Nigerian Journal of Technology (NIJOTECH)
Vol. 44, No. 2, June, 2025, pp.215 - 221
www.nijotech.com

Print ISSN: 0331-8443
Electronic ISSN: 2467-8821
https://doi.org/10.4314/njt.v44i2.5

ENHANCING ROAD CRASH PREDICTION: A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS AND SAFETY PERFORMANCE FUNCTIONS ON THE LAGOS-IBADAN EXPRESSWAY

AUTHORS:
O. Bayode[1,*], O. Aiyelokun[2], O. Osanyinlokun[3], and A. Adanikin[4]

AFFILIATIONS:
[1,2]Department of Civil and Environmental Engineering, Bells University of Technology, Ota, Ogun State, Nigeria
[2]Department of Civil Engineering, University of Ibadan, Oyo State, Nigeria
[3]Department of Civil Engineering, Bamidele Olumilua University of Education, Science and Technology, Ikere- Ekiti, Ekiti State, Nigeria
[4]Harry Building Concept, Okota-Isolo, Lagos State, Nigeria.

***CORRESPONDING AUTHOR:**
Email: ope.olowo.oo@gmail.com

ARTICLE HISTORY:
Received: 06 December, 2024.
Revised: 06 May, 2025.
Accepted: 13 May, 2025.
Published: 07 July, 2025.

KEYWORDS:
Road Traffic Crash Prediction, Safety Performance Function, Machine Learning, SVM, RF, XGBoost.

ARTICLE INCLUDES:
Peer review

DATA AVAILABILITY:
On request from author(s)

EDITORS:
Chidozie Charles Nnaji

FUNDING:
None

Abstract

Road traffic crash prediction (RTCP) is a critical aspect of transportation safety, enabling the identification of high-risk locations and informing the implementation of proactive measures. This study explores the comparative performance of Machine Learning (ML) algorithms and traditional Safety Performance Functions (SPFs) to predict road traffic crashes along the Lagos-Ibadan Expressway, a major highway in Nigeria known for its high crash rates. To achieve the objective, SPFs estimated using Negative Binomial Regression (NBR) and ML regression models mainly Support Vector Machine (SVM), Random Forest (RF) and Extreme Gradient Boosting (XGBoost) were developed using historical crash data collected from Federal Road Safety Commission (FRSC) of Nigeria for 10years duration between 2014 and 2023, traffic components and geometric design features as input variables. The study's findings indicate that ML algorithms outperform SPFs in terms of predictive accuracy and sensitivity to complex, non-linear relationships among crash-contributing factors with R^2 of 0.99, 097 and 0.84 for training and 0.93,0.9 and 0.76 for testing dataset in the three ML models. However, SPFs remain advantageous in interpretation and ease of implementation. The analysis also highlights the importance of feature selection, with variables such as traffic volume, traffic speed, road curvature and pavement width emerging as significant predictors. Furthermore, this study offers insights for policymakers, traffic engineers, and researchers seeking to improve road safety outcomes through data-driven crash prediction methods. The results emphasize the potential of integrating ML techniques with traditional methods to develop hybrid frameworks for enhanced crash prediction and prevention strategies on high-risk roadways.

1.0 INTRODUCTION

Road safety has become a paramount concern due to the increasing number of vehicles and the complexity of traffic systems. In Nigeria, a road network system connects all parts of the country, and 80 per cent of freight and passenger traffic is transported through this transportation system [1]. Moreover, road traffic crashes are a significant cause of injury and death worldwide, necessitating effective predictive models to improve road safety [2]. Traffic engineers have introduced a crash predictive model to analyse the safety performance of highways and provide countermeasures to mitigate crash occurrence. These models are used for several applications, such as identifying the factors influencing road crashes, including traffic volume, traffic control devices, lanes,

----
HOW TO CITE:
Bayode, O., Aiyelokun, O., Osanyinlokun, O., and Adanikin, A. "Enhancing Road Crash Prediction: A Comparative Study of Machine Learning Algorithms and Safety Performance Functions on the Lagos-Ibadan Expressway", *Nigerian Journal of Technology*, 2025; 44(2), pp. 215 – 221; https://doi.org/10.4314/njt.v44i2.5

shoulders, intersections, geometric features, and traffic flow, and detecting and ranking hotspots [3],[4]. The Highway Safety Manual (HSM) relays the Safety Performance Function (SPF) as the vital regression tool to predict crash frequency using the Poisson and Negative Binomial regression model over a set road and traffic survey [5]. SPF has been proven to be a reliable tool for crash prediction.

Safety Performance functions (SPFs) represent the traditional regression model proposed in the Highway Safety Manual (HSM) for highway crash prediction and estimating expected numbers of crashes over some time for a given roadway data set [5]. The function employs Poisson or Negative Binomial models to screen variables for predicting motor crash occurrence [3],[7]. Negative Binomial regression is an improved version of the Poisson regression model, with provision for fitting and over-dispersion of crash data where variance exceeds mean and significant discrepancies between the observed and predicted values. It also has inherent characteristics of crashes, which are random, discrete, and non-negative [10],[11]. Negative binomial regression has been demonstrated to be an acceptable statistical tool for the prediction of count data; several safety researchers have utilized it to determine the correlations between road vehicle crashes and factors influencing road crashes [1], [2], [12] and [13].

Nonetheless, researchers have identified significant limitations in SPFs as predictive tools [6],[7]. The analyst assumes probability distribution and pre-defined functional form, determining whether the crash parameters are linear or exponential; this might introduce potential bias into the model. Also, the difficulty of the proper trade-off of over-fitting, under-fitting and correlation analysis of independent variables in trained data is frequently disregarded, affecting output accuracy and leading to high prediction errors when results are transferred to other sites or locations [9].

More recently, Machine Learning (ML) algorithm has been introduced to highway safety to estimate and predict crash frequency, evaluate large volumes of data, identify accident trends, and identify viable routes for accident prevention and traffic management. Performance evaluation of ML algorithms are critical for ensuring prediction models are accurate, dependable, and applicable in real-world settings [8],[9]. Multiple metrics and techniques are employed in ML substantiation to evaluate the performance of ML models and establish their efficiency in forecasting road safety outcomes. It entails the assessment of the model's robustness, correctness, and application using several traffic characteristics and geographical locations [3],[7].

Machine learning (ML) models have served as an alternative method for crash prediction in recent years, and several researchers have compared the performance of MLs and SPFs to determine the accuracy of the models for crash prediction. Moreover, [4] and [15] investigated the efficiency of SVM in predicting crash frequency at traffic analysis zones for rural highway sections. The study compared the performance of SVM with fixed-effect and random-effect negative binomial models. The results indicate that SVM performed better in terms of prediction accuracy. [16] employed RF technique to analyze angular crashes at unsignalised intersections, effectively combining RF with multivariate adaptive regression splines. Also, [17] employed the XGBoost-Bayesian network model to predict traffic crash severity on freeways in China, understanding the interaction between features influencing severity in traffic crashes.

Considering this new study area, few existing approaches compare different ML algorithms to utilize accuracy in road crash prediction. As a result, it is necessary to investigate novel technologies to improve the prediction of road crashes.

The main objective of this study is to compare three ML algorithms, namely Support Vector Machine (SVM), Random Forest (RF) and Extreme Gradient Boosting (XGBoost) to SPF using Negative binomial regression model (NBRM) in the prediction of road traffic crash occurrence along Lagos- Ibadan highway in Nigeria. Ultimately, a general understanding of the relationship between highway geometric characteristics and traffic accident frequency is vital for transportation planners, engineers, and policymakers to implement effective measures to reduce the frequency and severity of roadway accidents and their general performance, thereby enhancing overall transportation safety and efficiency.

2.0 METHODOLOGY

The study methodology adopted a mixed-methods approach, incorporating both Secondary and primary data through data extraction. The method are subdivided into Data Collection, Data processing, Modelling, Analysis and Model Evaluation. Figure 1 shows the conceptual workflow of the study methodology.

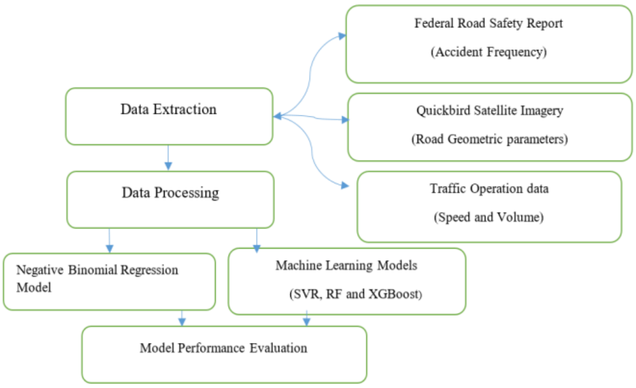

Figure 1: Conceptual workflow chart

2.1 Data Collection

The study focuses on multi-lane highways linking Nigeria's three major States (Lagos, Ogun and Oyo) Figure 2 shows the map of study area. Accident data for the highway under study was obtained from the Federal Road Safety Corps (FRSC) of the state command for a 10-year observation period from 2014 to 2023. The road geometric indices such as curve radius, length of horizontal curve, length of vertical curve and sight distance were downloaded and extracted from high-resolution satellite imagery known as Quickbird Satellite Imagery (QSI). The other road geometric parameters, such as pavement width, shoulder width, median width and elevation, were measured on-site to understand the road's characteristics. Stopwatch approach was utilised to compute the spot speed of passenger vehicles [19]. Traffic count was also employed to determine the number of vehicles using the facility. The numbers of vehicles were counted at each location for one week at 12 hours daily to compute the Average Daily Traffic (ADT) [20]. Table 1 illustrates the statistical description of independent variables.

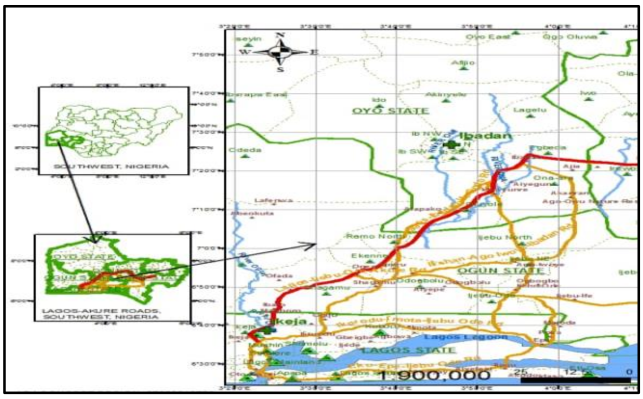

Figure 2: Map of study area

Table 1: Statistics attributes of independent variables

Variable	Min	Max	Mean	Variance	S.D
Average Daily Traffic (ADT)	8756	35423	18193.73	798193	8934.17
Speed Car (SC) (km/h)	60.82	120	97.67	181.2	13.45
Sight Distance (SD) (m)	550	2000	940.607	146083	382.208
Curve Radius (R)(m)	10	687179	6684.48	3.58e+09	59854.54
Horizontal Curve length (HCL)(m)	0.646	3123.05	859.645	403084.7	634.89
Vertical curve length (VCL) (m)	0	150	143.227	469.68	21.672
Pavement width (PW)(m)	8	12	9.5721	3.383	1.839
Shoulder Width (SW) (m)	1.8	4	2.947	0.55596	0.23578
Median Width (MS) (m)	1.5	6	5.5069	0.219	0.4683
Grade (G) (%)	-3.42	4.44	2.84	0.432	0.498
Speed Truck (ST) (km/h)	40.09	100	64.3	195.09	13.437

2.2 Data Processing

The data was analysed using python. The data samples were divided into training (80%) and testing (20%) for fitness and prediction. The data was normalised because of the dissimilar units, magnitudes, and nature of variables. Data normalisation was used to improves data fitting and prediction performance utilizing Equations 1 and the variables were denormalized using Equation 2 for easy interpretation of models output.

$$x_n = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (1)$$

$$\acute{y} = y_{ni}(\max(y_i) - \min(y_i)) + \min(y_i) \quad (2)$$

Where, x_n is the normalised variable, x_i is the initial ivariable, y is the denormalised variable, and y_i is the initial normalisedvariable.

2.3 Negative Binomial (NB) Regression

The negative binomial regression used to compute the probability function is expressed in Equations 3 and 4 [1] [3] [9].

$$Prob(Yi = yi) = \frac{\Gamma(yi+\Phi)}{\Gamma(yi+1)\Gamma(\Phi)}\left(\frac{\mu i}{\mu i+\Phi}\right)^{yi}\left(\frac{\Phi}{\mu i+\Phi}\right)^{\Phi} \quad (3)$$

The expectation of Yi is μi = g (xi)

Variance of Yi is

$$\mathrm{Var}\ (Yi) = \mu i + \frac{\mu i^2}{\Phi} \quad (4)$$

Where, Yi is the accident number collected at site I, Yi is the dependent variable following the NB distribution with the inverse dispersion parameter Φ, xi is a vector representing the accident-related variables at site I, and g(xi) is the functional form of the NB regression model [31].

2.4 Machine Learning Models

2.4.1 Support vector machine

Equations 5 and 6 describe the support vector machine (SVM) algorithm employed for analysis [4] and [9].

$$ý = f(x) = w^T \Phi(x) + b \quad (5)$$

$$\min Z(w, \varepsilon, \xi i, \xi ii) = \frac{1}{2} w^T + C\left[v\varepsilon + \frac{1}{N}\sum(\xi i + \xi ii)\right] \quad (6)$$

Subject to $w^T \Phi(x(i)) + b - y(i) \leq \varepsilon + \xi i$ vi= 1,……,n$w^T \Phi(x(i)) + b - y(i) \leq \varepsilon + \xi i$ vi= 1,……,n
y(i) - $w^T \Phi(x(i)) + b - y(i) \leq \varepsilon + \xi i$ vi= 1,……,n
$\xi i, \xi ii \geq 0$ vi= i……, n
Where, ξi and ξii are slack variables, C is a regularisation parameter, and v is a second parameter for each x(i), the allowable error is ε. Function Φ(x(i)) to linearise the nonlinear relationship between x(i) and y(i). Assuming the input training variables x(i) for 1,….,n and training variable y(i) for,…..,n.

2.4.2 Random forest

This study adopted the random forest (RF) algorithm for predicting road accident frequency. RF randomly combines the predictions of several decision tree techniques by estimating the node's average result [7]. The principle has been found to resist overfitting and provide approximate generalised error [21] [22]. The mathematical formation of RF is expressed in Equation 7.

$$R.F_{sub(x)} = \frac{1}{x}\sum_{x=1}^{x} f_x(R) \quad (7)$$

Where, $R.F_{sub(x)}$ is the importance of feature (x) calculated from all trees in the random forest model, x is the total number of trees, and R is the normalized feature importance for (x) trees.

2.4.3 Extreme gradient boosting

Extreme Gradient Boosting (XGBoost) algorithms, a type of machine learning, were also used in this research for road accident predictions. The algorithm was proposed by Chen and Guestrin 2016; this method combines many weak learning models to produce a robust learning model [23]. XGBoost uses a tree model for integration. It is an improved gradient lifting algorithm [17]. The mathematical objective function of the model is expressed in Equation 8, a function of the second derivative of Taylor expansion.

$$Obj^t = \sum_{t=1}^{n}(l(y_i, \bar{y}^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} l_i f_t^2(x_i)) + \Omega(f_t) + C \quad (8)$$

Where, i is the ith crash sample in the dataset, l is the model's loss function, which measures the difference between the real value y_i and the predicted value, Ω is the complexity function of the model, and C is the constant.

2.4.4 Model performance evaluation criteria

The model's performance was evaluated and compared using some criteria such as the coefficient of determination (R^2) expressed in Equation 9, root mean square error (RMSE) in Equation 10, and Mean Absolute error (MAE) in Equation 11.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(ý - y_i)^2}{\sum_{i=1}^{n}(ý - \bar{y})^2} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ý - y_i)^2} \quad (10)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |ý - y_i| \quad (11)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (12)$$

Where, n is the number of observations, y_i is the actual value of the ith observation, ý is the predicted value of the ith observed and ȳ is the average value of observations computed with Equation 12.

3.0 RESULT AND DISCUSSION

3.1 Negative Binomial Output

The model's coefficients are similar to the variable coefficients, as indicated in Table 2. Pavement width is significant with a p-value less than 0.05 and had a negative sign effect ($\beta = -3.322$), signifying that a unit increase in lane width will cause a decrease in crash frequency if other variables are kept constant. The shoulder width shows a negative sign effect ($\beta = -0.668$) with significant ($p < 0.05$) to accidents. Similarly, vertical curve length and radius of curvature have a significant association with road accidents ($p < 0.05$) with a negative coefficient of ($\beta = -0.002$ and 0.52, respectively). Additionally, variables such as speed of the vehicle, sight distance and Average daily traffic all have significant ($p < 0.05$) associations with roads with a positive sign indicating a unit increase in all these parameters will cause an increase in crash frequency while in contrasts horizontal curve length has no significant association with road crash from the model simulation. Based on the coefficient of estimation, the combined model produced a higher number of variables or predictors that have a significant association with road accidents [25] [26] [27][31].

Table 2: Estimated coefficients of the negative binomial regression model for the roads

Accident	Estimates	Std. Err.	Z	P-value	[95% Conf. Interval]	
Intercept	-21.513	8.3134	6.697	0.01	8.98E-16	1.1584E+11
Pavement width(PW) (m)	-3.322	1.0537	9.937	0.002	0.051	53.185
Shoulder Width (LW) (m)	-0.668	0.7487	0.797	<0.01	0.002	0.629
Median Width (MW) (m)	0.796	0.1854	18.423	0.32	1.338	8.169
Sight Distance (SD) (m)	32.532	9.3396	12.133	<0.01	7.10E-08	3.3285E+20

Speed Truck (ST) (km/h)	0.014	0.0029	21.562	<0.01	0.998	1.014
Speed Car (SC) (km/h)	0.018	0.0079	1.06	0.003	0.988	1.057
Average Daily Traffic (ADT)	2.225	0.3816	34.015	<0.01	17.443	877.304
Grade (G)	0.028	0.0129	4.618	0.032	0.918	1.002
Vertical curve Length (VCL)(m)	-0.002	0.0006	11.877	0.001	0.991	0.996
Radius of curve (R) (m)	-0.52	0.0008	0.015	0.001	0.997	1.003
Horizontal curve Length (HCL)(m)	0.035	0.019	3.406	0.065	0.906	1.033

3.2 ML Model Variable Correlation and Importance

The correlation and importance of each variable evaluated show the interplay of the features highlighted to contribute to the prediction of accident frequency. Figure 3(a) illustrates a thermal diagram of the correlation between the variables. The variables are sorted from large to minor according to the correlation coefficient of the independent variables to the dependent variable (accident frequency). The correlation sorted the variables in the following arrangement as indicated in the figure: pavement width, shoulder width, median width, speed of car, speed of truck, average daily traffic, sight distance, length of vertical curve, elevation, and grade. These are consistent with [27], [28].

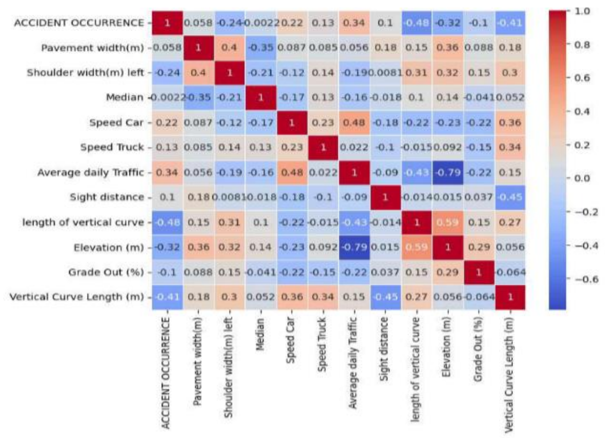

Figure 3(a): Variable Correlation

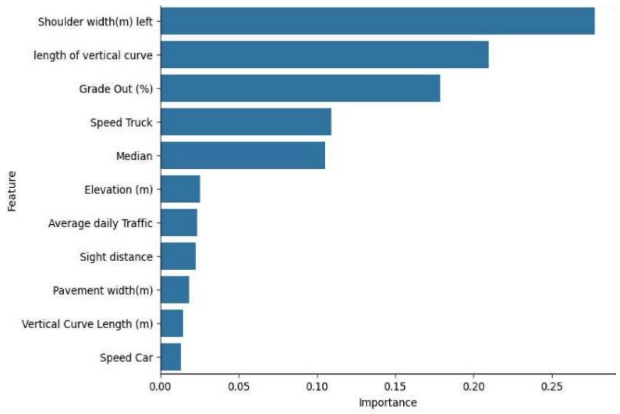

Figure 3(b): Variable Importance

Figure 3(b) indicates the degree of importance of the variables to accident frequency prediction. The importance was sorted according to the mean value from the absolute value obtained from the degree of impact on the target variable. The diagram indicates that shoulder width has the parameter with the highest global impact and contributing factor to accident occurrence along the highway. The findings are consistent [29], [30].

3.3 Model Performance Evaluation Comparisons

Table 3 shows the summary of the result of the performance evaluation conducted on all the models. Machine Learning (ML) model performance was compared to the Negative Binomial (NB) model for both the training and testing datasets. From the result, SVM and DT outperformed other models in training and testing datasets. Both models' coefficient of determination (R^2) were 0.9985 and 0.9769, respectively, tending to 1, which indicates 99.9% and 97.7% data fitting in the training dataset. At the same time, XGBoost has a slightly better data fitting R^2 value of 0.8453 compared to the NB model with 0.6775. SVM's root mean square error (RMSE) and Mean Absolute error (MAE) recorded the lowest value across all models, indicating better transferability and less prediction error in the training data. Similarly, for the testing data, SVM and RF displayed more accuracy in predicting accident frequency with the highest values of R^2; RF has the lowest RMSE and MAE, indicating better transferability and prediction error than other models. Overall, the performance evaluation suggests that ML models, especially SVM and DT models, are better suited for accident frequency predictions and have better goodness of fit and accuracy.

Table 3: Performance evaluation comparison

Training Set			
Performance Evaluation	**R^2**	**RMSE**	**MAE**
SVM	0.9985	0.9784	0.9702
RF	0.9769	3.5029	2.2635
XGBoost	0.8453	8.0947	6.1866
NB	0.6775	10.423	10.436
Testing Set			
SVM	0.9315	6.2399	4.9789
RF	0.9046	5.9952	4.9366
XGBoost	0.7696	9.4419	8.3223
NB	0.6544	12.435	13.423

4.0 CONCLUSION

This research considers the influence of road parameters and traffic conditions in predicting traffic accident frequency, implementing machine learning and traditional negative binomial regression models. The following findings are highlighted from the research;

- Negative binomial regression reveals that road parameters such as pavement width, shoulder width, and length of vertical curve contribute significantly to the occurrence of traffic accidents. The model also indicates traffic conditions, including car and truck speeds and average daily traffic, as an influential factor in traffic accident occurrence, with a significant coefficient of P-value less than 0.05.
- Similarly, the machine learning (ML) model correlation and importance of variables indicate pavement width, shoulder width, length of vertical curves, car and truck speeds, Average daily traffic, and sight distance as essential features contributing to traffic accident occurrence along highways.
- Finally, the performance evaluation comparison of all the models suggests that machine learning models are satisfactory and adequate for predicting traffic accident frequency. However, Support Vector Machine (SVM) has lesser data overfitting and better performance compared to Random Forest (RF) and Extreme Gradient Boosting (XGBoost).

This research will promote new ideas for developing innovative methods in highway safety. Although ML-based models provide positive results in predicting traffic accident frequency, further research should be explored to evaluate the models using new datasets across other jurisdictions. Evaluating and validating other ML models can also be explored to predict road traffic accidents and improve road safety.

REFERENCE

[1] Popoola, M. O., Abiola, O. S. and Odunfa, S. O. "Effect of Traffic and Geometric Characteristics of Rural Two Lane Roads on Traffic Safety: a case study of Ilesha-Akure-Owo road, South-West, Nigeria," *FUOYE Journal of Engineering and Technology*, vol. 3, no. 2, pp. 125-130, 2018. doi.org/10.46792/fuoyejet.v3i2.256.

[2] Bayode, O., Aderinola, O. S., Oluyemi-Ayibiowu, B. D., Okungade, O and Opeyemi, D. "Impact of Geometric Properties of Two Lane Highway on Traffic operation and Safety," *UI Journal of Civil Engineering and Technology*, vol. 6, no. 1, pp. 15-25, 2024.

[3] Li, X., Lord,D., Zhang, Y. and Xie, Y. "Predicting motor vehicle crashes using Support Vector Machine models.," *Accident Analysis and Prevention*, vol. 40, pp. 1611-1618, 2008. doi:10.1016/j.aap.2008.04.010

[4] Dong, N., Huang, H. and Zheng, I. "Support vector machine in crash prediction at the level of traffic analysis zones: Assessing the spatial proximity effects.," *Accident Analysis and Prevention*, vol. 82, pp. 192-198, 2015. doi.org/10.1016/j.aap.2015.05.018.

[5] Highway Safety Manual (HSM). American Association of State Highway and Transportation Officials, Washington, D.C., 2010.

[6] Lord, D. and Mannering, F. "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives.," Transportation research part A: policy and practice, vol. 44, no. 5, pp. 291-305, 2010.

[7] Tayebikhorami, S., and Sacchi, E. "Validation of Machine Learning Algorithms as Predictive Tool in the Road Safety Management Process: Case of Network Screening," *Journal of Transportation Engineering: Part A*, vol. 148, no. 9, pp. 1-30, 2022. doi/pdf/10.1061/JTEPBS.0000719.

[8] Chang, L. Y. "Analysis of freeway accident frequencies: Negative binomial 465 regression versus artificial neural network.," *Safety Science*, vol. 43, p. 541–557, 2005.

[9] Xie, Y., Lord, D., and Zhang, Y. "Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis.," *Accident Analysis and Prevention*, vol. 39, no. 2, pp. 922-933, 2007.

[10] Rezapour, M., Molan, A. M., and Ksaibati, K. "Analyzing Injury Severity of Motorcycle At-Fault Crashes Using Machine Learning Techniques, Decision Tree and Logistic Regression Models.," *International Journal of Transportation Science and Technology*, vol. 9, no. 2, pp. 89-99, 2020.

[11] Vitianingsih, A. V., Suryana, N., and Othman, Z. "Spatial Analysis Model for Traffic Accident-Prone Roads Classification: A Proposed Framework.," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, pp. 365-370, 2021.

[12] Pan, C. and Prakash, R. "Modeling Motorway Accidents using Negative Binomial Regression," *in Proceedings of the Eastern Asia Society for transportation Studies*, 2013.

[13] Omari-Sasu, O., and Adejei, O. "Statistical Models for Count Data with Applications to Road Accidents in Ghana.," *International Journal of Statistics and Applications*, vol. 6, no. 3, pp. 123-137, 2016.

[14] Shaik, M. E., Islam, M. M., and Hossain, Q. S. "A review on neural network techniques for the prediction of road traffic accident severity," *Asian Transport Studies*, vol. 7, no. 10, p. 1000040, 2021.

[15] Singh, G., Sachdeva, S. N., and Pal, M. "Support vector machine model for prediction of accidents on non-urban sections of highways.," *in Proceedings of the Institution of Civil Engineers Transport*, 2018.

[16] Abdel-Aty, M., and Haleem, K. "Analysing angle crashes at unsignalised intersections using machine learning techniques," *Accident Analysis and Prevention*, vol. 43, no. 1, pp. 461-470, 2011.

[17] Yang, Y., Wang, K., Yuan, Z., and Liu, D. "Predicting Freeway Traffic Crash Severity Using XGBoost-Bayesian Network Model with Consideration of Features Interaction," *Journal of Advanced Transportation*, pp. 1-16, 2022. doi: 10.1155/2022/4257865.

[18] Aderinola, O. S., and Laoye, A. A "Traffic Crashes Prediction of States in Nigeria using time series Analysis," *Global Journal of Engineering and Technology Advances*, vol. 3, no. 1, pp. 15-26, 2020.

[19] Oyedepo, O. J., and Makinde, O. O. "Accident Prediction Models for Akure-Ondo Carriageway, Ondo State Southwest Nigeria: Using Multiple Linear Regression," *African Research Review*, vol. 4, no. 2, pp. 30-49, 2010.

[20] Aderinlewo, O. O., and Afolayan, A. "Development of Road Accident Prediction Models for Akure-Owo Highway, Ondo State," *Journal of Engineering Science*, vol. 15, no. 2, pp. 53-70, 2019.

[21] Silva, P. B., Andrade, M., and Ferreira, S. "Machine Learning Applied to Road Safety Modelling: A Systematic Literature Review," *Journal of Traffic and Transportation Engineering*, vol. 7, no. 6, pp. 775-790, 2020.

[22] Vanitha, R and Swedha, M. "Prediction of Road Accident using Machine Learning Algorithms," *Middle-East Journal of Applied Science and Technology*, vol. 6, no. 2, pp. 64-75, 2023.

[23] Chen, T. and Guestrin, C. "XGBoost: A Scalable Tree Boosting System," San Francisco, CA, USA,, 2016.

[24] Li, L., Prato, C. G., and Wang, Y. "Ranking contributors to traffic crashes on mountainous freeways from an incomplete dataset: A sequential approach of multivariate imputation by chained equations and random forest classifier," *Accident Analysis and Prevention*, vol. 146, no. 1, p. 330–345, 2020.

[25] Alvarez, P., Fernandez, M. A., Gordaliza, A., Mansilla, A. and Molinero, A. "Geometric road design factors affecting the risk of urban run-off crashes," *A case-control study: PLoS One*, vol. 15, no. 6, p. e0234564, 2020.

[26] Madushani, J. S., Sandamal, R. K., Meddage, D. P., Pasindu, H. R., and Gomes, P. A. "Evaluating expressway traffic crash severity using logistic regression and explainable and supervised machine learning classifiers," *Transportation Engineering*, vol. 13, no. 1, pp. 13-20, 2023.

[27] Mohammadnazar, A., Mahdinia, I., Ahmad, N., Khattak, A. J., and Liu, J. "Understanding how relationships between crash frequency and correlates vary for multi-lane rural highways: Estimating geographically and temporally weighted regression models," *Accident Analysis and Prevention*, vol. 157, no. 1, p. 106146, 2021.

[28] Berhanu, Y., Schroder, D., Wodajo, B. T., and Aliimayehu, E. "Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing on accident and congestion-prone road networks," *Result in Engineering*, vol. 23, no. 102737, pp. 1-22, 2024. doi: 10.1016/j.rineng.2024.102737.

[29] Bayode, O., Aderinola, O. S., and Oluyemi-Ayibiowu, B. D. "Application of Machine Learning for Road Safety Modeling of Selected South-West Highway in Nigeria," *European Journal of Applied Science, Engineering and Technology*, vol. 3, no. 3, pp. 202-213, 2025. doi: 10.59324/ejaset.2025.3(3).13

[30] Zhang, X., Waller, S. T., and Jiang, P. "An Ensemble Machine Learning-Based Modeling Framework for Analysis of Traffic Crash Frequency.," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 3, pp. 258-276, 2020.

[31] Onuwurah, U. O., Ihueze, C. C., and Nwankwo, C. O. "Modelling Road Traffic Crash Variables in Anambra State, Nigeria: An Application of Negative Binomial Regression," *Journal of Multidisciplinary Engineering Science Studies*, vol. 7, no. 6, pp. 3942-3950, 2022.