Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Statistical and Machine Learning modelling of annual average daily traffic on South-Eastern Nigerian highways

Henry Kene Ugwuanyi

henrykene9@gmail.com

Michael Okpara University of Agriculture, Umudike

Fidelis Onyebuchi Okafor

University of Nigeria, Nsukka

Chidozie Charles Nnaji

University of Nigeria, Nsukka

Research Article

Keywords: annual average daily traffic, AADT estimation, traffic prediction, machine learning, Gradient Boosting, population density, rainfall, pothole volume, South-Eastern Nigeria, highway modelling

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11247632/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Statistical and Machine Learning modelling of annual average daily traffic on South-Eastern Nigerian highways

HENRY Kene Ugwuanyi[1], Fidelis Onyebuchi Okafor[2], Chidozie Charles Nnaji[3]
[1]Department of Civil Engineering, Michael Okpara University of Agriculture Umudike, Abia State, Nigeria, henrykene9@gmail.com, ORCID 0000-0002-8706-8176 (Corresponding Author), [2,3]Department of Civil Engineering, University of Nigeria Nsukka, Enugu State, Nigeria, [3]Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa.
[2]fidelis.okafor@unn.edu.ng,[3]chidozie.nnaji@unn.edu.ng.
Corresponding Author: henrykene9@gmail.com

Abstract

Reliable annual average daily traffic (AADT) estimates are essential for highway planning, pavement management, capacity assessment and infrastructure investment. This study develops and evaluates an integrated statistical and machine-learning framework for modelling AADT on five selected highways in South-Eastern Nigeria using demographic, economic, environmental and temporal predictors. The dataset comprises 100 state-year observations from 2001 to 2020 for the Umuahia–Ikot Ekpene, Onitsha–Owerri, Abakaliki–Ogoja, Enugu–Port Harcourt and Owerri–Port Harcourt highways. Predictors included passenger population density, gross domestic product (GDP), annual rainfall, pothole volume, year and highway effects. Pearson and Spearman correlations, multiple linear regression with HC3 heteroskedasticity-consistent inference, multicollinearity diagnostics and chronological machine-learning validation were applied. Mean pooled AADT was 1,313,110.42 PCU/day (SD = 569,491.41), with substantial spatial heterogeneity, ranging from 588,708.80 PCU/day on Abakaliki–Ogoja to 2,202,095.25 PCU/day on Enugu–Port Harcourt. Passenger population density showed the strongest Pearson association with AADT ($r = 0.791$), followed by GDP ($r = 0.676$), rainfall ($r = 0.501$), pothole volume ($r = 0.307$) and year ($r = 0.141$). The regression model explained 88.81% of in-sample variation ($R^2 = 0.8881$; adjusted $R^2 = 0.8769$). In chronological holdout testing for 2017–2020, Gradient Boosting achieved $R^2 = 0.5338$, RMSE = 396,916.31 PCU/day, MAE = 249,437.28 PCU/day and MAPE = 28.45%, outperforming the linear trend benchmark in RMSE. The findings demonstrate the value of combining interpretable statistical modelling with chronologically validated nonlinear prediction for AADT estimation in data-constrained highway systems.

Keywords: annual average daily traffic; AADT estimation; traffic prediction; machine learning; Gradient Boosting; population density; rainfall; pothole volume; South-Eastern Nigeria; highway modelling.

Highlights

- An integrated statistical and machine-learning framework was developed for AADT estimation on five South-Eastern Nigerian highways.
- The analysis combines demographic, economic, rainfall, pavement-condition and temporal information.
- Population density showed the strongest bivariate association with AADT.
- Gradient Boosting provided the strongest overall performance in the 2017–2020 chronological holdout.
- State effects were major contributors to predictive performance, demonstrating strong spatial heterogeneity.

1. INTRODUCTION

Annual average daily traffic (AADT) is a fundamental measure of roadway use and is routinely applied in highway planning, pavement design, capacity analysis, maintenance programming, safety assessment and transport investment (AASHTO, 2018). The accuracy of AADT estimation becomes particularly important where permanent counting infrastructure is limited and traffic data must be inferred from short-duration counts, historical observations or external explanatory variables (FHWA, 2022). Contemporary traffic-monitoring practice emphasizes systematic counting, temporal adjustment, quality assurance and appropriate treatment of spatial and temporal variation (Alourani et al. 2023).

Recent traffic-prediction research increasingly combines conventional statistical models with machine-learning methods (Adedeji et al. 2025). Linear models provide interpretable coefficients and hypothesis tests, whereas ensemble and kernel-based methods can represent nonlinearities and interactions (Afandizadeh 2024). However, predictive performance can be overstated when validation observations are randomly sampled from the same temporal distribution as the training data. For traffic applications, chronological holdout validation is therefore important because the intended application is generally prediction for later periods.

South-Eastern Nigeria presents a relevant setting for this problem. The region contains important inter-state highways linking commercial, industrial, administrative and agricultural centres. The selected highways are exposed to substantial differences in population concentration, economic activity, rainfall and pavement condition (Strien & Grêt-Regamey 2024). These characteristics provide a basis for examining whether AADT can be modelled more effectively by integrating external socioeconomic and environmental information rather than relying only on historical time trends (Nigam & Srivastava 2023).

1.1 RESEARCH GAP AND CONTRIBUTION

Three gaps motivate the study. First, evidence on long-term, multivariable AADT modelling for Nigerian inter-state highways remains limited relative to the extensive literature from Europe, North America and other highly instrumented transport systems (Imoh & Movahedi, 2025). Second, environmental variables such as annual rainfall and pothole volume have rarely been examined jointly with population density and GDP in a single long-term AADT framework. Third, model fit and predictive validation are often conflated. The present study addresses these gaps by integrating explanatory regression, robust inference, nonlinear machine learning, feature-importance analysis and chronological out-of-time validation within one locally calibrated dataset.

1.2 RESEARCH OBJECTIVES

1. characterize the spatial and temporal variation in AADT across the selected highways from 2001 to 2020;
2. quantify the distributions and temporal behaviour of AADT, GDP, passenger population density, rainfall and pothole volume;
3. estimate the strength and direction of associations between AADT and the explanatory variables;
4. estimate conditional relationships using multiple regression and evaluate the adequacy of regression assumptions;
5. compare conventional, regularized and nonlinear machine-learning models using chronological validation; and
6. identify the predictors that contribute most strongly to AADT prediction and discuss implications for highway planning and traffic management.

2. MATERIALS AND METHODS

3.1 Study Area and Highway Pavements

The study covers five selected highway pavements in South-Eastern Nigeria: Umuahia–Ikot Ekpene in Abia State; Onitsha–Owerri in Anambra State; Abakaliki–Ogoja in Ebonyi State; Enugu–Port Harcourt in Enugu State; and Owerri–Port Harcourt in Imo State. The highways were selected because of their strategic inter-state functions, traffic significance, data availability and exposure to pavement distress and environmental conditions relevant to the research questions.

Table 1 shows that the study covers five strategically selected highways across Abia, Anambra, Ebonyi, Enugu, and Imo States, representing important inter-state and commercial corridors in Southeastern Nigeria.

Table 1. Selected highway pavements and their roles in the study

States	Selected highway	Study role
Abia	Umuahia–Ikot Ekpene	Inter-state arterial highway
Anambra	Onitsha–Owerri	High-demand commercial/inter-state highway
Ebonyi	Abakaliki–Ogoja	Inter-state corridor with comparatively lower traffic scale
Enugu	Enugu–Port Harcourt	Major high-volume inter-state highway
Imo	Owerri–Port Harcourt	Major inter-state/commercial highway

From table 1, the selected highways differ in their functional roles and expected traffic intensity, ranging from the comparatively lower-traffic Abakaliki–Ogoja highway to the major high-volume Enugu–Port Harcourt highway. This variation provides an appropriate basis for examining spatial differences in AADT and assessing how demographic, economic, environmental, and temporal factors influence traffic demand across different highway settings.

Figure 1 is the map location of the Umuahia – Ikot Ekpene road. The map is presented on a scale of 1:1,000,000.

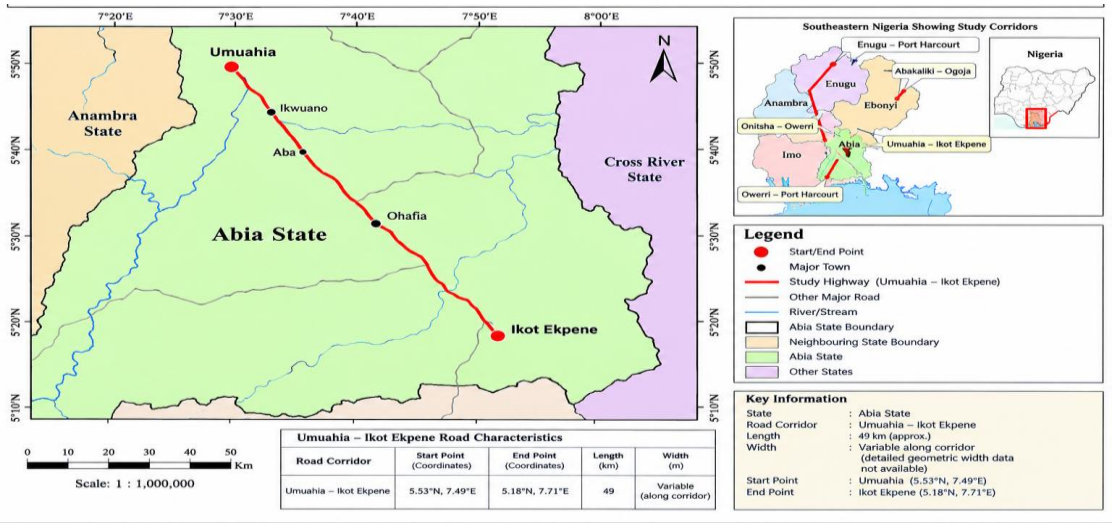


Figure 1. Map showing Umuahia – Ikot Ekpene road

The study highway is a critical federal arterial link (designated as part of the Trunk A network, A343) in southeastern Nigeria, physically connecting the capital of Abia State (Umuahia) to the commercial and historical transit hub of Akwa Ibom State (Ikot Ekpene). The highway extends approximately 49 kilometers in length.

Figure 2 shows the geospatial location of Onitsha – Owerri road on a scale of 1:1,000,000.

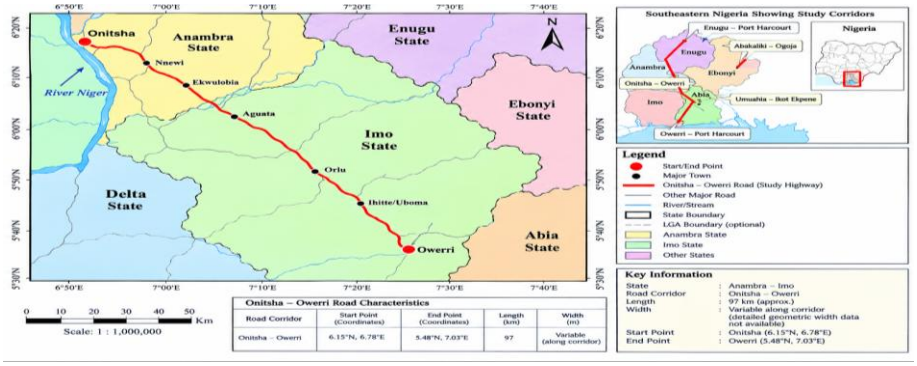


Figure 2. Map showing Onitsha – Owerri road

The study area is located on the Onitsha – Owerri Expressway (Trunk A highway). This is a critical economic artery spanning approximately 97 kilometers, physically connecting the primary commercial and industrial hub of Anambra State (Onitsha) to the capital city of Imo State (Owerri). The highway connects some important towns such as Oba, Nkwelle Ezunaka, Mgbidi, and Awo-Omamma.

Figure 3 presents the map of Abakaliki – Ogoja road on a scale of 1: 1,000,000

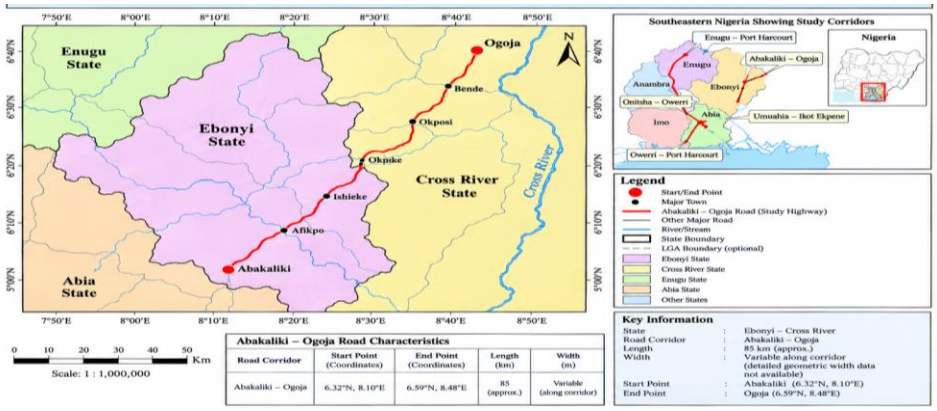


Figure 3. Map showing Abakaliki – Ogoja road

The study area embraces the Abakaliki – Ogoja Highway (designated as part of the Trans-African Highway network and the critical Federal Trunk A network in Nigeria). This major trans-regional pavement stretches approximately 85 kilometers, establishing an indispensable infrastructure link between the capital city of Ebonyi State (Abakaliki) and the key agricultural and regional hub of northern Cross River State (Ogoja).

Figure 4 shows road map of Enugu – Port Harcourt road at a scale of 1: 1,000,000.



Figure 4. Map of Enugu – Port Harcourt road

The Enugu – Port Harcourt Expressway is a vital Federal Trunk 'A' highway (Route A3 segment) that serves as a high-volume north-south logistics artery. It traverses four major Southeastern and Niger Delta states: Enugu, Abia, Imo, and Rivers.

Figure 5 presents the geospatial map of Owerri – Port Harcourt road at a plot scale of 1:1,000,000. This means that 1mm of a distance on the paper represents 1,000,000mm distance on the physical highway pavement.



Figure 5. Map showing Owerri – Port Harcourt road

The Owerri – Port Harcourt Road is a vital interstate arterial highway (Federal Route A3 segment) that serves as an economic lifeblood connecting southeastern Nigeria (Imo State capital, Owerri) directly to the Niger Delta oil-hub (Rivers State capital, Port Harcourt) (Mgbemena et al., 2024). The pavement transverses a flat-to-gently-undulating terrain within a tropical monsoon climate belt characterized by prolonged, high-intensity rainfall seasons (Mgbemena et al., 2024).

3.2 DATA SOURCES AND VARIABLES

The analytical dataset contains 100 observations, corresponding to five state/highway observations for each year from 2001 to 2020. Historical AADT and passenger population density data were obtained from the Federal Road Safety Corps (FRSC). Rainfall data were obtained from the Nigerian Meteorological Agency (NiMet) (WMO., 2024), while socioeconomic information was sourced from the National Bureau of Statistics (NBS). Pothole volume information was obtained from road-condition/maintenance records by the Federal Roads Maintenance Agency (FERMA). Table 2 presents the dependent variable and the various explanatory variables, their roles and meanings.

Table 2. Study data and their roles

Variable	Role	Source	Measurement/meaning
AADT	Dependent	FRSC	Annual average daily traffic in PCU/day
Passenger population density	Independent	FRSC	Demographic/travel-demand indicator
GDP	Independent	NBS	State economic activity indicator
Rainfall	Independent	NiMet	Annual precipitation amount
Pothole volume	Independent	FERMA	Indicator of pavement deterioration
Year	Temporal	temporal	Calendar year, 2001–2020
State	Spatial categorical factor	spatial	State/highway identity

Table 2 identifies AADT as the dependent variable, while passenger population density, GDP, rainfall, and pothole volume serve as explanatory variables representing demographic, economic, environmental, and pavement conditions, respectively. Year and state provide temporal and spatial dimensions, enabling the study to examine changes in traffic demand over time and differences among the selected highway pavements.

3.3 STATISTICAL MODELLING

The pooled regression model was specified with state effects, GDP, passenger population density, rainfall, pothole volume and year. In general form:

$$AADT_i = \beta_0 + \beta_1 GDP_i + \beta_2 PPD_i + \beta_3 Rainfall_i + \beta_4 Pothole_i + \beta_5 Year_i + State\ effects + \varepsilon_i \quad (1)$$

Pearson correlation was used to quantify linear association, while Spearman correlation was used to assess monotonic association without requiring a strictly linear relationship. Regression inference used HC3 heteroskedasticity-consistent standard errors because the diagnostic analysis identified non-constant residual variance.

3.4 REGRESSION DIAGNOSTICS

The analysis assessed residual normality using the Jarque–Bera test, heteroscedasticity using the Breusch–Pagan test, multicollinearity using variance inflation factors (VIF), serial dependence using the Durbin–Watson statistic, and residual behaviour using residual-versus-fitted and related diagnostic plots. These diagnostics were not treated as a search for perfect assumptions; rather, they were used to determine how conventional regression inference should be interpreted and why complementary predictive methods were required.

3.5 MACHINE-LEARNING MODELS

Traditional linear trend model as a transparent temporal benchmark.

Multiple Linear Regression and regularized linear models, including Ridge and Elastic Net.

Decision Tree Regression.

Random Forest Regression.

Extra Trees Regression.

Gradient Boosting Regression.

Support Vector Regression.

K-Nearest Neighbours Regression.

The data were divided chronologically into training observations from 2001–2016 and holdout observations from 2017–2020. A secondary 2017–2019 evaluation was retained to examine how model performance changed when the unusual 2020 observations were excluded. Model performance was evaluated using R^2, RMSE, MAE and MAPE, supplemented by median absolute error, explained variance and maximum error where appropriate.

3.6 MODEL INTERPRETATION

Feature importance and sensitivity analyses were used to complement predictive metrics. Impurity-based importance from Random Forest was used to quantify the relative contribution of predictors within the fitted ensemble. Holdout permutation importance was also examined to assess how predictive performance changed when information associated with a feature was disrupted. These measures were interpreted as predictive importance rather than causal effects.

4. RESULTS

4.1 DESCRIPTIVE STATISTICAL CHARACTERISTICS OF AADT ACROSS THE HIGHWAYS

The pooled AADT mean was 1,313,110.42 PCU/day, with a median of 1,264,432.50 PCU/day and a standard deviation of 569,491.41 PCU/day. The minimum and maximum values were 364,259 and 2,733,585 PCU/day, respectively. The distribution had a positive skew of approximately 0.56, indicating a moderately extended upper tail associated with higher-volume highway observations.

Figure 4.1 presents the temporal trends in AADT for each of the highways included in the study— Umuahia – Ikot Ekpene road**,** Onitsha – Owerri road**,** Abakaliki – Ogoja road**,** Enugu – Port Harcourt road**, and** Owerri – Port Harcourt road**.**

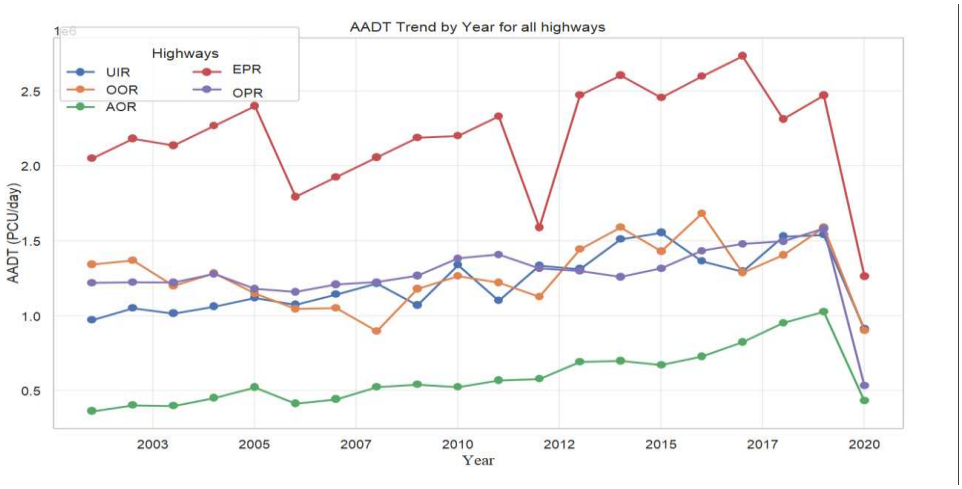


Figure 1. AADT trend of all the highways

One of the most prominent features of Figure 4.1 is the substantial decline in AADT between 2019 and 2020 across all five highways. The results demonstrate that all states experienced a sharp reduction in traffic volume during this period. However, the magnitude of the decline varied substantially across the pavements. Owerri – PHC road experienced the largest reduction of −66.3%, followed by Abakaliki – Ogoja road, −57.8%. Enugu – PHC road experienced a decline of −48.9%, while Onitsha – Owerri road and Umuahia – Ikot Ekpene road experienced reductions of −43.3% and −40.9%, respectively. These results indicate that the 2019–2020 disturbance was not proportional across all highways**.** If the decline had affected all states in the same manner, the percentage reductions would have been similar. Instead, the variation from approximately 40.9% on Umuahia – Ikot Ekpene road to 66.3% in Owerri – PHC road demonstrates considerable geographical heterogeneity in the response of traffic volumes. The differing magnitude of the 2019–2020 reductions has an important methodological implication for AADT modelling.

Table 3 indicates the statistical means of the AADT for the various highways under study, and the respective acronym for the highways.

Table 3. Mean values of the AADT on the highways

State/highway	Highway Acronym	Mean AADT (PCU/day)
Umuahia–Ikot Ekpene Road(Abia)	UIR	1,226,869.25
Onitsha–Owerri Road (Anambra)	OOR	1,273,647.00
Abakaliki–Ogoja Road (Ebonyi)	AOR	588,708.80
Enugu–Port Harcourt Road(Enugu)	EPR	2,202,095.25
Owerri–Port Harcourt Road (Imo)	OPR	1,274,231.80

EPR recorded the highest mean AADT, whereas AOR recorded the lowest. The contrast demonstrates substantial spatial heterogeneity in traffic demand. The 2019–2020 period also represented a major temporal break: AADT declined by approximately 40.9% on UIR, 43.3% on OOR, 57.8% on AOR, 48.9% on EPR and 66.3% in OPR. At the pooled level, mean AADT declined from 1,641,919 in 2019 to 808,602 in 2020, a reduction of approximately 50.8%.

4.1.1. Correlation Analysis.

Table 4 presents the results of the correlation analysis conducted using Pearson (r) and Spearman (ρ) relationships.

Table 4. Correlation Analysis

Predictor	Pearson r with AADT	Spearman ρ with AADT
Passenger population density	0.791	0.819
GDP	0.676	0.526
Rainfall	0.501	0.576
Pothole volume	0.307	0.450
Year	0.141	0.203

From table 4, passenger population density had the strongest simple association with AADT under both Pearson and Spearman measures. GDP was also positively associated with AADT. Rainfall displayed a moderate positive bivariate association, while pothole volume and year had weaker associations. The difference between Pearson and Spearman values indicates that some relationships are not adequately represented by a simple straight-line form.

The predictor correlation structure revealed substantial overlap. Rainfall and pothole volume had a correlation of approximately 0.897, while rainfall and GDP correlated at approximately 0.763 and rainfall and passenger population density at approximately 0.644. These relationships indicate that individual regression coefficients should be interpreted conditionally and with attention to multicollinearity.

Figure 2 presents the Pearson correlation heatmap for AADT and the selected explanatory variables, namely passenger population Density, Gross Domestic Product (GDP), rainfall, pothole volume, and year**.**

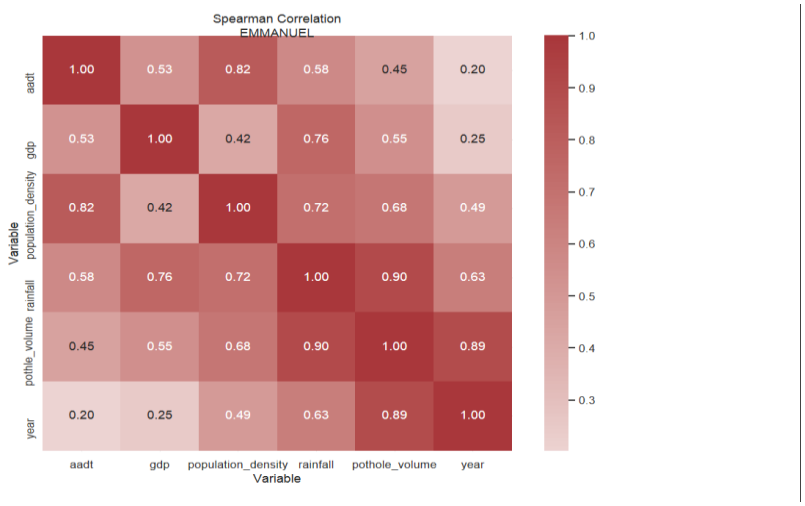


Figure 2 Pearson Correlation Heatmap for the variables

From figure 2, Pearson's linear coefficients with AADT are density 0.791, GDP 0.676, rainfall 0.501, pothole volume 0.307, and year 0.141. Passenger population density therefore supplies the strongest unadjusted straight-line association.

The largest predictor pairing is rainfall-pothole volume at 0.897; rainfall-GDP is 0.763 and rainfall-density 0.644. These coefficients show that adjusted OLS terms will compete for shared variation.

The heatmap is descriptive: a positive cell does not identify a mechanism or establish temporal precedence. Its main modelling implication is the need for Variance Inflation Factor (VIF) diagnostics, robust uncertainty, and out-of-time validation.

4.1.2. Multiple Linear Regression Result

Table 5 presents the multiple regression coefficients, HC3 robust standard errors, t-statistics, and significance levels for the effects of GDP, passenger population density, rainfall, pothole volume, and year on AADT.

Table 5. Multiple Regression Results

Term	**Coefficient**	**HC3 SE**	**t**	**p**
GDP	0.012518	0.0042	3.0139	0.0026
Passenger population density	0.021222	0.0105	2.0116	0.0443
Rainfall	−1,142.58	355.23	−3.2165	0.0013
Pothole volume	−115,424.25	146,087.97	−0.7901	0.4295
Year	65,104.58	19,088.18	3.4107	<0.001

Table 5 shows that GDP, passenger population density, and year have statistically significant positive effects on AADT, with p-values of 0.0026, 0.0443, and <0.001, respectively. Rainfall has a statistically significant negative effect on AADT (coefficient = −1,142.58; p = 0.0013), indicating an inverse conditional association after controlling for the other variables. Pothole volume has a negative but statistically insignificant coefficient (−115,424.25; p = 0.4295), suggesting that its independent effect is not statistically

distinguishable from zero in the fitted model. Generally, the results indicate that economic activity, passenger population density, and temporal growth are positively associated with AADT, while rainfall is negatively associated with AADT, with the estimates interpreted using HC3 robust standard errors.

4.1.3. Regression Results and Diagnostics.

The pooled OLS model produced $R^2 = 0.8881$ and adjusted $R^2 = 0.8769$, with an F-statistic of 38.3563 ($p = 2.168 \times 10^{-24}$). The overall model therefore explained a large proportion of the in-sample variation in AADT. GDP, passenger population density, rainfall and year were statistically distinguishable from zero at the 5% level under the reported HC3 inference, whereas pothole volume was not. The coefficients represent conditional associations, not causal effects

Table 6 presents the diagnostic tests used to assess the normality, homoscedasticity, serial dependence, and multicollinearity assumptions underlying the multiple regression model.

Table 6. Regression Diagnostics

Diagnostic	**Statistic/result**	**Interpretation**
Jarque–Bera	134.3623; $p = 6.662 \times 10^{-30}$	Residual normality rejected
Breusch–Pagan LM	29.5645; $p = 5.202 \times 10^{-4}$	Heteroscedasticity detected
Breusch–Pagan F	4.1974; $p = 1.493 \times 10^{-4}$	Heteroscedasticity detected
Durbin–Watson	1.7132	Possible mild positive serial dependence
Rainfall VIF	11.4987	Potentially serious multicollinearity
Pothole VIF	9.0925	High multicollinearity, near 10
GDP VIF	4.2246	Moderate multicollinearity
Year VIF	3.8475	Moderate multicollinearity
Population density VIF	1.8926	Low multicollinearity

From table 6, the diagnostic results demonstrate why the regression analysis should not be interpreted solely through classical OLS standard errors. Residual normality was rejected and heteroscedasticity was detected. The rainfall VIF exceeded 10 and pothole VIF was close to 10, confirming that the two environmental predictors share substantial information. HC3 robust standard errors were therefore used for inference, while machine-learning models provided a complementary predictive framework.

4.2. MACHINE-LEARNING MODEL PERFORMANCE

This validation approach is particularly appropriate for the research because it respects the temporal ordering of the data: earlier observations are used for model development, while later observations are reserved for evaluating predictive performance

Table 7 compares the predictive performance of the evaluated machine-learning and conventional regression models on the 2017–2020 chronological test set using R^2, RMSE, MAE, and MAPE. The results provide a basis for assessing the models' ability to generalize AADT predictions to later observations beyond the training period.

Table 7. Machine-Learning Performance

Model	**R^2 (2017–2020)**	**RMSE (PCU/day)**	**MAE (PCU/day)**	**MAPE (%)**
Gradient Boosting	0.5338	396,916.31	249,437.28	28.45
K-Nearest Neighbours	0.5198	402,801.03	298,293.81	32.33
Random Forest	0.5126	405,847.31	275,226.98	29.97
Extra Trees	0.5110	406,501.20	275,544.05	29.84
Decision Tree	0.4749	421,241.95	305,556.83	32.07
Traditional Linear Trend	0.4712	422,727.98	275,407.30	31.14
Ridge Regression	0.4287	439,362.57	289,089.58	31.90

On the primary 2017–2020 holdout period, the Gradient Boosting model achieved $R^2 = 0.5338$, RMSE = 396,916.31 PCU/day, MAE = 249,437.28 PCU/day and MAPE = 28.45%. The model therefore reduced RMSE relative to the traditional linear trend benchmark by approximately 25,812 PCU/day. Random Forest and Extra Trees were also competitive. KNN produced a comparatively strong holdout R^2 but substantially less favourable MAE and MAPE than Gradient Boosting.

A major validation finding was the difference between the 2017–2019 and 2017–2020 evaluation periods. Gradient Boosting achieved $R^2 = 0.9278$ and RMSE = 141,283.01 PCU/day over 2017–2019, but performance declined when 2020 was included. This indicates that the 2020 observations represented a difficult distributional condition for models trained on earlier years. The result reinforces the importance of preserving chronological order in validation.

Generally, the results indicate that the nonlinear machine-learning approaches, particularly Gradient Boosting, provided additional predictive capability for modelling AADT under the chronological holdout validation used in the study, as illustrated by Figure 3.

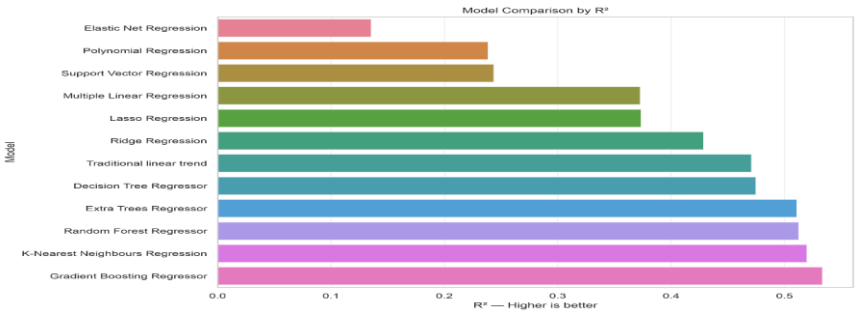


Figure 3. Model Comparison based on R^2

The results show a clear advantage for some of the machine-learning models over the weaker conventional and regularized regression approaches. The Gradient Boosting Regressor recorded the highest R^2 value of 0.534, followed by K-Nearest Neighbours (KNN) Regression with 0.520 and the Random Forest Regressor with 0.513. The close values of these three leading models indicate that they were all capable of capturing substantial predictive patterns in the AADT data.

Figure 4 presents a comparative assessment of the predictive performance of the models developed for this thesis using the Root Mean Square Error (RMSE).

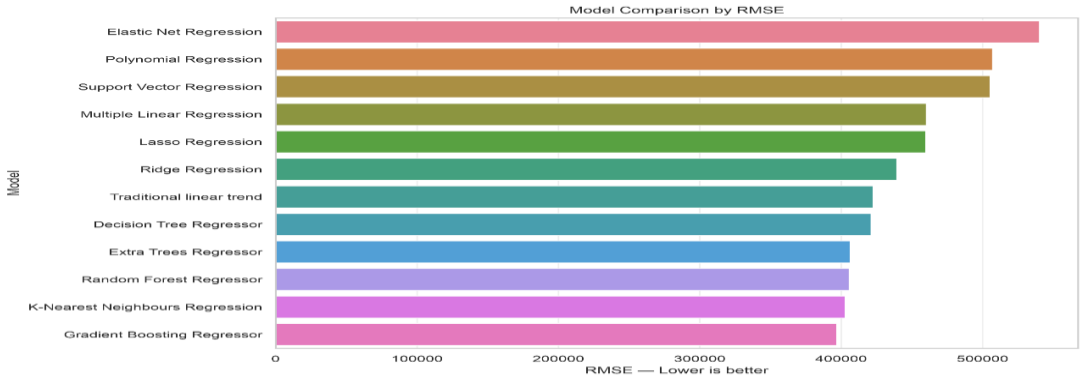


Figure 4. Model Comparison based on RMSE.

RMSE and MAE are important for evaluating AADT prediction because large prediction errors can have practical consequences for road capacity planning, pavement loading, maintenance, traffic management, and allocation of highway development resources, while substantial overestimation may lead to unnecessary infrastructure expenditure.

Figure 5 shows that the Gradient Boosting Regressor recorded the lowest MAE (approximately 249,437.3 PCU/day), indicating that its predicted AADT differed from the observed values by an average of about 249,437 PCU/day, regardless of whether the model overestimated or underestimated traffic.

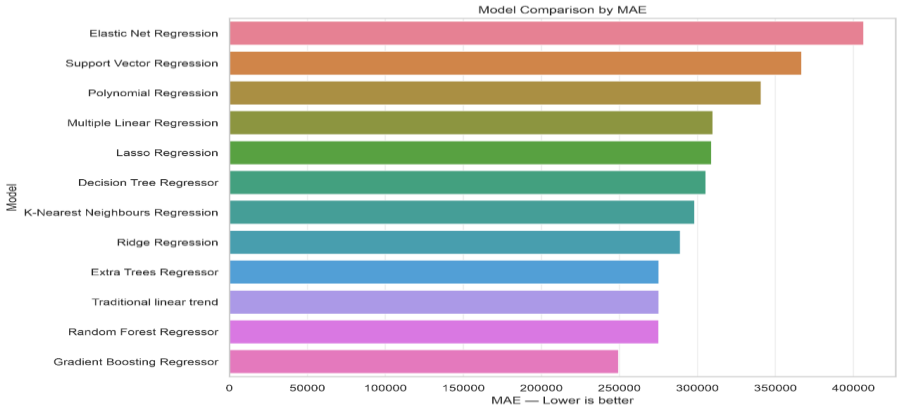


Figure 5 Model Comparison based on MAE

The most prominent result in Figure 4.34 is the superior performance of the GradientBoosting Regressor, which recorded the lowest MAE of approximately 249,437.3PCU/day. This means that, on average, the Gradient Boosting model's predicted AADT differed from the observed AADT by approximately 249,437PCU/day, without considering whether the prediction was an overestimation or an underestimation.

4.3. FEATURE IMPORTANCE AND SENSITIVITY ANALYSIS

Random Forest impurity importance measures how much each feature contributes to reducing prediction impurity across the decision trees in the Random Forest ensemble. A higher value indicates that the feature was used more extensively and/or more effectively to create splits that improved prediction within the fitted Random Forest model.

Table 8 presents the relative importance of the predictor variables and spatial indicators in the Random Forest model based on impurity-derived feature importance. The results indicate the extent to which each feature contributes to the model's prediction of AADT, thereby providing insight into the variables with greater or lesser predictive influence.

Table 8. Feature Importance and Sensitivity

Feature	**Random Forest impurity importance**
Highway/Enugu – Port Harcourt indicator	0.6150
Passenger population density	0.1443
Highway/Abakaliki - Ogoja indicator	0.1247
GDP	0.0721
Year	0.0212
Pothole volume	0.0149
Rainfall	0.0055

Table 8 shows that the Enugu–Port Harcourt highway indicator had the highest Random Forest impurity importance (0.6150), indicating that highway-specific differences contributed substantially to the model's prediction of AADT. Passenger population density (0.1443) and the Abakaliki–Ogoja indicator (0.1247) were the next most influential features, followed by GDP (0.0721), while year (0.0212), pothole volume (0.0149), and rainfall (0.0055) made comparatively smaller contributions. Conclusively, the results suggest that spatial highway characteristics and demographic conditions were more influential in the Random Forest model than the environmental and temporal variables, although impurity importance represents predictive contribution rather than a causal effect.

4.4. CHRONOLOGICAL OUT-OF-TIME VALIDATION OF GRADIENT BOOSTING

The Gradient Boosting model was evaluated using a chronological out-of-time validation, with observations from 2001–2016 used for training and 2017–2020 reserved as an unseen test period. The model achieved an R^2 of 0.5338, indicating that it explained approximately 53.38% of the variation in AADT in the test

period, while the RMSE of 396,916.31 PCU/day and MAE of 249,437.28 PCU/day indicate moderate prediction errors. The MAPE of 28.45% further shows that the model's average percentage prediction error was about 28%, reflecting reasonable but imperfect generalization to later observations. In general, the results demonstrate that Gradient Boosting captured meaningful nonlinear relationships in the AADT data, although its moderate out-of-time performance also indicates the need for caution when applying the model to future periods, particularly given the relatively small dataset and structural changes in traffic conditions.

4.7 HIGHWAY-LEVEL GENERALISATION

State-level evaluation showed that predictive performance varied considerably across the five highways. For the 2017–2020 holdout period, RMSE was approximately 239,797 PCU/day for UIR, 262,156 for OOR, 220,228 for AOR, 608,944 for EPR and 492,110 for OPR. The highway-level R^2 values were volatile because each state contributed only four holdout observations. Negative R^2 values for some states should therefore be interpreted cautiously and alongside RMSE, MAE, MAPE and sample size.

5. DISCUSSION

5.1 SPATIAL AND TEMPORAL HETEROGENEITY

The pronounced differences among the five corridors demonstrate that AADT in South-Eastern Nigeria is spatially heterogeneous. Enugu–Port Harcourt operates at a much higher traffic scale than Abakaliki–Ogoja, while Abia, Anambra and Imo occupy intermediate levels. Such differences are consistent with the differing functions of the corridors, urban concentration, economic activity, network connectivity and trip-generation characteristics. The common decline in 2020 across all states, combined with different magnitudes of reduction, indicates that a regional disturbance can interact with location-specific conditions.

5.2 DEMOGRAPHIC AND ECONOMIC RELATIONSHIPS

The strong positive association between passenger population density and AADT supports the transportation-demand interpretation that population concentration creates trip-generation and trip-attraction pressure. The GDP–AADT association was also substantial, and GDP remained statistically significant in the multivariable model. This is consistent with contemporary AADT research that incorporates socioeconomic indicators. However, these relationships should be interpreted as associations because the observational design does not establish causal mechanisms.

5.3 ENVIRONMENTAL RELATIONSHIPS

Rainfall produced a positive bivariate relationship with AADT but a negative conditional regression coefficient. This apparent reversal is statistically informative rather than contradictory. The bivariate coefficient measures unadjusted association, whereas the regression coefficient measures the relationship remaining after controlling for GDP, population density, pothole volume, year and state effects. The strong rainfall–pothole correlation further suggests that the environmental predictors share information. This finding supports the use of VIF diagnostics and regularized/nonlinear modelling.

The non-significant pothole coefficient in the pooled regression should not be interpreted as evidence that pavement condition is irrelevant to highway traffic. Pothole volume may be correlated with rainfall and traffic loading, and its direct coefficient can therefore be unstable in a multivariable model. Its role may also be nonlinear or corridor-specific. Machine-learning models provide a complementary way of testing whether pavement-condition information contributes to predictive accuracy even when a linear coefficient is not statistically distinguishable from zero.

5.4 REGRESSION VERSUS MACHINE LEARNING

The regression model and machine-learning models answer related but different questions. The HC3 regression provides interpretable conditional associations and formal statistical inference, while the machine-learning models focus on out-of-sample predictive performance. The high in-sample regression R^2 and the much lower holdout performance of several machine-learning algorithms demonstrate why explanatory fit should not be equated with forecasting reliability (Zhang et al. 2019).

Gradient Boosting provided the strongest overall holdout balance in the present dataset. Its sequential ensemble structure is capable of representing nonlinearities and interactions that may be difficult to specify in advance. The result is broadly consistent with the contemporary literature showing that ensemble methods

can perform strongly in traffic prediction. Nevertheless, the substantial deterioration in performance when 2020 was included demonstrates that even the best-performing model remains sensitive to distributional change.

5.5 COMPARISON WITH PREVIOUS RESEARCH

The findings are consistent with the wider movement toward machine-learning traffic prediction reported by Shaygan et al. (2022), Xing et al. (2023) and Yao et al. (2023), particularly in the recognition that nonlinear and multivariate relationships are important. The present results also agree with weather-sensitive traffic studies that identify rainfall as useful explanatory information, although the direction of the conditional annual relationship differs from short-term traffic settings. The importance of socioeconomic variables is consistent with recent AADT research incorporating GDP and population.

The study also extends the earlier South-Eastern Nigerian mathematical modelling foundation reported by Ugwuanyi et al. (2024) by incorporating a broader machine-learning comparison, chronological holdout validation and model-interpretation procedures. The contribution is therefore methodological as well as geographical: the framework links conventional statistical diagnostics with predictive machine learning within a locally relevant intercity-highway dataset.

6. IMPLICATIONS, LIMITATIONS AND FUTURE RESEARCH

6.1 PRINCIPAL CONTRIBUTIONS

The study provides a multi-highway AADT dataset framework covering 100 state-year observations across five important South-Eastern Nigerian corridors.

It integrates demographic, economic, environmental, pavement-condition and temporal information in one modelling framework.

It combines interpretable statistical regression with nonlinear machine-learning prediction rather than relying on either approach alone.

It demonstrates the importance of chronological holdout validation and shows the performance consequences of the unusual 2020 traffic conditions.

It provides feature-importance and sensitivity evidence to improve interpretation of predictive models.

It identifies substantial spatial/state heterogeneity, showing why state-aware modelling is important in the study area.

6.2 PRACTICAL IMPLICATIONS

Traffic agencies can use the framework as a supplementary tool where continuous AADT counts are unavailable, provided local validation is maintained.

Highway planning can incorporate demographic and economic indicators alongside traffic observations when assessing future demand.

Environmental and pavement-condition variables can be incorporated into monitoring systems because their information may improve prediction even when individual regression coefficients are unstable.

Model performance should be reported using both absolute and percentage error measures because highway traffic scales differ substantially among corridors.

Exceptional periods such as 2020 should be explicitly represented in validation and scenario analysis rather than treated as ordinary observations.

6.3 LIMITATIONS

The study is based on 100 state-year observations and five selected corridors; consequently, the findings should not automatically be generalised to all Nigerian highways. The predictors do not include every determinant of traffic demand, such as fuel prices, vehicle ownership, road capacity, land use, freight facilities, security conditions, major construction activity and route diversion. Some explanatory variables are strongly correlated, especially rainfall and pothole volume. The 2020 observations also represent a structural disturbance that is difficult to learn from a relatively short historical record. Finally, feature importance identifies predictive contribution rather than causality.

6.4 RECOMMENDATIONS FOR FURTHER RESEARCH

i. Extend the dataset to additional Nigerian highways and longer forecasting horizons.
ii. Integrate GIS-derived road-network attributes, land use, accessibility, road geometry and spatial autocorrelation.
iii. Use continuous counters, probe traces and sensor data to support seasonal and short-term traffic models.
iv. Evaluate hierarchical, spatial, spatiotemporal, Bayesian and graph-based models when sample size permits.
v. Develop explicit shock indicators for pandemics, fuel shortages, flooding, major construction and policy changes.
vi. Conduct independent external validation on additional corridors before operational deployment.

7. CONCLUSIONS

This study developed and evaluated a mathematical and machine-learning framework for modelling AADT on five selected South-Eastern Nigerian highways using demographic, economic, environmental and temporal information. The empirical evidence demonstrates substantial spatial and temporal heterogeneity in highway traffic demand. Enugu recorded the highest average AADT, Ebonyi the lowest, and all five states experienced a marked reduction between 2019 and 2020.

Passenger population density showed the strongest simple association with AADT, while GDP and rainfall also provided substantial explanatory information. The pooled regression model explained a large proportion of in-sample variation, but heteroscedasticity, non-normal residuals and environmental multicollinearity limited the appropriateness of uncorrected classical inference. HC3 robust standard errors therefore provided the basis for reported coefficient inference.

The chronological machine-learning evaluation demonstrated that Gradient Boosting produced the strongest overall holdout performance among the compared models, with $R^2 = 0.5338$, RMSE = 396,916.31 PCU/day, MAE = 249,437.28 PCU/day and MAPE = 28.45% over 2017–2020. Its performance advantage over the traditional linear trend was measurable but not sufficient to eliminate prediction uncertainty, particularly under the exceptional 2020 conditions. The results therefore support the use of machine learning as a supplementary, locally validated decision-support approach rather than a replacement for systematic traffic counting.

In general, the study contributes a context-specific framework for AADT estimation in South-Eastern Nigeria that integrates statistical inference, nonlinear prediction, chronological validation and model interpretation. The framework provides a basis for further research and for cautious application in highway planning, maintenance prioritization and traffic-management systems.

Declaration of competing interest

The author declares no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

AI Declaration.

AI assisted tool was used to refine and polish the grammatical arrangement in some paragraphs of the research.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration

1. Ethics: Not Applicable
2. Consent to Participate: Not Applicable
3. Consent to Publish: Not Applicable

REFERENCES

AASHTO. (2018). A policy on geometric design of highways and streets (7th ed.). American Association of State Highway and Transportation Officials.

Adedeji, W. O., Oyelami, S., Oyewo, A. T., Abioye, A. G., Ojerinde, B. J., & Adekoya, A. O. (2025). Design and development of machine learning-based traffic predictions. OAUSTECH Journal of Engineering and Intelligent Technology, 1(1).

Afandizadeh, S. (2024). Deep learning algorithms for traffic forecasting: A comprehensive review and comparison with classical ones. Journal of Advanced Transportation, 2024, Article 9981657.

Alourani, A., Ashfaq, F., Jhanjhi, N. Z., & Ali Khan, N. (2023). BiLSTM- and GNN-based spatiotemporal traffic flow forecasting with correlated weather data. Journal of Advanced Transportation, 2023, Article 8962283.

Federal Highway Administration. (2022). Traffic monitoring guide. U.S. Department of Transportation.

Imoh, U. U., & Movahedi Rad, M. (2025). Analysis and prediction of traffic conditions using machine learning models on Ikorodu Road in Lagos State, Nigeria. *Infrastructures, 10*(5), 122

Nigam, & Srivastava. (2023). Weather-sensitive traffic prediction using recurrent neural-network approaches. [Bibliographic details to be verified against the final thesis/reference database].

Shaygan, M., Meese, C., Li, W., Zhao, X., & Nejad, M. (2022). Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transportation Research Part C: Emerging Technologies, 145*, 103921.].

Ugwuanyi, H. K., Okafor, F. O., Uba, M. I., Chima, M. C., & Akwarandu, L. I. (2024). Mathematical modeling of the annual average daily traffic on South Eastern Nigerian highways using environmental and demographic factors. International Journal of Advances in Engineering and Management, 6(11), 128–144.

van Strien, M. J., & Grêt-Regamey. (2024). A global time series of traffic volumes on extra-urban roads. Scientific Data.

Wengsong, Z., Romghan, Y., Yao, X., Liu, X., Wang, R., & Wang, L. (2023). Traffic flow prediction under multiple adverse weather based on self-attention mechanism and deep learning models. Physica A: Statistical Mechanics and its Applications, 625, 128988. https://doi.org/10.1016/j.physa.2023.128988

World Meteorological Organization. (2024). Guide to instruments and methods of observation (WMO-No. 8), Volume I: Measurement of meteorological variables. WMO.

Xing, Z., Huang, M., & Peng, D. (2023). Overview of machine learning-based traffic flow prediction. Digital Transportation and Safety, 2(3), 164–175.

Yao, J.-F., He, R., Shi, T.-T., Wang, P., & Zhao, X.-M. (2023). Review on machine learning-based traffic flow prediction methods. Journal of Traffic and Transportation Engineering, 23(3), 44–67.

Zhang, Y., Haghani, A., & Hu, S. (2019). Machine learning models for AADT estimation on low-volume roads. Journal of Transportation Engineering, Part A: Systems, 145(6), 04019021.

Zeng, W., Sun, X., & Xing, H. (2024). Establishing resilience-targeted prediction models of rainfall for transportation infrastructures for three demonstration regions in China. Scientific Reports, 14, 15469.

Zou, X., Sun, L., Lan, T., Fan, C., Liu, S., Zhao, H., & Qiu, J. (2024). The effects of weather factors on road traffic casualties: Analysis on provincial panel data of China from 2006 to 2021. Heliyon, 10(17), e36788.