Tikrit Journal of Engineering Sciences (2025); 32(4): 2395
DOI: http://doi.org/10.25130/tjes.32.4.25

IRAQI Academic Scientific Journals

ISSN: 1813-162X (Print); 2312-7589 (Online)

Tikrit Journal of Engineering Sciences

available online at: http://www.tj-es.com

TJES Tikrit Journal of Engineering Sciences

Truck Accident Prediction and Risk Factors Analysis in Jordan: A Machine Learning Approach

Mu'ath Al-Tarawneh *

Civil and Environmental Engineering Department, College of Engineering, Mutah University, Mutah-Karak, 61710, P.O. BOX 7, Jordan.

Keywords:
Accident risk analysis; Accident prediction model; Machine learning; Traffic safety; Truck accidents.

Highlights:

- *Risk factor analysis for truck accidents in Jordan was considered.*
- *Accident severity prediction based on driver behavior was estimated.*
- *Machine learning was utilized in accident prediction.*

ARTICLE INFO

Article history:

Received	24 Oct. 2024
Received in revised form	15 Nov. 2024
Accepted	02 Dec. 2024
Final Proofreading	05 Dec. 2024
Available online	30 Aug. 2025

Citation: Al-Tarawneh M. **Truck Accident Prediction and Risk Factors Analysis in Jordan: A Machine Learning Approach**. *Tikrit Journal of Engineering Sciences* 2025; **32**(4): 2395.
http://doi.org/10.25130/tjes.32.4.25

***Corresponding author:**

Mu'ath Al-Tarawneh

Civil and Environmental Engineering Department, College of Engineering, Mutah University, Mutah-Karak, 61710, P.O. BOX 7, Jordan.

Abstract: In recent decades, Jordan has experienced significant population and urban development growth, which has been associated with a surge in traffic accidents and their negative impacts on individuals, properties, and the local economy. Trucks are more likely to be involved in fatal accidents due to a combination of factors, including driver behavior, traffic conditions, highway characteristics, and environmental conditions. Thus, the factors affecting the accident risk rate among truck drivers in Jordan have been studied, along with the ability to predict the number of accidents and the number of fatalities that the driver could contribute to their career as a truck driver. The results show that the proposed feature selection methodology successfully chooses the significant factors that affect the accident risk rate, indicating that the driver's behaviors and fatigue-related features affect the type and severity of truck accidents. The results show that using the overbalancing technique (SMOTE) enhances the prediction models' accuracy and decreases the false positives among minor classes as follows: The prediction models can predict the potential of the driver's involvement in an accident with an accuracy of 84.5% using balanced data compared to 75.4% using imbalanced data; the number of accidents with an accuracy of 85.5% using balanced data compared to 69.7 % using imbalanced data; and the number of fatalities with an accuracy of 85.6% using balanced data compared to 78.5 % using imbalanced data set.

التنبؤ بحوادث الشاحنات وتحليل عوامل خطرها في الأردن: باستخدام تعلم الآلة

معاذ أحمد الطراونة
قسم الهندسة المدنية والبيئة/ كلية الهندسة / جامعة مؤته / مؤتة – الاردن.

الخلاصة
في العقود الأخيرة، شهد الأردن نموا كبيرا في عدد السكان والتنمية الحضرية، والذي ارتبط بالارتفاع الكبير في حوادث المرور وآثاره السلبية على الأفراد والممتلكات والاقتصاد المحلي .من المرجح أن تساهم الشاحنات في حوادث مميتة بسبب مجموعة من العوامل، بما في ذلك سلوك السائق وظروف حركة المرور وخصائص الطريق السريع والظروف البيئية .وهكذا، تمت دراسة العوامل التي تؤثر على معدل مخاطر الحوادث بين سائقي الشاحنات في الأردن، إلى جانب القدرة على التنبؤ بعدد الحوادث وعدد الوفيات التي يمكن أن يساهم فيها السائق في حياته المهنية كسائق شاحنة. أظهرت النتائج أن منهجية اختيار الميزات المقترحة تختار بنجاح العوامل المهمة التي تؤثر على معدلات خطورة الحوادث، مما يشير إلى أن سلوكيات سائق الشاحنة اثناء القيادة والعوامل المتعلقة بالإرهاق لدى سائقي الشاحنات تؤثر على نوع ومدى خطورة حوادث الشاحنات. كما أظهرت النتائج أن استخدام تقنية (SMOTE) يعزز دقة نماذج التنبؤ ويقلل من الإيجابيات الخاطئة بين الفئات الثانوية على النحو التالي: يمكن لنماذج التنبؤ، التنبؤ بإمكانية مساهمة السائق في حادث بدقة ٨٤,٥٪ باستخدام بيانات متوازنة مقارنة بـ 75.4٪ باستخدام بيانات غير متوازنة؛ عدد الحوادث بدقة 85.5٪ باستخدام بيانات متوازنة مقارنة بـ 69.7٪ باستخدام بيانات غير متوازنة؛ وعدد الوفيات بدقة ٨٥,٦٪ باستخدام بيانات متوازنة مقارنة بـ 78.5٪ باستخدام مجموعة بيانات غير متوازنة.

الكلمات الدالة: مخاطر الحوادث، تعلم الآلة، نموذج التنبؤ بالحوادث، السلامة المرورية، حوادث الشاحنات.

1.INTRODUCTION

Traffic accidents in Jordan, including truck accidents, have posed a significant concern. Statistics released by the Directorate of Public Security (Traffic Management Department) for 2022 indicated a noticeable increment in the number of vehicles, where vehicle per person increased from one vehicle for every 69 people in 1970 to one vehicle for every six people in 2022 [1]. Thus, the type and rate of traffic accidents were impacted; the annual increase in the number of injury accidents amounted to 1.1 percent for the years from 2018 to 2022 per 10 thousand registered vehicles, and the yearly growth of the same period in the number of injures amounted to 0.5 percent per 10 thousand registered vehicles, traffic accidents contributed to raising the cost of accidents with an annual increase rate of 0.9 percent for the years 2018-2022 [1]. The increase in registered vehicles in Jordan has affected the repetition Rate of Road Traffic Accidents (RRTA), averaging 30.8 RRTA daily in 2021. The economic implications associated with RRTA in Jordan have grown significantly, with an estimated cost of 2.2 billion U.S. dollars, which amounted to RRTA from 2017-2021 [2]. Numerous factors affect the likelihood of accidents, including human factors such as reckless driving behaviors, violations associated with speeding, and distracted driving while using a smartphone [3-6]; severe weather conditions (dust, snow, and heavy rain); road conditions; road lighting levels; unprepared vehicles; and insufficient traffic laws [3-4, 7]. RRTA prediction could be sorted into several categories to cover all aspects related to RRTA [8], including RRTA risk prediction (which involves predicting the likelihood of traffic accident occurrence), RRTA frequency prediction (which consists of estimating the number of accidents within a specified period of time), and RRTA severity prediction (which involves predicting how severe the accident might be regarded to the number of fatalities and injuries). There are various techniques for traffic accident analysis and all RRTA prediction types mentioned above. These include machine-learning techniques such as Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machines (GBM) [9-12]; Deep learning techniques, for example, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [13-15]; statistical modeling such as the use of regression models and Poisson process [16-18]. Utilizing machine learning techniques for RRTA prediction has been effective. Studies conducted in Jordan have utilized ANN and SVM for RRTA prediction and classification [19-21]. Ordinal regression models have been utilized to assess the magnitude of accidents, considering weather conditions, road surface, speed limit, and illumination levels [22]. Additionally, machine learning techniques can be leveraged to examine crash datasets and identify the factors contributing to accidents and injuries. These studies offer valuable insights for formulating strategies to mitigate traffic accidents in Jordan, including identifying high-intensity accident-prone areas and implementing preventive measures. Machine learning techniques have been used worldwide for traffic accident prediction. Siam et al. [23] investigated different machine-learning techniques for traffic accident and severity prediction in Bangladesh. The results were generated using a decision tree (C5). A team from Concordia University [9] studied traffic accidents in Montreal using Random Forest (RF), Balanced Random Forest (BRF), and XG boost algorithms from three open datasets. Those models successfully predicted 85% of Montreal's accidents. Lin et al. [24] investigated using different machine learning algorithms (RF, Bayesian network, K-nearest neighbor). The best model prediction accuracy was 61%. Chang and Chen [25] used a CART (classification and regression trees) model to predict the number of accidents in Taiwan with

55% accuracy. However, none of the studies in the literature addressed truck accidents as a separate part of the work to determine the factors that may increase the percentage of accidents among truck drivers whose accidents are more severe compared to other vehicle classes. Therefore, in this study, the factors affecting the number of traffic accidents and their severity within the category of truck drivers in Jordan were studied to make this study a guide for the decision-makers and the employed sector of truck drivers that the road freight sector in Jordan depends on the most.

2.DATA COLLECTION AND STUDY AREA

Since there is no database to assess the factors that affect the crash rate among truck drivers in Jordan, the author formulated the Driver Behavior Questionnaire (DBQ) based on the literature to collect essential data needed to evaluate the relation between a truck's crash rate and driver behavior. Twenty-four questions were included in the questionnaire and categorized into three sections. The covered sections are demographic and job-related characteristics, driver behavior, and accident-related features (the accident's nature includes any operational vehicle rolling on the road, pedestrians, fixed objects, and properties). The driver behavior section covers all the distractions that might affect the driver while in operation, including drug use, in-vehicle distractions (music, texting, eating, etc.), and speeding. The questionnaires were distributed to 988 truck drivers in their gathering spaces and parking lot along Desert Highway (male drivers since it is not common for female drivers to work in such a job). The survey was conducted between October 2019 and January 2020 in three provinces (Karak, Ma'an, and Aqaba). The trucks that participated in the questionnaire included 2-axle commercial vehicles (5%), 3-axle vehicles (10 %), 4-axle vehicles (16%), 5-axle vehicles (29.5%), and six or more axle vehicles (39.5%). Despite that, all drivers who contributed to the study's questionnaire were located in the provinces mentioned above along Highway 15 (desert highway), which is considered the backbone of freight transport between Amman and the only port in Jordan located in Aqaba province, and for external imports and exports with Arab Gulf Countries via Saudi Arabia; the drivers descend from different regions in Jordan: North region: 22%, Central region: 37%, South region: 41%.

3.DBQ OUTPUT

The analysis of DBQ output was done using a statistical analysis program. Table 1 summarizes the most essential questions reflecting demographic and trip characteristics. From Table 1, some demographic data questions include age, driver height, and weight to study the effect of Body Mass Index (BMI) on the crash rate, total years of experience, and monthly gross income. However, the trip-related characteristics include average sleeping hours before travel, maximum travel hours, and distance for a one-way trip. It can be seen from Table 1 that the majority of truck drivers are between 30 and 50 years old, with an average of 59%, and average weights between 70 and 90 kg. Furthermore, most truck drivers have 5-10 years of experience, which accounts for 32% of the total drivers, but at the same time, only 3% receive income that is more significant than 1000 JOD. About pre-travel preparations, it is noted that the majority of sleep hours is more than 8 hours before traveling, with an average of 44%. Moreover, most drivers drive continuously for more than 6 hours, as they cover a distance of more than 300 km for one trip at a rate of 81%. Table 2 summarizes the essential practices that a truck driver may practice in Jordan, including distractions while driving. These include eating, smoking, using the phone, listening to the radio, taking steroids and medications for chronic diseases, accompanying the driver, and talking to him while driving. The degree of repetition of the practice was divided into four categories. Table 3 shows the type of damage caused by accidents, which can be classified as property damage only (PDO), injuries, fatalities, or a combination of accident classes. Furthermore, Table 3 summarizes the number of accidents that the driver anticipated while working as a truck driver and the causes of the accidents. From Table 3, it is evident that most of the damages occurred because of property damages (material), and 84% of total respondents have been involved in one accident at least. This reflects the enormous number of accidents in Jordan, which has already suffered from a high rate and severity of accidents compared to surrounding countries [1]. Moreover, excessive speeding and poor road geometry design are regarded as the most common causes of accidents among truck drivers, with percentages of contributions of 28% and 17%, respectively. Although most drivers have been involved in at least one accident, it does not reflect the severity of these accidents only by determining the number of injuries and fatalities for each accident. Table 3 shows the fatalities among those involved in total accidents per driver. The maximum number of deaths was two, with 5.9% among total fatalities compared to 9.5% for one fatality.

Table 1 Demographic data and trip-related characteristics (experience, income, travel distance, and travel time).

Age (year)	18-30 23%	31-40 29%	41-50 30%	> 50 18%	
Weight (kg)	< 70 31%	70-90 48%	91-110 17%	> 110 4%	
Height (cm)	160-175 54%		176-190 46%		
Year Of Experience	<5 17%	5-10 32%	11-15 20%	> 15 31%	
Avg Sleep hours	< 4 7%	4-6 17%	6-8 32%	>8 44%	
Monthly Income (JOD)	< 500 33%	500-750 49%	751-1000 15%	> 1000 3%	
Continuous Driving Hours	< 4 30%	4-6 26%	6-8 38%	>8 6%	
Maximum Distance for a one-way trip (km)	< 100 2%	100-200 7%	201-300 12%	301-400 49%	> 400 32%

Table 2 Driver behavior while driving.

Practice	Never	Sometimes	Frequently	Always
Eating	23%	39%	20%	18%
Smoking	10%	10%	20%	60%
Radio	8%	34%	23%	35%
Having a companion during travel	31%	44%	21%	4%
Using Phone	25%	24%	39%	12%
Taking steroids	71%	23%	4%	2%
Medicines	84%	10%	4%	2%
Sleeping	92%	6%	2%	0%
Driving within the posted speed	6%	24%	36%	34%

Table 3 Truck accident type, accident causes, and total numbers of accidents.

Accident Consequences	Property damage only (PDO) 40%	Property damage, injuries 17%	Property damage, Fatalities 16%	Property damage, injuries, Fatalities 6%	Injuries 5%	Never 16%
Causes of The Accident	Lack of concentration 8%	Poor road design 17%	Excessive speeding 28%	Vehicle Malfunction 6%	weather conditions 7%	Wrong direction 2%
	Human error 11%	Lack of sleep 6%	number of lanes 1%	Health problems 2%	never 10%	Poor lighting 2%
Total Number of Accidents	0 16%	1-2 48%	3-4 12%	5-6 9%	>6 15%	

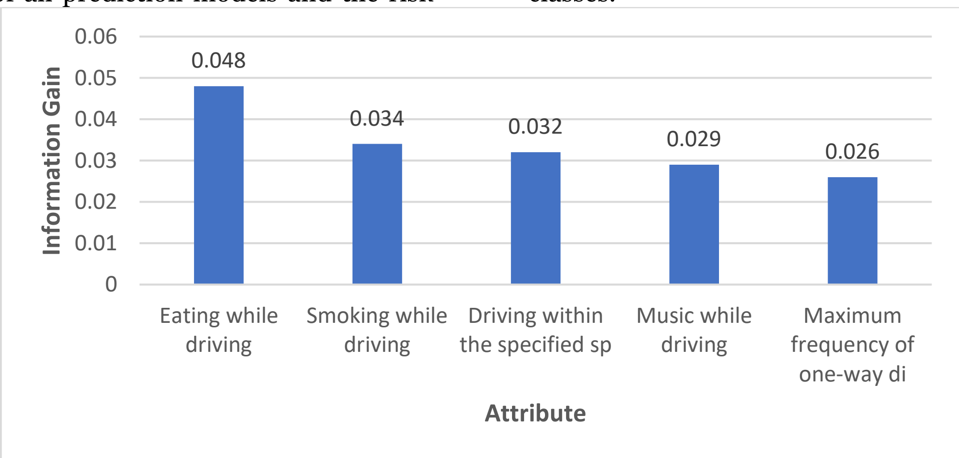
4.ACCIDENT RISK PREDICTION MODELS

One of the main objectives of this study is to evaluate the probability of involving a truck driver in an accident and select the most significant attributes affecting the accident involvement by a truck driver. This study used a database to construct a classification model to predict the probability of participating in an accident. Two target classes: (Yes: involved in an accident, No: not involved in an accident), the number of possible accidents (Three target classes: A: No accidents, B: 1-3 accidents, and C: ≥4 accidents), as well as the number of fatalities caused by accidents (Two target classes: a: Fatal accident, and b: Not fatal accident). Waikato Environment for Knowledge Analysis (WEKA) software is widely used in machine learning and data, and it was established in Java at the University of Waikato, New Zealand [26]. WEKA software used to select attributes and construct a classification model. The dataset is divided into a training and validation set (67% of the data) and a test set (33%), randomly selected instances.

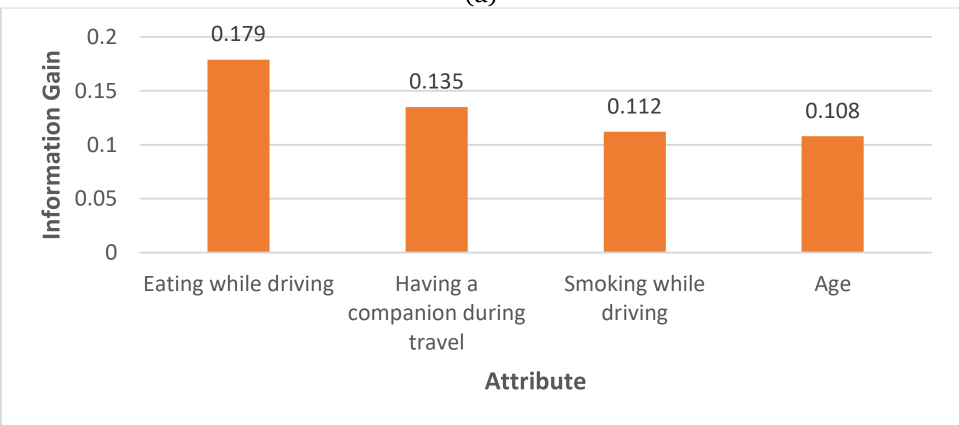
The data distribution for accident involvement, number of accidents, and number of fatalities from each accident indicated imbalanced data for multi-class prediction. To handle the imbalanced data, the Synthetic Minority Over-sampling Technique (SMOTE)was used to generate positive instances to balance the data. SMOTE was widely applied for imbalanced data and gave better results than simple oversampling [27]. Meanwhile, the imbalanced data set results were tested for the three prediction approaches.

4.1.Feature Selection Study

The feature selection study was conducted using the ranker search method in the attribute evaluator (InfoGainAttributeEval) by WEKA software. This method evaluates the worth of an attribute by measuring the information gained concerning the class. The information gain (IG) values vary from 0 (no information) to 1 (maximum information), which means the features with high information gain close to 1 will contribute the most to class prediction. In contrast, the features with low information gain close to 0 can be excluded based on appropriate arbitrary cutoff values. The cutoff values are determined using the standard deviation of the IG values for all features in each prediction model [28]. All the instances associated with accident causes and damage types have been excluded since they are correlated with the target classes for all prediction models. This highlights the importance of driver behavior and demographic and job-related instances in determining prediction models' performance. Figures 1(a), 2(a), and 3(a) show the information gain values using the proposed selection methodology. They use an imbalanced data set to predict accident involvement, number of accidents, and number of fatalities using the following cutoff values: 0.025 for accident involvement, 0.04 for the number of accidents, and 0.03 for the number of fatalities, respectively. Also, Figures 1(b), 2(b), and 3(b) show the information gain values using the balanced data set using SMOTE for accident involvement, number of accidents, and number of fatalities, respectively. The cutoff values using balanced data are as follows: 0.1 for accident involvement, 0.15 for the number of accidents, and 0.1 for the number of fatalities. Figures 1, 2, and 3 show that the information gain values using the balanced data set are higher than those generated using the imbalanced dataset. This will be reflected in the accuracy of all prediction models and the risk factor analysis. Fig. 1 (a) shows that the driver's behavior features (Eating, smoking, listening to the radio) contribute to predicting if the driver is involved in an accident, in addition to the maximum distance and driving within the posted speed. However, the information gain values are considerably low. Thus, the factor risk analysis affecting the accident involvement of truck drivers will be concluded using a balanced data set. Fig. 1 (b) explains the importance of the features affecting the driver's fatigue in predicting the accident potential, such as (maximum distance traveled, age, and average number of sleep hours). Fig.1 (b) also shows the impact of the driver's behavior (eating and smoking while driving) on the involvement in the accident. Fig. 2 (a) shows the effects of driver's behavior on predicting the number of accidents using an imbalanced data set, including eating, listening to the radio, using a phone, speeding, and smoking. Fig. 2 (b) summarizes the most significant features affecting the number of accident predictions using a balanced data set. Fig. 2 (b) shows that the features related to distractions during driving are still compelling (eating, using a phone, listening to the radio). In contrast, age and monthly income features remarkably correlate to the number of accidents in target classes.

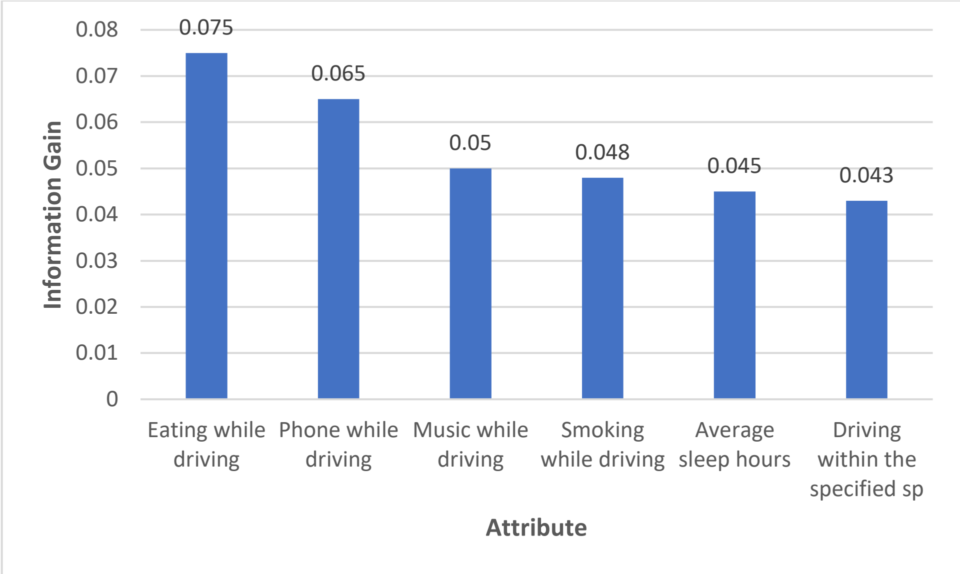


(a)

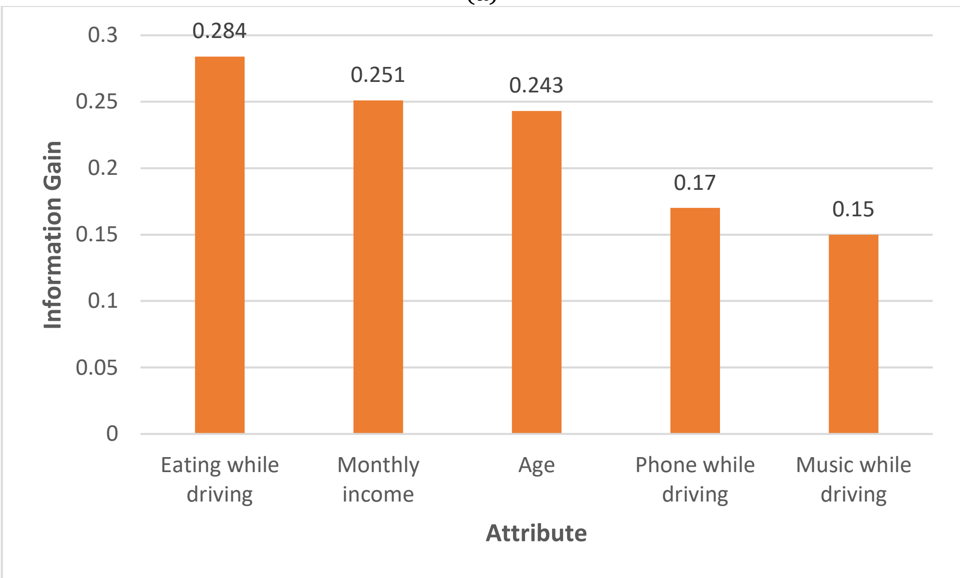


(b)

Fig. 1 Information Gain Values for Features Affecting Accident Involvement Prediction Using (a) Imbalanced Data Set and (b) Balanced Data Set.



(a)



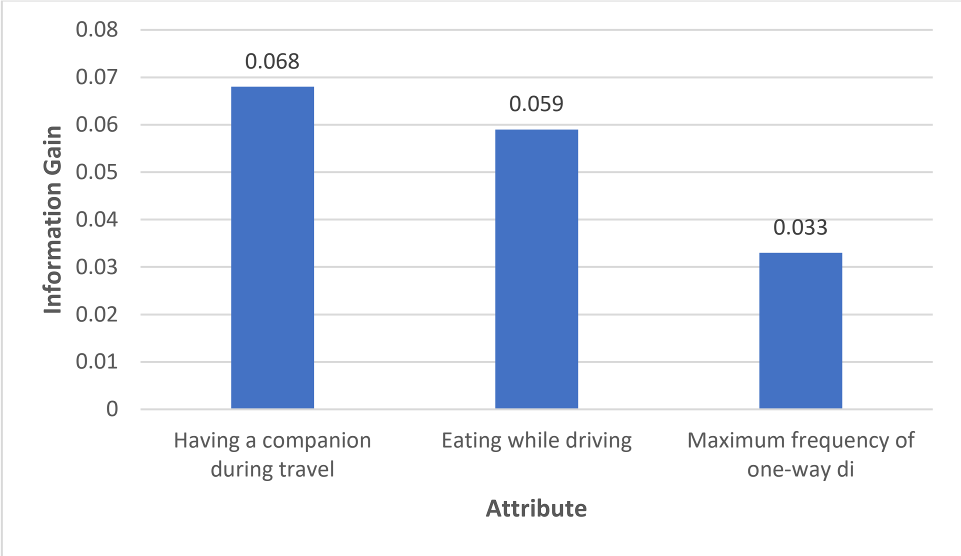
(b)

Fig. 2 Information Gain Values for Features Affecting the Number of Accidents Prediction Using (a) Imbalanced Data Set and (b) Balanced Data Set.

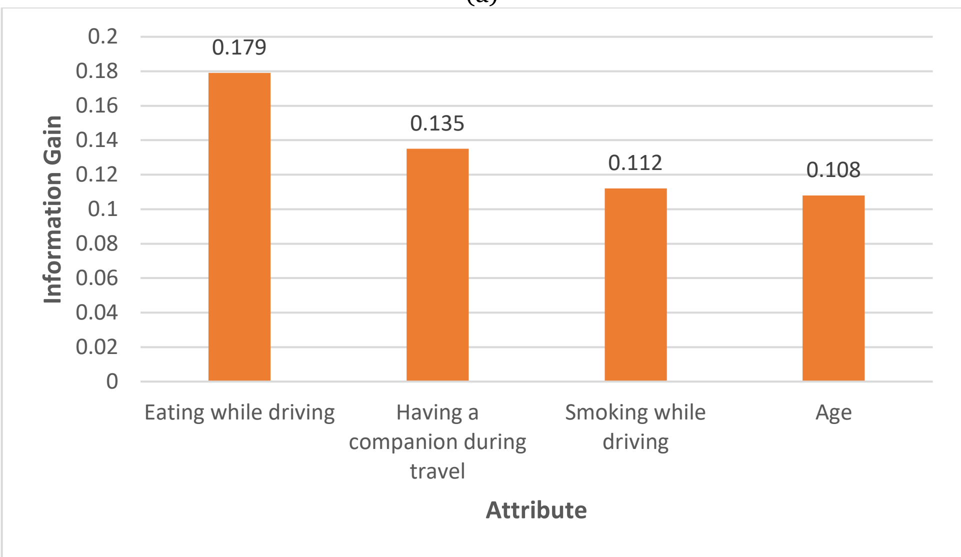
Figures 3 (a) and 3 (b) show that distractions (chatting with a trip companion, eating) are significant in predicting fatalities from both imbalanced and balanced datasets. Also, the age feature affects the fatality prediction using a balanced data set, as shown in Fig. 3 (b). Figures 1 (b), 2(b), and 3(b) show that the driver's behavior while driving is significant in both the quantity and severity of truck accidents, in addition to fatigue-related features (maximum traveled distance, average sleeping hours, age).

4.2.Prediction Results

The experimental procedures are listed below. In this section, prediction models were constructed using the WEKA software classification tool for the involvement of truck drivers in accidents, the number of accidents, and the number of expected fatalities from an accident. All models were built using the features from the proposed feature selection methodology, as shown in Figures 1, 2, and 3, for both imbalanced and balanced datasets. For accident involvement classification, the reduced imbalanced dataset was used in 663 configurations using Auto-Weka to define the best classifier based on the error rate. Based on the training dataset, the Classifier Decision Table is the most accurate classifier, with an average accuracy of 75.4 % (percentage of correctly predicted classes among all classes). Table 4 (a) shows the confusion matrix of the predicted class versus the actual class for the 33% test set. Table 4 (b) shows the confusion matrix for the balanced dataset, which was also used in 820 configurations to define the best classifier using the reduced dataset. The Random forest classifier was the most accurate classifier, with an average accuracy of 84.5%.



(a)



(b)

Fig. 3 Information Gain Values for Features Affecting the Number of Fatalities Prediction Using (a) Imbalanced Data Set and (b) Balanced Data Set.

Table 4 Confusion Matrix for Accident Involvement Prediction Using (a) Imbalanced Dataset and (b) Balanced Dataset.

	Yes	No	TPR	FNR
Yes	218	48	81.9%	18.1%
No	32	28	46.7%	53.3%

(a)

	Yes	No	TPR	FNR
Yes	227	40	85.1%	14.9%
No	42	219	83.9%	16.1%

(b)

To define the best classifier based on the error rate for the number of possible accidents, 678 configurations were used to find the best classifier. Classifier Random Subspace was the most accurate classifier based on the training dataset, with an average accuracy of 69.7%. Table 5 (a) shows the confusion matrix of the predicted class for the number of accidents versus the actual class for the 33% test set. The results from the balanced data set show that the Naive Bayes classifier is the best classifier for predicting the number of accidents with an average accuracy of 85.5%, as shown in Table 5 (b).

Table 5 Confusion Matrix for the Number of Accidents Prediction Using (a) Imbalanced Dataset and (b) Balanced Dataset.

	A	B	C	TPR	FNR
A	32	13	7	61.5%	38.5%
B	32	147	18	74.7%	25.3%
C	10	19	48	62.3%	37.7%

(a)

	A	B	C	TPR	FNR
A	137	21	5	84.1%	15.9%
B	12	133	7	87.5%	12.5%
C	11	15	149	85.1%	14.9%

(b)

To predict the number of fatalities, the reduced dataset has been used in 821 configurations using Auto-Weka to define the best classifier based on the error rate. Classifier OneR is the best classifier based on the training dataset, with an average accuracy of 78.5%. Table 6 (a) shows the confusion matrix of the predicted number of fatalities classes versus the actual class for the 33% test set. The results from the oversampled data set show that the IBK classifier is the most accurate classifier for predicting the number of fatalities, with an average accuracy of 85.6%, as shown in Table 6 (b).

Table 6 Confusion Matrix for the Number of Fatalities Prediction Using (a) Imbalanced Dataset and (b) Balanced Dataset.

	a	b	TPR	FNR
a	228	46	83.1%	16.9%
b	24	28	53.8%	46.2%

(a)

	a	b	TPR	FNR
a	187	35	84.2%	15.8%
b	27	183	87.1%	12.9%

(b)

Further discussion could be made based on the confusion matrix tables for all models. The model's accuracy is not the only criterion for measuring the effectiveness of any classifier. Thus, it is essential to take into account the source of the error inside each model based on the target class size. This can be executed using False Negative Rates (FNR) for each target class, which is the percentage of incorrectly predicted instances over all instances for each actual target class. On the other hand, True Positive Rates (TPR) determine the percentage of correctly classified instances. Table 4 (a) shows that class (NO) is the minor class for accident involvement prediction using the imbalanced data set. The FNR value was 53.3 %, which decreased to 16.1% due to SMOTE, as shown in Table 4 (b). As for the number of accidents, the target classes (A) and (C) were the minor classes with FNR values of 38.5% and 37.7%, respectively, as shown in Table 5 (a). Using the balanced dataset, these values decreased to 15.9% for class (A) and 14.9% for class(C). In addition, the value of FNR for the number of fatalities decreased for the minor class (b) from 46.2% to 12.9% after applying the oversampling technique, as shown in Table 6. In summary, the results from Tables 4, 5, and 6 show that using SMOTE has an advantage over the results from an imbalanced dataset regarding the accuracy and FNR and TPR values.

5.CONCLUSION

This research developed a model for predicting accident severity for truck drivers in Jordan, including the probability of the driver's involvement in an accident, the number of potential accidents, and the number of fatalities. Also, factors affecting the accident severity among truck drivers are discussed. The conclusions of this paper are:

1) The features selected from the balanced dataset have higher information gain scores than those selected using the imbalanced dataset.
2) The driver's driving behaviors, such as eating, smoking, using phones, listening to the radio, and chatting, affected the number of accidents and their severity rate.
3) The fatigue-related features (average sleep hour, age) significantly impacted the accuracy of the prediction models.
4) The results showed that developing prediction models using balanced datasets had an advantage over the results from imbalanced datasets regarding accuracy and FNR values.
5) The best classifiers for prediction models were the Random Forest for accident involvement, with an average accuracy of 84.5%; the Naive Bayes classifier for the number of accidents, with an average accuracy of 85.5%; and the IBK classifier for predicting the number of fatalities, with an average accuracy of 85.6%.

The approach introduced in this study, including feature selection methodology and prediction models, could be implemented for crash risk analysis and prediction for different vehicle categories.

6.LIMITATIONS

This study contained several limitations as follows:

- According to the Performance indicators, passengers, and cargo report (second quarter\2024) issued by LTRC (Land Transport Regulatory Commission) [29], the percentage of trucks owned by individuals is 74.4% and only 25.6% as a company's fleet. One of the significant limitations is that the trucks are mainly owned by individuals, which means less database about drivers through organized companies, considering that some employees considered the questionnaires a violation of their privacy.
- The proposed methodology used SMOTE to balance the dataset, which might contain some overfitting. However, SMOTE is considered less vulnerable to overfitting than other balancing data techniques because SMOTE generates new synthetic samples instead of original data duplication. Also, the result showed better accuracy after applying SMOTE, which reduced the likelihood of overfitting.

ACKNOWLEDGEMENTS

The author is grateful for everyone who helped during the data collection stage. This study is fully funded by the author.

REFERENCES

[1] Public Security Directory in Jordan. **Annual Report of Traffic Accidents in Jordan for the Year 2022**. 2022. Available from: https://www.psd.gov.jo/media/jfopsjca/final-annual-report-2022-2.pdf.

[2] Al-Tammemi AB, Altarawneh SA, Ghazaleh LZA, Jaradat TM, Abujudeh R, Almbaidin M, Hasan H. **Navigating Road Traffic Accidents in Jordan: A Retrospective Exploration of the Health and Economic Impacts**. *Cureus* 2024; **16**(5): e61385.

[3] Ghuzlan KA, Al-Khateeb GG, Haddad M. **Evaluation of 2008 Traffic Safety Policies in Jordan**. *Jordan Journal of Civil Engineering* 2012; **6**(1): 12–25.

[4] Abuaddous M, Al-Hares A, Albtoush AF, Al-Btoosh JFAA. **Identification and Ranking of Accident Black Spots in Jordan**. *Civil Engineering and Architecture* 2022; **10**(4): 1661–1674.

[5] Jadaan K, Abojaradeh M, Shaqadan A, Abojaradeh D, Alshalout I. **Analysis of Human and Cultural Factors Causing Risk of Accidents in Jordanian Roads**. *International Journal of Injury Control and Safety Promotion* 2020; **27**(2): 234–243.

[6] Ahmed SK, Mohammed MG, Abdulqadir SO, El-Kader RGA, El-Shall NA, Chandran D, Ur Rehman ME, Dhama K. **Road Traffic Accidental Injuries and Deaths: A Neglected Global Health Issue**. *Health Science Reports* 2023; **6**(5): e1240.

[7] Al-Khateeb GG. **Analysis of Accident Data and Evaluation of Leading Causes for Traffic Accidents in Jordan**. *Jordan Journal of Civil Engineering* 2010; **4**(2): 76–94.

[8] Behboudi N, Moosavi S, Ramnath R. **Recent Advances in Traffic Accident Analysis and Prediction: A Comprehensive Review of Machine Learning Techniques**. *arXiv preprint* 2024; arXiv:2406.13968.

[9] Hébert A, Guédon T, Glatard T, Jaumard B. **High-Resolution Road Vehicle Collision Prediction for the City of Montreal**. *2019 IEEE International Conference on Big Data (Big Data)*; 2019; Los Angeles, CA, USA; pp. 1804–1813.

[10] Feng M, Zheng J, Ren J, Liu Y. **Towards Big Data Analytics and Mining for UK Traffic Accident Analysis, Visualization & Prediction**. *Proceedings of the 2020 12th International Conference on Machine Learning and Computing*; 2020; Shenzhen, China; pp. 225–229.

[11] Krishna UV, Sudhakaran S, Sanju S, Vignesh E, Kaladevi R. **Accident Prediction and Analysis Using Machine Learning Models**. *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*; 2023; Uttarakhand, India; pp. 37–40.

[12] Almanie T. **Quantitative Study of Traffic Accident Prediction Models: A Case Study of Virginia Accidents**. *International Journal of Advanced Networking and Applications* 2023; **14**(5): 5582–5589.

[13] Wu YW, Hsu TP. **Mid-Term Prediction of At-Fault Crash Driver Frequency Using Fusion Deep Learning with City-Level Traffic Violation Data**. *Accident Analysis & Prevention* 2021; **150**: 105910.

[14] Wakatsuki Y, Tatebe J, Xing J. **Improving the Accuracy of Traffic Accident Prediction Models on Expressways by Considering Additional Information**. *International Journal of Intelligent Transportation Systems Research* 2022; **20**(1): 309–319.

[15] Koramati S, Mukherjee A, Majumdar BB, Kar A. **Development of Crash Prediction Model Using Artificial Neural Network (ANN): A Case Study of Hyderabad, India**. *Journal of The Institution of Engineers (India): Series A* 2023; **104**(1): 63–80.

[16] Macedo MR, Maia ML, Rabbani ERK, Neto OCL, Andrade M. **Traffic Accident Prediction Model for Rural Highways in Pernambuco**. *Case Studies on Transport Policy* 2022; **10**(1): 278–286.

[17] Thapa D, Paleti R, Mishra S. **Overcoming Challenges in Crash Prediction Modeling Using Discretized Duration Approach: An**

Investigation of Sampling Approaches. *Accident Analysis & Prevention* 2022; **169**: 106639.

[18] Thapa D, Mishra S, Velaga NR, Patil GR. **Advancing Proactive Crash Prediction: A Discretized Duration Approach for Predicting Crashes and Severity**. *Accident Analysis & Prevention* 2024; **195**: 107407.

[19] Jadaan KS, Al-Fayyad M, Gammoh HF. **Prediction of Road Traffic Accidents in Jordan Using Artificial Neural Network (ANN)**. *Journal of Traffic and Logistics Engineering* 2014; **2**(2): 104–108.

[20] Al-Masaeid HR, Khaled FJ. **Performance of Traffic Accidents' Prediction Models**. *Jordan Journal of Civil Engineering* 2023; **17**(1): 34–44.

[21] Shatnawi N, Al-Omari AA, Alkhateeb S. **Prediction of Risk Factors Influencing Severity Level of Traffic Accidents Using Artificial Intelligence**. *International Review of Civil Engineering* 2023; **14**(1): 1–7.

[22] Darwish W, Darwish HH. **Ordinal Regression Analysis of Traffic Collision Accidents in Jordan 2021: Factors and Severity Assessment**. *Acta Logistica* 2023; **10**(2): 241–250.

[23] Siam ZS, Hasan RT, Anik SS, Dev A, Alita SI, Rahaman M, Rahman RM. **Study of Machine Learning Techniques on Accident Data**. *Advances in Computational Collective Intelligence: 12th International Conference, ICCCI 2020*; 2020; Da Nang, Vietnam; pp. 25–37.

[24] Lin L, Wang Q, Sadek AW. **A Novel Variable Selection Method Based on Frequent Pattern Tree for Real-Time Traffic Accident Risk Prediction**. *Transportation Research Part C: Emerging Technologies* 2015; **55**: 444–459.

[25] Chang LY, Chen WC. **Data Mining of Tree-Based Models to Analyze Freeway Accident Frequency**. *Journal of Safety Research* 2005; **36**(4): 365–375.

[26] Singhal S, Jena M. **A Study on WEKA Tool for Data Preprocessing, Classification and Clustering**. *International Journal of Innovative Technology and Exploring Engineering* 2013; **2**(6): 250–253.

[27] Blagus R, Lusa L. **SMOTE for High-Dimensional Class-Imbalanced Data**. *BMC Bioinformatics* 2013; **14**(106): 1–16.

[28] Ebiaredoh-Mienye SA, Swart TG, Esenogho E, Mienye ID. **A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease**. *Bioengineering* 2022; **9**(8): 350.

[29] Land Transport Regulatory Commission (LTRC). **Performance Indicators, Passengers, and Cargo Report (Second Quarter 2024)**. 2024. Available from: https://www.ltrc.gov.jo/sites/default/files/01.pdf.