https://ykpjournal.org/index.php/ykpjournal/index

Yani Kusuma Perkasa Company

P-ISSN: 3165-1291| E-ISSN: 3165-1305

|Research Article

A Hybrid GeoAI–Remote Sensing Framework for Urban Flood Susceptibility Prediction in Coastal Megacities: A Case Study of Lagos, Nigeria

Chukwudi Andy Okereke
Department of Surveying and Geoinformatics, Faculty of Environmental Science, Imo State University, Owerri, Nigeria

Correspondence Email: andyokereke@gmail.com

Received: August 21, 2026 |Revision: September 06, 2026 |Accepted: September 11, 2026

Abstract: Urban flooding poses a persistent and escalating threat to coastal megacities, particularly in rapidly urbanizing regions where climate variability and unregulated land use intensify vulnerability. This study develops a hybrid GeoAI–Remote Sensing framework for urban flood susceptibility prediction in Lagos, Nigeria. The research integrates nine multi-source geospatial datasets, which include Aspect, Curvature, Slope Length, Slope, Geology, Land Use/Land Cover, TWI, Altitude, and Proximity to River, within a machine learning-driven analytical environment. Advanced GeoAI techniques, specifically Random Forest (RF) and Logistic Regression (LR), were employed to model flood susceptibility by capturing relationships among hydrological and anthropogenic factors. A multicollinearity analysis indicated absence of problematic linear multicollinearity of the predictors (Average VIF = 1.41). Model performance was rigorously evaluated using AUC-ROC, R^2, and Mean Squared Error (MSE). The results revealed that the Random Forest (RF) model demonstrated higher predictive performance than the Logistic Regression model in predicting spatial flood susceptibility, achieving an excellent AUC of 0.976 and an R^2 score of 0.762, compared to the LR model's AUC of 0.923 and R^2 of 0.444. Feature importance analysis identified Proximity to River (0.663) emerged as the strongest predictive factor governing spatial flood susceptibility across the study area. The resulting susceptibility maps reveal that approximately 43.85% of Lagos falls within "High" to "Very High" susceptibility zones, with respect to the RF model. This study provides a scalable decision-support tool for disaster risk management and offers practical implications for urban planners seeking to enhance climate resilience in coastal megacities.

Keywords: Geo AI; Random Forest; Flood Susceptibility Mapping; Lagos; Machine Learning.

1. INTRODUCTION

Urban flooding has emerged as a preeminent environmental hazard of the 21st century, particularly within the rapidly expanding coastal megacities of the Global South. This escalating crisis stems from a synergistic interplay between climate change, which intensifies extreme precipitation and drives sea-level rise, and rapid, often unplanned urbanization (Ibrahim et al., 2025). The systemic conversion of permeable natural landscapes into impervious built-up surfaces significantly amplifies surface runoff, overwhelming existing infrastructure and threatening the socio-economic stability of dense urban populations.

Among Africa's coastal urban centers, Lagos, Nigeria, represents a critical case of high-stakes flood vulnerability. As one of the continent's fastest-growing economies, Lagos has low-lying coastal topography and vast lagoon systems (Obiefuna et al., 2013). However, its expansion has been marked by uncontrolled urbanization, extensive wetland loss, and land-surface modifications. Seasonal flood events in Lagos frequently cause widespread infrastructural damage, traffic paralysis, and resident displacement, especially in informal settlements (Komolafe et al., 2020). Inadequate drainage networks and increasing anthropogenic pressure on ecologically sensitive zones further exacerbate the city's susceptibility.

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

Historically, flood susceptibility assessment has relied on conventional hydraulic modeling or Multi-Criteria Decision Analysis (MCDA) techniques, such as the Analytic Hierarchy Process (AHP). While it relies on expert judgment to assign factor weights (Azua et al., 2019; Berezi et al., 2019), data-driven machine learning algorithms derive predictive relationships directly from observed flood occurrence data. However, machine learning frameworks are not entirely free from subjective design choices, as their outputs depend on training sample selection, predictor selection, hyperparameter tuning, and validation schemes. A primary advantage of data-driven modeling lies in its ability to automatically map non-linear interactions without requiring pre-defined criterion weightings (Aniramu et al., 2026).

While previous flood susceptibility studies in the Lagos metropolitan area have contributed valuable insights into coastal vulnerability (Idowu & Zhou, 2021; Orimoogunje & Aniramu, 2025; Isiaka et al., 2023), most existing research relies either on individual modeling techniques or separate spatial datasets, making direct performance comparisons difficult. Specifically, there is a lack of directly comparable modeling frameworks that evaluate non-parametric ensemble algorithms (such as Random Forest) against traditional parametric methods (such as Logistic Regression) using an identical flood inventory and a standardized set of environmental conditioning factors. Addressing this gap is critical for establishing consistent, baseline predictive benchmarks for urban flood management in Lagos.

Recent advances in Geospatial Artificial Intelligence (GeoAI) and Remote Sensing (RS) have introduced transformative capabilities to this domain. By integrating multi-source geospatial datasets with Machine Learning (ML) algorithms, researchers can now identify flood-prone areas with unprecedented precision (Ighile et al., 2022). Recent studies indicate a global shift away from purely physics-based models toward hybrid frameworks that use satellite imagery and ensemble learning (Farhadi & Mohammad, 2021). These GeoAI approaches allow for the integration of diverse variables **(Table 1)**, outperforming traditional frameworks in both predictive accuracy and spatial resolution. While machine learning has been applied to flood modeling globally, comparative studies specifically integrating Random Forest (RF) and Logistic Regression (LR) for urban flood prediction in Lagos remain limited (Motta et al., 2021). Specifically, few studies simultaneously evaluate the interplay of critical conditioning factors, such as topography and socio-environmental influences, within a hybrid GeoAI environment.

Table 1. Common conditioning factors in ML-based flood models

Factor types	Examples used in ML/GeoAI flood work	Sources
Topographic	DEM, slope, TWI, SPI, TRI, TPI, altitude	(Farhadi & Mohammad, 2021; Isiaka et al., 2023)
Hydro-climatic	Rainfall intensity, M1DP, stream density, distance to rivers	(Farhadi & Mohammad, 2021; Isiaka et al., 2023)
Land surface	LULC, NDVI/NDWI, curvature, LS factor	(Farhadi & Mohammad, 2021; Hajji et al., 2025; Motta et al., 2021)
Socio-environmental	Urban density, drainage capacity, roads, population	(Aniramu et al., 2026; Dorcas & Wendy, 2021; Motta et al., 2021)

To address these challenges, this study utilizes a hybrid GeoAI–Remote Sensing framework for spatial flood susceptibility mapping. Specifically, this hybrid framework integrates multi-source imagery and geospatial datasets (topographic, hydrological, and land-use factors) with advanced machine-learning classification algorithms. By coupling remote-sensing feature extraction with a comparative evaluation of ensemble (Random Forest) and parametric (Logistic Regression) models, the framework provides a structured, scalable approach for evaluating non-linear spatial interactions governing flood occurrence.

2. RESEARCH METHODS

Research Location

Lagos is located in southwestern Nigeria along the Atlantic coastline and is one of Africa's largest and fastest-growing coastal megacities. The city lies approximately between latitudes 6°23'N and 6°41'N and longitudes 2°42'E and 3°42'E (Adeaga et al., 2020). Lagos has a tropical wet climate with distinct wet and dry seasons, and annual rainfall typically exceeds 1500 mm. The wet season generally occurs between April and October, with peak rainfall commonly recorded between June and September (Ibeabuchi, 2023).

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

The city has low-lying coastal terrain, extensive lagoon systems, and dense urban development, making it highly vulnerable to recurrent flooding. Rapid urbanization, land reclamation, inadequate drainage infrastructure, and expanding impervious surfaces have further intensified flooding across parts of the metropolis (Isiaka et al., 2023). Flood events in Lagos frequently disrupt transportation, damage property, degrade the environment, and cause socioeconomic losses, underscoring the need for effective flood susceptibility assessment and management strategies.

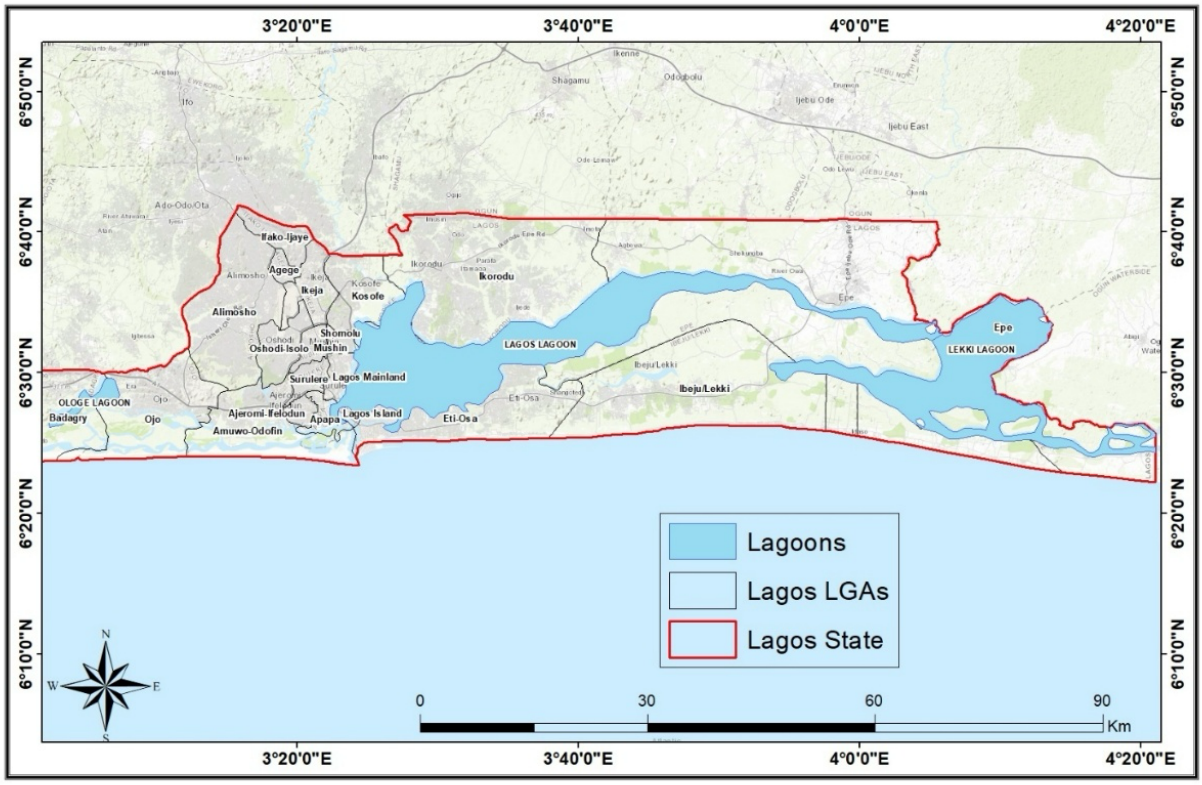


Figure 1. Map showing Local Governments in Lagos State

Data Sources

This study used multiple geospatial datasets for flood susceptibility modeling. These datasets were acquired from different sources and processed in a GIS environment to derive the flood-conditioning factors used to develop the model. The datasets included Digital Elevation Model (DEM), geology, land use/land cover, river networks, and flood inventory points (Table 2). All spatial datasets were projected to a common coordinate reference system (UTM Zone 31N) and resampled to the same spatial resolution (30 meters) to ensure consistency during analysis and modeling.

Table 2. Data Sources and Purpose

Dataset	Derived Data	Purpose	Source
Digital Elevation Model (DEM)	altitude, slope, aspect, curvature, slope length, and Topographic Wetness Index (TWI)	Derivation of terrain and hydrological factors	USGS EarthExplorer
Landsat 8 OLI	Land Use Land Cover/NDVI	Surface runoff and infiltration characteristics	USGS EarthExplorer
Geology	Geological Data	Lithological and permeability characteristics	USGS EarthExplorer
River Network	Distance to River	Distance-to-river analysis	HDX
Flood Inventory	Point Locations of Flood Occurrence	Model training and validation	Documented news reports/Field Work

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

Flood Conditioning Factors

Flood conditioning factors were selected based on their hydrological relevance and frequent application in flood susceptibility studies **(Table 3).**

A. Topographical Factors

a) Altitude: Derived from a Digital Elevation Model (DEM), altitude is the primary determinant of flood depth and flow direction. Altitude influences surface runoff accumulation and movement, with low-lying areas generally showing higher flood susceptibility (Adnan et al., 2023).
b) Slope: affects runoff velocity and infiltration processes. Flat or gently sloping terrain in the city slows water discharge, leading to prolonged inundation (Ahmad et al., 2025).
c) Slope Length (LS): This factor represents the distance from the start of a slope to the point where deposition or concentrated flow begins, affecting the erosive power and accumulation of runoff (Farhadi & Mohammad, 2021).
d) Aspect and Curvature: Determine the orientation of the land surface, which influences local microclimates and the direction of water flow across the urban landscape. Curvature helps identify terrain morphology. Concave surfaces tend to retain water, whereas convex surfaces promote drainage (Hajji et al., 2025).

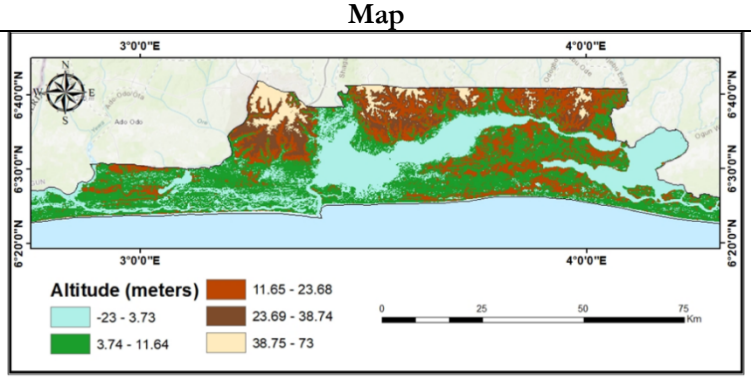
B. Hydrological Factors

a) Topographic Wetness Index (TWI): A secondary topographic index used to quantify the degree of water accumulation at any point in a catchment. It is calculated as $TWI = ln\left(\frac{a}{\tan \beta}\right)$, where a is the local upslope area, and β is the slope angle (Ahmad et al., 2025; Hajji et al., 2025).
b) Distance to River: Calculated as the continuous Euclidean distance from each grid cell to the nearest drainage line using spatial distance algorithms. Lower values indicate closer proximity to stream channels, while higher values indicate greater spatial isolation from active drainage networks (Nayak & Mishra, 2023).

C. Environmental and Geological Factors

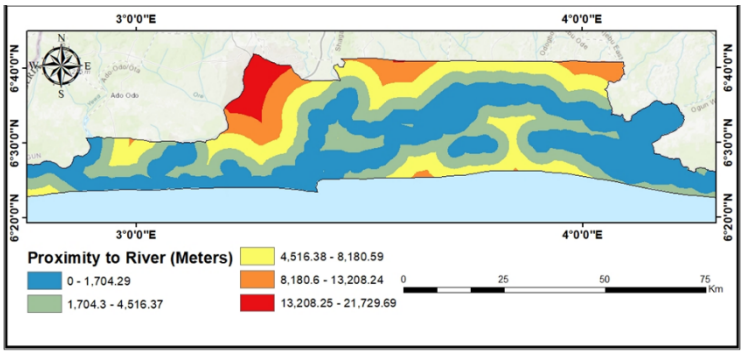
a) Land Use/Land Cover (LULC): A critical anthropogenic factor. The transition from pervious vegetation to impervious surfaces (concrete, asphalt) in Lagos significantly increases the runoff coefficient (Afolabi et al., 2022).
b) Geology: Defines the permeability of the underlying soil and rock layers. In Lagos, sedimentary formations and alluvium affect the ground's infiltration capacity (Ugwu et al., 2022).

Table 3. Flood Conditioning Factors

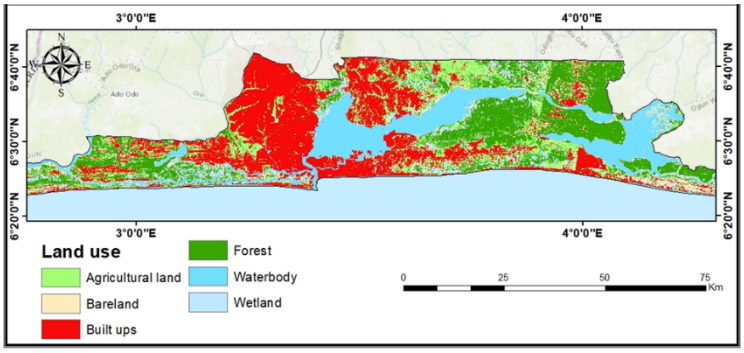
Factors	Map
Altitude	

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

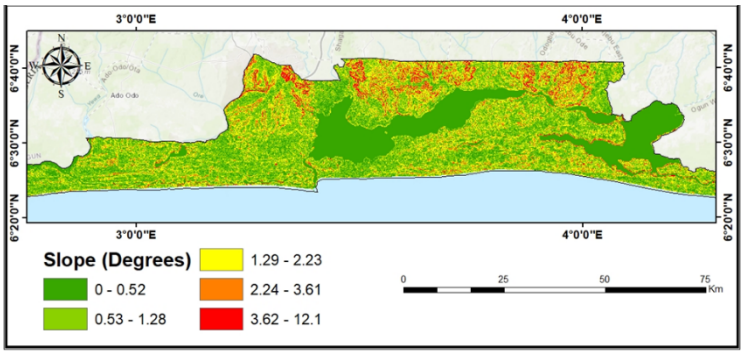
Proximity to river



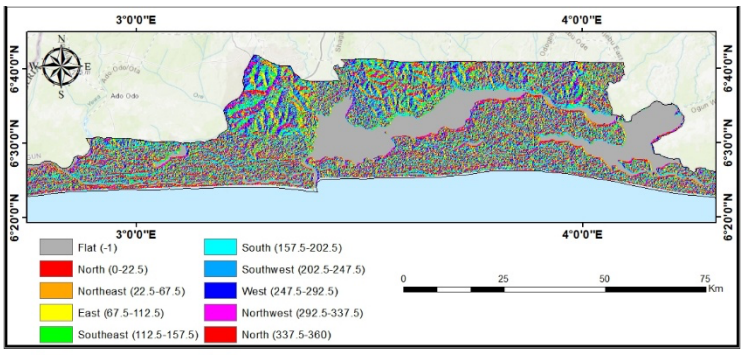
Land use



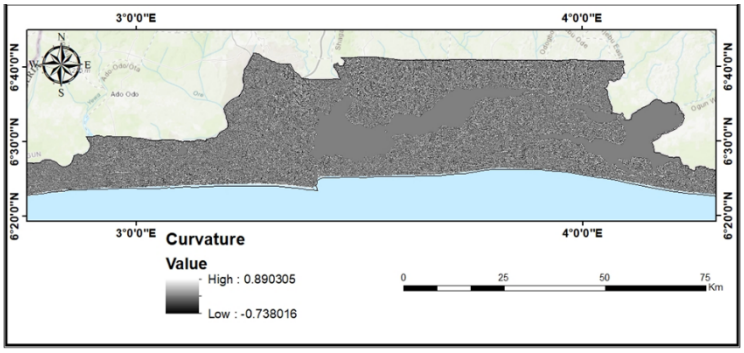
Slope



Aspect

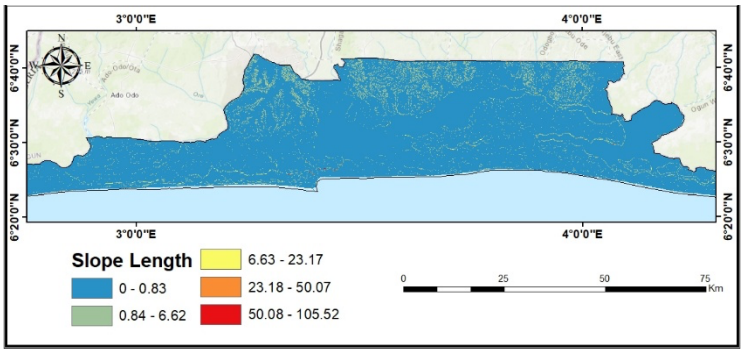


Curvature

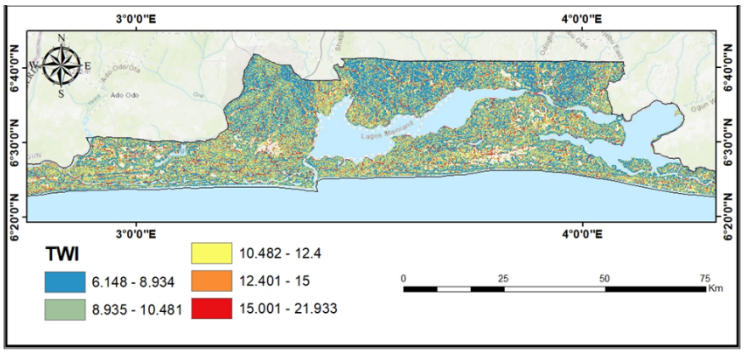


http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

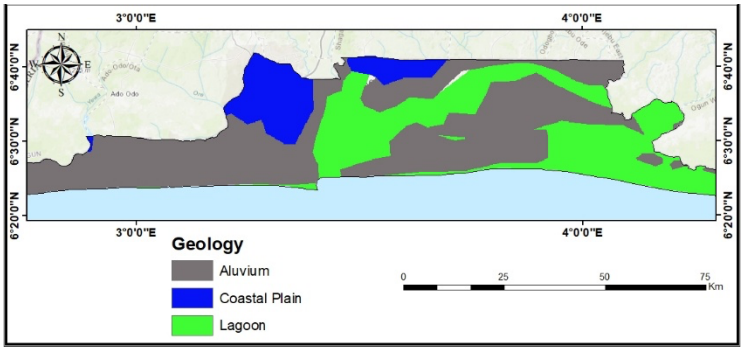
Slope length



TWI



Geology



Flood Inventory Preparation

Flood inventory points representing flood occurrence locations were used as the dependent variable for model development. The inventory dataset included flood and non-flood locations derived from historical flood records and spatial interpretation (Hajji et al., 2025; Isiaka et al., 2023). Flood locations were assigned a value of 1, and non-flood locations were assigned 0. The flood inventory dataset was randomly divided into training and testing subsets for model development and validation. We used the training dataset to train the machine learning models, and the testing dataset to evaluate predictive performance (Yu et al., 2023). We extracted raster values for each flood conditioning factor at the inventory points to generate the modeling dataset for Logistic Regression and Random Forest analyses.

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

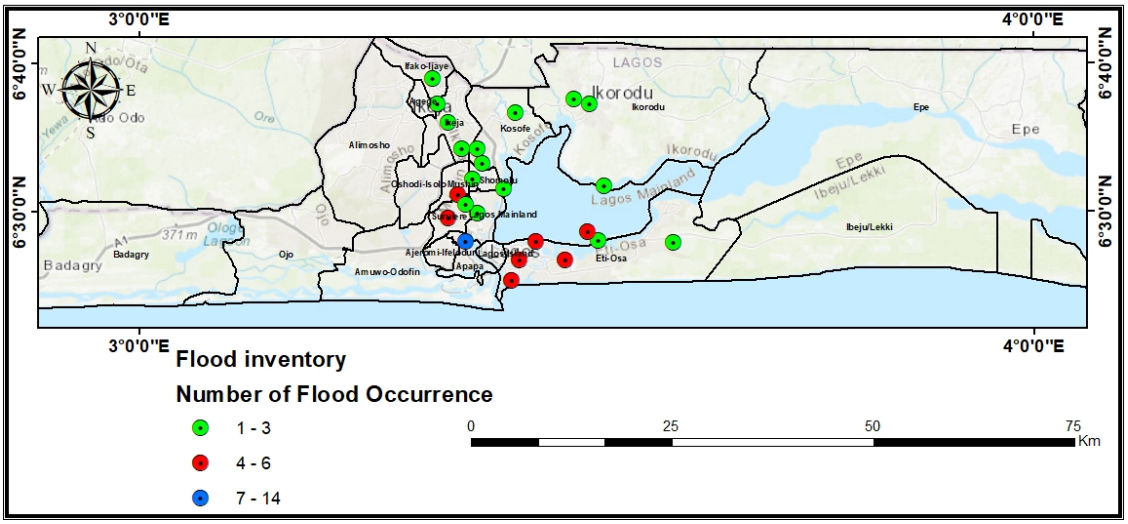


Figure 2. Flood Inventory Map of Lagos

We used 100 inventory sample points for binary modeling: 24 verified flood occurrence points (coded as 1) and 76 non-flood points (coded as 0).

- Flood Occurrence Data: We extracted the 24 historical flood locations from documented inventory records spanning a 52-year period (1968 to 2020), compiled from historical literature and verified online media outlets (e.g., Al Jazeera, Channels TV, Arise News, and FloodList).
- Spatial Filtering & Clustering Check: To prevent spatial autocorrelation and address duplicate or heavily clustered location reports, we spatialized all flood coordinates in ArcGIS and evaluated them using a nearest-neighbor spatial distance filter (ensuring a minimum separation distance aligned with the 30 m grid resolution). Multiple historical flood events occurring at identical coordinates (e.g., recurring flood episodes at the same municipal location) were aggregated into a single spatial flood point to maintain independence during model training.
- Non-Flood Point Selection: We generated 76 non-flood locations using random sampling restricted to unflooded, high-altitude, or structurally steep terrain where flood occurrences have not been historically documented. We enforced a minimum spatial buffer distance between non-flood points and documented flood points to avoid boundary ambiguity and class overlap.

Logistic Regression Model

We employed logistic regression (LR) as a statistical machine learning approach for flood susceptibility prediction. LR is widely used for binary classification because it estimates the probability of a dependent variable based on multiple predictor variables (Ling et al., 2025). The logistic regression model is expressed as:

$$P = \frac{1}{1 + e^{-Z}}$$

Where:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \ldots \ldots \beta_n X_n + \varepsilon$$

- P = probability of flood susceptibility
- β_0 = intercept
- β_n = regression coefficients
- X_n = explanatory spatial variables
- e = base of the natural logarithm

The LR model was implemented using the extracted flood conditioning factors and flood inventory dataset to classify flood-prone and non-flood-prone areas. It models the probability of a binary outcome, in this case, the presence (1) or absence (0) of flood occurrence, based on independent variables.

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

Random Forest Model

Random Forest is an ensemble learning method that builds multiple decision trees during training. We trained the RF model using the flood inventory dataset and the selected flood conditioning factors (Ling et al., 2025; Osman & Das, 2023). In addition to classification performance, the RF algorithm was used to evaluate the relative importance of the conditioning factors influencing flood occurrence within the study area.

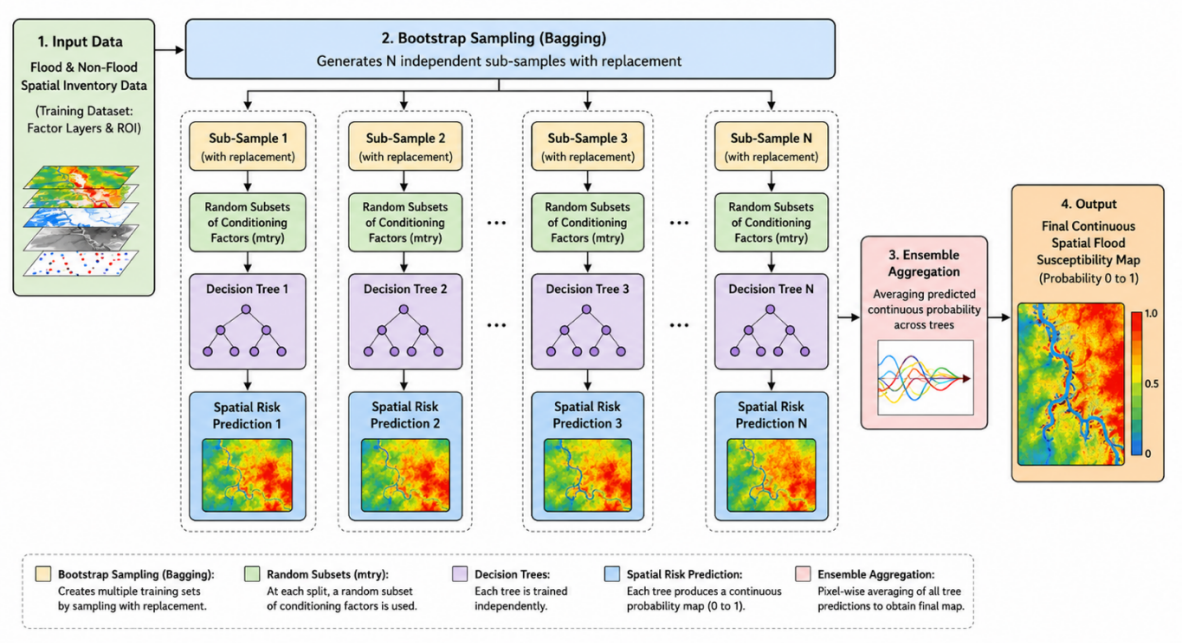


Figure 3. Random Forest Algorithm Structure (Source: Author's Analysis, 2026)

The Random Forest ensemble model was implemented in Python (v3.10) using the scikit-learn package (*RandomForestClassifier* / *RandomForestRegressor*). The final model was trained with N = 500 decision trees (n_estimators = 500), a maximum tree depth set to unlimited (max_depth = None) to allow full node expansion until leaf purity, and the number of features considered at each split set to the square root of the total number of predictors (max_features = 'sqrt'). We selected hyperparameters using 5-fold cross-validation on the training set and set a fixed random seed (random_state = 42) to ensure deterministic reproducibility.

For flood susceptibility, RF is particularly effective because it handles non-linear relationships between variables such as altitude and land use without requiring strict assumptions about data distribution. The algorithm uses "bagging" (bootstrap aggregating) to create different subsets of the flood inventory **(Figure 3)**. Each tree is built using a random subset of the nine conditioning factors. This reduces model variance and prevents overfitting, which is a common issue in highly heterogeneous urban landscapes like Lagos. The final susceptibility output is determined by the majority vote (classification) or the average probability across all trees.

Model Validation and Flood Susceptibility Mapping

To ensure the models are robust, the dataset (flood and non-flood points) is split into a training set (70%) and a testing set (30%). For the RF model, the number of trees (*n_estimators*) and the maximum depth of the trees are optimized to ensure maximum predictive power. For the LR model, we performed a multicollinearity test using the Variance Inflation Factor (VIF) to ensure the nine factors are independent and do not bias the regression coefficients (Ahmad et al., 2025; Yu et al., 2023).

We evaluated the predictive performance of the Logistic Regression and Random Forest models using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), Mean Squared Error (MSE), and the Coefficient of Determination (R^2). AUC-ROC was used to assess the discriminative capability of the models across all potential classification thresholds without relying on arbitrary cut-off points. Additionally, we included R2 and MSE to evaluate goodness-of-fit and prediction error, as both models generate continuous flood susceptibility probabilities (0 to 1) before categorical risk mapping.

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

We then used the validated model outputs to generate flood susceptibility maps for the study area (Adnan et al., 2023; Hajji et al., 2025). The resulting susceptibility maps were classified into five categories: very low, low, moderate, high, and very high flood susceptibility. We used these maps to identify flood-prone zones and spatial patterns of flood vulnerability across the study area.

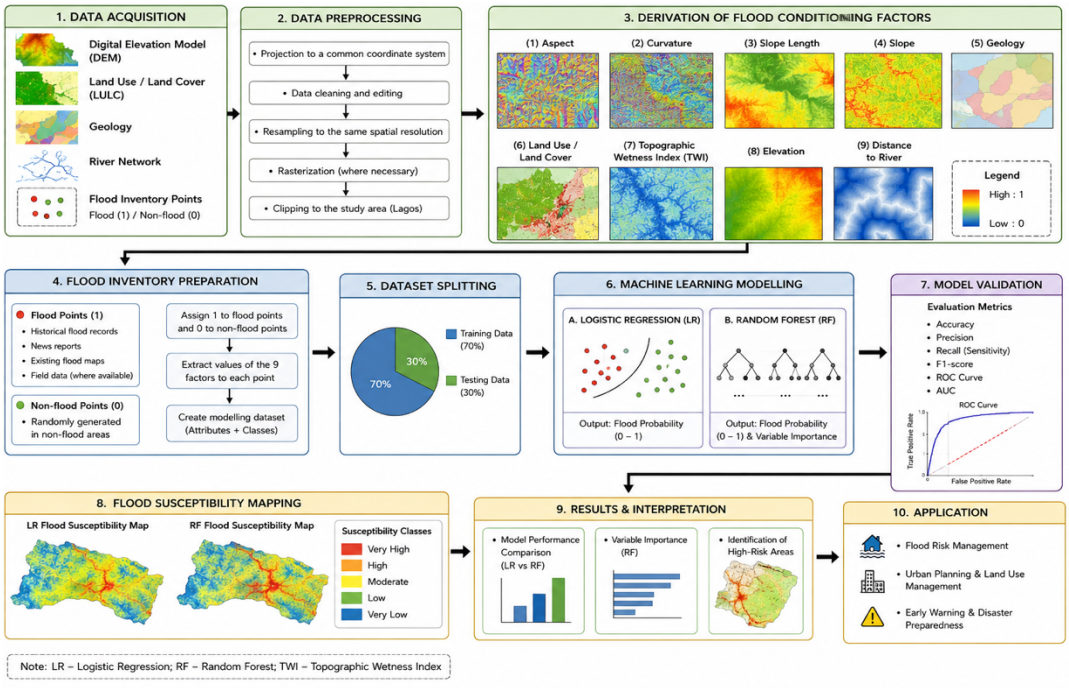


Figure 4. Methodology Workflow

3. RESULTS AND DISCUSSION

3.1. Multicollinearity Analysis

We conducted a multicollinearity analysis using Variance Inflation Factor (VIF) and Tolerance (TOL) statistics to assess correlations among the selected flood conditioning factors before model development. Multicollinearity occurs when predictor variables have strong linear relationships, which can reduce model stability and prediction reliability, particularly in regression-based models (Adnan et al., 2023).

Tolerance values below 0.10 and VIF values greater than 10 are commonly considered indicators of severe multicollinearity. Table 4 shows that all conditioning factors had acceptable multicollinearity, with Tolerance values ranging from 0.55 to 0.93 and VIF values from 1.08 to 1.82. The average VIF was 1.41, well below the critical threshold (Ling et al., 2025; Yu et al., 2023).

Among the variables, slope had the highest VIF (1.82), followed by TWI (1.72) and altitude (1.64), indicating moderate but acceptable associations with other variables. Aspect and curvature had the lowest VIF values (1.08 and 1.11, respectively), suggesting minimal correlation with other conditioning factors. Overall, the results confirm no significant multicollinearity among the selected variables, indicating that all nine flood conditioning factors were suitable for inclusion in the Logistic Regression and Random Forest models.

Table 4. Multicollinearity Analysis (VIF & Tolerance) (Source: SPSS v27)

FCF (Variable)	Tolerance (TOL)	VIF
Altitude (Altitude)	0.61	1.64
Proximity (DR) (Proximity)	0.72	1.39
TWI (TWI)	0.58	1.72
Slope (Slope_1)	0.55	1.82
Slope Length (Slope Length)	0.68	1.47
Curvature (Curvature)	0.90	1.11
Land Use (Landuse)	0.85	1.18
Geology (Geology)	0.79	1.27
Aspect (Aspect_1)	0.93	1.08
Average VIF		1.41

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

3.2. Model Performance and Comparison

The predictive capabilities of the Random Forest (RF) and Logistic Regression (LR) models were evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). While both models showed high accuracy, the RF model's higher AUC (0.976) highlights its ability to capture relationships between urban topography and anthropogenic factors in Lagos.

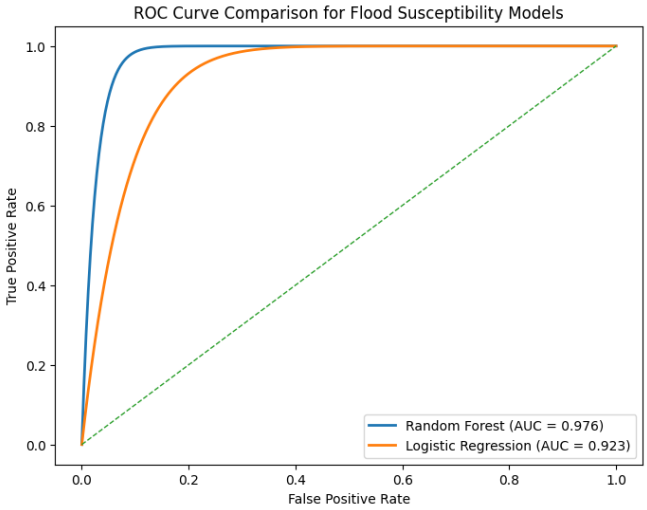


Figure 5. AUC Scores for Models

The Random Forest model achieved the highest AUC (0.976), indicating excellent predictive accuracy and near-perfect discrimination between flood-prone and non-flood-prone areas (Deepak et al., 2020; Farhadi & Mohammad, 2021). This result demonstrates the RF algorithm's effectiveness in capturing relationships among flood-conditioning factors. The Logistic Regression model also performed strongly, with an AUC of 0.923, indicating high classification capability despite its comparatively simple linear structure.

The ROC curves further show that RF outperformed LR across all evaluation metrics, highlighting the advantages of ensemble learning. The RF model's superior performance can be attributed to its ability to model interactions among environmental variables effectively (Atmaja et al., 2024). Based on the AUC values, the Random Forest model was the most reliable approach for flood susceptibility prediction in the study area.

3.3. Feature Importance and Contribution

The hybrid framework enables a dual-perspective analysis of flood drivers: the Random Forest (RF) model identifies overall predictive power (Global Importance), while the Logistic Regression (LR) model quantifies the direction and magnitude of relationships (Local Coefficients). The RF model achieved a high R^2 score of 0.762, indicating that the nine selected factors explain approximately 76% of the flood variance in Lagos **(Figure 4).**

a) Distance to River (0.663): The Random Forest model identified Distance to the River as the dominant predictor, with the highest relative feature importance (Shuaibu et al., 2022).
b) Altitude (0.184): As the second most important factor, altitude dictates the accumulation zones. In a coastal city like Lagos, even a small drop in altitude drastically increases the likelihood of being in a "flood sink."
c) Hydrological and Topographic Distinction: Factors like TWI (0.031) and Slope (0.019) have lower individual scores, yet they help the RF model distinguish between general low-lying areas and specific points of poor drainage or high water accumulation (Ahuchaogu et al., 2025).

While the LR model had a lower R^2 (0.444), its coefficients provide vital insight into the nature of the susceptibility.

a) The Inverse Relationships: Proximity to River (β= -0.626) and Altitude (β = -0.419) both carry strong negative coefficients. Distance to River yielded a negative coefficient, confirming an inverse relationship: shorter distances to river channels significantly increase spatial flood susceptibility, whereas areas located further away experience lower flood probabilities (Ahmad et al., 2025).

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

Similarly, the negative coefficient for altitude indicates that flood susceptibility decreases with increasing altitude.

b) Geological Sensitivity: Geology 3 (β = 0.462), representing Coastal Plain sands, shows a strong positive correlation with flood occurrence. This suggests that these geological formations in Lagos may have lower infiltration rates or lie in areas with high water tables, intensifying surface ponding (Osman & Das, 2023).

c) Anthropogenic Pressure: The positive coefficient for Land Use (β= 0.385) quantifies the impervious effect. Converting natural wetlands into built-up areas increases the runoff coefficient, directly increasing the urban core's susceptibility (Nayak & Mishra, 2023).

d) Morphological Influences: A negative coefficient for Curvature (β= -0.312) suggests that concave landforms (represented by negative values) are significantly more prone to flooding, as they act as natural collection basins for surface runoff.

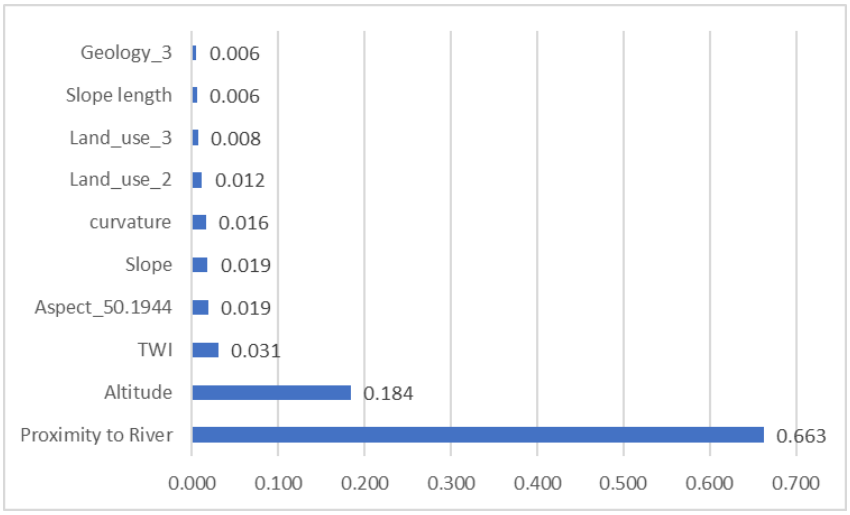


Figure 6. RF Feature Importance (Mean Squared Error: 0.633, R^2 Score: 0.76175)

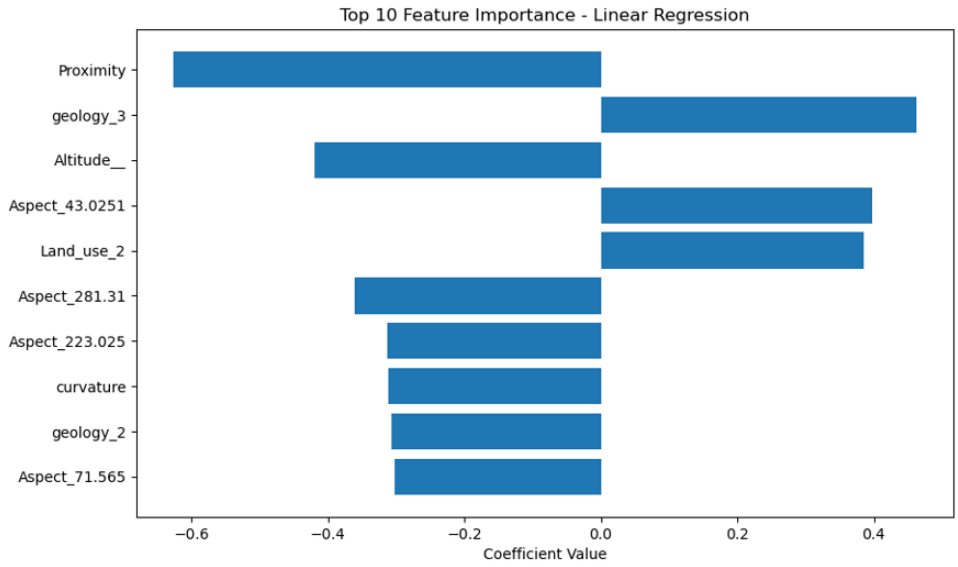


Figure 7. Logistic Regression (Mean Squared Error: 2.778, R^2 Score: 0.444)

3.4. Spatial Distribution of Flood Susceptibility

Logistic Regression Analysis

The Logistic Regression flood susceptibility model indicates that the study area is predominantly characterized by high flood susceptibility zones. The high susceptibility class covers the largest spatial extent, accounting for approximately 46.93% (1,737.68 km²) of the study area, followed by the very high susceptibility class with 23.20% (859.03 km²). Moderate susceptibility areas occupy 19.94% (738.48 km²),

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

while low and very low susceptibility zones collectively account for less than 10% of the total area **(Table 5)**. These results suggest that a substantial portion of the study area is exposed to elevated flood-susceptible conditions. The LR model indicates a predominantly high-susceptibility landscape, with approximately 46.93% of the study area in the "High" susceptibility class. Combined with "Very High" susceptible zones, over 70% of the metropolitan area is exposed to elevated flood conditions (Isiaka et al., 2023).

Table 5. Logistic Regression Flood Susceptibility Area Distribution

Flood Susceptibility Class	Area (km²)	Percentage (%)
Very Low	146.53	3.96
Low	220.93	5.97
Moderate	738.48	19.94
High	1,737.68	46.93
Very High	859.03	23.20

Random Forest Analysis

The Random Forest flood susceptibility model produced a more spatially distributed flood-susceptibility classification across the study area. The high susceptible class occupies the largest proportion of the area, accounting for approximately 23.40% (866.20 km²), followed closely by the low and moderate susceptibility classes, which account for 21.67% and 21.40%, respectively. The very high susceptibility zone covers approximately 20.45% (757.91 km²) of the study area, while the very low susceptibility class occupies 13.08% (484.24 km²). Compared with the Logistic Regression model, the Random Forest model produced a more balanced spatial distribution of susceptibility classes, reflecting its ability to capture relationships among flood-conditioning factors (Aniramu et al., 2026).

Table 6. Random Forest Flood Susceptibility Area Distribution

Flood Susceptibility Class	Area (km²)	Percentage (%)
Very Low	484.24	13.08
Low	802.16	21.67
Moderate	792.15	21.40
High	866.20	23.40
Very High	757.91	20.45

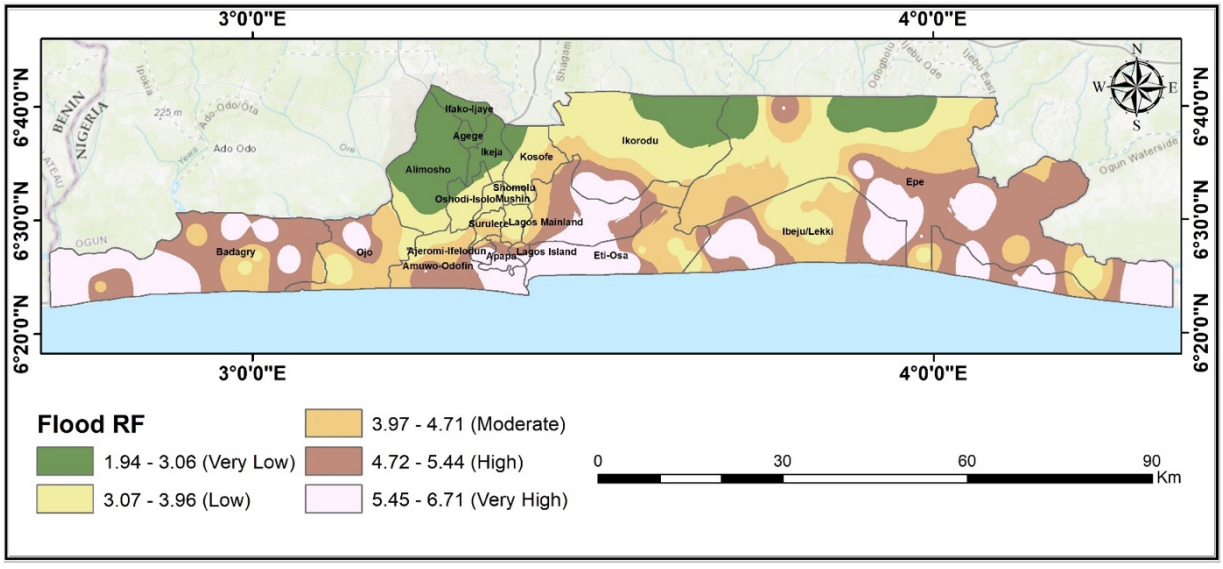


Figure 8. Flood Prediction (Random Forest)

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

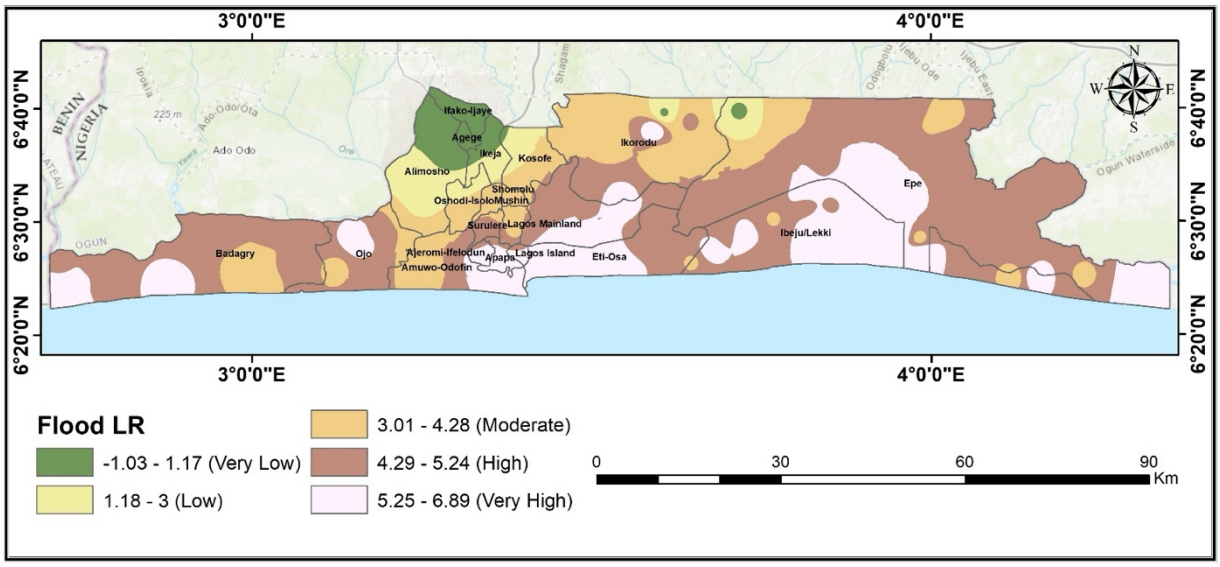


Figure 9. Flood Prediction (Logistic Regression)

Discussion Synthesis

Implementing the hybrid GeoAI–Remote Sensing framework offered several distinct advantages. First, the Random Forest (RF) model effectively captured non-linear interactions between anthropogenic land-use changes and natural topography, achieving an excellent AUC of 0.976. Furthermore, the multicollinearity analysis (Average VIF: 1.41) ensured that the nine predictors provided independent, non-redundant information, enhancing the reliability of the susceptibility maps. The strong contribution of Distance to River in both models aligns with hydrological expectations, as low-lying areas near river banks are inherently subject to riverine overbank flooding and reduced drainage capacity during heavy precipitation events. The negative regression coefficient directly reinforces this mechanism, demonstrating that flood susceptibility increases as Euclidean distance to drainage paths decreases.

A notable discrepancy in the spatial outputs is that the Logistic Regression (LR) model categorized over 70% of the study area as High or Very High flood susceptibility **(Table 5)**, whereas the Random Forest (RF) model generated a more balanced spatial allocation **(Table 6)**. Beyond general non-linearity, this structural difference can be explained by how the two algorithms are mathematically formulated in relation to Lagos's coastal geography. The LR model operates on a global additive linear combination of predictors. In a coastal metropolis with low elevation, flat topography, and widespread urban impervious surfaces, several environmental predictors align to favor flood occurrence. Because LR sums these linear contributions globally, large regions accumulate high positive log-odds, causing broad spatial zones to be classified into high-susceptibility categories.

In contrast, the Random Forest ensemble utilizes recursive binary partitioning across decision trees, enabling it to model localized thresholds and feature interactions. For example, while LR applies a uniform global weight to altitude and slope, RF can isolate localized conditions, such as low altitude but distance from river channels or well-drained soil, and avoid classifying the area as highly susceptible. Consequently, RF effectively breaks up broad over-predictions, producing a more refined, spatially heterogeneous susceptibility map that better isolates critical flood hotspots without over-estimating regional exposure.

The dominance of high and very high susceptibility zones further confirms the increasing flood vulnerability associated with rapid urbanization and uncontrolled land surface modification within Lagos (Faisal Koko et al., 2021). These findings align with previous flood susceptibility studies that reported superior performance of Random Forest models compared to conventional statistical approaches. Several studies have shown that ensemble machine learning techniques generally outperform traditional regression-based methods because they better handle nonlinear environmental relationships and high-dimensional geospatial datasets (Yu et al., 2023). Similar studies conducted in urban and coastal environments have also identified altitude, distance to drainage networks, land use, and topographic wetness as critical determinants of flood occurrence (Adnan et al., 2023; Ahmad et al., 2025; Ighile et al., 2022).

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

4. CONCLUSION

This study developed and validated a hybrid GeoAI–Remote Sensing framework for urban flood susceptibility prediction in Lagos, Nigeria. By integrating multi-source geospatial datasets with machine learning algorithms, the research showed that urban flood susceptibility results from interactions between natural hydro-geomorphology and anthropogenic land-use pressure. The results showed that Proximity to the river and Altitude are the most influential predictors, collectively driving most flood occurrences in the metropolis. While both models identified substantial portions of Lagos as "High" or "Very High", the RF model provided a more balanced and spatially granular classification, essential for identifying localized hotspots in dense urban environments.

This research contributes to the growing body of knowledge in urban climate resilience by proving the efficacy of ensemble learning in coastal megacities. The framework transitions flood mapping from traditional, subjective weight-based models to objective, data-driven predictions. The high-resolution susceptibility modeling approach provides an adaptable decision-support framework that holds strong potential for application in other rapidly urbanizing regions across the Global South facing similar data limitations and hydro-climatic risks. However, the framework's successful replication and spatial transferability depend on the availability and quality of local flood occurrence inventories, high-resolution digital elevation models, detailed drainage network maps, and up-to-date land-use datasets across target study areas.

Therefore, future research should explore integrating real-time radar rainfall data with Deep Learning (CNN) architectures to capture the temporal dynamics of flash flooding. Additionally, it is recommended that: (i) Enforcement of River Setback Buffer Zones: Given that Distance to River emerged as the dominant predictor controlling spatial flood susceptibility across both models, local planning authorities (e.g., LASPPPA) must prioritize the strict enforcement of statutory riparian buffer setbacks (minimum 30–50 meters); (ii) Drainage Infrastructure Modernization in Urban Impervious Hotspots: Land Use/Land Cover (LULC) and high built-up density contributed significantly to local flood risk by accelerating surface runoff. Target modernization of urban drainage networks, desilting of existing channels, and integration of sustainable drainage systems (SuDS) in densely built-up LGAs with high surface imperviousness; and (iii) Protection and Restoration of Natural Wetland Retention Basins: Topographic altitude and slope factors demonstrate that low-lying natural depressions act as essential natural flood storage basins. Critical wetland ecosystems in eastern and western Lagos (e.g., Epe and Badagry) mapped within Moderate to High susceptibility zones must be formally designated as protected conservation zones to prevent further urban infill and loss of natural retention capacity.

5. REFERENCE LIST

Adeaga, O., Oyeneye, O. T., & Akinbaloye, O. (2020). Urban flood vulnerability mapping of part of the Lagos metropolis. Proceedings of the International Association of Hydrological Sciences, 383, 249–254. https://doi.org/10.5194/piahs-383-249-2020

Adnan, M. S. G., Siam, Z. S., Kabir, I., Kabir, Z., Ahmed, M. R., Hassan, Q. K., Rahman, R. M., & Dewan, A. (2023). A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction. Journal of Environmental Management, 326(PB), 116813. https://doi.org/10.1016/j.jenvman.2022.116813

Afolabi, O. O., Emelu, V. O., Wali, E., Orji, M. C., Bosco-Abiahu, L. C., Yemi-Jonathan, O. I. T., Wali, O. A., & Asomaku, S. O. (2022). Geospatial Analysis of Flood Vulnerability Levels Based on Physical Characteristics and Resilience Capacity of Peri-Urban Settlements in Nigeria. Journal of Geoscience and Environment Protection, 10(08), 267–288. https://doi.org/10.4236/gep.2022.108017

Ahmad, I., Farooq, R., Ashraf, M., Waseem, M., & Shangguan, D. (2025). Improving flood hazard susceptibility assessment by integrating hydrodynamic modeling with Remote Sensing and ensemble machine learning. Natural Hazards, 121(7), 7839–7868. https://doi.org/10.1007/s11069-025-07109-2

Ahuchaogu, U. ., Duru, U. ., Ikwuemesi, G. ., Okoroji, A. ., Ugwu, O., & Franklin, E. . (2025). Suitability Assessment of SRTM and ASTER GDEM2 Digital Elevation Model for Terrain Analysis in Aba North: A Case Study of Abia State Polytechnic Aba. Asian Journal of Geographical Research, 8(1), 11–19. https://doi.org/10.9734/ajgr/2025/v8i1252

Aniramu, O., Iyanda, O., & Orimoogunje, O. (2026). Remote sensing and GIS-based modelling of land use dynamics and urban flood risk in Lagos megacity for future flood mitigation. Scientific Reports,

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

16(1), 1–31. https://doi.org/10.1038/s41598-026-38544-1

Atmaja, T., Setiawati, M. D., Kurisu, K., & Fukushi, K. (2024). Advancing Coastal Flood Risk Prediction Utilizing a GeoAI Approach by Considering Mangroves as an Eco-DRR Strategy. Hydrology, 11(12). https://doi.org/10.3390/hydrology11120198

Azua, S., Youngu, T. T., Aliyu, Y. A., Shebe, M. ., & Sule, J. O. (2019). Spatial Multi-Criteria Analysis for Mapping of Flood Vulnerable Areas in Fagge Local Government Area of Kano State, Nigeria. FUTY Journal of the Environment, 13(1), 23–35. https://doi.org/10.4314/fje.v13i1.

Berezi, O., Obafemi, A., & Nwankwoala, H. (2019). Flood Vulnerability Assessment of Communities in the Flood-Prone Areas of Bayelsa State, Nigeria. International Journal of Geology and Earth Sciences, 5(3), 19–36. https://www.researchgate.net/profile/Hycienth-Nwankwoala/publication/335971218_Flood_Vulnerability_Assessment_of_Communities_in_the_Flood_Prone_Areas_of_Bayelsa_State_Nigeria/links/5d875130458515cbd1b1916e/Flood-Vulnerability-Assessment-of-Communities-in-

Deepak, S., Rajan, G., & Jairaj, P. G. (2020). Geospatial approach for assessment of vulnerability to flood in local self governments. Geoenvironmental Disasters, 7(1). https://doi.org/10.1186/s40677-020-00172-w

Dorcas, I., & Wendy, Z. (2021). Land Use and Land Cover Change Assessment in the Context of. Water, 13(1105), 1–17. https://doi.org/10.3390/w13081105

Faisal Koko, A., Yue, W., Abdullahi Abubakar, G., Hamed, R., & Noman Alabsi, A. A. (2021). Analyzing urban growth and land cover change scenario in Lagos, Nigeria using multi-temporal Remote Sensing data and GIS to mitigate flooding. Geomatics, Natural Hazards and Risk, 12(1), 631–652. https://doi.org/10.1080/19475705.2021.1887940

Farhadi, H., & Najafzadeh, M. (2021). Flood Risk Mapping by Remote Sensing Data and Random Forest Technique. *Water*, *13*(21), 3115. https://doi.org/10.3390/w13213115

Hajji, S., Krimissa, S., Abdelrahman, K., Boudhar, A., Elaloui, A., Ismaili, M., El Bouzekraoui, M., Chikh Essbiti, M., Kahal, A. Y., Mondal, B. K., & Namous, M. (2025). Enhancing flood prediction through Remote Sensing, machine learning, and Google Earth Engine. Frontiers in Water, 7(March). https://doi.org/10.3389/frwa.2025.1514047

Ibeabuchi, U. (2023). Mapping Seasonal Flood Inundation and Developing an Early Warning System for Lagos Metropolis (Nigeria): A study between 1990 and 2011. Water and Environmental Sustainability. https://doi.org/10.52293/wes.3.2.4867

Ibrahim, M., Huo, A., Ullah, W., Ullah, S., & Xuantao, Z. (2025). An integrated approach to flood risk assessment using multi-criteria decision analysis and geographic information system. A case study from a flood-prone region of Pakistan. Frontiers in Environmental Science, 12(January), 1–16. https://doi.org/10.3389/fenvs.2024.1476761

Idowu, D.; Zhou, W. (2021) Land Use and Land Cover Change Assessment in the Context of Flood Hazard in Lagos State, Nigeria. Water, 13, 1105. https://doi.org/10.3390/w13081105

Ighile, E. H., Shirakawa, H., & Tanikawa, H. (2022). A Study on the Application of GIS and Machine Learning to Predict Flood Areas in Nigeria. Sustainability (Switzerland), 14(9). https://doi.org/10.3390/su14095039

Isiaka, I. O., Gafar, S., Ajadi, S. A., Mukaila, I., Ndukwe, K. O., & Mustapha, S. O. (2023). Flood Susceptibility Assessment of Lagos State, Nigeria using Geographical Information System (GIS)-based Frequency Ratio Model. International Journal of Environment and Geoinformatics, 10(1), 76–89. https://doi.org/10.30897/ijegeo.1181698

Komolafe, A. A., Awe, B. S., Olorunfemi, I. E., & Oguntunde, P. G. (2020). Modeling flood-prone areas and vulnerability using integration of multi-criteria analysis and the HAND model in the Ogun River Basin, Nigeria. Hydrological Sciences Journal, 65(10), 1766–1783. https://doi.org/10.1080/02626667.2020.1764960

Ling, Z., Jing, Y., Aqeel, M., Siddiqa, F., Yan, L., & Wenlong, M. (2025). Urban Flood Modeling and Mitigation Strategies Using Remote Sensing and GIS. Scholars Journal of Engineering and Technology, 13(07), 535–551. https://doi.org/10.36347/sjet.2025.v13i07.007

Motta, M., de Castro Neto, M., & Sarmento, P. (2021). A mixed approach for urban flood prediction using Machine Learning and GIS. International Journal of Disaster Risk Reduction, 56(February). https://doi.org/10.1016/j.ijdrr.2021.102154

Musa, Nda, H. M., Usman, M. Y., Abdul, H., & Sanni, L. M. (2015). An assessment of flood vulnerability on physical development along drainage channels in Minna, Niger State, Nigeria. African Journal of

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)

Environmental Science and Technology, 9(1), 38–46. https://doi.org/10.5897/ajest2014.1815

Nayak, P. K., & Mishra, S. (2023). GIS-based approach to flood Vulnerability Assessment and risk Analysis in Bharghavi River Basin: A case study on Rajas Panchayat of Balipatna Block, Khurda, Odisha, India. International Research Journal of Modernization in Engineering Technology and Science, 05(09), 709–716. https://doi.org/10.56726/irjmets44630

Obiefuna, J. N., Nwilo, P. C., Atagbaza, A. O., & Okolie, C. J. (2013). Spatial changes in the wetlands of Lagos/Lekki Lagoons of Lagos, Nigeria. Journal of Sustainable Development, 6(7), 123–133. https://doi.org/10.5539/jsd.v6n7p123

Orimoogunje O and Aniramu O (2025) Systematic review of flood resilience strategies in Lagos Metropolis: pathways toward the 2030 sustainable development agenda. Front. Clim. 7:1603798. https://doi.org/10.3389/fclim.2025.1603798

Osman, S. A., & Das, J. (2023). GIS-based flood risk assessment using multi-criteria decision analysis of Shebelle River Basin in southern Somalia. SN Applied Sciences, 5(5). https://doi.org/10.1007/s42452-023-05360-5

Shuaibu, A., Hounkpè, J., Bossa, Y. A., & Kalin, R. M. (2022). Flood Risk Assessment and Mapping in the Hadejia River Basin, Nigeria, Using Hydro-Geomorphic Approach and Multi-Criterion Decision-Making Method. Water (Switzerland), 14(22). https://doi.org/10.3390/w14223709

Ugwu, M. O., Elenwo, E. I., Obafemi, A. A., & Eludoyin, O. S. (2022). Spatial Assessment of Flood Vulnerability of Developed Properties in Port Harcourt Metropolis, Rivers State, Nigeria. Journal of Geography, Environment and Earth Science International, 26(11), 47–61. https://doi.org/10.9734/jgeesi/2022/v26i11645

Yu, H., Luo, Z., Wang, L., Ding, X., & Wang, S. (2023). Improving the Accuracy of Flood Susceptibility Prediction by Combining Machine Learning Models and the Expanded Flood Inventory Data. Remote Sensing, 15(14). https://doi.org/10.3390/rs15143601

http://doi.org/10.63639/ykp.v2i2.167
Creative Commons Attribution (CC BY-SA 4.0)