, which handles class imbalance and utilises Optuna for hyperparameter tuning to predict road accident severity in Amman, the capital of Jordan. The data included 14,972 fatalities and injury accidents with ten variables capturing accident, vehicle, driver, road, and environmental characteristics.

In the methodology, sixteen model configurations were evaluated, encompassing XGBoost, LightGBM, Random Forest, and CatBoost, under both oversampling and non-oversampling conditions, with and without Optuna-based optimization. Data was preprocessed using Interquartile Range (IQR)-based outlier removal, Label Encoding, Standard Scaling and Synthetic Minority Over-sampling Technique (SMOTE) in the preprocessing stage. The IQR is a powerful non-parametric method, which uses upper and lower quantiles to remove outlier data. Label Encoding is chosen to converts categorical features into unique numerical values without increasing the dimensions of the data (i.e., One-hot encoding). The class imbalance-SMOTE is used to avoid the bias of the target variable (High severity < 8% and Low severity = 59%) by synthetically propagate minority data to balance with majority data. Then, a six-fold stratified cross-validation method was performed over the training data to avoid overfitting. Moreover, 60 trials of Optuna Bayesian optimisation were also performed for all the models.

Comparative evaluation of the sixteen model configurations revealed that class balancing had a significantly influence on the predictive performance than hyperparameter optimisation, when the sixteen models were compared. The best model was CatBoost with SMOTE, which has a Macro-F1 score of 0.431, whereas the Macro-F1 score of CatBoost without SMOTE was 0.328, which is around 31% improvement. Likewise, Macro-F1 for XGBoost has risen from 0.355 to 0.420, and Macro-F1 for Random Forest has increased from 0.350 to 0.413 after applying SMOTE. The findings indicate the need to tackle class imbalance to ensure good identification of severe accidents in the Jordanian data set. The effect of Optuna optimisation was comparatively small, on the other hand. In the models examined, the variations between the tuned and untuned models were typically small (0.02-0.03 Macro-F1), with limited practical significance. This suggests that the model performance was constrained more by the available accident-related information than by setting the hyperparameters of the model. The limited contribution of Optuna can be attributed to the characteristics of the dataset. Although Bayesian hyperparameter optimization efficiently identified suitable parameter combinations, the available explanatory variables contain predominantly categorical information and only a few numerical variables with limited variability. Moreover, the investigation revealed that vehicles involved in accidents, collisions (between vehicles), and speed are the most significant factors affecting the severity of accidents in Jordan.

The integrated approach is a powerful and reproducible approach for predicting accident severity in multi-class scenarios. The findings show that class balancing with SMOTE has a greater impact on predictive performance than hyperparameter optimisation alone. Future studies should empirically evaluate reduced and adapted hyperparameter search spaces on similar unbalanced tabular datasets. This may help to determine whether stricter, more order-oriented boundaries consistently restore the missing generalization performance when broad spaces are combined with synthetic oversampling. Moreover, collecting more comprehensive and accurate data in Jordan would certainly improve the accuracy of the outputs and will enhance the applicability of machine learning models for traffic safety decision-making and resource allocation.

Acknowledgment

The authors received no financial support for the research, authorship and/or publication of this article.

Conflicts of interest

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

References

[1] GHADI, M., TOROK, A., TANCZOS, K. Study of the economic cost of road accidents in Jordan. *Periodica Polytechnica Transportation Engineering* [online]. 2018, **46**(3), p. 129-134. ISSN 0303-7800, eISSN 1587-3811. Available from: https://doi.org/10.3311/PPtr.10392

[2] KHAWALDAH, H., ALZBOUN, N. Socio-economic and environmental impacts of Syrian Refugees in Jordan: a Jordanians' perspective. *Heliyon* [online]. 2022, **8**(8), e10005. eISSN 2405-8440. Available from: https://doi.org/10.1016/j.heliyon.2022.e10005

[3] GHADI, M. Investigating the asymmetric effects of economic and demographic factors on traffic accident severity - evidence from a nonlinear ARDL approach. *Promet - Traffic and Transportation* [online]. 2025, **37**(3), p. 648-664. ISSN 0353-5320, eISSN 1848-4069. Available from: https://doi.org/10.7307/ptt.v37i3.814

[4] GHADI, M. Q. Multilevel analysis of road accident frequency: the impact of the road category. *International Journal of Transport Development and Integration* [online]. 2023, **7**(2), p. 123-130. ISSN 2058-8305, eISSN 2058-8313. Available from: https://doi.org/10.18280/ijtdi.070207

[5] ABELLAN, J., LOPEZ, G., DE ONA, J. Analysis of traffic accident severity using Decision Rules via Decision Trees. *Expert Systems with Applications* [online]. 2013, **40**(15), p. 6047-6054. ISSN 0957-4174, eISSN 1873-6793. Available from: https://doi.org/10.1016/j.eswa.2013.05.027

[6] GHADI, M. Q., TOROK, A. Evaluation of the impact of spatial and environmental accident factors on severity patterns of road segments. *Periodica Polytechnica Transportation Engineering* [online]. 2020, **49**(2), p. 146-155. ISSN 0303-7800, eISSN 1587-3811. Available from: https://doi.org/10.3311/PPtr.14692

[7] JIMA, D., SIPOS, T. The impact of road geometric formation on traffic crash and its severity level. *Sustainability* [online]. 2022, **14**(14), 8475. eISSN 2071-1050. Available from: https://doi.org/10.3390/su14148475

[8] ISLAM, M. H., TEIK HUA, L., HAMID, H., AZARKERDAR, A. Relationship of accident rates and road geometric design. *IOP Conference Series: Earth and Environmental Science* [online]. 2019, **357**, 012040. ISSN 1755-1315. Available from: https://doi.org/10.1088/1755-1315/357/1/012040

[9] GORZELANCZYK, P., GROBELNY, P. An estimate of the number of accidents on polish highways based on the kind of road. *Communications - Scientific Letters of the University of Zilina* [online]. 2024, **26**(1), p. F13-22. ISSN 1335-4205, eISSN 2585-7878. Available from: https://doi.org/10.26552/com.C.2024.012

[10] WU, W., JIANG, S., LIU, R., JIN, W., MA, C. Economic development, demographic characteristics, road network and traffic accidents in Zhongshan, China: gradient boosting decision tree model. *Transportmetrica A: Transport Science* [online]. 2020, **16**(3), p. 359-387. ISSN 2324-9935, eISSN 2324-9943. Available from: https://doi.org/10.1080/23249935.2020.1711543

[11] SOHAEE, N., BOHLULI, S. Nonlinear analysis of the effects of socioeconomic, demographic, and technological factors on the number of fatal traffic accidents. *Safety* [online]. 2024, **10**(1), 11. eISSN 2313-576X. Available from: https://doi.org/10.3390/safety10010011

[12] GHADI, M. Q. Investigating the nexus between traffic accidents and tourism in Jordan: evidence from the ARDL approach. *Journal of Engineering, Project, and Production Management* [online]. 2025, **15**(3), 0021. ISSN 2221-6529, eISSN 2223-8379. Available from: https://doi.org/10.32738/JEPPM-2025-0021

[13] GHADI, M. Q. Investigating the impact of climate change on traffic accidents in Jordan. *Sustainability* [online]. 2025, **17**(5), 2161. eISSN 2071-1050. Available from: https://doi.org/10.3390/su17052161

[14] AL-MASAEID, H. R., BAZLAMIT, S. M., AL-ZEDANEEN, A. E., AL-MOFLEH, H. Impact of fasting on traffic accidents. 2020.

[15] BHUIYAN, H., ARA, J., HASIB, K. M., SOURAV, M. I. H., KARIM, F. B., SIK-LANYI, C., GOVERNATORI, G., RAKOTONIRAINY, A., YASMIN, S. Crash severity analysis and risk factors identification based on an alternate data source: a case study of developing country. *Sci Rep* [online]. 2022, **12**, 21243. eISSN 2045-2322. Available from: https://doi.org/10.1038/s41598-022-25361-5

[16] SANTOS, K., DIAS, J. P., AMADO, C. A literature review of machine learning algorithms for crash injury severity prediction. *Journal of Safety Research* [online]. 2022, **80**, p. 254-269. ISSN 0022-4375, eISSN 1879-1247. Available from: https://doi.org/10.1016/j.jsr.2021.12.007

[17] STANTON, N. (Ed.). Advances in human aspects of transportation [online]. Vol. 786. Cham: Springer International Publishing, 2019. ISSN 2194-5357, eISSN 2194-5365, ISBN 978-3-319-93884-4. Available from: https://doi.org/10.1007/978-3-319-93885-1

[18] ALPALHAO, N., SARMENTO, P., JARDIM, B., DE CASTRO NETO, M. Assessing the risk of traffic accidents in Lisbon using a gradient boosting algorithm with a hybrid classification/regression approach. *Transportation Research Interdisciplinary Perspectives* [online]. 2025, **32**, 101495. eISSN 2590-1982. Available from: https://doi.org/10.1016/j.trip.2025.101495

[19] YU, B., WANG, Y. T., YAO, J. B., WANG, J. Y. A comparison of the performance of ann and svm for the prediction of traffic accident duration. *Neural Network World* [online]. 2016, **26**, p. 271-287. ISSN 1210-0552, eISSN 2336-4335. Available from: https://doi.org/10.14311/NNW.2016.26.015

[20] FARHART, Z., KAROUNI, A., DAYA, B., CHAUVET, P., HMADEH, N. Traffic accidents severity prediction using support vector machine models. *International Journal of Innovative Technology and Exploring Engineering* [online]. 2020, **9**(7), p. 1345-1350. eISSN 2278-3075. Available from: https://doi.org/10.35940/ijitee.F4393.059720

[21] KHALEEL, A., FARFOURA, M. E., ALQABANI, H., BARHAM, K. Y. A Hybrid learning neuro-shallow architecture obesity detection for Syrian refugees in Jordan. In: 2025 1st International Conference on Computational Intelligence Approaches and Applications ICCIAA: proceedings [online]. IEEE. 2025. ISBN 979-8-3315-2366-4, eISBN 979-8-3315-2365-7, p. 1-6. Available from: https://doi.org/10.1109/ICCIAA65327.2025.11013717

[22] ABDULJAWAD, M. Comparative analysis of machine learning algorithms for intrusion detection in IoT networks. *International Journal of Advances in Soft Computing and its Applications* [online]. 2025, **17**(3), p. 259-274. ISSN 2710-1274, eISSN 2074-8523. Available from: https://doi.org/10.15849/IJASCA.251130.15

[23] JAMAL, A., ZAHID, M., TAUHIDUR RAHMAN, M., AL-AHMADI, H. M., ALMOSHAOGEH, M., FAROOQ, D., AHMAD, M. Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study. *International Journal of Injury Control and Safety Promotion* [online]. 2021, **28**(4), p. 408-427. ISSN 1745-7300, eISSN 1745-7319. Available from: https://doi.org/10.1080/17457300.2021.1928233

[24] LIANG, Y., YUAN, H., WANG, Z., WAN, Z., LIU, T., WU, B., CHEN, S., TANG, X. Nonlinear effects of traffic statuses and road geometries on highway traffic accident severity: a machine learning approach. *PLoS One* [online]. 2024, **19**, e0314133. eISSN 1932-6203. Available from: https://doi.org/10.1371/journal.pone.0314133

[25] BOO, Y., CHOI, Y. Comparison of mortality prediction models for road traffic accidents: an ensemble technique for imbalanced data. *BMC Public Health* [online]. 2022, **22**, 1476. eISSN 1471-2458. Available from: https://doi.org/10.1186/s12889-022-13719-3

[26] CHEN, F., LIU, X. Q., YANG, J. J., LIU, X. K., MA, J. H., CHEN, J., XIAO, H. Y. Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model. *Sci Rep* [online]. 2025, **15**, 25729. eISSN 2045-2322. Available from: https://doi.org/10.1038/s41598-025-00797-7

[27] SUDHESHNA, P., ALJAWAHRY, H. M. M., PHADKE, R., SIVASUBRAMANIAN, S., KANTH TATIRAJU V. R. An effective prediction modelling of traffic accidents based on categorical boosting with grid search. In: 2024 International Conference on Integrated Intelligence and Communication Systems ICIICS: proceedings [online]. IEEE. 2024. ISBN 979-8-3315-0497-7, eISBN 979-8-3315-0496-0, p. 1-5. Available from: https://doi.org/10.1109/ICIICS63763.2024.10859490

[28] EL FEROUALI, S., ELAMRANI ABOU ELASSAD, Z., QASSIMI, S., ABDALI, A. From baseline to best practice: an advanced feature selection, feature resampling and grid search techniques to improve injury severity prediction. *Applied Artificial Intelligence* [online]. 2025, **39**(1). ISSN 0883-9514, eISSN 1087-6545. Available from: https://doi.org/10.1080/08839514.2025.2452675

[29] Traffic accidents in Jordan - Jordan Traffic Institute [online] [accessed 2024-08-02]. 2024. Available from: https://www.psd.gov.jo/ar-jo

[30] AWAD, M., MOUSSA, G. S., WAHABALLA, A. M., YOUNES, H. A state-of-the-art review of injury severity analysis in traffic crashes: toward a generalized modeling framework. *Innovative Infrastructure Solutions* [online]. 2026, **11**, 19. ISSN 2364-4176, eISSN 2364-4184. Available from: https://doi.org/10.1007/s41062-025-02409-9

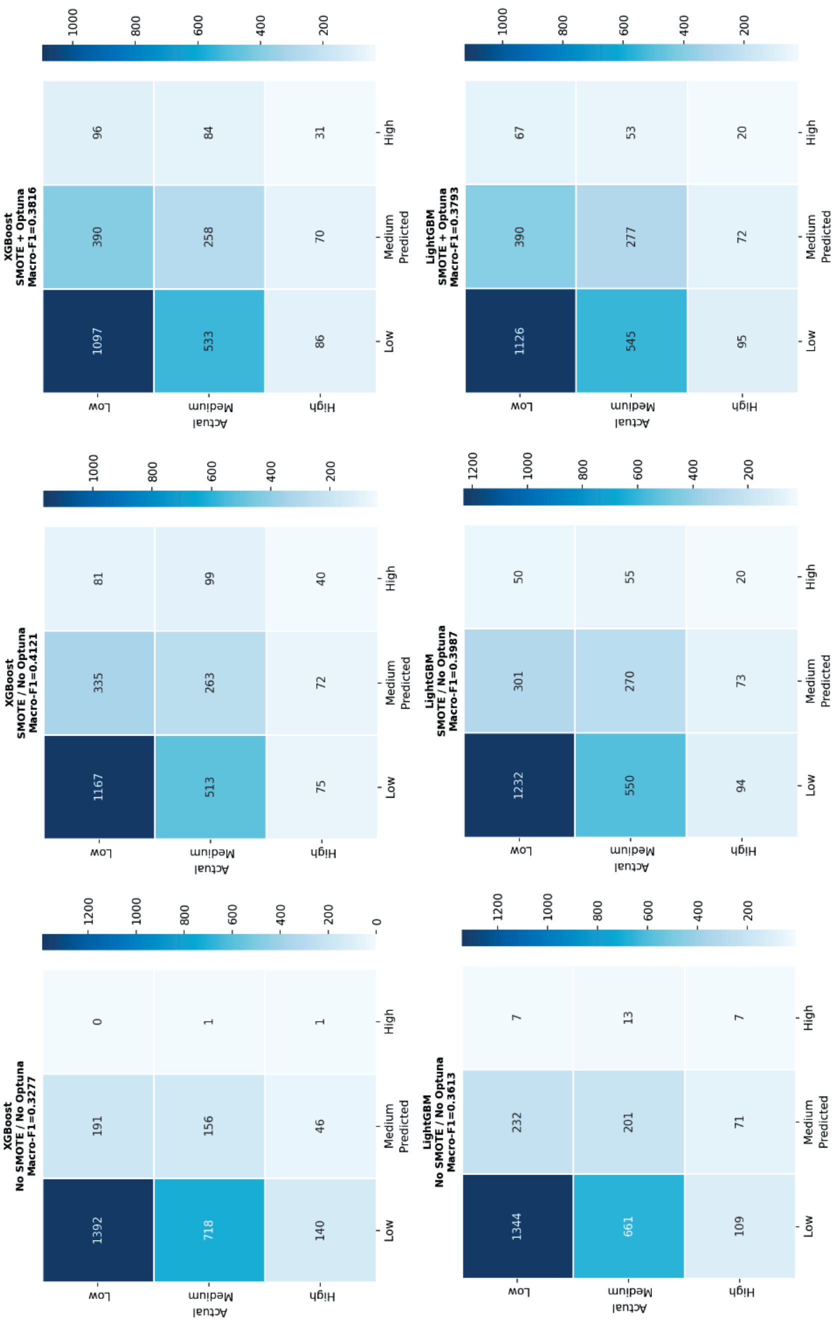
[31] CHAKRABORTY, M., GATES, T. J., SINHA, S. Causal analysis and classification of traffic crash injury severity using machine learning algorithms. *Data Science for Transportation* [online]. 2023, **5**, 12. ISSN 2948-135X, eISSN 2948-1368. Available from: https://doi.org/10.1007/s42421-023-00076-9

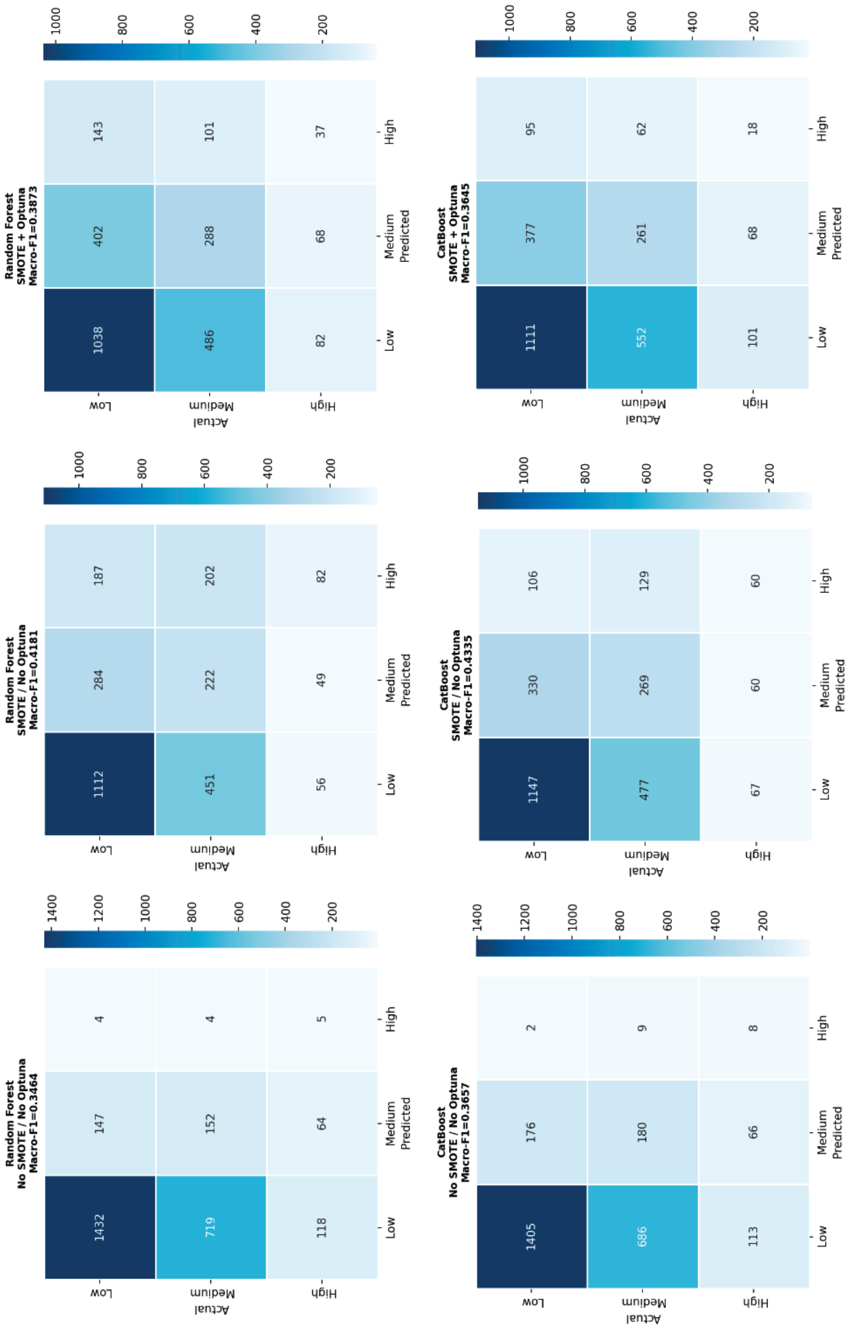
[32] AARTS, L., COMMANDEUR, J. J. F., WELSH, R., NIESEN, S., LERNER, M., HOMAS, P., BOS, N., DAVIDSE, R. J. Study on serious road traffic injuries in the EU. 2016.

[33] CHAWLA, N. V., BOWYER, K. W., HALL, L. O., KEGELMEYER, W. P. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* [online]. 2002, **16**, p. 321-357. ISSN 1076-9757. Available from: https://doi.org/10.1613/jair.953

Appendix 1

Confusion matrices for XGBoost, LightGBM, Random Forest, and CatBoost across three training conditions (No SMOTE/No Optuna, SMOTE/No Optuna, and SMOTE+Optuna), for three-class road accident severities (Low, Medium, High).




Random Forest
No SMOTE / No Optuna
Macro-F1=0.3464
Random Forest
SMOTE / No Optuna
Macro-F1=0.4181
Random Forest
SMOTE + Optuna
Macro-F1=0.3873
CatBoost
No SMOTE / No Optuna
Macro-F1=0.3657
CatBoost
SMOTE / No Optuna
Macro-F1=0.4335
CatBoost
SMOTE + Optuna
Macro-F1=0.3645
Actual
Predicted
Low
Medium
High