The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 26, Pages 357-365

IConTES 2023: International Conference on Technology, Engineering and Science

Modeling Crashes Severity Using Ensemble Techniques

Taqwa Alhadidi
Al-Ahliyya Amman University

Mohammed Elhenawey
Queensland University of Science and Technology

Abstract: Traffic crashes are modelled using different techniques and contributing factors. In this work, several ensemble machine learning algorithms were used to model crash severity at urban roundabouts using data from 15 roundabouts in Jordan. The original dataset covers four years, from 2017 to 2021. A total of 15 variables were collected and used in this work. Results indicated that ten variables are important. The various models show their ability to classify traffic crash severity with a high overall accuracy range from 96% to 98%. Results indicated that driver fault and age are the most significant contributing factors for crash severity.

Keywords: Machine learning, K-nearest neighborhood, Support vector machine, Safety, Driver age, driver fault.

Introduction

The World Health Organization (WHO) aims to reduce traffic crash deaths and injuries by 2030 significantly. According to WHO, approximately 1.3 million people die yearly from traffic crashes; more than 93% are from low- and middle-income countries. Even though low- and middle-income countries have 60% of world vehicles. Non-fatal traffic crashes lead to suffering around 35 million people from different types of injuries, which may result in disability(WHO, 2022). According to Jordan Traffic Institute (JTI), in the last two years, Jordan witnessed 170,000 and 160,000 road crashes, with 562 and 589 fatalities in 2021 and 2022, respectively. Jordanian traffic crashes cost increases from $415 million in 2020to $454 million in 2022 (JTI, 2022). Come to crisp traffic crashes; several studies have been conducted to investigate the different contributing factors to crash occurrence in Jordan. According to JTI, about 97% of the total accidents that occurred in Jordan in 2021 happened due to various driver faults, including not taking proper pre-cautious actions (i.e., not using seatbelt, distracted drivers), going on wrong way direction, and violation of driving priorities (JTI, 2022). In recent years, researchers have paid close attention to traffic accident analysis to identify the elements that substantially impact traffic accidents. However, most research methodology is based on aggregated tabular data analysis using different statistical techniques including multiple linear regression, curve estimation, spatial analysis, or machine learning (Al-Mistarehi et al., 2022; Alomari et al., 2019; Edries & Alomari, 2022; Hazaymeh et al., 2022; R. Mujalli, 2018; R. O. Mujalli et al., 2017, 2023) Although the majority of traffic accidents occur in urban areas [1]; however, most of studies have been done at rural areas (Al-Rousan et al., 2021; R. Mujalli, 2018; Peng et al., 2019) . Several factors were identified as significant factors affecting crashes severity including lightning, roadway surface, holiday, driving speed, and roadway geometry (Almannaa et al., 2023; Yahaya et al., 2021) .Although roundabouts are a type of intersection that aims to improve traffic safety by converting the crossed movement to circular movement, traffic crashes at roundabouts still occur but with the least severity compared to other roadway elements (Hariri Asli, 2022; Mamlouk & Souliman, 2019; Polders et al., 2015; Qawasmeh et al., 2023). Yet, none of these studies consider driver age and driver gender in the severity of traffic crashes at roundabout. In this work, we focus on investigating and analyzing traffic crashes at several roundabouts. In this study we used several machine learning algorithms to identify the most important

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

factors that affect the occurrence of traffic crash occurrence. Subsequently, various machine learning algorithms were implemented and tested their performance to classify different crash severity types. The overall performance for each model was reported using different evaluation metrics discussed later in this paper. To sum up, the main contribution of this work is to identify and understand the contributing factors for traffic crashes at urban roundabouts in Jordan and to model traffic crashes severity at urban roundabouts considering driver age and gender.

Literature Review

Using different modelling techniques, machine learning has been done on various elements of transportation network including roadway, intersection and freeway segments. Modeling accident severity at freeways in Hebei, China was done using data from 2018. The gradient boosting algorithm was used to select the best feature for classifying traffic crashes severity based on 23 variables that were collected. After choosing the most important features, Bayesian network analysis was used to predict crash severity. Study results indicated that the gradient boosting is able to predict traffic crash severity with an accuracy of 89.05% (Yang et al., 2022). Using several machine learning algorithms to predict crash severity in Bangladesh was done in 2019, the researchers used Decision Tree, K-Nearest Neighbors (KNN), Naïve Bayes and AdaBoost these four supervised learning techniques, to classify the severity of accidents. Results indicated that the accuracy of model increases as the number of classes increase, also results indicated that the adaboost and Naïve Bayes have the highest accuracy (Labib et al., 2019). Also, a study on Saudi Arabia investigated that machine learning algorithm can be implemented to classify traffic accident severity at different types of network segment. Results indicated that the various used algorithms have different accuracy rate at different roadway type. Most importantly, their study showed that holidays affect the crash severity (Almannaa et al., 2023).Several works were conducted to model traffic crashes globally ((Almamlook et al., 2019; Almannaa et al., 2023; Al-Mistarehi et al., 2022; Al-Moqri et al., 2020; Alrumaidhi & Rakha, 2022; Anderson & Hernandez, 2017; Azhar et al., 2022) Most of these studies used machine learning algorithms including K-nearest neighborhoods (KNN), support vector machine, adaptive boosting tree or other machine learning algorithms (Ahmed et al., 2021; Almamlook et al., 2019; Almannaa et al., 2023a; Al-Mistarehi et al., 2022; Al-Moqri et al., 2020). Researchers indicated that providing driver sociodemographic attributes may improve the model prediction ((Almannaa et al., 2023; Jamal et al., 2021; R. O. Mujalli et al., 2017; Rahim & Hassan, 2021)

Settings

Data

In this research, crash data from the Jordanian traffic institute (JTI) is requested for the most critical roundabouts in the capital of Jordan, Amman. The dataset involves traffic accidents from 2017 to 2021, with 30,486 crashes from 15 roundabouts collected. The collected data have 23 variables describing roadway characteristics and conditions, weather condition, vehicle, driver, and crash type with the number of casualties. To avoid any misleading results, data was carefully screened and processed based on data completeness, irrelevant or redundant variables were disregarded. As such, in this study, the original dataset was reduced to 12,971 datapoints out of 30,486 were validated. This reduction was done according to previous stat-of-art, as following:

- Any missing value in driver age was removed as driver age is a significant attribute in crash severity prediction (Mujalli, 2018). The total missing values in driver age are 3,770 points.
- Only driver at fault records were included in this work and a total of 13,201 data points were deleted using this condition.
- Driver age an error in recording for instance it is not allowed to drive before 15 years old in Jordan nor the age can be a negative value a total of 922 data points were deleted.
- Drop any duplicate data point, following the previous steps were not keep any duplicate records.

Variable names, types and statistics are shown in Table 1. Table 1 shows that the collected data is imbalanced. This imbalance reduces the prediction model accuracy, for every new observation the prediction model tends to classify the new data point to the major class. Several researchers indicated that the traffic crash dataset suffers from an imbalance problem among all the studied factors (Jiang et al., 2020a, 2020b; Mohammadpour et al., 2023)

Balancing Techniques

The commonly used sampling technique was utilized to handle the imbalance problem, in the sampling technique a balance dataset is established from the imbalance data be either add more sample to the minor classes, in this case this process is called "over-sampling", or removing datapoints from the major class this process is called "under-sampling". Various balancing techniques were suggested to balance the data including Synthetic Minority Oversampling Technique (SMOTE) and ADASYN. SMOTE creates “synthesis” from the minor class by choosing k nearest neighbors without duplicates or replacements which effectively increases the overall accuracy and ability to generalize and saves the data from overfitting. Herein, the crash data was balanced before model development (Jiang et al., 2020a).

Machine Learning

In this section, a brief description of the used algorithm is presented, followed by evaluation metrics. It starts with KNN, Support Vector Machine and Adaptive Boosting Techniques. Then, a set of evaluation metrics was defined.

K-Nearest Neighborhood (KNN)

The k Nearest Neighbors (KNN) algorithm is one of the simplest methods of non-parametric modeling techniques (Zhang et al., 2017). The concept of the KNN is that similar data points that belong to the cluster have high probability. In essence, KNN starts by finding the K nearest neighborhood of training dataset them it predicts within the major class in the k nearest neighbors. Due to its simplicity and ability to predict with less time it has been selected as one of the top algorithms in data mining (Wu et al., 2008). KNN accuracy depends on choosing the best cluster size. The optimum K was selected based on the accuracy of prediction. Afterward, the response (i.e., accident severity in our problem) is classified by considering the majority vote of the K nearest points within the class as shown in equation 1.

$$y_j^{test} = \frac{1}{R}\sum_{X_j^{train} \in R_k} y_j^{train} \quad (1)$$

In equation 1, R is the number of assigned classes based on checking the model accuracy for each value. y_j^{test} is the test observation which is assigned to class R based on the majority of class R voting after train the model using X_j^{train} as input varaibles and y_j^{train} as the response variable (Friedman et al., 1975).

Support Vector Machines (SVMs)

The SVM algorithm is a supervised learning technique that classifies data based on the difference between classes. The algorithm, as shown in Equation 2, seeks the hyperplane (i.e., splitter) with the greatest minimum distance to the training data. The SVM seeks the weight (w) with the greatest margin around the hyperplane while satisfying the two constraints (see Equations (3) and (4)) (Hsu & Lin, 2002)

$$\min_{w,b,\xi}\left(\frac{1}{2}\, w^T\, w + \, c \sum_{n=1}^{N} \xi_n\right) \qquad (1)$$

subjected to:

$$y_n(w^T \phi(X_n) + b) \geq 1 - \xi_n \;\; , n = 1, \dots \dots , N \qquad (2)$$

$$\xi_n \; \geq 0 \, , n = 1, \dots N \qquad (3)$$

Where: W is the set of parameters used to define class boundaries. C penalty parameter, ξ_n parameter to express the margin error. b intercept associated with the hyperplanes function to transform data from X space. ϕ(X_n) function to transform data from X space into Z space. y_n target value. The sum of the two terms in Equation 2 minimizes the objective function. Essentially, the first term seeks to define the gap between different classes.

Minimizing this term is equivalent to increasing the margin between classes. The second term seeks to reduce the error term multiplied by the penalty (regularization) parameter.

The penalty term is designed to deal with overfitting, while term C is designed to optimize the model's performance. Where n is the data observation's index, w is the decision boundary between classes, C is the regularization (or penalty) parameter, and jn is the margin violation error parameter. K is the number of observations x space that uses the $\phi(X_n)$ function to be converted to another space. In fact, the transformation is done to create a Z space which can be used to ease defining classes boundaries. In other hand, some functions can be used directly (i.e., Kernels) to create the transformation easier like the work in this paper. Meanwhile, equation 2 can be solved either by using the Kernels or using the $\phi(X_n)$ to transform data to Z plane. In fact, before constructing the model, the kernel type should be determined (e.g., linear, polynomial, Gaussian). One kernel may outperform the other depending on the problem. Based on data size, some practical considerations suggest using different kernels for different problems (Hsu & Lin, 2002).

Adaptive Boosting Algorithm

The adaptive boosting (AdaBoost) algorithm is a machine learning algorithm that is based on the incremental contribution concept. AdaBoost was created in response to the question of whether it was possible to combine a group of "weak" learner algorithms with low accuracy to create a learning algorithm with high accuracy. The traditional approach in machine learning prior to the introduction of AdaBoost was based on selecting the most discriminating class of features. In other words, algorithms must be thought of as a class. AdaBoost makes use of a collection of weak classifiers, each of which is trained using the same training dataset but with a different weight distribution. Each of the weak learners concentrates on the instances where the previous learner made a mistake. The output of AdaBoost is the weighted average of all weak learner outputs. AdaBoost is likely to have lower misclassification error than a sum of weak learners, and it also has a bound-on generalization error.

Evaluation Metrics

In classification problem, the output could classify the output in 4 possible values, true positive prediction (TP), true negative prediction (TN), false positive prediction (FP), and false negative prediction (FN). These different possible outcomes were used to compute the different evaluation metrics. The evaluation metrics are precision, recall, f1-score and support. These metrics are computes as following equations.

$$Precision = \frac{TP}{TP + NP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$f1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Results and Discussion

In this section, different modeling techniques results are presented and discussed.

Explanatory Data Analysis Results

After pre-processing the data according to section 3 and summarizing their values as in Table I. Results showed that more than 97.5% of the total accidents were property damage only, most of accidents occurred at a flat dry roadway surface during midday. Also, about 95% of the total accidents occurred in clear weather conditions. Interestingly, more than 75% of the accidents caused by male drivers with age less than 40 years old

Feature Engineering

This step was the first step in modeling data. It started with checking the data type, reporting the original data correlation matrix, and checking the most relevant variables in modeling the problem. The most relevant variables were selected based on using the P-value and the F-score. The selected features were chosen based on P-value and F-score greater than 0.05 and 5 respectively. According to this threshold, the most important variables were number of casualties for slight injuries, intermediate injuries, death and major injuries, number of vehicles, driver fault, lightening, hour, lanes, driver age and day of week.

Table 1.Explanatory Variable Descriptive statistics

Variable Name	Variable *Type*	*Categories*	*Mean*	*Standard Deviation*	*Percentage*
Crash type	Categorial	Collision Pedestrian Run Off Road	1.99	0.16	99.40 % 0.41% 0.19%
Day of week	Categorial	Weekday Weekend	0.23	0.48	76.98% 23.02%
Time of Day	Categorial	Morning Midday Evening Night	1.49	0.49	12.84% 39.09% 32.38% 15.69%
Season	Categorial	Summer Fall Winter Spring	1.59	1.18	26.10 % 24.71% 24.48% 24.71%
Holiday	Categorial	Yes No			2.86 % 97.14%
Driver gender	Categorial	Male Female	0.8	0.4	82% 18%
Driver Age	Continous	[13,76]	27	18	
Speed	Continous	[10,120]	48.4	9.8	
Direction and Location of Occurrence	Categorial	Divided two-lanes Undivided two-lanes One-way Inside a bus stop In the circular area	3.59	0.76	70.97 % 18.80 % 8.36 % 1.63 % 0.24 %
Alignments (Horizontal and Vertical)	Categorial	Uphill Curve Uphill Straight Downhill Curve Flat Curve Downhill	2.98	0.24	0.03% 2.35% 0.05% 96.82% 0.14% 0.61%
Vehicle Classification	Categorial	Passenger car Light truck Medium riding Heavy truck Bike	0.61	1.09	0.05% 96.82% 0.14% 0.61% 89.19%
Vehicle involved in the accident	Continous	1 2 3 ≥4	2.03	0.34	8.18% 1.87% 0.55% 0.16%
Surface Condition	Categorial	Dry Not dry	0.12	0.4	94.96% 5.04%
Lightening	Categorial	Day Night and road with sufficient lighting Insufficient night lighting	2.31	0.96	71.06% 22.59% 4.01%

		Sunset			2.17%
		Sunrise			0.12%
		Darkness			0.05%
Weather	Categorial	Clear Not clear	0.21	0.63	97.77% 2.23%
Severity	Categorial	Fatal Non-Fatal	0.2	0.5	1.88% 98.12%

Machine Learning Algorithms Results

This section discusses the results of the machine learning techniques used in this study, which were developed in Python.

KNN

In this study, KNN was used to identify the traffic severity. A 10-fold cross-validation was used to select the best model for each value of K, and the average highest accuracy among the 10 folds was chosen. After comparing different numbers of K to overall classification accuracy, the optimal K was determined.Table 2 presents the overall classification report of using the KNN method for different classes. In this table, 4 metrics are presented. Precision, recall, f1-score and support. While the confusion matrix is shown in Table 3.

Table 1. KNN Classification report

	Precision	Recall	F1-score	Support
Non-Fatal	0.97	1	0.98	4433
Fatal	1	0.96	0.98	4382

Table 2.KNN confusion Matrix

	Non-Fatal	Fatal
Non-Fatal	4420	13
Fatal	155	4227

According to Table 2 and 3, the accuracy of predicting fatal accidents was the highest and the precision for the non-fatal accidents was the lowest. The overall model accuracy was 98%.

SVM

In this study, SVM was used to identify the traffic accidents severity out of the four possible types. A 10-fold cross-validation was used to select the best model for each value of K, and the average highest accuracy among the 10 folds was chosen. Moreover, different kernels were used to train the model, the best kernel was rfb with gamma of 0.001. Table 4 presents the overall classification report of using the SVM method. In this table, 4 metrics are presented. Precision, recall, f1-score and support. While the confusion matrix is shown in Table 5.

Table 3. SVM Classification report

	Precision	Recall	F1-score	Support
Non-Fatal	1	0.99	0.99	4433
Fatal	0.99	1	0.99	4382

Table 4. SVM confusion Matrix

	Non-Fatal	Fatal
Non-Fatal	4386	47
Fatal	1	4382

According to Table 4 and 5, non-fatal accidents were the highest and the precision for the fatal accidents was the lowest. The overall model accuracy was 99%.

Adaboost

The AdaBoost classifier was used in this study to identify the crash severity of the wo possible types. Table 6 presents the overall classification report of using the Adaboost method. In this table, 4 metrics are presented. Precision, recall, f1-score and support. While the confusion matrix is shown in Table 7.

Table 5. Adaboost Classification report

	Precision	Recall	F1-score	Support
Non-Fatal	0.99	1	0.99	4433
Fatal	1	0.99	0.99	4382

Table 6. Adaboost confusion Matrix

	Non-Fatal	Fatal
Non-Fatal	4428	5
Fatal	57	4325

According to Table 6 and 7, the accuracy of predicting the fatal accident only was the highest. The variable importance in the adaboost is shown in Fig1.

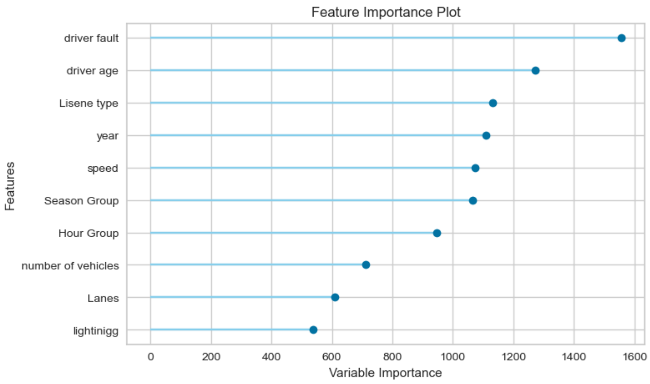

Figure 1. Adaboost variables importance.

The most relevent factors that significantly impact the crash severity prediction are (in order): driver fault, driver age, license type, year, speed, season, time of day, number of vehicles, lanes and lightning.

Conclusion

In this study, several machine learning algorithms were used to predict crashes severity at 15 urban roundabouts in Jordan. Given that roundabouts are created to improve traffic safety, yet in Jordan most of urban roundabouts are labeled as hotspots. Since 2014, traffic police central department has made several improvements in crash severity recording, they now collect several attributes including driver age, gender, spatial location characteristics even they use GPS to provide accident occurrence location exactly. In essence, the work in this study shows that there are ten significant variables to model crash severity including driver fault, driver age, license type, speed, year of occurrence, season, time of day, number of vehicles, geometric characteristics, and lightening. The study indicates that driver gender, holiday and roadway surface characteristics are not significant important features. The various machine learning algorithms were tested on balanced data after using SMOTE technique to balance the dataset. The various algorithms showed a consistent prediction precision for the different crash severity, while the overall accuracy for all of them are not identical, but there is not statistically significant difference between them as they are 98%, 99% and 99% for KNN, SVM and adaboost respectively. The results of this study help the various Jordanian agencies to create a long-term plan to improve traffic safety by create new regulations to control driver behavior as it is the main reason behind the traffic accidents.

Recommendations

For future research, larger, more comprehensive, and reliable samples could be tested using the implemented models. Moreover, test different models and check variables effect on the models.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Technology, Engineering and Science (www.icontes.net) held in Antalya/Turkey on November 16-19, 2023.

*Authors would like to thank the Jordanian traffic institute for providing the data used in this research. Also, first author would like to thank Al-Ahliyya Amman University for the financial support in joining the ICONTES conference.

References

Almamlook, R. E., Kwayu, K. M., Alkasisbeh, M. R., & Frefer, A. A. (2019). Comparison of machine learning algorithms for predicting traffic accident severity. *IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology JEEIT 2019*, 272–276.

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.

Al-Mistarehi, B. W., Alomari, A. H., Imam, R., & Mashaqba, M. (2022). Using machine learning models to forecast severity level of traffic crashes by R studio and ArcGIS. *Frontiers in Built Environment*, *8*.

Al-Moqri, T., Haijun, X., Pierre Namahoro, J., Naji Alfalahi, E., & Alwesabi, I. (2020). Exploiting machine learning algorithms for predicting crash injury severity in Yemen: Hospital case study. *Applied and Computational Mathematics*, *9*(5), 155.

Alomari, A. A., Khasawneh, M. A., Mohammad, P., & Ganam, B. (2019). Evaluation of traffic accidents in Jordan using accident hazard scale. *Jordan Journal of Civil Engineering, 13*(1), 12-20.

Al-Rousan, T. M., Umar, A. A., & Al-Omari, A. A. (2021). Characteristics of crashes -caused by distracted driving on rural and suburban roadways in Jordan. *Infrastructures*, *6*(8), 107.

Alrumaidhi, M., & Rakha, H. A. (2022). Factors affecting crash severity among elderly drivers: A multilevel ordinal logistic regression approach. *Sustainability (Switzerland)*, *14*(18), 11543..

Anderson, J., & Hernandez, S. (2017). Roadway classifications and the accident injury severities of heavy-vehicle drivers. *Analytic Methods in Accident Research*, *15*, 17–28.

Azhar, A., Ariff, N. M., Bakar, M. A. A., & Roslan, A. (2022). Classification of driver injury severity for accidents involving heavy vehicles with decision tree and random forest. *Sustainability (Switzerland)*, *14*(7).

Edries, B., & Alomari, A. H. (2022). Forecasting the fatality rate of traffic accidents in Jordan: Applications of time-series, curve estimation, and multiple linear regression models. *Journal of Engineering Science and Technology Review*, *15*(6), 70–77.

Friedman, J. H., Baskett, F., & Shustek, L. J. (1975). An algorithm for finding nearest neighbors. *IEEE Transactions on Computers*, *100*(10), 1000–1006.

Hariri Asli, H. (2022). Investigation of the factors affecting pedestrian accidents in urban roundabouts. *Computatıonal Research Progress In Applıed Scıence & Engıneerıng*, *8*(1), 1–4.

Hazaymeh, K., Almagbile, A., & Alomari, A. H. (2022). Spatiotemporal analysis of traffic accidents Hotspots Based on Geospatial Techniques. *ISPRS International Journal of Geo-Information*, *11*(4), 260.

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 415–425.

Jamal, A., Zahid, M., Tauhidur Rahman, M., Al-Ahmadi, H. M., Almoshaogeh, M., Farooq, D., & Ahmad, M. (2021). Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study. *International Journal of Injury Control and Safety Promotion*, *28*(4), 408–427.

Jiang, L., Xie, Y., Wen, X., & Ren, T. (2020a). Modeling highly imbalanced crash severity data by ensemble methods and global sensitivity analysis. *Journal of Transportation Safety and Security*, 1–23.

Jiang, L., Xie, Y., Wen, X., & Ren, T. (2020b). Modeling highly imbalanced crash severity data by ensemble methods and global sensitivity analysis. *Journal of Transportation Safety and Security*, 1–23.

JTI. (2022). *Annual accidents report.* Retrieved from https://www.jti.com/

Labib, M. F., Rifat, A. S., Hossain, M. M., Das, A. K., & Nawrine, F. (2019). Road accident analysis and prediction of accident severity by using machine learning in Bangladesh. *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 1–5.

Mamlouk, M., & Souliman, B. (2019). Effect of traffic roundabouts on accident rate and severity in Arizona. *Journal of Transportation Safety and Security*, *11*(4), 430–442.

Mohammadpour, S. I., Khedmati, M., & Zada, M. J. H. (2023). Classification of truck-involved crash severity: Dealing with missing, imbalanced, and high dimensional safety data. *PloS One*, *18*(3), e0281901.

Mujalli, R. (2018). Modeling risk of road crashes using aggregated data analysis of traffic accidents on two-lanes rural highways using data mining view project. *Jordan Journal of Civil Engineering, 12*(1), 45-60.

Mujalli, R. O., Al-Masaeid, H., & Alamoush, S. (2023). Modeling traffic crashes on rural and suburban highways using ensemble machine learning methods. *KSCE Journal of Civil Engineering*, *27*(2), 814–825.

Mujalli, R. O., López, G., & Garach, L. (2017). Modeling injury severity of vehicular traffic crashes. *ACM International Conference Proceeding Series*, 51–55.

Peng, Y., Zhu, S., & Jiang, Y. (2019). Examining the crash severity on divided rural multilane highway segments using multilevel ordinal logistic models. *Advances in Mechanical Engineering*, *11*(4).

Polders, E., Daniels, S., Casters, W., & Brijs, T. (2015). Identifying crash patterns on roundabouts. *Traffic Injury Prevention*, *16*(2), 202–207.

Qawasmeh, B., Kwigizile, V., & Oh, J. S. (2023). Performance and safety effectiveness evaluation of mini-roundabouts in Michigan. *Journal of Engineering and Applied Science*, *70*(1), 36.

Rahim, M. A., & Hassan, H. M. (2021). A deep learning based traffic crash severity prediction framework. *Accident Analysis and Prevention*, *154*, 106090.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., & Yu, P. S. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*, 1–37.

Yahaya, M., Guo, R., Fan, W., Bashir, K., Fan, Y., Xu, S., & Jiang, X. (2021). Bayesian networks for imbalance data to investigate the contributing factors to fatal injury crashes on the Ghanaian highways. *Accident Analysis and Prevention*, *150*, 105936.

Yang, Y., Wang, K., Yuan, Z., & Liu, D. (2022). Predicting freeway traffic crash severity using XGBoost-Bayesian network model with consideration of features interaction. *Journal of Advanced Transportation*. Article ID 4257865. https://doi.org/10.1155/2022/4257865

Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for knn classification. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *8*(3), 1–19.

Author Information

Taqwa Alhadidi
Al-Ahliyya Amman University
Amman, Jordan
Contact e-mail: *t.alhadidi@ammanu.edu.jo*

Mohammed Elhenawey
Queensland University of Science and Technology
Queensland, Australia

To cite this article:

Alhadidi, T., & Elhenawey, M. (2023). Modeling crashes severity using ensemble techniques. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 26,* 357-365.