Journal of Spatial Information Sciences

... JSIS



MACHINE LEARNING-BASED PREDICTION AND SHAP INTERPRETATION OF NO2 CONCENTRATIONS IN KADUNA METROPOLIS, NIGERIA

JACOB REUBEN JOBIEN, BELLO ABDULFATAI ABIODUN, EKELE JOHN IFENE



JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 243–258, 2026 DOI: https://doi.org/10.5281/zenodo.20763914
PUBLISHED 19-06-2026

www.journals.unizik.edu.ng/jsis

MACHINE LEARNING-BASED PREDICTION AND SHAP INTERPRETATION OF NO_2 CONCENTRATIONS IN KADUNA METROPOLIS, NIGERIA

[1]Jacob Reuben Jobien, [2]Bello Abdulfatai Abiodun, [3]Ekele John Ifene

[1]Department of Surveying and Geoinformatics, Confluence University of Science and Technology, Osara, Kogi State, Nigeria
[2]Department of Geography and Environmental Science, Faculty of Environmental Sciences, University of Calabar, Nigeria
[3]School of Postgraduate Studies, Nasarawa State University, Keffi, Nigeria
[1]Corresponding Author's Email: jobienjacob@gmail.com, 08067186726

DOI: https://doi.org/10.5281/zenodo.20763914

Abstract

Nitrogen dioxide (NO_2) pollution remains a major environmental concern in rapidly urbanizing cities, particularly where ground-based air quality monitoring is limited. This study modelled and interpreted the spatial distribution of NO_2 concentrations across Kaduna Metropolis, Nigeria, using a Random Forest (RF) machine learning approach. Predictor variables, including proximity to industrial sites, proximity to infrastructure, proximity to roads, NDWI, NDVI, land surface temperature, NDBAI, and NDBI, were processed and partitioned into 70% training and 30% independent testing subsets. The RF model showed strong predictive performance, with R^2 values of 0.964 for training, 0.773 for spatial cross-validation, and 0.817 for independent testing, indicating good model robustness and generalization. SHAP analysis showed that proximity to industrial sites, infrastructure, and roads were the main drivers of NO_2 concentrations in Kaduna Metropolis, highlighting the strong influence of anthropogenic activities, while environmental variables such as NDWI, NDVI, LST, NDBAI, and NDBI had comparatively smaller effects. SHAP dependence plots showed that areas closer to industrial sites and roads had higher NO_2 contributions, while NDVI generally reduced predicted NO_2. LST showed a positive influence. The study offers useful evidence for air quality monitoring, industrial emission control, transport planning, and sustainable urban environmental management in Kaduna Metropolis.

Keywords: Nitrogen dioxide (NO_2), Random Forest, SHAP analysis, Environmental and Anthropogenic Factors, Kaduna Metropolis

1.0 INTRODUCTION

Nitrogen dioxide (NO_2) is a key atmospheric pollutant in urban environments and is widely recognized for its adverse effects on human health, atmospheric chemistry, and climate processes [1], [2]. It is primarily emitted from combustion-related activities, including vehicular traffic, industrial operations, and energy generation, all of which are prevalent in rapidly expanding cities in developing countries [3], [4]. In sub-Saharan Africa, and particularly in Nigerian cities, rapid urbanisation, population growth, and increasing anthropogenic activities have intensified concerns regarding air quality, yet monitoring networks remain sparse and insufficient to capture the spatial variability of pollutants such as NO_2 [5], [6]. Kaduna metropolis, a major urban and industrial centre in northern Nigeria, exemplifies these challenges due to its

complex mix of residential, commercial, industrial, and transportation land uses (Kafi *et al*., 2024; Musa&Abubakar, 2024).

The spatial distribution of NO_2 in urban areas is shaped by both environmental and anthropogenic factors [4], [9]. Environmental variables, including land surface temperature (LST) [10], [11] and satellite-derived spectral indices such as NDBI, NDVI, NDWI, and NDBaI, reflect urban heat, land cover, vegetation, moisture, and bare surfaces, all of which influence pollutant dispersion and removal [12], [13], [14], [15], [16]. Anthropogenic factors, particularly proximity to roads, industrial areas, and infrastructure corridors, represent concentrated emission sources and are strongly linked to elevated NO_2 levels [17], [18].

Recent advances in remote sensing and machine-learning techniques have enabled more robust modelling of air pollutants in data-scarce regions [19], [20]. Among these techniques, the Random Forest (RF) algorithm has gained prominence due to its ability to model complex, non-linear relationships, manage multicollinearity among predictors, and provide reliable prediction accuracy without strict parametric assumptions (Keith, 2020; Salman *et al*., 2024). RF has been widely applied in environmental studies [23], [24], [25], [26], [27], including air-quality prediction, where it integrates diverse datasets such as satellite-derived variables and spatial predictors to estimate pollutant concentrations.

Despite these advancements, there remains a need for location-specific studies that integrate both environmental indicators and proximity-based variables to better understand urban air-pollution patterns in Nigerian cities. In Kaduna metropolis, limited studies have explored the combined influence of land surface characteristics and spatial proximity to emission sources on NO_2 distribution using machine-learning approaches. Therefore, this study employs a Random Forest model to predict NO_2 concentrations using LST, NDBI, NDVI, NDWI, NDBaI, proximity to industrial areas, proximity to roads, and proximity to infrastructure as predictor variables. By doing so, it seeks to provide a comprehensive assessment of the spatial determinants of NO_2 in Kaduna and contribute to improved air-quality monitoring and urban environmental management strategies.

2.0 MATERIALS AND METHODS

2.1 Study Area

The study area is bounded by Latitudes 10°04′N to 10°13′N and Longitudes 7°10′E to 7°45′E, with a total area of about 1727 km^2. Kaduna Metropolis is the administrative and commercial hub of Kaduna State. It consists of Kaduna North, Kaduna South and some parts of Chikun and Igabi Local Government Areas (LGAs) (See Figure 1). The region receives annual rainfall between 1500mm and 2000mm, with the wet season lasting six to seven months and averaging 1400mm [28]. According to [29], land surface

temperature in Kaduna Metropolis ranged from 14.44°C to 41.19°C. The vegetation is predominantly Northern Guinea Savanna, characterized by grasses and scattered trees [28]. The study area is well-connected by major highways and has essential amenities like healthcare centres, water, and electricity [30].

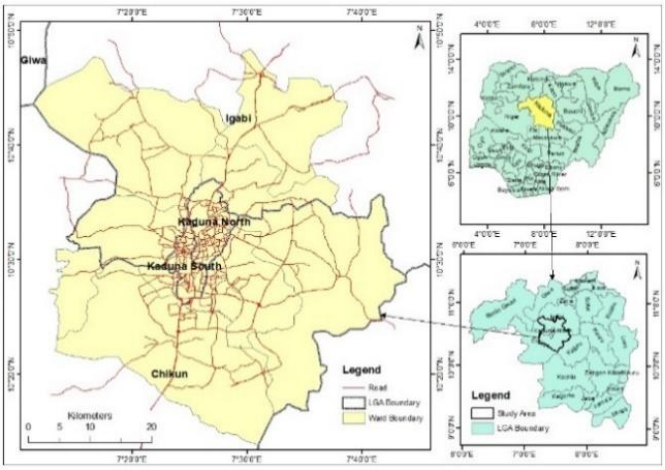


Figure 1: The Study Area

2.2 Types and Sources of Data

The types and sources of data used for this study are presented in Table 1.

Table 1: Types and Sources of Data

SN	Name	Resolution	Source	Processing Platform	Purpose
1	NO_2 (per mol/m²)	3.5x7km	S5P/TROPOMI (2025)	Google Earth Engine (GEE)	Response variable
2	LST, NDVI, NDBI, NDWI and NDBaI	30m	Landsat 8 (2025)	GEE	Predictors
3	Road, infrastructure and industrial sites shapefile	Area coverage	Geofabrik (2026)	ArcGIS Pro	Predictors
4	Admin. boundary shapefile	Area coverage	Geofabrik (2026)	ArcGIS Pro	For clipping and mapping;

Nitrogen dioxide (NO_2) column concentrations for 01 January to 31 December 2025 were obtained from the Sentinel-5P (S5P) TROPOspheric Monitoring Instrument (TROPOMI) satellite. NO_2 served as the response variable, with environmental and anthropogenic predictors represented as raster layers, including Land Surface Temperature (LST), Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Water Index (NDWI), Normalised Difference

Bareness Index (NDBaI), and proximities to roads, infrastructure, and industrial areas. The proximity layers were generated in ArcGIS using the Euclidean distance tool. The study did not use 2026 NO_2 data because the analysis was designed around the 2025 observation period. Using 2026 NO_2 data would have introduced temporal inconsistency with the Landsat-derived environmental variables, particularly LST, NDVI, NDBI, NDWI, and NDBaI, which were prepared for the 2025 analysis year.

2.3 Estimation of Predictors

2.3.1 Estimation of LST

Land Surface Temperature (LST) data were obtained from Climate Engine, which provides pre-processed satellite-derived temperature products. The LST values were derived from thermal infrared bands using radiative transfer equations incorporating spectral radiance, brightness temperature, and surface emissivity corrections. LST estimation was done using the following formula (Equation1 1-7) [31]:

$$LST = \left[\frac{\tau}{1 + \omega \left(\frac{\tau}{\rho} \right) x \ln(\varepsilon)} \right] \quad (1)$$

where,

τ is at-sensor brightness temperature;

ω is the wavelength of emitted radiance (Landsat 8 Band 10 is approximately 11.5 μm)

ρ = h × c/s (1.438×10^{-2} m.K);

h being the Plank's constant (6.626×10^{-34} J·s);

s is the Boltzmann Constant (1.38×10^{-23} J/K);

c is the velocity of light (2.988×10^{8} m/s);

ε is the land surface emissivity; and

ρ = 14380

The land surface emissivity (ε) was calculated using the following Equation:

$$\varepsilon = 0.004 * P_v + 0.986 \quad (2)$$

The vegetation proportion (P_v) was obtained from the following Equation:

$$P_v = \left[\frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} \right]^2 \quad (3)$$

where,

$NDVI_{min}$ and $NDVI_{max}$ are the minimal and the maximal values of the NDVI (calculated according to the following Equation:

$$NDVI = \left[\frac{NIR - R}{NIR + R}\right] \quad (4)$$

where,

NIR and R are the infrared and red bands of Landsat 8.

The temperature value at the sensor (brightness) was extracted using the following Equation:

$$T_b = \left[\frac{K_2}{\ln\left(\frac{K_1}{L} + 1\right)}\right] \quad (5)$$

where,

K_1 and K_2 are the thermal conversion constants provided in the Landsat metadata. Radiance for Landsat 8 TIR band will be obtained from Equation:

$$L = MLxDN + AL \quad (6)$$

where,

L is the top-of-atmosphere radiance, ML is the radiance multiplicative scaling factor, and AL is the radiance additive scaling factor (these are found in the metadata of the Landsat image).

Convert the temperature from Kelvin to Celsius by subtracting 273.15 from the result

$$LST_C = LST_K - 273.15 \quad (7)$$

2.3.2 Computation of Normalized Difference Water Index (NDWI)

The NDWI of the study area was calculated using the Google Earth Engine. It is represented mathematically using the formula as follows [32]:

$$NDWI = \frac{G - NIR}{G + NIR} \quad (8)$$

where, G and NIR are the green and infrared bands of Landsat 8, respectively

2.3.3 Computation of Normalized Difference Builtup Index (NDBI)

The NDBI of the study area was calculated using the Google Earth Engine. It is represented mathematically using the formula as follows [32]:

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad (9)$$

where, SWIR and NIR are the short wave and infrared bands of Landsat 8.

JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 243–258, 2026 DOI: https://doi.org/10.5281/zenodo.20763914
PUBLISHED 19-06-2026
www.journals.unizik.edu.ng/jsis

2.3.4 Computation of Normalised Difference Bareness Index (NDBaI)

The NDBaI of the study area was calculated using the Google Earth Engine. It is represented mathematically using the formula as follows [31]:

$$NDBaI = \frac{SWIR - TIR}{SWIR + TIR} \quad (10)$$

2.4 Mapping of Variables

The response variable (NO_2) and the predictors were mapped in ArcGIS as shown in Figure 2

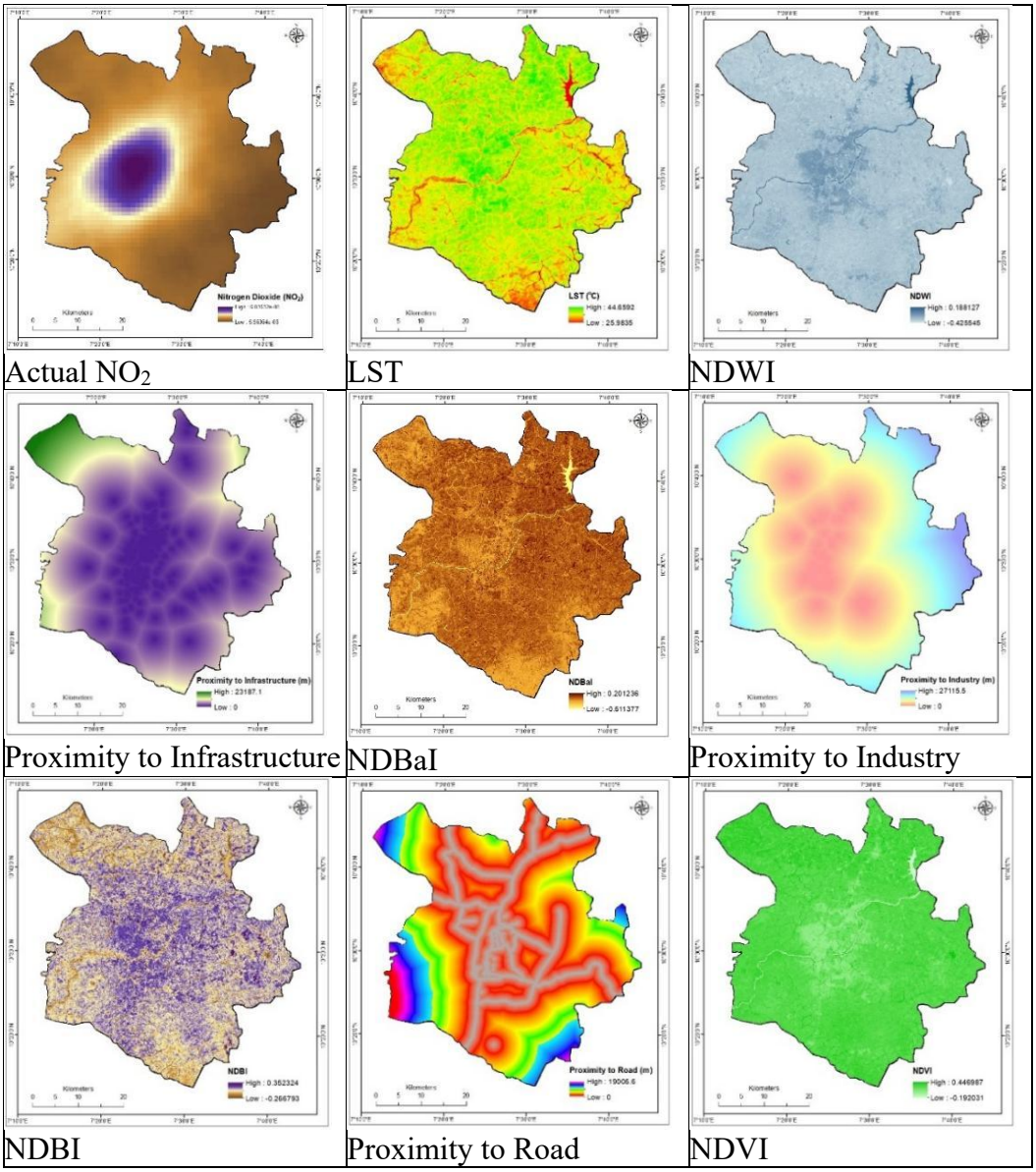


Figure 2: Maps of Dependent and Independent Variables

JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 243–258, 2026 DOI: https://doi.org/10.5281/zenodo.20763914
PUBLISHED 19-06-2026
www.journals.unizik.edu.ng/jsis

2.5 Data Preprocessing and Sampling

All raster datasets were projected to WGS 1984 UTM Zone 32N, resampled to a uniform 100 m resolution using bilinear interpolation, and clipped to the Kaduna Metropolis boundary in ArcGIS to ensure spatial consistency. Only raster cells with complete data were retained, converted to tabular format, and cleaned of missing values in R. A random sample of up to 10,000 points was selected using stratified sampling based on NO_2 quantiles, and the dataset was split into 70% training and 30% testing subsets to maintain the distribution of NO_2 and improve model generalization.

2.6 Spatial Cross-Validation, Random Forest Modelling and Hyperparameter Tuning

Spatial cross-validation was applied using hexagonal blocks of 1000m to partition the training data into five spatially independent folds. For each fold, the model was trained on distinct training blocks and evaluated on the corresponding validation blocks, providing a realistic assessment of predictive performance while accounting for spatial autocorrelation. A Random Forest (RF) regression model was employed using the ranger implementation due to its computational efficiency and ability to handle high-dimensional data. RF is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to improve accuracy and reduce overfitting. Hyperparameter tuning was conducted using a grid search approach. The following parameters were optimised: number of variables tried at each split (mtry): 2, 3, 4; minimum node size: 3, 5, 8; sample fraction: 0.75, 0.90; number of trees: 400, 700.

2.7 Final Model Training and Evaluation

The optimal Random Forest model was trained on the full training dataset using the selected hyperparameters. Model performance was evaluated using three datasets: training dataset, spatial cross-validation dataset and independent test dataset. Performance metrics included Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). Scatter plots of observed versus predicted values were generated to visually assess model accuracy and bias. RMSE, MAE, and R^2 were calculated using the following formula [33]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (2)$$

$$R^2 = \left(\frac{Cov\left(y, \hat{y}\right)}{\sigma_y \sigma_{\hat{y}}}\right)^2 \quad (3)$$

where,

n is the total number of observations/data points, i is the index of each observation (from 1 to n), y_i is the actual value of NO_2 for observation i, $\hat{y}_i$ is the predicted value of NO_2 for observation i, $\left(y_i - \hat{y}_i\right)$ is the

prediction error (residual) for observation i, $|y_i - \hat{y}_i|$ is the absolute prediction error, σ_y is the standard deviation of the actual values of NO_2, $\sigma_{\hat{y}}$ is the standard deviation of the predicted values of NO_2.

2.8 SHAP-Based Model Interpretation

To interpret the contribution of predictor variables, SHapley Additive exPlanations (SHAP) were computed using the fastshap and shapviz packages. A subset of training data was used for SHAP analysis to reduce computational cost. SHAP values quantify the marginal contribution of each predictor to model predictions. Global feature importance was derived from the mean absolute SHAP values, while dependence plots were generated to examine the relationship between individual predictors and NO_2 predictions. The SHAP value for feature *j* is defined as [34]:

$$\phi_j(v) = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} \left[v(S \cup \{j\}) - v(S) \right], \quad (4)$$

where, N is the set of all available features.

S is a subset of features excluding feature *j*.

v(S) is the model trained on feature subset S.

v(S U {*j*}) - v(S) isolates the marginal contribution that feature *j* adds to that specific sub-coalition.

3.0 Results and Discussion

3.1 Model Performance Metrics

The predicted NO_2 is presented in Figure 3. The performance metrics of the Random Forest (RF) model presented in Figure 4 demonstrated strong performance in predicting NO_2 concentrations in Kaduna Metropolis, explaining 96.4% of the variability during training ($R^2 = 0.964$) with minimal prediction errors. Predictions closely matched observed values, capturing complex nonlinear relationships between NO_2 and environmental predictors. While performance declined in spatial cross-validation ($R^2 = 0.773$) and independent testing ($R^2 = 0.817$), the model still demonstrated good generalization to unseen locations. Higher uncertainty at elevated NO_2 levels reflects localized emission sources such as traffic and industrial activities. These results confirm the spatial heterogeneity of NO_2 and support the suitability of RF for urban air quality estimation, The findings are consistent with [27] that reported strong RF performance in $PM_{2.5}$ prediction with an R^2 of 0.73. highlighting its potential for environmental management and urban planning in Nigerian cities.

JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 243–258, 2026 DOI: https://doi.org/10.5281/zenodo.20763914
PUBLISHED 19-06-2026
www.journals.unizik.edu.ng/jsis

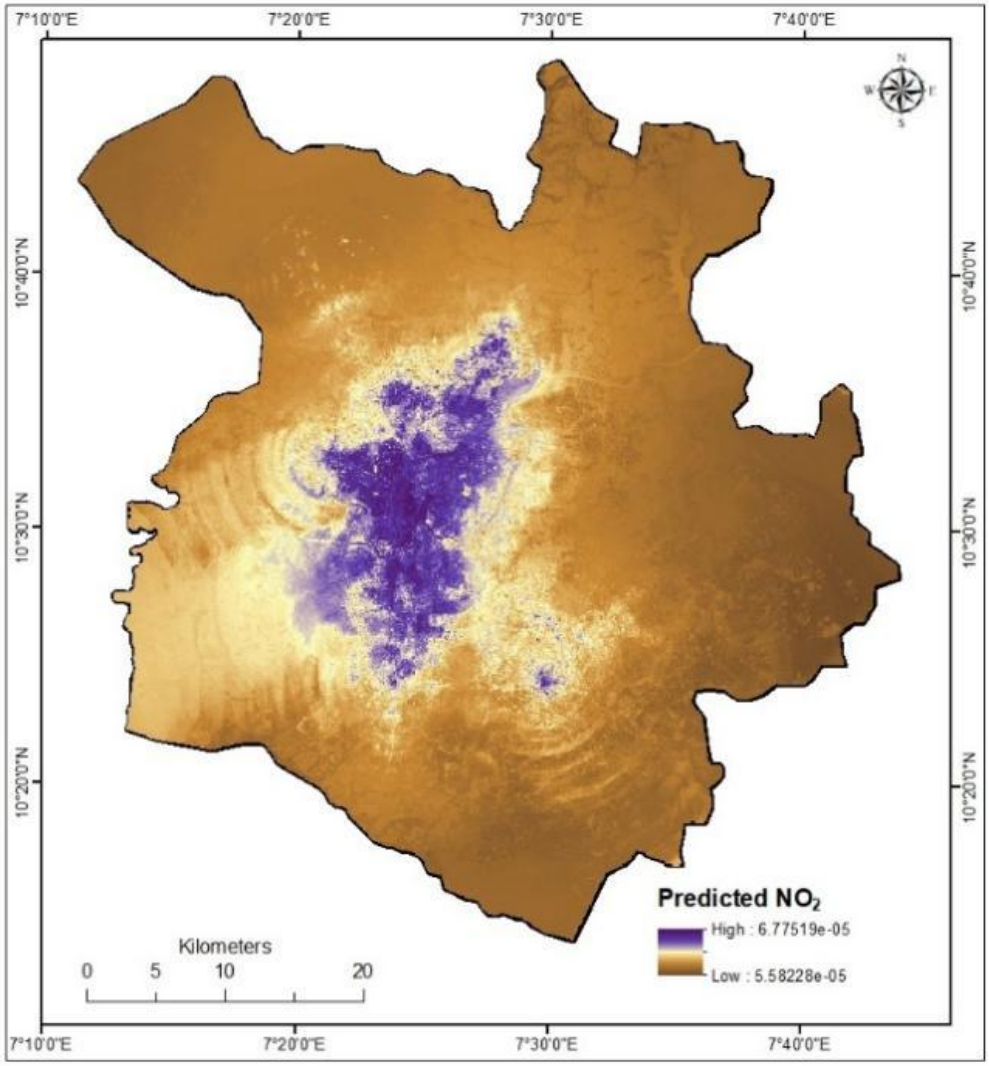


Figure 3: Predicted NO2

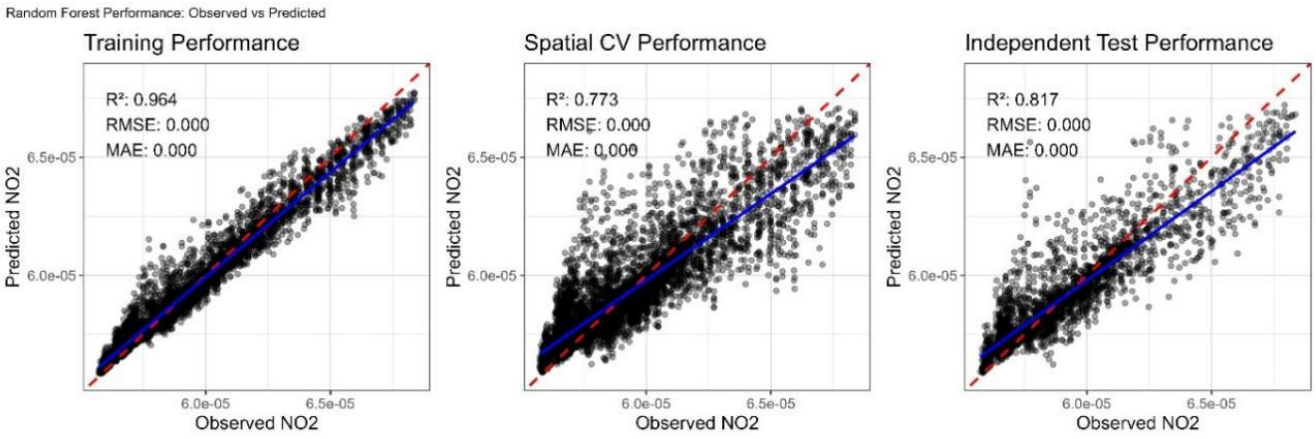


Figure 4: Scatter Plots of Performance Metrics

3.2 Results of Global SHAP Importance

The global SHAP importance for each factor was calculated in R and the results are presented in Table 2.

Table 2: Global SHAP Importance

Feature	MeanAbsSHAP	Percentage
Proximity to industrial site	0.000001288	47.97
Proximity to infrastructure	0.000000397	14.79
Proximity to road	0.000000388	14.45
NDWI	0.000000225	8.38
NDVI	0.000000106	3.95
LST	0.000000105	3.91
NDBaI	0.000000094	3.50
NDBI	0.000000082	3.05
Total	0.000002685	100

The SHAP analysis in Table 2 indicates that proximity to industrial sites is the most influential predictor of NO_2 concentrations in Kaduna Metropolis, contributing nearly 48% to the model's overall impact and highlighting the central role of industrial emissions from manufacturing, fuel combustion, and heavy machinery. Proximity to infrastructure (≈15%) and roads (≈14%) were the next most influential factors, highlighting the significant impact of transportation corridors and traffic on urban air quality, consistent with [4]. Among environmental variables, NDWI (≈8%) showed a noticeable effect, suggesting that surface moisture and water bodies can influence pollutant dispersion. NDVI (≈4%) indicated that vegetation provides some mitigation of NO_2 through absorption and improved air circulation, although its effect is comparatively small relative to anthropogenic sources. Land Surface Temperature (LST, ≈4%) contributed modestly, reflecting interactions between urban heat and pollutant accumulation, in line with findings by [10]. Built-up indices (NDBaI ≈3.5%, NDBI ≈3%) had weaker direct effects but remain relevant for characterizing urban expansion and its indirect influence on pollution through increased human activity and altered microclimates.

3.3 Results of SHAP Dependence Plots

The SHAP dependence plots presented in Figure 5 reveal that proximity to industrial sites is the strongest contributor, with a mean absolute SHAP value of 1.288×10^{-6}, displaying a clear nonlinear negative trend in which predicted NO_2 levels decrease as distance from industrial areas increases. Similarly, proximity to roads with a SHAP value of 3.88×10^{-7} exhibited a positive association with NO_2, reflecting the impact of traffic-related emissions, while the spread of SHAP values at higher road influence suggests interactions with nearby industrial activity. Proximity to infrastructure with a SHAP value of 3.97×10^{-7} also contributed significantly, confirming that anthropogenic sources, including industrial operations and transportation, are dominant drivers of NO_2 pollution in the metropolis. These patterns are consistent with

JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 243–258, 2026 DOI: https://doi.org/10.5281/zenodo.20763914
PUBLISHED 19-06-2026
www.journals.unizik.edu.ng/jsis

[4], who reported strong positive correlations between NO_2 concentrations, industrial density, and road density in urban areas.

Among environmental variables, NDWI with a SHAP value of 2.25×10^{-7} showed a positive response, indicating that wetter surfaces or areas near water bodies may facilitate pollutant accumulation or influence dispersion. In contrast, NDVI with a SHAP value of 1.06×10^{-7} exhibited a negative effect, highlighting vegetation's mitigating role through pollutant absorption and enhanced urban ventilation. Land Surface Temperature with a SHAP value of 1.05×10^{-7} displayed a positive nonlinear trend, particularly beyond 37 to 38 degrees Celsius, suggesting that warmer urban surfaces amplify NO_2 accumulation via urban heat island effects. These findings align with [10], who observed interactions between urban temperature dynamics and air pollution in machine learning-based predictions. Built-up indices NDBaI with a SHAP value of 9.4×10^{-8} and NDBI with a SHAP value of 8.2×10^{-8} showed weaker, more complex nonlinear effects. NDBI largely had a negative relationship with SHAP values, indicating that dense built-up areas only increase NO_2 in combination with intense anthropogenic activities.

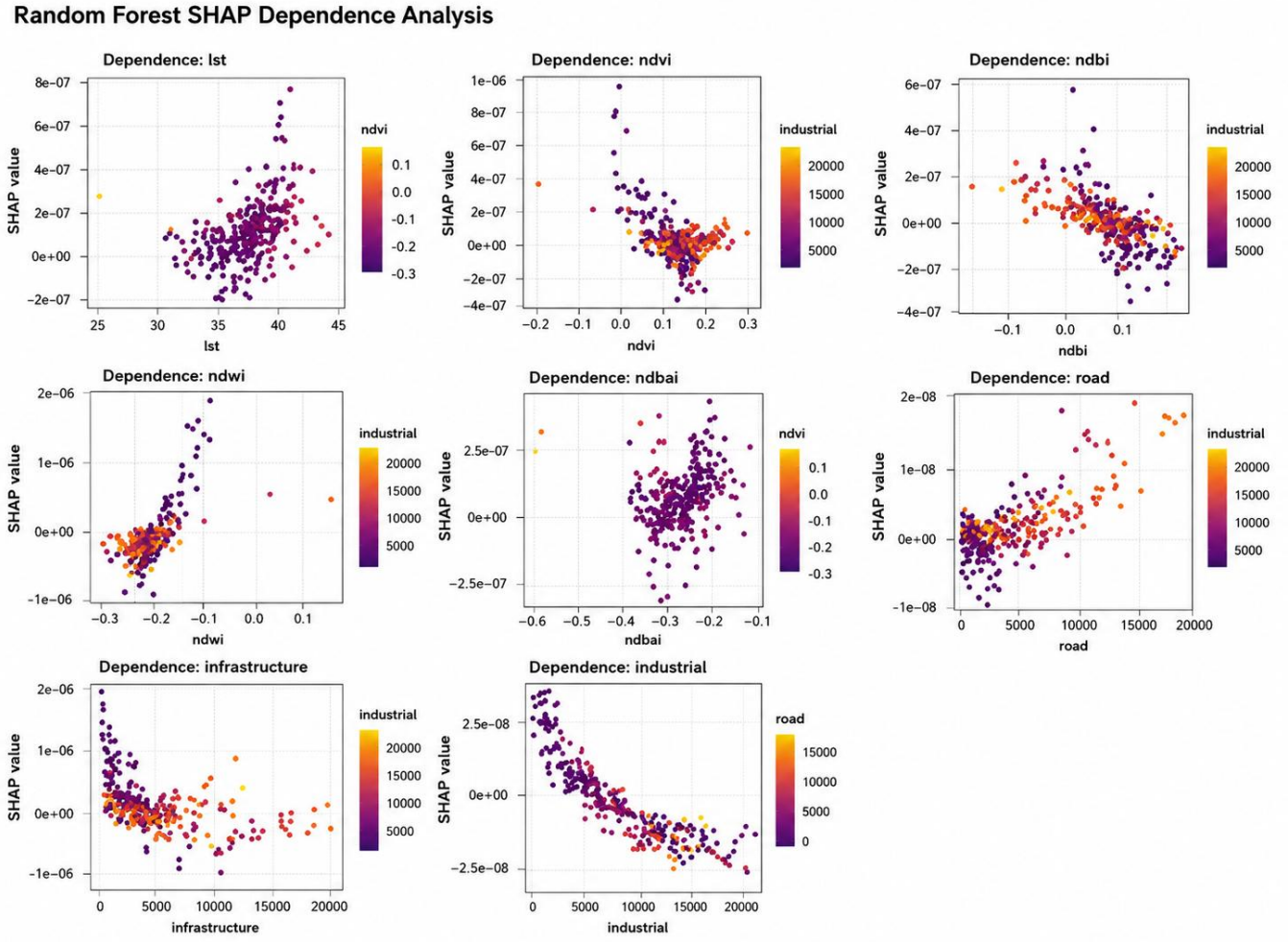


Figure 5: SHAP Dependence Plots

Conclusion

This study demonstrated that the Random Forest model can effectively predict NO_2 concentrations in Kaduna Metropolis, capturing complex nonlinear relationships between air pollution and environmental predictors. Proximity to industrial sites, infrastructure, and major roads emerged as the primary drivers of NO_2, underscoring the dominant role of human activities in shaping urban air quality. Environmental factors such as vegetation, water, land surface temperature, and built-up indices also influenced NO_2 levels, though to a lesser extent. These findings highlight the need for strengthened industrial emission controls, improved transportation management, and the integration of green infrastructure in urban planning. Implementing such measures can support more effective mitigation of air pollution and enhance public health outcomes in Kaduna Metropolis. Furthermore, future studies should integrate seasonal meteorological variables, higher-resolution emission datasets, and ground-based monitoring observations to improve model performance and support more effective air quality management strategies in Kaduna Metropolis, Nigeria.

References

[1] De Vries W., Posch M., Simpson D., De Leeuw F., Van Grinsven H., Schulte-Uebbing L., Sutton M., & Ros G. (2024). *Trends and geographic variation in adverse impacts of nitrogen use in Europe on human health, climate, and ecosystems: A review*. Earth-Science Reviews. https://doi.org/10.1016/j.earscirev.2024.104789

[2] Bashir M. R., Alomran M., Gill A. N., Khan Q. A. T., Jameel S., Rafiq M. & Khan F. (2024). *Rising Nitrogen Dioxide (No2): A Growing Threat to Environment and Plant Health*. Journal of Agriculture and Biology. https://doi.org/10.55627/agribiol.002.01.01023

[3] Agudelo-Castañeda D., Arellana J., Morgado-Gamero W., De Paoli F. & Portz L. C. (2023). *Linking of built environment inequalities with air quality: A case study*. Transportation Research Part D: Transport and Environment. https://doi.org/10.1016/j.trd.2023.103668

[4] Jubaer A., Hossain R., Ahmed A. & Hossain M. (2025). *Factors influencing spatiotemporal variability of NO_2 concentration in urban area: a GIS and remote sensing–based approach*. Environmental Monitoring and Assessment, 197. https://doi.org/10.1007/s10661-024-13531-z

[5] Akanji A. R., Francis M. O. & Akintola, A. (2024). *Air Quality Trends and Pollution Analysis in Nigerian Cities Using Time Series Methods*. International Journal of Advanced Statistics and Probability. https://doi.org/10.14419/w5rj1f64

[6] Omokpariola D. O., Nduka J. N. & Omokpariola P. L. (2024). *Short-term trends of air quality and pollutant concentrations in Nigeria from 2018–2022 using tropospheric sentinel-5P and 3A/B satellite data*. Discover Applied Sciences, 6(4), 182. https://doi.org/10.1007/s42452-024-05856-8

[7] Musa K. & Abubakar M. L. (2024). *Monitoring urban growth and landscape fragmentation in Kaduna, Nigeria, using remote sensing approach*. Journal of Degraded and Mining Lands Management, 12(1), 6757–6769. https://doi.org/10.15243/jdmlm.2024.121.6757

[8] Kafi K. M., Ibrahim S., Aliyu F. U., Usman M. A. & Olugbodi K. H. (2024). *Impact of LULC spatial dynamics on incompatible mixed land use in Kaduna: A remote sensing and GIS risk analysis*. Eco Cities, 5(2), 2818. https://doi.org/10.54517/ec.v5i2.2818

[9] Amritha S., Varikoden H., Patel V., Kuttippurath J. & Gopikrishnan, G. (2024). *Global, regional and city scale changes in atmospheric NO_2 with environmental laws and policies*. Sustainable Cities and Society. https://doi.org/10.1016/j.scs.2024.105617

[10] Suthar G., Kaul, N., Khandelwal S. & Singh, S. (2024). *Predicting land surface temperature and examining its relationship with air pollution and urban parameters in Bengaluru: A machine learning approach*. Urban Climate, 53, 101830. https://doi.org/10.1016/j.uclim.2024.101830

[11] Kazemi A., Cirella G., Hedayatiaghmashhadi, A. & Gili M. (2025). *Temporal-Spatial Distribution of Surface Urban Heat Island and Urban Pollution Island in an Industrial City: Seasonal Analysis*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 7100–7116. https://doi.org/10.1109/jstars.2025.3541406

[12] Pandey A., Mondal A., Guha S., Upadhyay P. K. & Singh D. (2023*). A Long-Term Analysis of the Dependency of Land Surface Temperature on Land Surface Indexes*. Papers in Applied Geography, 9(3), 279–294. https://doi.org/10.1080/23754931.2023.2187314

[13] Lin C.-Y., Tsai S., Wang C.-Y. & Lin C.-Y. (2021). *Delineation of potential hot spots of aeolian dust using normalized difference water index*. Theoretical and Applied Climatology, 145, 63–76. https://doi.org/10.1007/s00704-021-03611-2

[14] Martinez A., de la I. & Labib S. M. (2022). *Demystifying Normalized Difference Vegetation Index (NDVI) for Greenness Exposure Assessments and Policy Interventions in Urban Greening*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4207665

[15] Mondal S. & Gavsker K. K. (2024). *Investigating the urban eco-environmental quality utilizing remote sensing based approach: evidence from an industrial city of Eastern India*. Discover Applied Sciences, 6(12), 666. https://doi.org/10.1007/s42452-024-06345-8

[16] Goswami A., Sen S. & Chakraborty P. K. (2022). *Examination of Vegetation Health and Its Relation with Normalized Difference Built-Up Index. Advanced Sensing in Image Processing and IoT*. https://doi.org/10.1201/9781003221333-14

[17] Restrepo, C. (2021). *Nitrogen Dioxide, Greenhouse Gas Emissions and Transportation in Urban Areas: Lessons from the Covid-19 Pandemic*. Volume 9. https://doi.org/10.3389/fenvs.2021.689985

[18] Sillapapiromsuk S., Koontoop G. & Bootdee S. (2022). *Health Risk Assessment of Ambient Nitrogen Dioxide Concentrations in Urban and Industrial Area in Rayong Province, Thailand*. Trends in Sciences. https://doi.org/10.48048/tis.2022.4476

[19] Tang D., Zhan Y. & Yang F. (2024). *A review of machine learning for modelling air quality: Overlooked but important issues*. Atmospheric Research. https://doi.org/10.1016/j.atmosres.2024.107261

[20] Anggraini T., Irie H., Sakti A. & Wikantika K. (2023). *Novel Global Air Quality Index Development through Integration of Remote Sensing Data, Ground-Based Stations, and Machine Learning Techniques*. Environmental Advances. https://doi.org/10.1016/j.envadv.2023.100456

[21] Keith, M. (2021). *Random Forest*. Machine Learning with Regression in Python. https://doi.org/10.1007/978-1-4419-9863-7_612

[22] Salman H. A., Kalakech A. & Steiti A. (2024). *Random Forest Algorithm Overview*. Babylonian Journal of Machine Learning, 2024, 69–79. https://doi.org/10.58496/BJML/2024/007

[23] Wu H., Yang T., Li, H. & Zhou Z. (2023). *Air quality prediction model based on mRMR–RF feature selection and ISSA–LSTM*. Scientific Reports, 13. https://doi.org/10.1038/s41598-023-39838-4

[24] Huo S., Lu T., Deng P. & Wei H. (2024). *Based on RF-SARMIR Algorithm Predict Air Quality*. 2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), 393–398. https://doi.org/10.1109/raiic61787.2024.10671172

[25] Zhang L., Chen Y., Dong H., Wu D., Chen S., Li X., Liang B. & Yang Q. (2024). *Improving the construction and prediction strategy of the Air Quality Health Index (AQHI) using machine learning: A case study in Guangzhou, China*. Ecotoxicology and Environmental Safety, 287, 117287. https://doi.org/10.1016/j.ecoenv.2024.117287

[26] Musleh F., Taha R. & Musleh A. R. (2025). *Predicting Air Quality Using Machine Learning: A Comparative Analysis*. 2024 International Conference on IT Innovation and Knowledge Discovery (ITIKD), 1–6. https://doi.org/10.1109/itikd63574.2025.11005241

[27] Gündoğdu S. & Elbir T. (2024). *A data-driven approach for PM2.5 estimation in a metropolis: random forest modeling based on ERA5 reanalysis data*. Environmental Research Communications, 6. https://doi.org/10.1088/2515-7620/ad352d

[28] Abubakar M. L., Thomas D., Ahmed M. S. & Abdussalam A. F. (2024). *Assessment of the Relationship between Land Surface Temperature and Vegetation using MODIS NDVI and LST Time Series Data in Kaduna Metropolis, Nigeria*. FUDMA Journal of Sciences, 8(2), 137–148. https://doi.org/10.33003/fjs-2024-0802-2305

[29] Mande H. K. & Abashiya O. D. (2021). *Assessment of Urban Heat Island in Kaduna Metropolis Between 2000 and 2018*. FUDMA Journal of Sciences, 4(4), 166–174. https://doi.org/10.33003/fjs-2020-0404-450

[30] Tini, N. H. & Light, B. J. (2020). *Impacts of Urban Sprawl on Livability in Kaduna Metropolis, Nigeria*. International Journal of Scientific Research in Science and Technology, 334–343. https://doi.org/10.32628/IJSRST207644

[31] Mansourmoghaddam M., Rousta I., Ghafarian Malamiri H., Sadeghnejad M., Krzyszczak J. & Ferreira C. S. S. (2024). *Modelling and Estimating the Land Surface Temperature (LST) Using Remote Sensing and Machine Learning (Case Study: Yazd, Iran)*. Remote Sensing, 16(3), 454. https://doi.org/10.3390/rs16030454

[32] Purwanto P., Astuti I. S., Rohman F., Utomo K. S. B. & Aldianto Y. E. (2022). *Assessment of the dynamics of urban surface temperatures and air pollution related to COVID-19 in a densely populated City environment in East Java*. Ecological Informatics, 71, 101809. https://doi.org/10.1016/j.ecoinf.2022.101809

[33] Singh S., Kumar M., Sengar V., Kumar A., Abhishek K. & Shafeeq B. M. A. (2026). *Ensemble learning for air quality index prediction: integrating gradient boosting, XGBoost, and stacking with SHAP-based interpretability*. Scientific Reports, 16(1), 8544. https://doi.org/10.1038/s41598-026-39232-w

[34] Qin L., Zhu Y., Liu S., Zhang X. & Zhao Y. (2025). *The Shapley Value in Data Science: Advances in Computation, Extensions, and Applications*. Mathematics, 13(10), 1581. https://doi.org/10.3390/math13101581