JSTT JOURNAL OF SCIENCE AND TRANSPORT TECHNOLOGY

Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

UNIVERSITY OF TRANSPORT TECHNOLOGY

Vehicle and time specific crash modelling on selected rural highway curves using geometric and speed parameters: A transformed linear regression approach

Raghavendra S Sanganaikar[1,2*], Raviraj H Mulangi[2] and Yatish R G[2]

[1]Department of Civil Engineering, Vidyavardhaka College of Engineering Mysuru, India

[2]Department of Civil Engineering, National Institute of Technology Karnataka, India

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.1.15-28

[*]Corresponding author:
Email address:
raghavendras.227cv502@nitk.edu.in

Received: 20/07/2025
Received in Revised Form: 28/10/2025
Accepted: 27/01/2026

Abstract: This study develops vehicle and time-specific crash rate prediction models for rural highway curves using high-resolution geometric and speed data. A 30km segment of State Highway-1 in Karnataka, India, encompassing 32 horizontal curves, served as the study site. Detailed data collection included 10 years of crash records, traffic volume count, LiDAR-based geometric features, and spot speeds recorded from laser speed cameras. Distinct models were built for motorized two-wheelers (MTW), passenger cars (CAR), heavy commercial vehicles (HCV), and for both daytime and nighttime conditions. The study offers a novel contribution by incorporating nighttime crash rate modelling rarely addressed due to challenges in data availability, and by developing disaggregated models for multiple vehicle classes. A backward stepwise regression (BSR) approach with square root transformation was employed, ensuring model transparency and interpretability. Sight-distance deficiency consistently emerged as the most influential predictor of crash rate, highlighting the critical role of visibility on curved segments. Validation through Leave One Out Cross Validation (LOOCV) confirmed acceptable predictive performance (R^2 = 0.43-0.80), with residuals exhibiting normal distribution. The findings underscore the importance of curve geometry and visibility in crash risk and provide actionable insights for design audits and safety interventions on rural highways.

Keywords: crash prediction models, daytime-nighttime crashes, vehicle specific analysis, backward stepwise regression, LiDAR-based geometry, sight distance deficiency.

1. Introduction

The United Nations' Sustainable Development Goal (SDG) 3.6 calls for reducing global road traffic deaths by 50% by 2030. However, this goal remains far from reach. Currently, road crashes cause approximately 1.3 million deaths globally each year, with India accounting for nearly 11% of these fatalities. Road traffic injuries have emerged as the 12th leading cause of death globally, and about 92% of these deaths occur in low- and middle-income countries. Rapid urbanization and infrastructure expansion, alongside a near doubling of the global vehicular population in the last decade, have outpaced safety measures [1]. In India alone, nearly 4 lakh accidents were recorded in 2021, resulting in approximately 1.5 lakh deaths and 3.85 lakh injuries. Of all crashes, nearly 80% occur on straight and curved sections, with about 12% reported specifically on horizontal curves [2]. Given these alarming trends, road user safety must be recognized as a national and global priority. Crashes are now among the leading causes of hospitalization, disability, and mortality, with profound socio-economic consequences. Crash prediction models are widely recognized as systematic tools to identify hazardous roadway conditions and estimate the impact of safety interventions [3]. These models are also useful in estimating crash frequency at locations with limited or no crash data [4]. Therefore, it is essential to establish mathematical relationships between crash occurrences and contributing variables. Operational attributes and geometric features such as curvature, speed, and sight distance have shown strong correlation with crash rates, especially when true latent variables are considered [5]. Conventional crash models generally employ single-regression frameworks, using geometry and speed-based predictors. However, the rare and random nature of crashes introduces challenges in predicting trends over short timeframes [6], [7]. Additionally, the limited availability and spatial resolution of crash data often reduce model accuracy and transferability [8]. Studies employing generalized linear modelling such as negative binomial and logistic regression have demonstrated improved accuracy when multiple variables particularly speed-related measures are included [9], [10]. A consistent finding across studies is the positive correlation between speed reduction and crash likelihood, particularly on horizontal curves [11], [12]. Road cross-section features, traffic volume, and design parameters collectively influence safety and are easily quantifiable [13]. Notably, crash risk has been found to be significantly higher at night due to visibility constraints and driver limitations [14], [15]. Despite the availability of traffic volume, geometry, and speed data, crash events remain sparse and unpredictable. Models using alignment indices, operating speed, and volume have been developed with varying success [16], [17], [18]. However, temporal variations especially day versus night comparisons remain underexplored, despite growing evidence of their importance. Most past studies conducted data collection only during the daytime and under fair weather conditions [19], [20], [21], [22], [23].

These limitations underscore the need for a more detailed understanding of how time of day and vehicle class interact with geometric and operational conditions to affect crash risk. This study aims to address these gaps by developing vehicle- and time-specific crash prediction models using precise, high-resolution data collected across multiple horizontal curves in a rural highway setting with following specific tasks:

To develop and validate time-specific and vehicle-specific crash rate prediction models for selected rural highway curves using high resolution geometric and speed data.

To identify and quantify the influence of geometric parameters and vehicle operating speed characteristics on crash occurrence.

To compare the performance of multiple modelling techniques and evaluate the trade-off between predictive accuracy and interpretability.

2. Methodology

The section deals with the detailed methodology adopted in the current study. Fig. 1 gives the work flowchart adopted.

2.1. Study area and site selection

The study was conducted along a 30-kilometer stretch of State Highway 1 (SH-1) in Karnataka, India, extending between Padubidri and Karkala. Following a detailed reconnaissance survey, a suitable corridor was identified based on consistent pavement conditions, standardized road markings, and the presence of horizontal curves exhibiting diverse radii. The selected segment lies in largely flat terrain, with an average roadway gradient of 5%, with 32 horizontal curves. To ensure accurate measurement of free-flow speeds, sections influenced by intersections or speed-calming devices such as humps were deliberately excluded. This approach provided a uniform and controlled environment conducive to reliable data collection for geometric and safety-related analysis. The 30 km stretch refers to single travel direction each way which adds up and make the dataset to 60 in total.

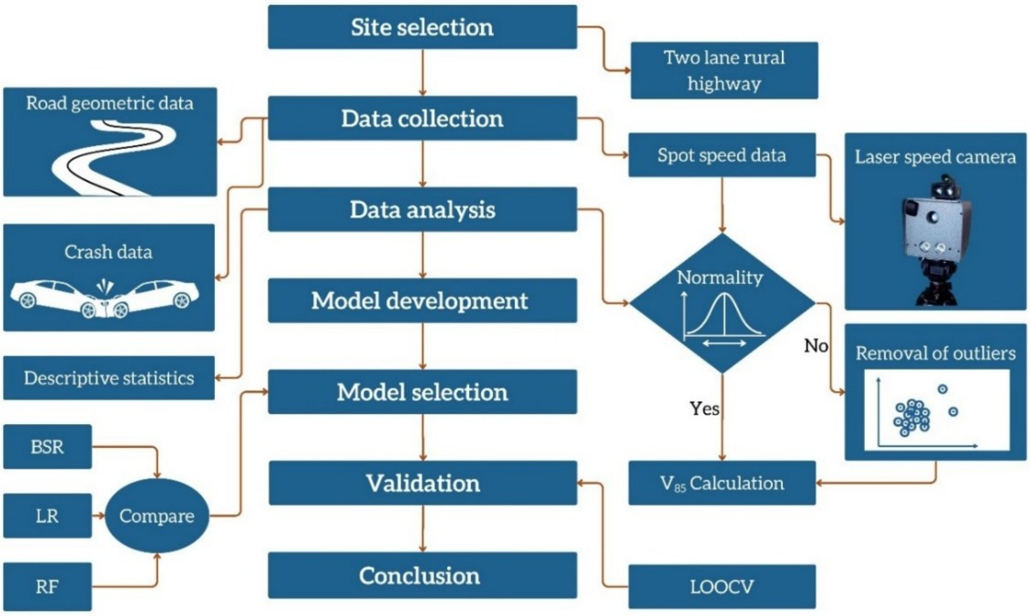

Fig. 1. Research outline for the present study

2.2. Data collection and preliminary analysis

The study utilized a comprehensive dataset comprising road crash records, traffic volume counts, spot speed measurements, and geometric road characteristics. Ten years of crash data between 2014 and 2024 were sourced from police records and classified by vehicle type, focusing on three categories: motorized two-wheelers (MTW), passenger cars (CAR), and heavy commercial vehicles (HCV). Crash incidents were georeferenced to specific horizontal curves based on the First Information Reports (FIRs) provided by the local authorities. Each crash was further categorized by time of occurrence as daytime or nighttime. Although the original dataset included gender-based driver involvement, records indicated that female drivers contributed to less than 1% of crash cases. As a result, gender data were excluded from further analysis. Crash frequencies were normalized by converting them into crash rates, expressed as the number of crashes per 100,000 kilometers traveled.

Traffic volume data were collected over a continuous 7-day period from 6:00 AM to 10:00 PM daily. Due to significantly lower vehicle flow beyond 10:00 PM and logistical limitations, data collection was restricted to this timeframe. The observed counts were used to compute Average Daily Traffic (ADT), which was subsequently converted into Annual Average Daily Traffic (AADT) using Indian Roads Congress (IRC) recommended factors [24], [25]. Spot speed data were recorded using high-precision laser speed cameras with an accuracy of ±1 km/h and a 30° angular tolerance. Speed measurements were taken at three critical points along each horizontal curve: the point of curvature (PC), the mid-point (MC), and the point of tangency (PT). To ensure measurement reliability, equipment was aligned tangentially to the travel path and kept within the allowable angular limit. Data were collected during both daytime and nighttime under free flow conditions, maintaining a minimum 5 second headway between vehicles to eliminate the influence of vehicle interaction on speed selection [26]. Fig. 2 depicts the spot speed data collection and output during daytime and nighttime.

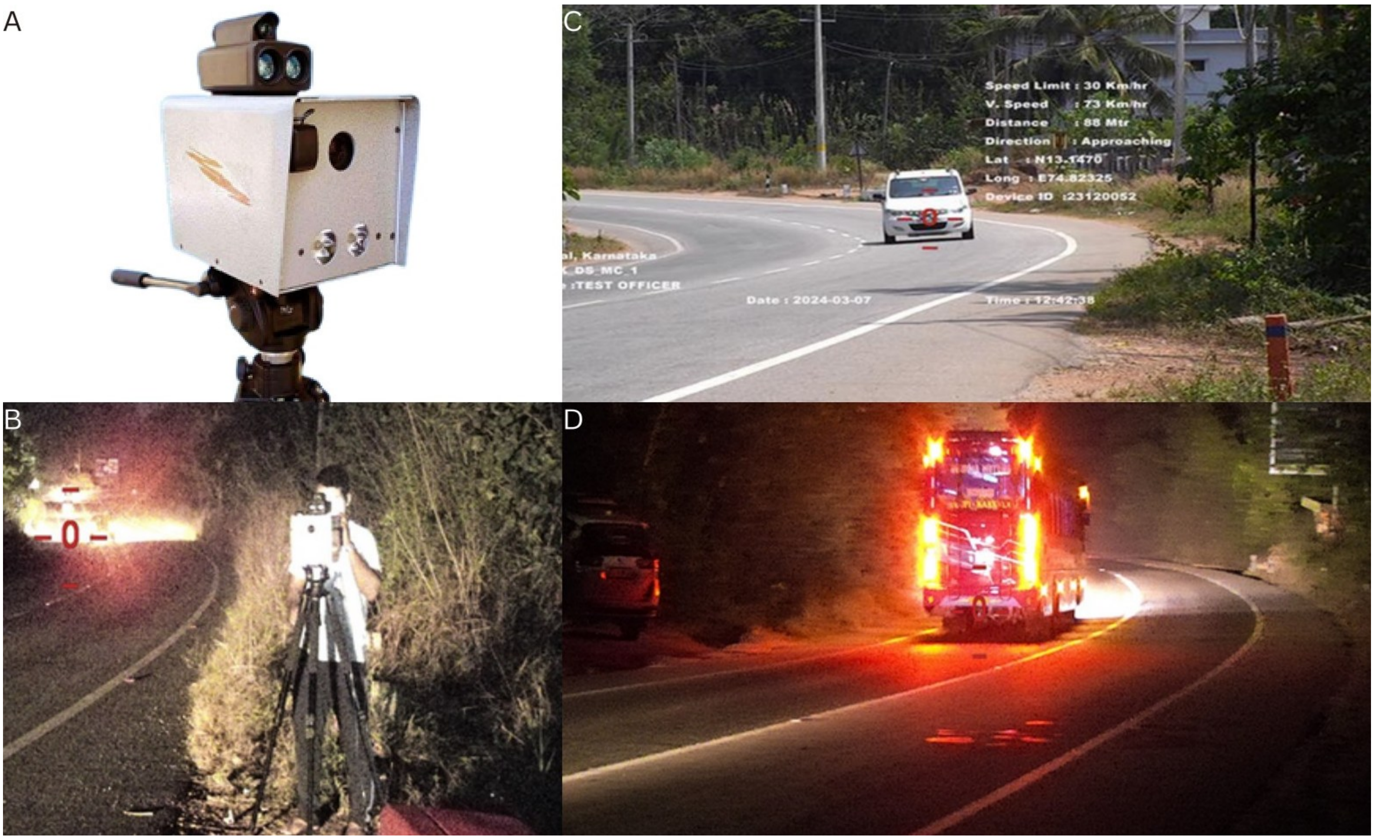

Fig. 2. A: Laser speed camera used for spot speed data collection. B: Data collection during night under progress C: Sample output of daytime speed data on curve D: Sample output of nighttime speed data on curve

A detailed topographic survey using Light Detection and Ranging (LiDAR) technology was conducted to extract geometric attributes of the roadway. Known for its high spatial accuracy up to 3 mm, LiDAR proved highly effective in capturing fine geometric details [27]. The initial point cloud dataset included over 2.6 billion data points, which was refined to approximately 1 billion points for analysis. Using 'CloudCompare version.2.14.alpha', key geometric features were extracted, including superelevation (e), average gradient (G), shoulder width (ShW), and lane width (LnW). 'CloudCompare' is an open source 3D point cloud processing software used in highway engineering studies for extracting geometric features, slope analysis, and surface modelling. Its capability to handle huge LiDAR datasets and perform distance computation and cross section profiling has made it a popular tool in transportation research [28], [29]. The software enables precise extraction of roadway alignment and curvature parameters critical for safety and design consistency analyses. Since lane width was found to be consistent across all curves, it was excluded from further statistical modelling. A centerline polyline was created along the roadway and imported into Autodesk Civil 3D (2025) to derive curve-specific parameters such as horizontal curve radius (R), curve length (CL), deflection angle (DA), approach tangent length (Atl), and departure tangent length (Dtl). To compute available sight distance, the trimmed LiDAR point cloud was converted into a surface model in Civil 3D. Calculations were carried out in accordance with IRC guidelines, by providing driver eye height of 1.2 m and object height of 0.15 m [30]. The process is depicted in Fig. 3.

For speed analysis, a minimum of 50 spot speed observations were collected per vehicle category at each of the three curve points. The Kolmogorov - Smirnov (K-S) test was used to assess the normality of raw speed datasets.

Outliers identified through this process were excluded, reducing the minimum sample size at each point to no less than 30 observations [17], [23]. The 85th percentile operating speed (V_{85}), representing the speed below which 85% of vehicles travel under free flow conditions, was determined using cumulative frequency distribution plots [31]. All V_{85} datasets exhibited near-normal distributions. Summary statistics for the variables used in the study are presented in Table 1.

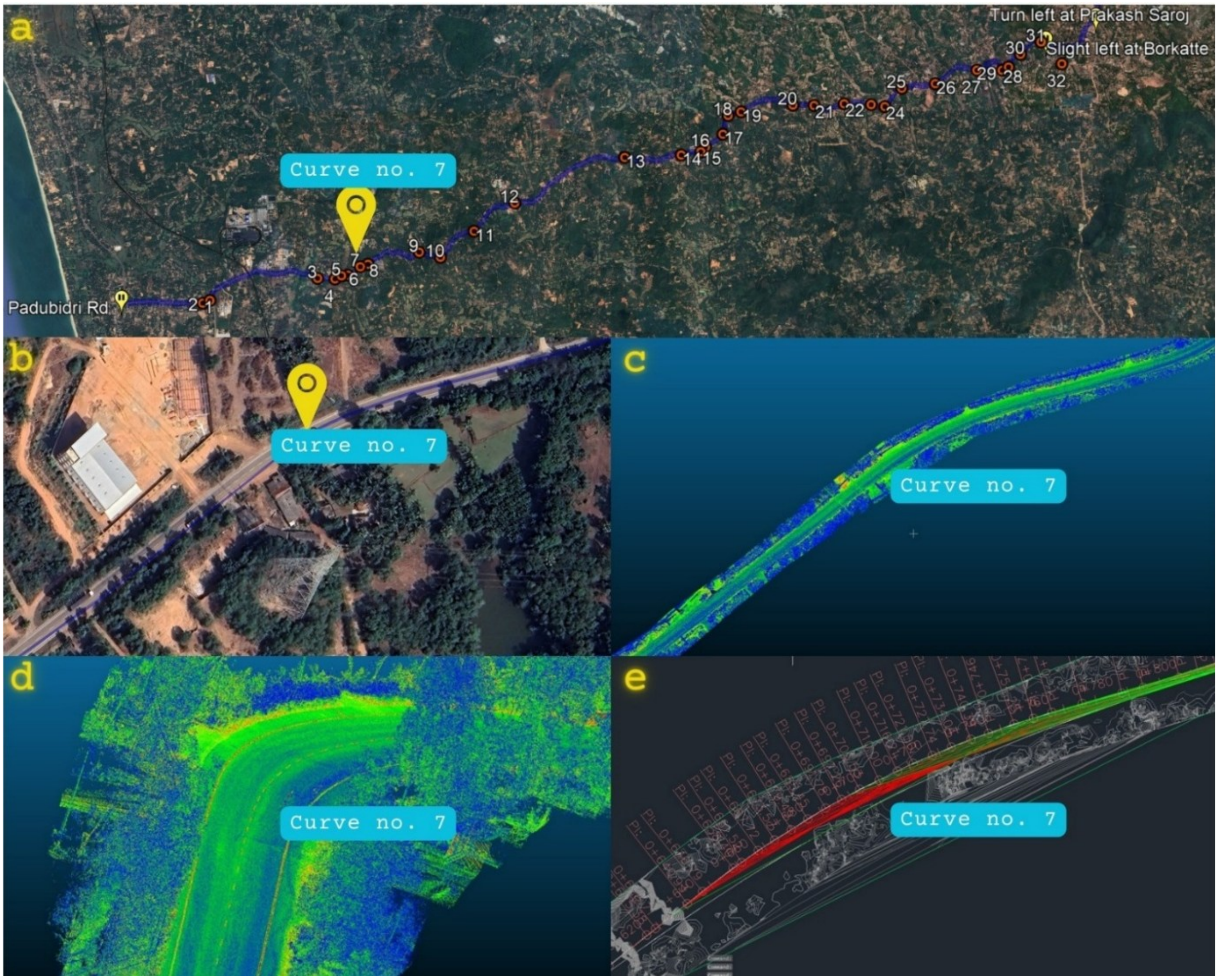

Fig. 3. (a) Study stretch showing all curve locations (b) Google earth image from of one of the curves (Curve 7). (c) LiDAR data output for curve 7. (d) 3D zoomed in version of curve 7. (e) Stopping sight distance (SSD) analysis output

Table 1. Descriptive statistics of geometric and speed data used in crash prediction model

Variable, unit	Mean	Std Dev	Min	Max
R, m	197.58	80.94	42.33	336.85
DA, degree	33.1	16.37	11	75
CL, m	102.14	46.54	43.08	264.06
Curt, degree/m	0.35	0.21	0.17	1.13
e_Def , %	1.14	2.09	-2	5.7
Atl, m	43.42	22.98	10.84	101.98
Dtl, m	46.65	27.83	10.84	122.36
G, %	0.9	2.82	-4.83	6.28

Table 1. (continued)

Variable, unit	Mean	Std Dev	Min	Max
S_Def, m	48.93	25.89	0	96
ShW, m	1.57	0.2	1.1	2
IL, m	392.3	75.38	290	631
∆VMD, km/h	0.53	5.7	-24	9
∆VCD, km/h	2.63	3.12	-3	12
∆VHD, km/h	2.17	2.04	-2	9
∆MCHD, km/h	13.63	8.02	-8	27
∆VMN, km/h	1.87	2.06	-1	7
∆VCN, km/h	3.03	3.6	-2	17
∆VHN, km/h	5.3	4	-4	13
∆MCHN,km/h	10.87	5.68	0	21

∆VMD- Change in MTW speed during daytime, ∆VCD- Change in CAR speed during daytime, ∆VHD- Change in HCV speed during daytime, ∆MCHD- Change in speed at mid point of curve between CAR and HCV during daytime, ∆VMN- Change in MTW speed during nigthttime, ∆VCN- Change in CAR speed during nigthttime, ∆VHN- Change in HCV speed during nigthttime, ∆MCHN- Change in speed at mid point of curve between CAR and HCV during nigthttime.

3. Model results and discussions

This study focused on developing vehicle-specific and time-specific crash rate prediction models to better understand road user safety on rural highway curves. Separate models were formulated for motorized two-wheelers (2W), passenger cars (CAR), and heavy commercial vehicles (HCV), along with distinct models for daytime and nighttime crash occurrences, and a comprehensive model for overall crash rate. Among these, the nighttime crash model stands out as a key contribution, given the limited existing literature and the inherent challenges in acquiring reliable nighttime traffic data. Multiple approaches were explored to assess the model performance including Random Forest and Lasso regression, both demonstrated high predictive capability. Recent studies have similarly employed advanced data-driven methods such as Artificial Neural Networks (ANN) for pavement and safety modelling, demonstrating their capability to capture complex nonlinear relationships between geometric and performance parameters [32]. But, since the sample size was small, these methods raised potential concerns of overfitting. Consequently, a square-root transformed backward stepwise regression (BSR) model was adopted as the primary modelling technique. This method was selected for its transparency, ease of interpretation, and suitability for evidence-based design reviews and policy formulation related to rural road safety. To ensure the statistical robustness of the models, Variance Inflation Factor (VIF) values were used to detect multicollinearity among predictors. A VIF threshold of 5 was used [33]; any variable exceeding this limit was excluded during the initial modelling phase. The models were then re-estimated using predictors that exhibited no significant multicollinearity.

The finalized crash rate model results are presented in Table 2. Given the modest dataset size (n=30), the models were validated using Leave-One-Out Cross-Validation (LOOCV). Model performance was evaluated using standard error metrics including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Validation results are provided in Table 3, while comparative outputs from different modelling techniques are summarized in Table 4. All model development and validation procedures were conducted using the open-source platform RStudio (Version 2025.05.01).

Table 2. Crash prediction model summary based on stepwise backward elimination with transformed dependent variables

Response variable	Parameter	Estimate	t-Stat	p-value	VIF	R^2	AIC
$\sqrt{Cr\_T}$	Curt	-0.329	-2.39	0.02	1.85	0.685	-40.88
	Atl	-0.003	-2.62	0.01	1.19		
	Dtl	-0.002	-2.09	0.04	1.19		
	S_Def	0.007	6.79	<0.001	1.46		
$\sqrt{Cr\_D}$	DA	-0.005	-3.24	0.004	2.44	0.797	-59.93
	CL	+0.002	3.63	0.002	3.13		
	e_Def	-0.022	-2.64	0.015	1.40		
	Atl	-0.003	-4.01	0.001	1.57		
	Dtl	-0.003	-4.28	<0.001	1.75		
	S_Def	+0.005	6.79	< 0.001	1.67		
	∆MCHD	+0.0051	2.65	0.0145	1.10		
$\sqrt{Cr\_N}$	DA	-0.0042	-3.3487	0.0027	2.40	0.71	-68.49
	CL	+0.0012	2.5038	0.0195	2.85		
	Atl	-0.0014	-2.0768	0.0487	1.45		
	Dtl	-0.0013	-2.2457	0.0342	1.43		
	S_Def	+0.0047	7.3766	< 0.0001	1.59		
$\sqrt{Cr\_M}$	CL	+0.0035	2.66	0.014	5.28	0.744	-25.69
	Curt	–0.8707	-3.78	0.001	3.27		
	S_Def	+0.0074	5.56	0.000	1.70		
	Sh_W	+0.3974	2.15	0.043	1.92		
	Log(IL)	–3.9909	-4.43	<0.001	7.20		
	∆VMD	–0.0220	-2.30	0.031	1.50		
$\sqrt{Cr\_C}$	e_Def	–0.0374	–2.310	0.030	1.27	0.43	-18.91
	Atl	–0.0067	–2.942	0.007	3.03		
	Dtl	–0.0048	–2.543	0.018	3.02		
	Log(S_Def)	+0.1791	+2.737	0.011	1.07		
	Log(IL)	+2.2848	+2.681	0.013	5.03		
$\sqrt{Cr\_H}$	Curt	–0.6040	–3.711	0.001	1.33	0.55	-22.64
	S_Def	+0.0074	+5.661	0.000	1.33		

Cr_T- Total Crash rate (number of crashes per 1 lakh kms), Cr_D - Daytime crash rate , Cr_N - Nighttime crash rate, Cr_M- crash rate for MTW, Cr_C- Crash rate for CAR, Cr_H- Crash rate for HCV, DA- Deflection Angle, CL- Curve Length, Curt- Curvature, e_Def- Superelevation deficiency, Atl- Approach Tangent Length, Dtl- Departure Tangent Length, S_Def- Sight distance deficiency, Sh_W- Shoulder width, IL- Influence length, ∆_VMD- Change in operating speed from entry point to mid-point of curve MTW during day, ∆MCHD- Change on operating speed at mid point between CAR and HCV during daytime, VIF- Variance inflation factor

With an in sample R^2 value of 0.685, the crash rate prediction model for total crashes, 'Curt', 'Atl', 'Dtl' and 'S_Def' emerged as significant variable predictors. Among these 'S_Def' emerged as dominant factor indicating higher crash rates for higher deficiency of stopping sight distance on curevs. The geometric variables 'Curt', 'Atl' and 'Dtl' showed inverse relation with crash occurrence indicating flatter curves and approach geometry plays important role in crash mitigation. These results are in consistent with recent works on road user safety [34], [35], [36]. The daytime crash rate prediction model explains 79.7% variability in the response variable. CL, S_Def and ΔMCHD had positive impact and DA, e_Def, Atl, Dtl showed inverse relation on daytime crash rate. This indicates the provision of flatter curves, limiting the operating speed of cars and HCVs and provision of adequate sight distance and superelevation on the horizontal curves reduced occurrence of road crashes during daytime. In contrast to daytime crash rate model, the nighttime model with 71% of variation explaining power, did not show dependency on the length of curve, speed differential and superelevation. However, the turning angle, transition lengths and sight distance had significant effect on crash occurrence. Deficiency in stopping sight distance on curves emerged as strong predictor of crash rate during nighttime. The model also implied that the flatter and longer transitions before and after the curve reduced crash rate risk. Crash rate model for MTW indicate that the smoother speed transition, wider shoulders, longer influence lengths lower that crash occurrence and Sharper curves, deficiency in sight distance worsens the chances of crash occurrence. Even with lower predictive power of 43% variance explanation, the crash rate model for CARs has shown dependency on expected geometric and sight distance related variables. Higher influence lengths and transition lengths tend to reduce crash occurrence and greater deficiency in stopping sight distance tend to increase the crash occurrence. Surprisingly, superelevation has contrasting response of increased crash occurrence with reduced deficiency. With an R^2 value of 0.551, crash rate model for HCV emerged out as a good predicting model. It involved just two variables giving an advantage to the policymakers in understanding their effect on possible crash occurrences and decide on implementing necessary changes. Although some models, especially for passenger cars resulted moderate R^2 values, such levels are typical for crash-based analyses where randomness introduce high variability. In road user safety modelling, R^2 values between 0.4 and 0.8 are considered acceptable, since crashes are influenced by several unobserved behavioural and environmental factors that are difficult to quantify directly. Thus, the model performance achieved here is consistent with the inherent unpredictability of crash phenomena [37].

To evaluate the predictive performance of the developed models, validation was carried out using LOOCV in terms of MAE and RMSE. This approach ensures an unbiased assessment of model generalizability by systematically leaving out one observation at a time for testing. RMSE penalizes larger errors more heavily than smaller ones, thereby providing a sensitive measure of prediction accuracy, particularly in the presence of outliers. It is particularly useful for identifying whether the model produces a few large deviations from actual values. On the other hand, MAE offers a more balanced evaluation by measuring the average magnitude of the errors in prediction, regardless of their direction. It is less sensitive to outliers than RMSE and represents the typical size of the prediction errors. General equations of RMSE and MAE are given in Eq. (1) and (2).

$$MAE=\frac{1}{M}\sum_{i=1}^{M}|A_i-Pr_i| \tag{1}$$

$$RMSE=\sqrt{\frac{1}{M}\sum_{i=1}^{M}(A_i-Pr_i)^2} \tag{2}$$

Where,

M= Number of observations

A_i= Actual values

Pr_i= Predicted values

The LOOCV was carried out using the function ‘caret::trainControl(method = "LOOCV" caret::trainControl (method = "LOOCV") in RStudio. The results for all the six BSR models demonstrate acceptable results with lower RMSE and MAE values. Also, the Kolmogorove- Smirnov normality test results for residual distributions validate the BSR model approach. In addition, actual vs predicted values were also plotted as shown in Fig. 4 to supplement the validation. Among all the predicted models, the crash rate prediction model showed poor results for CAR. However, with almost same validation R^2 value for validation of the model, overfitting is not expected. Also, a separate Quantile-Quantile (Q-Q) plot was plotted to check the normality of residuals. The Q-Q plot depicted normal distribution of residuals as shown in Fig. 5.

The results of BSR were also compared with advanced machine learning techniques to understand improvisations, if any (Table 4). Results of BSR were compared with Lasso regression and Random Forest regression techniques. Except for total crash rate, Lasso and Random Forest regression techniques showed better predictability than BSR technique. Although Lasso and Random Forest techniques showed marginal improvements in some models, the final selection of backward elimination was guided by considering the limited dataset and a balance of predictive adequacy, statistical robustness, model transparency and normalisation of skewed dataset [38]. This guarantees the results are not only technically valid but also practically interpretable and implementable in roadway safety assessments and design practices.

Table 3. Model validation results based on LOOCV test

Response variable	RMSE	MAE	Residual normality Check (p-value)
$\sqrt{Cr\_T}$	0.121604	0.101	0.994
$\sqrt{Cr\_D}$	0.091	0.078	0.927
$\sqrt{Cr\_N}$	0.076	0.058	0.355
$\sqrt{Cr\_M}$	0.156	0.128	0.936
$\sqrt{Cr\_C}$	0.178	0.145	0.729
$\sqrt{Cr\_H}$	0.164	0.127	0.670

Table 4. Model comparison results

Model	Model type	R^2	RMSE	MAE
Total crash rate model	BSR	0.686	0.055	0.041
	LR	0.118	0.155	0.114
	RF	-0.124	0.175	0.142
Daytime crash rate model	BSR	0.797	0.068	0.058
	LR	0.931	0.040	0.034
	RF	0.971	0.026	0.020
Nighttime crash rate model	BSR	0.710	0.076	0.058
	LR	0.778	0.065	0.050
	RF	0.814	0.059	0.048
Crash rate model for MTW	BSR	0.744	0.125	0.100
	LR	0.747	0.124	0.102
	RF	0.920	0.070	0.060
Crash rate model for CAR	BSR	0.43	0.145	0.118
	LR	0.49	0.137	0.113
	RF	0.92	0.053	0.044
Crash rate model for HCV	BSR	0.55	0.150	0.115
	LR	0.57	0.147	0.119
	RF	0.91	0.067	0.053

BSR- Backward stepwise regression; LR- Lasso regression; RF- Random Forest

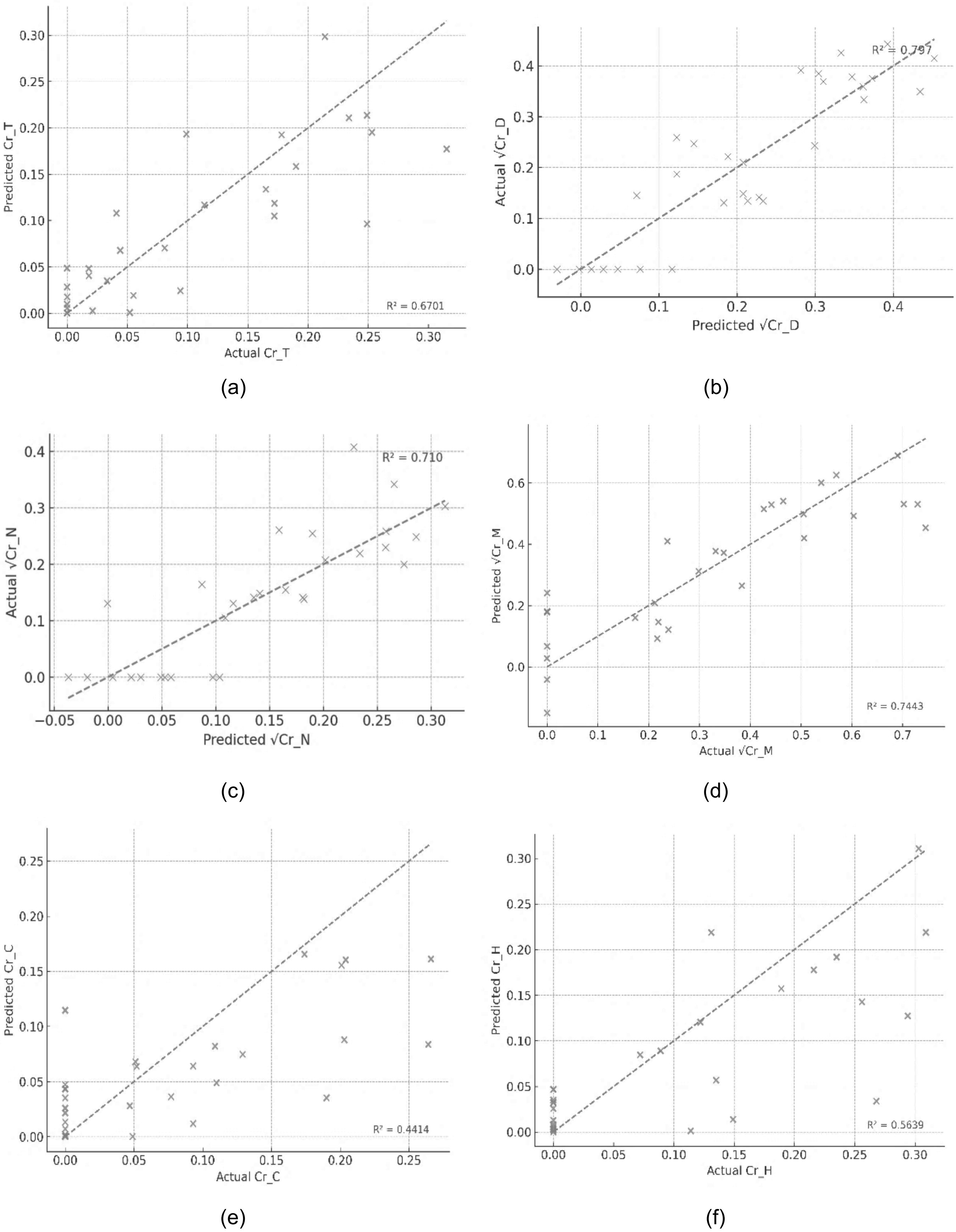

Fig. 4. Predicted vs Actual crash rate results for (a) Total, (b) Daytime, (c) Nighttime, (d) MTW, (e) CAR and (f) HCV

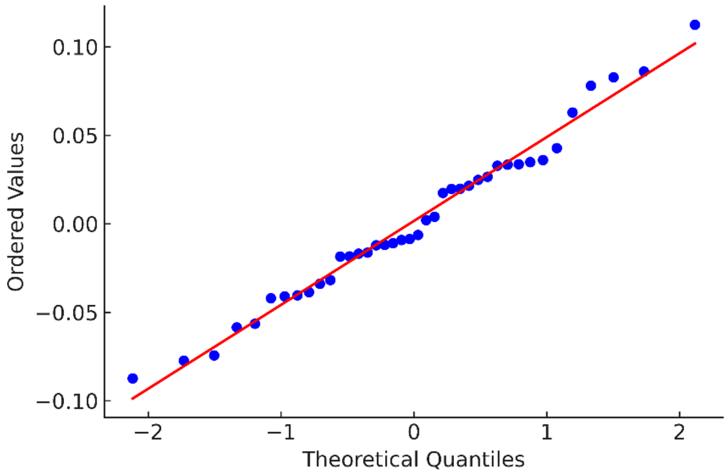

Fig. 5. QQ plot for checking normality of CAR model

There is high chance of LASSO selecting variables randomly from correlated group without considering its importance in real-time condition [39]. Also, random forest method emphasizes accuracy optimisation rather than considering actual on-field relation between significant and dependant variables [40]. Due to these limitations, and specific advantage of BSR, it was adopted in the current study. The study's findings support the use of backward stepwise regression as a robust and interpretable approach in small-sample safety modelling.

4. Conclusions

This study developed and validated crash rate prediction models tailored to vehicle type (MTW, CAR, HCV) and time of day (daytime, nighttime) using a 30-kilometer stretch of rural highway in Karnataka, India. Through comprehensive data collection by leveraging high-resolution LiDAR for geometric profiling and laser speed cameras for operating speed data, key geometric and operational parameters influencing crash risk were identified. The main conclusions drawn from current study are:

Vehicle-specific and time-specific crash rate models achieved reasonable predictive strength, with R^2 values ranging from 0.43 to 0.80, demonstrating the feasibility of disaggregated crash modelling on rural horizontal curves.

Sight Distance Deficiency (S_Def) was found to be the most consistent and influential predictor across all models, reinforcing the role of visibility in crash mitigation. Other significant factors included curvature (Curt), tangent lengths (Atl and Dtl), and operating speed differentials between vehicle classes and along curve segments. Notably, the influence of these factors varied by vehicle type. Curvature and sight-distance deficiency were dominant for two-wheelers and heavy vehicles, while transition lengths had greater effect on passenger-car crashes indicating the need for vehicle-specific design considerations.

While Random Forest and Lasso offered superior predictive accuracy, BSR mothod was selected for its interpretability, transparency, and alignment with evidence-based road safety design practices. Square-root transformation of the dependent variable improved linearity and residual behaviour. The models demonstrated strong internal validity, with R^2 values ranging from 0.43 (for cars) to 0.80 (daytime crashes), and Kolmogorov-Smirnov tests confirming residual normality ($p>0.05$) in all cases.

The developed models provide a practical tool for infrastructure designers, policymakers, and safety auditors to assess curve-related crash risk and prioritize geometric improvements.

Future scope

While the crash rate model for CAR exhibits statistical significance and acceptable predictive ability, the relatively low R^2 value highlights the need to incorporate additional influencing factors to improve its predictive strength.

Given the constraints of a limited dataset in the current study, model predictability could be significantly enhanced with a larger and more comprehensive data pool. Despite the challenges, future research should aim to collect extensive datasets to enable more refined and accurate crash prediction on rural highways.

Acknowledgement

We greatly appreciate the contributions in data collection and segregation by staff National Institute of Technology Karnataka, India. The authors appreciate the technical assistance provided by the Design Studio Laboratory at National Institute of Technology Karnataka, India.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability: Data will be made available on request

References

[1] T. Toroyan. (2009). Global status report on road safety. *Inj Prev*, 15(4), 286. doi: 10.1136/ip.2009.023697

[2] MoRTH. (2022). Road Accidents in India 2022. Ministry of Road Transport and Highways, Government of India.

[3] G. Yannis et al. (2016). Use of Accident Prediction Models in Road Safety Management - An International Inquiry. *Transportation Research Procedia*, 14, 4257-4266. doi: 10.1016/j.trpro.2016.05.397

[4] F.J. Camacho-Torregrosa, A.M. Pérez-Zuriaga, J.M. Campoy-Ungría, A. García-García. (2013). New geometric design consistency model based on operating speed profiles for road safety evaluation. *Accident Analysis & Prevention*, 61, 33-42. doi: 10.1016/j.aap.2012.10.001

[5] M.R.O.B.C. Macedo, M.L.A. Maia, E.R.K. Rabbani, O.C.C.L. Neto, M. Andrade. (2022). Traffic accident prediction model for rural highways in Pernambuco. *Case Studies on Transport Policy*, 10(1), 278-286. doi: 10.1016/j.cstp.2021.12.009

[6] R. Avelar, P. Carlson. (2014). Link between pavement marking retroreflectivity and night crashes on Michigan two-lane highways. *Transportation Research Record: Journal of the Transportation Research Board*, 2404(1), 59-67. doi: 10.3141/2404-07

[7] K. Kolody, D. Perez-Bravo, J. Zhao, T.R. Neuman. (2014). NCHRP 17-50: Highway Safety Manual User Guide.

[8] S. Cafiso, G. La Cava, A. Montella. (2007). Safety index for evaluation of two-lane rural highways. *Transportation Research Record: Journal of the Transportation Research Board*, 2019(1), 136-145. doi: 10.3141/2019-17

[9] J.C.W. Ng and T. Sayed. (2004). Effect of geometric design consistency on road safety. *Canadian Journal of Civil Engineering*, 31, 218-227. doi: 10.1139/l03-090

[10] M.O. Igene, O.R. Ogirigbo. (2018). Evaluating the geometric design consistency and road safety on two-lane single carriageways using operating speed criteria. *International Journal of Engineering and Emerging Scientific Discovery*, 3(2), 69-81.

[11] S.M. Gemechu and G.S. Tulu. (2021). Safety Effects of Geometric Design Consistency on Two-lane Rural Highways: The Case of Ethiopia. *American Journal of Traffic and Transportation Engineering*, 6(4), 107-115. doi: 10.11648/j.ajtte.20210604.11

[12] X. Wang, Q. Zhou, M. Quddus, T. Fan, S. Fang. (2018). Speed, speed variation and crash relationships for urban arterials. *Accident Analysis & Prevention*, 113, 236-243. doi: 10.1016/j.aap.2018.01.032

[13] E. Choueiri, R. Lamm, J.H. Kloeckner, T. Mailaender. (1994). Safety aspects of individual design elements and their interactions on two-lane highways: international perspective. *Transportation Research Record*, 1445, 34-46.

[14] X. Yan, J. He, C. Zhang, C. Wang, Y. Ye, P.

Qin. (2023). Temporal instability and age differences of determinants affecting injury severities in nighttime crashes. *Analytic Methods in Accident Research*, 38, 100268. doi: 10.1016/j.amar.2023.100268

[15] S. Regev, J.J. Rolison, S. Moutari. 2018. Crash risk by driver age, gender, and time of day using a new exposure methodology. *Journal of Safety Research*, 66, 131-140. doi: 10.1016/j.jsr.2018.07.002

[16] A. Charly and T. V. Mathew. (2024). Identifying risky driving behavior: a field study using instrumented vehicles. *Transportation Letters*, 16(7), 688-702. doi: 10.1080/19427867.2023.2233782

[17] A. Jacob and M.V.L.R. Anjaneyulu. (2013). Operating speed of different classes of vehicles at horizontal curves on two-lane rural highways. *Journal of Transportation Engineering*, 139(3), 287-294. doi: 10.1061/(ASCE)TE.1943-5436.0000503

[18] R. Bhavsar, A. Amin, and L. Zala. (2021). Development of Model for Road Crashes and Identification of Accident Spots. *International Journal of Intelligent Transportation Systems Research*, 19, 99-111. doi: 10.1007/s13177-020-00228-z

[19] J. Goyani, S. Arkatkar, G. Joshi, and S. Easa. (2024). Modified Geometric Design Consistency Criteria for Two-Lane Rural Highways Based on Crash Severity. *Transportation Research Record*, 2678(3), 279-291. doi: 10.1177/03611981231179469

[20] P.M. Chaudhari, J. Goyani, S. Arkatkar, G. Joshi, and S.M. Easa. 2022. Design Consistency Evaluation of Two-Lane Rural Highways in Hilly Terrains. *Transportation Research Procedia*, 62, 75-82. doi: 10.1016/j.trpro.2022.02.010

[21] J. Goyani, P. Chaudhari, S. Arkatkar, G. Joshi, and S.M. Easa. (2022). Operating Speed Prediction Models by Vehicle Type on Two-Lane Rural Highways in Indian Hilly Terrains. *Journal of Transportation Engineering, Part A: Systems*, 148(3), 04022001. doi: 10.1061/jtepbs.0000644

[22] V. Malaghan, D.S. Pawar, and H. Dia. (2020). Modeling Operating Speed Using Continuous Speed Profiles on Two-Lane Rural Highways in India. *Journal of Transportation Engineering, Part A: Systems*, 146(11), 04020124. doi: 10.1061/jtepbs.0000447

[23] V. Martinelli, R. Ventura, M. Bonera, B. Barabino, and G. Maternini. (2023). Estimating operating speed for county road segments – Evidence from Italy. *International Journal of Transportation Science and Technology*, 12(2), 560-577. doi: 10.1016/j.ijtst.2022.05.007

[24] IRC. (1972). Traffic Census on Non-Urban Roads (First Revision). IRC:9-1972. THE INDIAN ROADS CONGRESS.

[25] IRC. (2015). Guidelines For Traffic Forecast on Highways (First Revision). IRC:108-2015. THE INDIAN ROADS CONGRESS.

[26] P. Misaghi and Y. Hassan. (2005). Modeling operating speed and speed differential on two-lane rural roads. *Journal of Transportation Engineering*, 131(6), 408-418. doi: 10.1061/(ASCE)0733-947X(2005)131:6(408)

[27] J. You, S. Oh, J. Park, H. Song, and Y. Kim. (2021). A novel LiDAR sensor alignment inspection system for automobile productions using 1-D photodetector arrays. *Measurement*, 183, 109817. doi: 10.1016/j.measurement.2021.109817

[28] B. Yu et al. (2024). A review of road 3D modeling based on light detection and ranging point clouds. *Journal of Road Engineering*, 4(4), 386-398. doi: 10.1016/j.jreng.2024.04.009

[29] Y.-C. Lee, J.W. Ma, and F. Leite. (2023). A parametric approach towards semi-automated 3D as-built modeling. *Journal of Information Technology in Construction*, 28, 806-825. doi: 10.36680/j.itcon.2023.041

[30] IRC. (1976). IRC:66-1976. Recommended Practice for Sight Distance on Rural Highways. The Indian Road Congress.

[31] AASHTO. (2010). HIGHWAY SAFETY MANUAL.

[32] A. Bellary and S.N. Suresha. (2022). ANN Model to Predict Joint Stiffness of White-topped Pavements Using Falling Weight Deflectometer (FWD) Data. *International Journal of Pavement Research and Technology*, 16, 370-384. doi: 10.1007/s42947-021-00137-8

[33] K. Zhang and M. Hassan. (2019). Crash severity analysis of nighttime and daytime highway work zone crashes. PLoS One, 14(8), e0221128. doi: 10.1371/journal.pone.0221128

[34] D.T. Godumula and K.V.R.R. Shankar. (2023). Safety evaluation of horizontal curves on two lane rural highways using machine learning algorithms: A priority-based study for sight distance improvements. *Traffic Injury Prevention*, 24(4), 331-337. doi: 10.1080/15389588.2023.2184203

[35] Q. Li, H. Zhang, G. Sun, Y. Han, H. Liu. (2025). The impact of curve radius variations on vehicle driving states on 2-lane roads. *Traffic Injury Prevention*, 26(6), 703-710. doi: 10.1080/15389588.2024.2447088

[36] X. Zhou, B. Pan, and Y. Shao. (2021). Evaluating the impact of sight distance and geometric alignment on driver performance in freeway exits diverging area based on simulated driving data. *Sustainability*, 13(11), 6368. doi: 10.3390/su13116368

[37] L. Skaug, M. Nojoumian, N. Dang, and A. Yap. (2025). Road Crash Analysis and Modeling: A Systematic Review of Methods, Data, and Emerging Technologies. *Applied Sciences*, 15(13), 7115. doi: 10.3390/app15137115

[38] J. Pek, O. Wong, and A.C. Wong. (2017). Data Transformations for Inference with Linear Regression: Clarifications and Recommendations. *Practical Assessment, Research and Evaluation*, 22(1), 9. https://doi.org/10.7275/2w3n-0f07

[39] L. Freijeiro-González, M. Febrero-Bande, and W. González-Manteiga. (2022). A Critical Review of LASSO and Its Derivatives for Variable Selection Under Dependence Among Covariates. *International Statistical Review*, 90(1), 118-145. doi: 10.1111/insr.12469

[40] J.L. Speiser, M.E. Miller, J. Tooze, and E. Ip. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 134, 93-101. doi: 10.1016/j.eswa.2019.05.028