Count Data Modeling for Predicting Crash Severity on Indian Highways

Krantikumar V. Mhetre
Department of Civil Engineering, COEP TECH University, Pune, Maharashtra, India
kvm18.civil@coeptech.ac.in (corresponding author)

Aruna D. Thube
Department of Civil Engineering, COEP TECH University, Pune, Maharashtra, India
adt.civil@coeptech.ac.in

Received: 5 July 2023 | Revised: 18 August 2023 | Accepted: 9 September 2023

 DOI: https://doi.org/10.48084/etasr.6172

ABSTRACT

This study collected data on road accidents for the years 2016-2020 for the NH-48 highway in Maharashtra, India to model their conditions. Road crash data models were developed using 70% of actual data for training and 30% for testing purposes. Negative binomial regression modeling was used to predict crash fatalities. The results showed that the factors that affected the fatality of road crashes were head-on-collision, friction, time zone, and weather conditions of the crash. The developed models were validated and tested using log-likelihood, AIC, BIC, MAD, MSE, RMSE, and MAPE values. Head-on-collision, AM, PM, light rain, mist/fog, heavy rain, fine, and cloudy were positively associated with the fatality of road crashes, while friction was negatively associated. The developed models can be used to predict the fatality/non-fatality of road crashes and implement road safety strategies on highways to reduce them.

Keywords-crash; data; modeling; NH-48; road safety; India

I. INTRODUCTION

Road crashes result from a variety of factors, such as road geometry, road traffic, vehicle conditions, human faults, etc. As these factors include a wide variety of subfactors, road crash analysis becomes a complex phenomenon to achieve a single-point solution. India is ranked first in the world in terms of road crash deaths [1]. Between 2020 and 2021, there was an increase of 12.6% in road crashes, 16.9% in deaths, and 10.39% in injuries [2].

A. Road Crash Data Modeling in India

In [3], a correlation was discovered between highway geometric parameters, traffic parameters, and crash rates for two-lane non-urban highways, showing a positive connection between the volume of heavy vehicles, speed, and crash rates, while shoulder and lane widths showed a negative relationship with crash rates. In [4], a random parameter model was developed for rural highways in India using accident data for 3 years on a 200 km highway segment. In [5], a method was proposed for road safety audits on four-lane national highways using multiple and nonlinear regression analysis. Additionally, an ANN model was developed with MATLAB R2009b neural network toolbox to identify safety parameters that minimize accidents in selected sections of the highway. In [6], road geometry and traffic variables were investigated in motorcycle crashes using a statistical technique called zero-inflated negative binomial regression for the NH-6 highway. The accident prediction model used accidents per year per km as a response variable and other data as input variables. In [7], eight national and state highways in Haryana were selected to develop accident prediction models with the help of 2-6 years of accident data, road geometry, and traffic data. Two modeling approaches were used, the Fixed/Random Effect Negative Binomial (FENB/RENB) and the M5 model, which is a binary decision tree learner used to predict a response variable and has linear regression functions at the terminal (leaf) nodes. The M5 model tree method was suggested as an alternative to RENB.

In [8], the effect of geometric design characteristics on the frequency of crashes on divided highways under various flow conditions was studied, using statistical modeling approaches such as Poisson and negative binomial regression. This study concluded that the negative binomial is an appropriate tool for predicting road crashes in this context. In [9], a crash prediction model was developed to identify the impact of accidents on the national economy. Crash data from a national highway during 2009-2012 were collected and linear regression analysis was used to suggest remedial measures to improve road safety. In [10], the influential parameters that resulted in fatal crashes were studied on three national highways of 2, 4, and 6 lanes, using logistic regression modeling. Influential variables were the number of lanes, the type of collision, and the number of vehicles. In [11], a regression model was developed to determine the relationship between the deaths of all road users by different modes of travel. The results showed that cycling, walking, and public transport systems had a low probability of fatal injuries, whereas two-wheelers, cars, and buses were at higher risk of fatalities. In [12], a generalized linear model was developed for crash prediction and accident hotspot detection on rural highways. The model showed that average daily traffic and average spot speed were the key factors responsible for road accidents.

B. Road Crash Data Modeling in Other Countries

In [13], crash-prediction models were developed for four-lane median-divided roads in Italy using accident data from 1999 to 2003, and statistical analysis was performed with the help of a negative multinomial regression model. The findings of this study were separate for both curves and tangents. Length, Annual Average Daily Traffic (AADT), and 1/radius were the significant variables for the curves, while for the tangents it was the junctions along with length and AADT. In [14], a generalized linear modeling method was used to develop crash prediction models for Italy's rural motorways, using 2001-2005 crash data and developing separate models for total and severe crashes using a negative binomial distribution. This study considered traffic volume, speed, alignment, design, sight distance, cross-section, and interchange ramps as explanatory variables. In [15], the parameters that contributed to fatal hit-and-run crashes in California, USA, were identified, using a logistic regression approach. The parameters were traffic control devices, speed limits, roadway profiles, and lighting conditions to reduce hit-and-run cases. In [16], differences in the severity of driver injuries between single and multiple-vehicle accidents of trucks were investigated, using 10-year accident data from a rural highway and a mixed logit modeling technique. The study concluded that there was a substantial difference between impacts on the severity of driver injuries in single/multiple vehicle accidents due to a variety of variables, such as roadway, driver, and environmental and vehicle characteristics. In [17], a crash data analysis was performed in Riyadh City, Saudi Arabia, considering a logit modeling approach, showing that a crash involving multiple vehicles is less severe than a crash involving a single vehicle.

In [18], Generalized Linear Modeling (GLM) and Geographically Weighted Poisson Regression (GWPR) were compared using crash data from 58 counties in California, showing that GWPR performed better in the specific context. In [19], the severity of road accidents per vehicle type was investigated in Greece, using lognormal regression analysis and concluding that weather conditions and night increased the risk of accidents. In [20], the impact of variable speed limits on road safety was studied using decision trees. In [21], the Negative Binomial-Multinomial Logit Fractional Split (NB-MNLFS) method was presented as an alternative to the most common modeling technique for unobserved heterogeneity. This method was compared to the multivariate random-parameter negative binomial model, showing excellent potential. In [22], accidents involving driving under the consumption of drugs/alcohol and resulting in fatalities were studied, using data from the forensic toxicology databases of the Norwegian road traffic crash registries in 2005–2015 and logistic regression, and finding that risk factors, such as not using seat belts or helmets, driving with invalid licenses, and overspeeding, were important.

In [23], time series modeling was used to study the relationship between police enforcement, traffic accidents, and traffic violations. The results showed that traffic violations and crashes had weekly variations and were affected by weather and holiday conditions, while the police patrol time was not affected. In [24], the combined effect of lighting and weather conditions was studied in single-vehicle accidents in Scotland, using data on the severity of injuries from 2016 and 2017. The results showed that all the parameters had different effects on the severity of the crashes. In [25], a mobile application system was proposed to report road accidents and driver overspeed awareness to improve road safety. In [26], a study was conducted on driver characteristics and their impact on road safety. The driver's characteristics were assessed by an actual driving experiment with a safety expert and responses to a questionnaire. A simple linear regression model was developed, showing that driving experience and skill, headway, and defensive driving affected driving performance. In [27], a survey was conducted to determine the local public opinion on crashes, showing that human and road factors have the most significant impact, with negligent driving and overspeeding being the main causes of road traffic accidents.

II. METHOD

Figure 1 shows the method used to achieve the study objectives, presenting its 3 major stages: Road crash data collection, modeling fatality/non-fatality of crashes, and validation of the models.

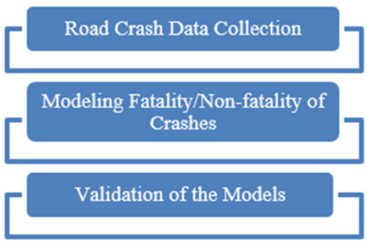

Fig. 1. Method.

A. Road Crash Data Collection

Data were collected from the National Highways Authority of India (NHAI) between 2016 and 2020 for the NH-48 rural national highway, which passes through the state of Maharashtra and has a length of approximately 265 km, considering both sides of the highway.

B. Modeling the Fatality/Non-fatality of the Crashes

The road crash data collected were statistically studied and, based on their distribution, a negative binomial regression model was identified to predict the fatality/non-fatality of crashes in the study area. Negative binomial regression is similar to regular multiple regression, except that the dependent variable (y) is an observed count that follows the negative binomial distribution. Thus, the possible values of y were the non-negative integers: 0, 1, 2, 3, and so on. Negative binomial regression is a generalization of Poisson regression, which loosens the restrictive assumption that the variance is equal to the mean made by the Poisson model. The mathematical form of the negative binomial regression is as follows:

$$\ln(y) = \beta_0 + \beta_1 X_i + \cdots + \beta_m X_{im} \quad (1)$$

C. *Validation of the Models*

The developed models were validated based on log-likelihood, Bayesian Information Criteria (BIC), Akaike's Information Criteria (AIC), Mean Absolute Deviation (MAD), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) values. Table I shows the criteria used for the validation of the models.

TABLE I. MODEL VALIDATION CRITERIA

No.	Parameter	Significance
1	Log-likelihood	Highest
2	AIC	Lowest
3	BIC	Lowest
4	MAD	Lowest
5	MSE	Lowest
6	RMSE	Lowest
7	MAPE	0-10 %= High precision
		10-20%= Good
		20-50%= Reasonable
		More than 50%= Inaccurate

III. RESULTS AND DISCUSSION

The data were divided into 2 parts, namely training and validation sets. The training data were the 70% of the actual data and were used to train the models, which were then validated on the 30% of the actual data. Four models were developed, which are discussed in detail below.

A. *Model 1*

The independent variable selected here was the nature of the crash, which was classified as head-on-collision, overturning, rear-end-collision, friction, etc. Figure 2 shows the parameter estimates of Model 1. It can be seen that intercepts B = 2 (head-on-collision) and B = 6 (friction) are the only significant parameters of the nature of the crash. The equation of the model is:

$$\ln(y) = -2.067 + (0.474 \times HeadOnCollision) - (2.263 \times Friction) \quad (2)$$

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval: Lower	95% Wald Confidence Interval: Upper	Hypothesis Test: Wald Chi-Square	Hypothesis Test: df	Hypothesis Test: Sig.	Exp(B)	95% Wald Confidence Interval for Exp(B): Lower	95% Wald Confidence Interval for Exp(B): Upper
(intercept)	-2.067	0.1906	-2.441	-1.694	117.602	1	000.000	0.127	000.087	000.184
[B=1]	0.545	0.3713	-0.183	1.273	002.153	1	000.142	1.724	000.833	003.570
[B=2]	0.474	0.2355	0.012	0.935	004.050	1	000.044	1.606	001.012	002.548
[B=3]	0.433	0.2352	-0.028	0.894	3.386	1	000.066	1.541	000.972	002.444
[B=4]	0.768	0.6787	-0.562	2.098	1.281	1	000.258	2.155	000.570	008.151
[B=5]	-26.138	941165.0442	-1844675.726	1844623.455	000.000	1	001.000	4.461E-12	000.000	a
[B=6]	-2.263	1.0244	-4.271	-0.256	4.882	1	000.027	0.104	000.014	000.774
[B=7]	-26.136^b							4.461E-12	000.000	000.000
[B=8]	0^c							1		
(Scale)	1^d									
(Negative binomial)	1^d									

Dependent Variable: Fatal or Non Fatal
Model: (Intercept), B

Fig. 2. Model 1 parameter estimates.

B. *Model 2*

The independent variable selected here was the time zone of the crash, which was classified as AM and PM. Figure 3 shows the parameter estimates of Model 2. It can be seen that the intercept Zone = 0 (AM) is the only significant parameter from the time zone of the crash. The equation of this model is:

$$\ln(y) = -1.993 + [0.535 \times TimeZone\ (AM/PM)] \quad (3)$$

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval: Lower	95% Wald Confidence Interval: Upper	Hypothesis Test: Wald Chi Square	Hypothesis Test: df	Hypothesis Test: Sig.	Hypothesis Test: Exp(B)	96%Wald Confidence Interval for Exp(B): Lower	96%Wald Confidence Interval for Exp(B): Upper
(intercept)	-1.99	0.1111	-2.211	-1.775	321.572	1	0.00	0.136	0.110	0.169
[Zone=0]	535	0	0.210	1	10.388	1	0.001	1.708	1.233	2.365
[Zone=1]	0^a							1		
(Scale)	1^b									
(Negative binomial)	1^b									

Dependent Variable: Fatal or Non Fatal
Model: (Intercept), Zone

Fig. 3. Model 2 parameter estimates.

C. *Model 3*

The independent variable selected here was the weather condition of the crash, which was classified as fine, cloudy, light rain, heavy rain, etc. Figure 4 shows the parameter estimates for the model. It can be seen that the intercepts H = 1 (fine), H = 2 (mist/fog), H = 3 (cloudy), H = 4 (light rain), and H = 5 (heavy rain), were the only significant parameters of the weather condition of the crash. This model is described as:

$$\ln(y) = -27.122 + 25.180 \times Fine + 25.823 \times Mist/Fog + 25.550 \times Cloudy + 24.963 \times LightRain + 25.513 \times HeavyRain \quad (4)$$

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval: Lower	95% Wald Confidence Interval: Upper	Hypothesis Test: Wald Chi-Square	Hypothesis Test: df	Hypothesis Test: Sig.	Exp(B)	95% Wald Confidence Interval for Exp(B): Lower	95% Wald Confidence Interval for Exp(B): Upper
(intercept)	-27.122	0.3757	-27.859	-26.386	5212.358	1	000.000	1.663E-12	7.965E-13	3.473E-12
[H=1]	25.180	0.3965	24.403	26.958	4032.531	1	000.000	8.624E+10	3.96E+10	1.876E+11
[H=2]	25.823	0.4140	25.011	26.634	3889.730	1	000.000	1.640E+11	7.283E+10	3.691E+11
[H=3]	25.550	0.4216	24.724	26.376	3673.130	1	000.000	1.248E+11	5.461E+10	2.851E+11
[H=4]	24.963	0.4642	24.053	25.873	2891.755	1	000.000	6.937E+10	2.793E+10	1.723E+11
[H=5]	25.513	0.6642	24.211	26.815	1475.537	1	000.000	1.202E+11	3.27E+10	4.420E+11
[H=10]	25.075^a	.	.	.	.	.	.	7.158E+10	000.000	000.000
[H=11]	0^b	.	.	.	.	.	.	1		
(Scale)	1^c									
(Negative binomial)	1^c									

Dependent Variable: Fatal or Non Fatal
Model: (Intercept), H

Fig. 4. Model 3 parameter estimates.

D. *Model 4*

The independent variables selected here were the nature of the crash, which was classified as head-on-collision, overturning, rear-end-collision, friction, etc., and its time zone, which was divided into AM and PM. Figure 5 shows the parameter estimates of the model. It can be seen that the intercepts Zone = 0 (AM) and B = 6 (Friction) are the only significant parameters from the time zone and the nature of the crash, respectively. The equation of the model is:

$$\ln(y) = -2.239 + [0.496 \times TimeZone(AM/PM)] + (-2.266 \times Friction) \quad (5)$$

Parameter Estimates										
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Interept)	-2.239	0.2020	-2.635	-1.843	122.821	1	0.000	0.107	0.072	0.158
[Zone=0]	0.496	0.1677	0.167	0.824	8.740	1	0.003	1.642	1.182	2.281
[Zone=1]	0^a	.	.	.	.	.	.	1.000	.	.
[B=1]	0.475	0.3735	-0.257	1.207	1.617	1	0.204	1.608	0.773	3.344
[B=2]	0.443	0.2365	-0.021	0.906	3.509	1	0.061	1.557	0.980	2.475
[B=3]	0.397	0.2363	-0.066	0.860	2.287	1	0.093	1.488	0.936	2.364
[B=4]	0.674	0.6827	-0.664	2.012	0.975	1	0.323	1.962	0.515	7.480
[B=5]	-26.235	937898.1100	-1838272.752	1838220.281	0.000	1	1.000	4.04E-12	0.000	b
[B=6]	-2.266	1.0249	-4.275	-0.258	4.890	1	0.027	0.104	0.014	0.773
[B=7]	-25.964^c							5.3E-12	0.000	0.000
[B=8]	0^a							1.000		
(Scale)	1^d									
(Negative binomial)	1^d									
Dependent Variable: Fatal orNon Fatal										
Model: (Intercet), Zone, B										

Fig. 5. Model 4 parameter estimates.

The results of this study state that the nature of the crash and weather conditions are the most important factors associated with the crashes. The nature of the crash and weather conditions were also associated in [10, 21, 23, 24]. Negative binomial regression modeling is the best-suited tool for crash prediction modeling [8].

IV. MODEL VALIDATION

The four developed models to predict the fatality or non-fatality of road crashes, shown in (2)-(5), were validated using the criteria shown in Table I. Table II shows the validation parameters for the four models. Model 4 showed the highest log-likelihood and the lowest AIC and MAD values among the others, being -477.904, 973.807, and 0.1597, respectively. Model 2 showed the lowest BIC, MSE, and MAPE values among all models, being 992.261, 0.132, and 0.1303, respectively. Model 1 had the lowest RMSE value among all models, being 0.359.

TABLE II. MODEL VALIDATION PARAMETERS

Model	Log-Likelihood	AIC	BIC	MAD	MSE	RMSE	MAPE
1	-482.220	980.441	1019.884	0.2515	0.1289	0.359	0.1306
2	-489.200	982.400	992.261	0.269	0.132	0.363	0.1303
3	-487.591	989.181	1023.695	0.256	0.1342	0.3664	0.1367
4	-477.904	973.807	1018.182	0.1597	0.1592	0.3999	0.1593

Finally, it can be seen that there is not much significant difference between the values of all models. All four models are valid, while Models 2 and 4 are the best-fitted to predict the fatality or non-fatality of road crashes.

V. CONCLUSIONS

Road crash data from the NH-48 rural national highway in Maharashtra, India, were studied and modeled to predict the fatality of road crashes using the negative binomial regression technique. Four models were developed incorporating variables such as the nature, time zone, and weather conditions of the crash. Two important parameters were identified under the nature of the crash, namely head-on-collision and friction. For the time zone of the crash, the two parameters (AM and PM) were identified. Five important parameters were identified under the weather conditions of the crash, namely light rain, mist/fog, heavy rain, fine, and cloudy. All four developed models were valid, while Models 2 and 4 were the best-fitted to predict the fatality or non-fatality of road crashes. Head-on collision, AM, PM, light rain, mist/fog, heavy rain, fine, and cloudy were positively associated with the fatality of road crashes, while friction was negatively associated. The developed models can be used to predict the fatality/non-fatality of road crashes and design further road safety strategies on the highway.

ACKNOWLEDGMENT

The authors wish to acknowledge the National Highways Authority of India for providing the crash data, Savitribai Phule Pune University, and the Department of Civil Engineering, COEP TECH University for permitting them to carry out this research work. The authors have not received any funding for this study.

REFERENCES

[1] "Road Accidents in India 2019," Ministry of Road Transport & Highways, India, 2019.

[2] "Road Accidents in India 2021," Ministry of Road Transport & Highways, India, 2021.

[3] M. Parida, S. S. Jain, and V. S. Landge, "Stochastic Modeling for Traffic Crashes on Non Urban Highways in India," presented at the Research into Practice: 22nd ARRB Conference, Canberra, Australia, 2006.

[4] R. R. Dinu and A. Veeraragavan, "Random parameter models for accident prediction on two-lane undivided highways in India," *Journal of Safety Research*, vol. 42, no. 1, pp. 39–42, Feb. 2011, https://doi.org/10.1016/j.jsr.2010.11.007.

[5] S. S. Jain, P. K. Singh, and M. Parida, "Road Safety Audit for Four-lane National Highways," presented at the 3rd International Conference on Road Safety and Simulation - Purdue University Transportation Research Board, Indianapolis, IN, USA, 2011.

[6] A. K. Sharma, V. S. Landge, and N. V. Deshpande, "Modeling motorcycle accident on rural highway," *International Journal of Chemical, Environmental & Biological Sciences*, vol. 1, no. 2, pp. 313–317, 2013.

[7] G. Singh, S. N. Sachdeva, and M. Pal, "M5 model tree based predictive modeling of road accidents on non-urban sections of highways in India," *Accident Analysis & Prevention*, vol. 96, pp. 108–117, Nov. 2016, https://doi.org/10.1016/j.aap.2016.08.004.

[8] P. Vayalamkuzhi and V. Amirthalingam, "Development of Comprehensive Crash Models for Four-lane Divided Highways in Heterogeneous Traffic Condition," *Transportation Research Procedia*, vol. 17, pp. 626–635, Jan. 2016, https://doi.org/10.1016/j.trpro.2016.11.117.

[9] G. Athipathi, S. Nagan, and T. Baskaran, "Accident Prediction on national highways in India," *International journal of theoretical and applied mechanics*, vol. 12, no. 1, pp. 153–165, 2017.

[10] H. M. Naqvi and G. Tiwari, "Factors Contributing to Motorcycle Fatal Crashes on National Highways in India," *Transportation Research Procedia*, vol. 25, pp. 2084–2097, Jan. 2017, https://doi.org/10.1016/j.trpro.2017.05.402.

[11] R. Goel, "Modelling of road traffic fatalities in India," *Accident Analysis & Prevention*, vol. 112, pp. 105–115, Mar. 2018, https://doi.org/10.1016/j.aap.2017.12.019.

[12] R. Bhavsar, A. Amin, and L. Zala, "Development of Model for Road Crashes and Identification of Accident Spots," *International Journal of Intelligent Transportation Systems Research*, vol. 19, no. 1, pp. 99–111, Apr. 2021, https://doi.org/10.1007/s13177-020-00228-z.

[13] C. Caliendo, M. Guida, and A. Parisi, "A crash-prediction model for multilane roads," *Accident Analysis & Prevention*, vol. 39, no. 4, pp. 657–670, Jul. 2007, https://doi.org/10.1016/j.aap.2006.10.012.

[14] A. Montella, L. Colantuoni, and R. Lamberti, "Crash Prediction Models for Rural Motorways," *Transportation Research Record*, vol. 2083, no. 1, pp. 180–189, Jan. 2008, https://doi.org/10.3141/2083-21.

[15] R. Tay, U. Barua, and L. Kattan, "Factors contributing to hit-and-run in fatal crashes," *Accident Analysis & Prevention*, vol. 41, no. 2, pp. 227–233, Mar. 2009, https://doi.org/10.1016/j.aap.2008.11.002.

[16] F. Chen and S. Chen, "Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1677–1688, Sep. 2011, https://doi.org/10.1016/j.aap.2011.03.026.

[17] S. Altwaijri, M. Quddus, and A. Bristow, "Analysing the Severity and Frequency of Traffic Crashes in Riyadh City Using Statistical Models," *International Journal of Transportation Science and Technology*, vol. 1, no. 4, pp. 351–364, Dec. 2012, https://doi.org/10.1260/2046-0430.1.4.351.

[18] Z. Li, W. Wang, P. Liu, J. M. Bigham, and D. R. Ragland, "Using Geographically Weighted Poisson Regression for county-level crash modeling in California," *Safety Science*, vol. 58, pp. 89–97, Oct. 2013, https://doi.org/10.1016/j.ssci.2013.04.005.

[19] Y. George, T. Athanasios, and P. George, "Investigation of road accident severity per vehicle type," *Transportation Research Procedia*, vol. 25, pp. 2076–2083, Jan. 2017, https://doi.org/10.1016/j.trpro.2017.05.401.

[20] M. Z. Hasanpour, M. R. Ahadi, A. S. Moghadam, and G. A. Behzadi, "Variable Speed Limits: Strategies to Improve Safety and Traffic Parameters for a Bottleneck," *Engineering, Technology & Applied Science Research*, vol. 7, no. 2, pp. 1535–1539, Apr. 2017, https://doi.org/10.48084/etasr.831.

[21] T. Bhowmik, S. Yasmin, and N. Eluru, "A joint econometric approach for modeling crash counts by collision type," *Analytic Methods in Accident Research*, vol. 19, pp. 16–32, Sep. 2018, https://doi.org/10.1016/j.amar.2018.06.001.

[22] A. Valen *et al.*, "Driver-related risk factors of fatal road traffic crashes associated with alcohol or drug impairment," *Accident Analysis & Prevention*, vol. 131, pp. 191–199, Oct. 2019, https://doi.org/10.1016/j.aap.2019.06.014.

[23] M. Feng, X. Wang, and M. Quddus, "Developing multivariate time series models to examine the interrelations between police enforcement, traffic violations, and traffic crashes," *Analytic Methods in Accident Research*, vol. 28, Dec. 2020, Art. no. 100139, https://doi.org/10.1016/j.amar.2020.100139.

[24] G. Fountas, A. Fonzone, N. Gharavi, and T. Rye, "The joint effect of weather and lighting conditions on injury severities of single-vehicle accidents," *Analytic Methods in Accident Research*, vol. 27, Sep. 2020, Art. no. 100124, https://doi.org/10.1016/j.amar.2020.100124.

[25] I. J. Mrema and M. A. Dida, "A Survey of Road Accident Reporting and Driver's Behavior Awareness Systems: The Case of Tanzania," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 6009–6015, Aug. 2020, https://doi.org/10.48084/etasr.3449.

[26] D. I. Tselentis, K. Folla, V. Agathangelou, and G. Yannis, "Investigating the Correlation between Driver's Characteristics and Safety Performance," *Transportation Research Procedia*, vol. 48, pp. 1254–1262, Jan. 2020, https://doi.org/10.1016/j.trpro.2020.08.147.

[27] M. Angın and S. I. A. Ali, "Analysis of Factors Affecting Road Traffic Accidents in North Cyprus," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7938–7943, Dec. 2021, https://doi.org/10.48084/etasr.4547.