Sigma J Eng Nat Sci, Vol. 43, No. 3, pp. 808−815, June, 2025

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00072

Research Article

Machine learning approach for predicting truck drivers' involvement in an injury accident

Ibrahim Khalil UMAR[1,*]

[1]*Department of Civil Engineering, Federal University of Technology, Babura, 340252, Nigeria*

ARTICLE INFO

Article history
Received: 14 January 2024
Revised: 02 March 2024
Accepted: 10 June 2024

Keywords:
Injury Accidents, Kano, Naïve Baye; Truck Drivers

ABSTRACT

Truck drivers' involvement in road traffic accidents increases the accident's severity. Predicting and identifying the significant parameters responsible for truck drivers' involvement in accidents is one of the sustainable measures in reducing accident severity. The study compares the performance of three machine learning models (Naïve Bayes, support vector machine, and K-nearest neighbor) for the prediction of truck drivers' involvement in an injury accident using 248 datasets obtained through questionnaire survey. The models' input includes driver's demographics (age, education), involvement in an injury accident, distance traveled in a week, driving experience, type of truck driving, presence of co-driver, sleeping on the wheel, and average daily driving hours. The models were evaluated using accuracy, F1-score, and AUC parameters. The Naïve Bayes model outperformed both the K-NN and SVM models by almost 10.5% and 6.9%, respectively. The Naïve Bayes model classifies the injury accidents with a moderate accuracy with kappa value of 0.4748 higher than K-NN (0.2628) and SVM (03390). Three different algorithms were also used to rank the relevance of the parameters in increasing the severity of truck-involved accidents. The study shows sleeping on wheels and distance traveled per week are the most significant factors contributing to truck drivers' involvement in injury accidents. Gender, age, and driving hours were found to be the least influencing factors.

Cite this article as: Umar IK. Machine learning approach for predicting truck drivers' involvement in an injury accident. Sigma J Eng Nat Sci 2025;43(3):808−815.

INTRODUCTION

Road traffic accidents claim millions of lives every year. According to a World Health Organization report, a road traffic accident is responsible for the death of 1.35 million people globally each year and injuring another 50 million people [1]. Thus, road accident has now become one of the prevalent determinants of death. About 518 billion dollars are lost annually as a result of such accidents. Consequently, several people become disable, losing their body parts [2]. Motor vehicle traffic fatalities are an important public health problem in both developed and developing countries [3]. Nigeria, like many other developing countries, has a large number of people killed as a result of road traffic accidents. A total of 13,656 cases of traffic accidents involving 89,143 people and 21,407 vehicles were recorded in the year 2022. Out of the total vehicles involved in the traffic 4,417 are trucks (trailers, tanker, luxurious bus, trucks) [4].

***Corresponding author.**
*E-mail address: ikdotumar@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

Accidents involving trucks are mostly associated with higher severity. A significant proportion of truck involved accidents are either injury or fatal accidents with truck drivers been at fault in more than 80% of the cases [5]. The probability of truck driver involvement in a fatal accident is higher than that of other drivers due to the heavy weight they carry that makes braking time longer. The fatality and property damage in truck related accidents surpassed that of non-truck related accidents as reported in several studies [6]. 45% of truck drivers in Tanzania were reported to have been involved in fatal accident with 40% of the drivers been involved in at least minor-traffic accident [7]. Studies found truck type, experience, driving hours and driving time as the most significant parameters affecting truck drivers' involvement in injury accidents [8]. Other parameters influencing truck divers' involvement in accidents are fatigue, distance travelled, and age [6]. Payment method, scheduling practice employment type, drivers training was also identified as factors influencing truck drivers' involvement in injury accidents [9].

Road traffic accident prediction models are essential tools for providing safer roads by identifying the causative factors and offering measures for improvement. Several regression and empirical models have been developed over the years but face many limitations, such as poor prediction accuracy due to the complex nature of traffic accidents [10]. The models try to understand the factors associated with accident occurrence by developing statistical relationships correlating various risk factors with the number of accidents occurring on a road section over some time. Accident prevention models are usually used to monitor the effectiveness of different road safety policies introduced to minimize accident occurrences. They also give transportation planners and engineers an idea to determine new approaches and strategies for road safety [11]. Therefore, a dependable model for accident prediction is essential for providing safer roads. [12] compared the performance of the ANN model and classical linear regression in predicting the severity of road traffic accidents in Nigeria. The ANN model outperformed the regression model by 18.7% and 2.5%, respectively, for fatality and injury models. [13] compared the performance of two statistical and four machine learning techniques. The machine learning techniques outperformed the statistical models in predicting the accident severity with random forest having the highest accuracy of 53.9%. A hybrid machine learning model for the classification of injury accident was developed by integrating Boruta algorithm into four machine learning techniques (random forest, naïve Bayes, K-Nearest Neighbor and binary logistic regression). The extreme gradient boosting model (XGBoost) performed better than all the four machine learning techniques with an accuracy of 82.1% and AUC-ROC value of 88% [14]. The performance of XGBoost model for the prediction of traffic accident was compared with other machine learning techniques (random forest, decision tree and logistic regression). The XGBoost demonstrated higher prediction capability with an overall accuracy of 0.93% and kappa value of 87% [15]. [16] used fuzzy logic, entropy approach and integration cluster for the analysis of different blackspots using traffic accidents data of Denizli city. Traffic characteristics, average speed and geometry were found to be the major contributing accident factors in the area. Classical k-means and fuzzy c-means clustering techniques were used for the analysis of traffic accident in Turkey. It was found that province in rural areas were characterized with higher fatal and injury accidents [17]. [18] analyzed accident blackspots using k-means and fuzzy clustering method. The machine learning approach proved to provide good results in analysis of road traffic accidents.

The suitability of machine learning techniques in handling complex problems makes them appropriate for predicting truck drivers' involvement in injury accidents. Since there is no single machine learning model generally accepted to suit all accident prediction problems. This study employs and compares the performance of three machine learning techniques, Naïve Bayes (NVB), Support Vector Regression (SVR), and K-nearest neighbor (KNN), for modeling truck drivers' involvement in injury accidents. This way, the most appropriate model for the study area could be determined. The study investigates the parameters responsible for truck drivers' involvement in injury accidents.

MATERIALS AND METHODS

Data

The data used in the study was obtained through a questionnaire survey conducted in Kano State Nigeria. A total of 248 truck drivers were interviewed on their involvement in injury accidents based on some driver's behaviors. The questions were asked through the short-structured questionnaire that includes the driver's demographics (age, education), involvement in an injury accident, distance traveled in a week (Km), driving experience, type of truck driving, presence of co-driver, sleeping on the wheel, and average daily driving hours. The drivers were randomly selected from Tipper garages, trailer parks, Nigerian National Petroleum Corporation depot, major markets in the state and the Luxurious Bus Park at Sarkin Yaki road. A summary of the survey is presented in Table 1. Half of the drivers interviewed (52.8%) were involved in an injury accident at least once in their career. Drivers driving an average of 5-9 hours were more involved in injury accidents (41%) than those with more and lesser driving hours. 44% of the drivers in injury accidents traveled over 4,000km weekly. Tanker drivers were also found to be more involved in accidents (51.1%) than other truck drivers. Almost half of the drivers had secondary school education with only 1.6% having degrees.

Table 1. Summary of questionnaire survey

Variables	Label	% Respondents	% Involved in injury accident
Involvement in an accident (*Y*)	Yes	52.82	
	No	47.18	
Age (X_1)	18 To 25	18.55	20.61
	26 To 30	23.39	20.61
	31 To 35	25.40	25.19
	36 To 40	24.19	22.14
	>40	8.47	11.45
Education (X_2)	Informal Education	22.98	20.61
	Primary	24.19	19.08
	Secondary	46.77	53.44
	Diploma	4.44	5.34
	Degree	1.61	1.53
License (X_3)	No	18.55	12.98
	Yes	81.45	87.02
Driving Experience (X_4)	0 To 5	30.65	29.77
	6 To 10	33.47	29.01
	11 To 15	22.18	21.37
	16 To 20	7.66	11.45
	>20	6.05	8.40
Average distance/week (X_5)	<2000km	27.42	22.90
	2000-4000km	42.34	32.83
	>4000km	30.24	44.27
Average Driving Hours/day (X_5)	< 5hrs	35.48	35.88
	5-9hrs	37.50	41.22
	>9hrs	27.01	22.90
Number of Days Driving/Week (X_6)	≤ 3 Days	9.68	6.87
	3-4 Days	30.65	22.90
	≥ 5 Days	59.68	70.99
Type of Vehicle (X_7)	Tanker	40.73	51.15
	Tipper	22.98	16.79
	Trailer	25.81	22.14
	Luxurious Bus	10.48	9.92
Sleep on Wheel (X_8)	No	66.13	49.62
	Yes	33.87	50.38
Sleeping Hours (X_9)	≤3hrs	13.31	13.74
	4hrs	24.60	28.24
	≥6hrs	62.10	58.02
Co-Driver (X_{10})	No	25.40	11.45
	Yes	74.60	88.55

Machine Learning Techniques

Naïve bayes

Naive Bayes is a probabilistic approach used for modelling classification issues, particularly text classification. It's formulated on the assumption that the individual variables used for classification are independent, contrary to real-world problems. It is built on the Naïve theorem, which reveals the chance of an event established on the erudition of the circumstances that might have been associated with it in the past [19]. The idea behind the Naïve Bayes algorithm is the posterior probability of a data instance t_i in a class c_j of the data model. The posterior probability $P(t_i|c_j)$ is the possibility of that t_i can be labeled c_j. $P(t_i|c_j)$ can be computed by multiplying all probabilities of all attributes of the data instance in the data model:

$$P(t_i|c_j) = \prod_{k=1}^{p} P(x_{ik}|c_j) \tag{1}$$

where p represents the number of elements in each data instance. The posterior probability for all classes is determined, and the class with the maximum probability will be the instance's label.

SVM algorithm

Support vector machine (SVM) is one of the most dominant ML algorithms due to its robustness in handling data uncertainty [20]. The approach is usually employed to establish the optimal decision boundary (hyperplane) separating different data sets. SVM seeks to identify the ideal hyperplane by maximizing the distance (known as the margin) between these data clusters [21]. In SVM, the margin signifies the separation between the nearest data points, also referred to as support vectors, and the hyperplane itself. Consequently, the primary objective in SVM is to locate the hyperplane that offers the most significant margin value, as this effectively reduces classification errors. The proposed SVM algorithm of the study is implemented using equation 2.

$$W^1 x + b = \emptyset \tag{2}$$

Where w is the weight vector of the orthogonal hyperplane, x represents the input in the dataset, and b is the bisector, and denote the dataset's null set. The study's proposed SVM algorithm is presented in Figure 1.

K-nearest neighbor (K-NN)

The K-NN is clear-cut, effectual, and straightforward technique used by researchers for modelling both regression and classification problems [22]. Some of the benefits of the K-NN techniques includes but not limited to its non-complicated nature that makes it easy to apply and comprehend. The K-NN technique was known for its strong ability in providing robust performance for both regression and classification. The only parameter that is tuned in the K-NN modelling is the K-parameter. The K-parameter is essential for obtaining good result using the K-NN model. The main principle of the technique is to pinpoint a group of "k" samples in the calibration data which are statistically similar to the nonentity samples. Cluster search is one of the best ways of finding the indefinite samples. The nonentity samples could also be obtained by averaging the response and contrasting it with to "k" samples K-NN used simple [23]. The three distance functions that determine the distance between neighboring points, as shown in Equations (3)–(5), were used for the regression problem:

$$F(e) = \sqrt{\sum_{i=0}^{f} \left(X_{obs_i} - X_{pre_i}\right)^2} \tag{3}$$

$$F(ma) = \sum_{i=0}^{f} |X_i - y_i| \tag{4}$$

$$F(mi) = \left(\sum_{i=0}^{f} |X_i - y_i|^q\right)^{1/q} \tag{5}$$

Whereas $F(e)$ indicates Euclidean function, $F(ma)$ indicates Manhattan function, $F(mi)$ is Minkowski function,

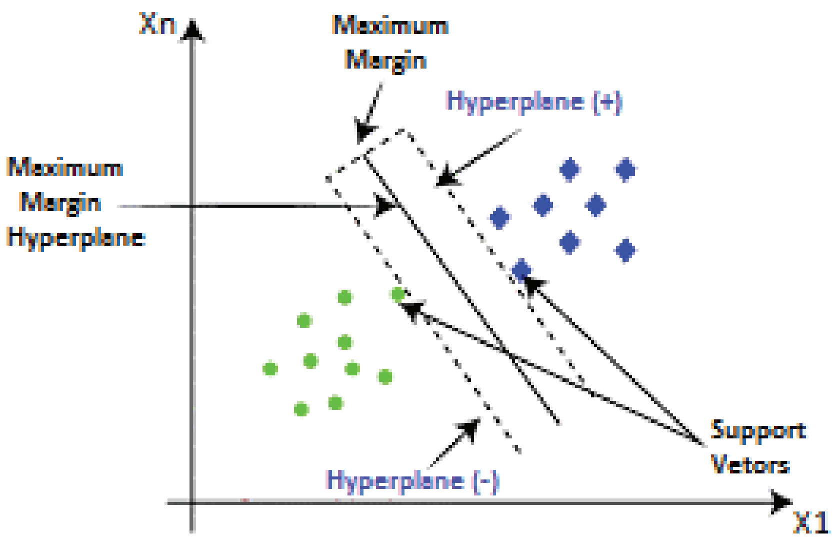

Figure 1. Flow diagram of the SVM algorithm.

xi and *yi* are ith dimensions, and *q* represents the order between the points *x* and *y*.

Performance Evaluation

The performance of the classifiers was evaluated using Four different metrics. The evaluation metric used are accuracy, F-1 factor, AUC-value and kappa value. The most common used metric for evaluating the overall performance of classifiers is the overall accuracy. However, the metric could be deceptive especially when the data is inclined to either true or false. This is owing to the fact that the majority class will outweigh and overshadow the minority class due to the data imbalance [24]. Kappa value could be defined as the amount of precision in the data due to the congruence amongst the data collectors. It is a form of standardized correlation coefficient, ranging from-1 to +1 that is generally used to determine for testing interclass reliability. An ideal fit between the actual and estimated data will have a kappa value of 1. The kappa-value can be used for interpretation the efficiency of the classifier. Mild classifiers have values between 0.01 and 0.2, regular classifiers have a value of 0.21-0.4, moderate classifiers have kappa value ranging from 0.41-0.60, 0.61-0.80 represent a substantial classifier and 0.81–1.00 are considered almost perfect [15]. The area under the curve (AUC) value of classifiers ranges between 1 and 0.5. It is obtained from the receiver operating characteristics (ROC) curve. A perfect model contains an AUC value of 1, while a model that cannot differentiate between true and false data will have an AUC value of less than 0.5. The optimum value will have an AUC value closer to 1 [25]. The equations for obtaining the models performance are:

$$\text{Accuracy} = \frac{TP + TN}{TP+TN+FN+FP} \tag{6}$$

$$\text{F1-Score} = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \tag{7}$$

$$Kappa = \frac{\Pr(a)-\Pr(e)}{1-\Pr(e)} \tag{8}$$

$$\Pr(a) = \frac{TP+TN}{TP+FP+FN+TN} \tag{9}$$

$$\Pr(e) = \frac{(TP+FP)(TP+FN)+(FN+TN)(TN+FP)}{(TP+FP+FN+TN)^2} \tag{10}$$

In which, TP is recorded when the model correctly classifies the fatal accidents as fatal accident, FP is recorded when the models incorrectly identify Injury accident as Fatal. FN is when the model incorrectly classifies a Fatal accident as Injury accident and TN is when the model correctly classifies correctly an injury accident.

RESULTS AND DISCUSSION

Feature Selection

Feature selection is an essential aspect of machine learning modelling be it regression or classification model. One of the major advantages of the feature selection process is that, it reduces the complexity of the model by removing redundant parameters. Reducing the model's complexity help improves the model's performance and reduces the computational cost and time. For ensuring that appropriate parameters were selected for the modelling, several feature selection algorithms (maximum relevance minimum redundancy, ANOVA, Chi-square) were applied in this study. The results show sleeping on wheels, presence of co-driver, the average distance traveled in a week, type of vehicle, driver's license, and driving experience as the major parameters influencing truck drivers' involvement in injury accidents. Sleepiness and long hours of driving increase fatigue, hence resulting in sleeping at the wheel, and when asleep, the driver loses his conscience, which

Table 2. Ranking of features

Parameter	Ranking			
	Kruskal Wallis	**Chi-square**	**ANOVA**	**MRMR**
Sleep on Wheel	1	1	1	1
Co-Driver	2	2	2	3
Average distance/week	3	3	3	7
Type of Vehicle	4	4	4	9
License	5	5	5	5
Education	6	7	7	6
Driving Experience	7	6	6	2
Number of Days Driving/Week	8	11	8	8
Sleeping Hours	9	8	9	4
Average Driving Hours/day	10	8	10	11
Age	11	9	11	10

will result in an accident. The findings of this study corroborate the findings by [8] and [6], where distance traveled and driving experience vehicle type were identified as the factors responsible for truck drivers' involvement in injury accidents. This is because those with higher mileage are more exposed to danger than those with lower mileage. Surprisingly, the presence of co-drivers, which is aimed at reducing fatigue and minimizing involvement in accidents, was ranked among the positive parameters influencing truck drivers' involvement in traffic accidents. This is understandable since most of the co-drivers of the truck drivers are novices with little or no experience learning to drive the trucks; hence, they could end up in an injury accident. In addition, most co-drivers are young less than 30 years of age. Duke et al. studied age-related safety of truck drivers and found young drivers (< 27 years) and old drivers (>63) to be involved in accidents than middle-aged drivers.

Classification Models

Three machine-learning techniques were employed to classify the accident severity in this study. Several models were developed through the Bayesian search optimization algorithm using each modeling technique, but only the optimum models were reported in this study. The performance of the models was evaluated using four evaluation metrics. The modeling results are presented in Table 3. From the result, it can be seen that the NVB models perform better than the other two machine learning techniques in terms of accuracy, F1-score, AUC, and Kappa values. The NVB outstripped the SVM and K-NN models by 6.9% and 10.5%, respectively, in the validation phase. The kappa values obtained which indicates show that K-NN and SVM have regular classifiers with kappa values between 0.2 and 0.4,

Table 3. Modelling results at the validation phase

Model	Accuracy (%)	F1-score	AUC	Kappa
K-NN	63.3	0.6061	0.6312	0.2628
SVM	66.9	0.6611	0.7352	0.3390
NVB	73.8	0.7257	0.7874	0.4748

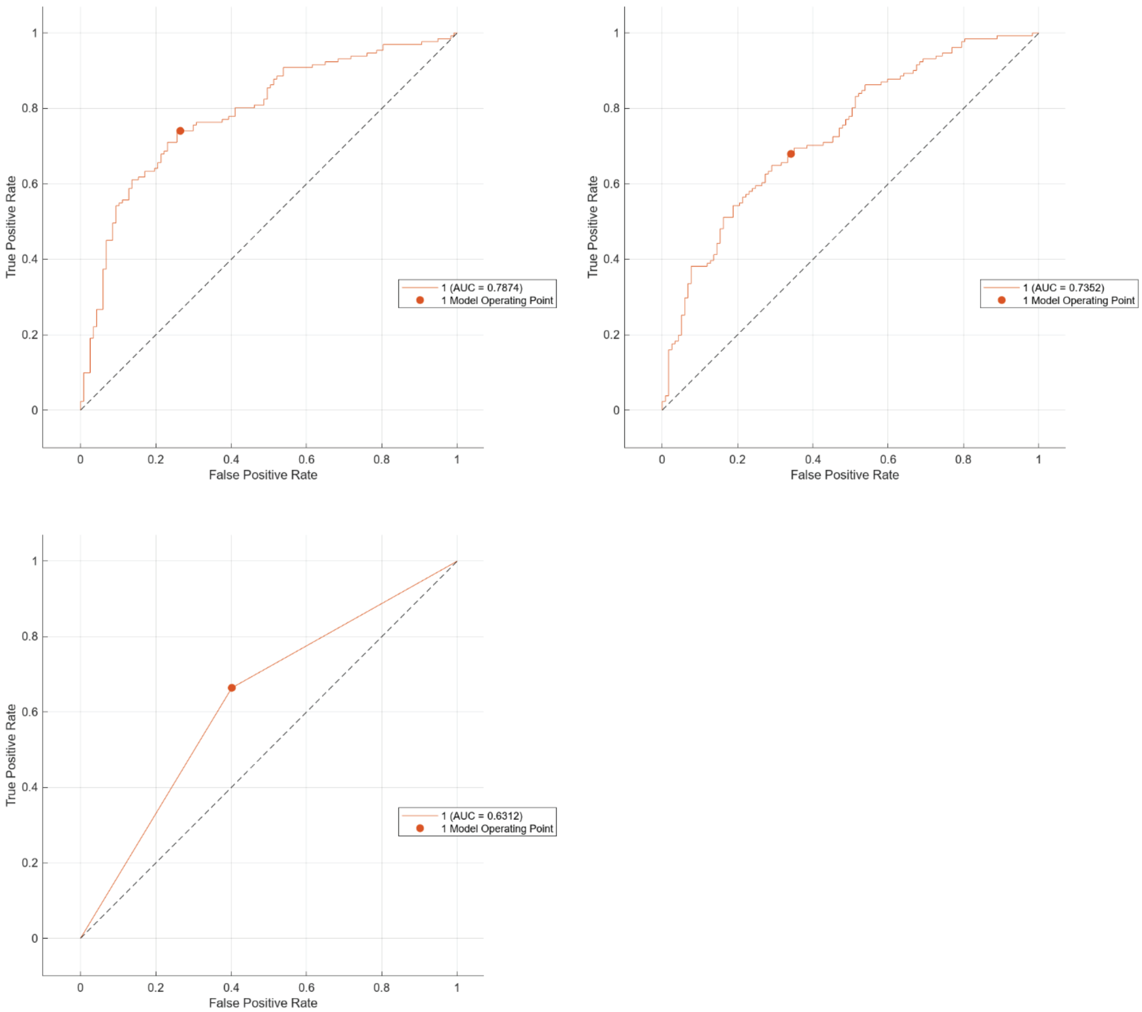

Figure 2. ROC curve a) NVB b) SVM c) K-NN.

while NVB models have moderate performance (0.4748), indicating higher classification ability of NVB over both K-NN and SVM. The F1-score, considered one of the most influential metrics measures for classification models as it represents the harmonic mean of recall and precision, also indicates better classification ability of the NVB model. The ROC curve of the models is presented in Figure 2. The AUC values were deduced from the ROC curve. The AUC value of classifiers ranges between 1 and 0.5. A perfect model has an AUC value of 1, while a model that could not differentiate between true and false data will have an AUC value of less than 0.5. The optimum value will have an AUC value closer to 1 [25]. All three models have AUC values greater than 0.5, with NVB having an AUC value of 0.7874. The AUC values show that the three models could classify the accident severity, with NVB being the best classifier.

Finally, the performance of the classifiers was compared with some studies in the literature using the overall accuracy and AUC measure. [26] modelled the severity of truck accident using logistic regression model, the logistic regression model classifies the accident severity with an accuracy of 70.9%. [27] achieved an AUC value of 76% using machine learning techniques for predicting driving risk among commercial truck drivers. RF, decision tree, and Instance-Based learning with parameter k model the severity of injury in a motorcycle crash with an accuracy of 73.91%, 73.64%, and 73.71%, respectively [28]. [13] compared the performance of two statistical and four machine learning techniques. The machine learning techniques outperformed the statistical models in predicting the accident severity with random forest having the highest accuracy of 53.9%. The NVB in the present study classifies the accidents severity with a higher accuracy and could hence be used as a decision-making tool by stakeholders.

CONCLUSION

A questionnaire survey was used to study the truck driver's involvement in an injury accident in Kano state, Nigeria. From the study, it can be seen that 100% of the truck drivers in the state are male, and 52.8% of the drivers were involved in injury accidents, making gender a non-significant factor for truck drivers' involvement in injury accidents. Sleeping on wheels, presence of co-driver, the average distance traveled in a week, type of vehicle, driver's license, and driving experience are the major factors responsible for drivers' involvement in an injury accident in the study area. The NVB model was the most suitable machine learning model for predicting the severity of truck-involved accidents in the state, surpassing the K-NN and SVM models. The NVB model could be used by the stakeholders as a tool for decision-making. The major limitation of the study was the use of the questionnaire survey due to the absence of detailed data from the relevant road safety agencies. Further studies could incorporate more parameters and employ other machine modeling techniques, such as random forests and decision trees.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] World Health Organization. Global status report on road safety. 2019.

[2] Sohail A, Cheema MA, Ali ME, Toosi AN, Rakha HA. Data-driven approaches for road safety: A comprehensive systematic literature review. Saf Sci 2023;158:105949. [CrossRef]

[3] Gökdağ M, Kaya MD, Atalay A, Haşiloğlu AS. Injuries and fatalities in Turkish road traffic accidents. Proc Inst Civ Eng Transp 2004;157:231–237. [CrossRef]

[4] Federal Road Safety Corps. Federal Road Safety Corps annual report. 2016.

[5] Talmage JB, Hudson TB, Hegmann KT, Thiese MS. Consensus criteria for screening commercial drivers for obstructive sleep apnea: Evidence of efficacy. J Occup Environ Med 2008;50:324–329. [CrossRef]

[6] Chen GX, Amandus HE, Wu N. Occupational fatalities among driver/sales workers and truck drivers in the United States, 2003–2008. Am J Ind Med 2014;57:800–809. [CrossRef]

[7] Kircher K, Andersson J. Truck drivers' opinion on road safety in Tanzania—a questionnaire study. Traffic Inj Prev 2013;14:103–111. [CrossRef]

[8] Meuleners L, Fraser ML, Govorko MH, Stevenson MR. Determinants of the occupational environment and heavy vehicle crashes in Western Australia: A case–control study. Accid Anal Prev 2017;99:452–458. [CrossRef]

[9] Edwards JR, Davey J, Armstrong KA. Profiling contextual factors which influence safety in heavy vehicle industries. Accid Anal Prev 2014;73:340–350. [CrossRef]

[10] Obasi IC, Benson C. Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents. Heliyon 2023 Jul 29;9:e18812. [CrossRef]

[11] Ahmed S, Hossain MA, Ray SK, Bhuiyan MMI, Sabuj SR. A study on road accident prediction and contributing factors using explainable machine learning models: Analysis and performance. Transp Res Interdiscip Perspect 2023;19:100814. [CrossRef]

[12] Umar IK, Gokcekus H. Modeling severity of road traffic accident in Nigeria using artificial neural network. J Kejuruter 2019;31:221−227. [CrossRef]

[13] Zhang J, Li Z, Pu Z, Xu C. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. IEEE Access 2018;6:60079−60087. [CrossRef]

[14] Zhang S, Khattak A, Matara CM, Hussain A, Farooq A. Hybrid feature selection-based machine learning classification system for the prediction of injury severity in single and multiple-vehicle accidents. PLoS One 2022;17:e0262941. [CrossRef]

[15] Jamal A, Zahid M, Tauhidur Rahman M, Al-Ahmadi HM, Almoshaogeh M, Farooq D, et al. Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study. Int J Inj Control Saf Promot 2021;28:408−427. [CrossRef]

[16] Murat YS, Şekerler A. Modelling traffic accident data by cluster analysis approach. İMO Teknik Dergi 2009;3:4759−4777.

[17] Tortum A, Atalay A. Clustering analysis of traffic accident risk in Turkey. Iranian J Public Health 2015;44:425.

[18] Murat YS, Cakici Z. An integration of different computing approaches in traffic safety analysis. Transp Res Procedia 2017;22:265−274. [CrossRef]

[19] Gao H, Zeng X, Yao C. Application of improved distributed naive Bayesian algorithms in text classification. J Supercomput 2019;75:5831−5847. [CrossRef]

[20] Wang WC, Xu DM, Chau KW, Chen S. Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. J Hydroinformatics 2013;15:1377−1390. [CrossRef]

[21] Adamu M, Haruna SI, Malami SI, Ibrahim MN, Abba SI, Ibrahim YE. Prediction of compressive strength of concrete incorporated with jujube seed as partial replacement of coarse aggregate: A feasibility of Hammerstein–Wiener model versus support vector machine. Model Earth Syst Environ 2021;1−11. [CrossRef]

[22] Akbulut Y, Sengur A, Guo Y, Smarandache F. NS-k-NN: Neutrosophic set-based k-nearest neighbors classifier. Symmetry 2017;9:179. [CrossRef]

[23] Hussain S, Muhammad Khan N, Emad MZ, Naji AM, Cao K, Gao Q, et al. An appropriate model for the prediction of rock mass deformation modulus among various artificial intelligence models. Sustainability 2022;14:15225. [CrossRef]

[24] Parikh R, Mathai A, Parikh S, Sekhar GC, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 2008;56:45−50. [CrossRef]

[25] Kaur H, Kumari V. Predictive modelling and analytics for diabetes using a machine learning approach. Appl Comput Inform 2022;18:90−100. [CrossRef]

[26] Umar IK, Bashır S. Investigation of the factors contributing to truck driver's involvement in an injury accident. Pamukkale Univ Eng Sci J 2020;26:402−408. [CrossRef]

[27] Mehdizadeh A, Yazdi MAA, Cai M, Hu Q, Vinel A, Rigdon SE, et al. Predicting unsafe driving risk among commercial truck drivers using machine learning: Lessons learned from the surveillance of 20 million driving miles. Accid Anal Prev 2021;159:106285. [CrossRef]

[28] Wahab L, Jiang H. A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. PLoS One 2019;14:e0214966. [CrossRef]