

Integrating Morphometric Controls for Runoff Dynamics in Bayelsa State, Nigeria: Enhancing Flood Susceptibility Mapping with Machine Learning

Okes Imoni

jennyimoni@gmail.com

Niger Delta University <https://orcid.org/0009-0007-3548-2889>

Research Article

Keywords: Flood Risk Management, Machine Learning, Catchments, Geospatial Morphometry, XGBoost, SRTM, Remote Sensing

Posted Date: March 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8993426/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

**Integrating Morphometric Controls for Runoff Dynamics in Bayelsa State, Nigeria:
Enhancing Flood Susceptibility Mapping with Machine Learning**

**Okes Imoni^{1*}; Prince Chukwuemeka²; Elijah Ige Ohimain³; Ayebawanaemi G. Winston⁴;
Franklin N. Okeke⁵; Desmond R. Eteh⁶; Tarinabo William ⁷; Ebingiye Nelvin Agbozu⁸;
Oyinkuro Emmanuel Kalama ⁹; Meremu Dogiye Amos¹⁰; Olasuodo Inearrow Ogidi¹¹; Becky
Peremoboere Bamiekumo¹²; Obinna Ihekona¹³**

¹Department of Biological Sciences,
Niger Delta University,
Wilberforce Island,
Bayelsa State, Nigeria
ORCID: 0009-0007-3548-2889

²Department of Computing and Informatics,
Bournemouth University,
Poole United Kingdom
ORCID: 0009-0005-1850-9870

³Department of Microbiology,
Niger Delta University,
Wilberforce Island,
Bayelsa State, Nigeria,

⁴Department of Geology,
Niger Delta University,
Wilberforce Island,
Bayelsa State, Nigeria

⁵Department of Computing and Informatics,
Bournemouth University,
Poole United Kingdom

⁶Department of Geology,
Niger Delta University,
Wilberforce Island,
Bayelsa State, Nigeria
ORCID: 0000-0003-0660-6743

⁷Department of Physics,
Niger Delta University,
Wilberforce Island,
Bayelsa State, Nigeria
ORCID: 0009-0005-7082-2485

⁸Department of Micro Biology,
Niger Delta University,
Wilberforce Island,
Bayelsa State, Nigeria

⁹Department of Civil Engineering,
Niger Delta University,
Wilberforce Island,
Bayelsa State, Nigeria

¹⁰Department of Petroleum Engineering,
Rivers State University,
Port Harcourt, Rivers State, Nigeria

¹¹Geosoft Global Innovation Limited
Egale Island, Port Harcourt
Port Harcourt, Rivers State, Nigeria

¹²Department of Geology,
Niger Delta University,
Wilberforce Island,
Bayelsa State, Nigeria

¹⁰Geosoft Global Innovation Limited
Eagle Island,
Port Harcourt, Rivers State, Nigeria

Abstract

Flooding remains one of the most frequent and destructive hazards in Nigeria's Niger Delta, where flat terrain and intense rainfall combine with rapid land-use change to heighten risk. This study integrates morphometric basin analysis with supervised machine learning to enhance flood-susceptibility mapping across four representative catchments in Bayelsa State Forcados, Nun, Ekole, and Seibri. Using a 30 m SRTM digital elevation model, fourteen morphometric parameters were derived to quantify drainage efficiency, infiltration potential, and relief characteristics that influence runoff generation. These parameters served as predictors for Random Forest, Support Vector Machine, and XGBoost classifiers trained with flood and non-flood samples extracted from Sentinel-1 SAR data (2018–2024). Among the tested models, XGBoost achieved the highest performance (accuracy = 93.1%, AUC = 0.95), reliably delineating high-risk sub-catchments. The resulting probability maps revealed micro-zones localized clusters of elevated flood potential within generally moderate basins such as Ekole, emphasizing the spatial heterogeneity of flood processes. Drainage density (Dd), relief ratio (Rh), and infiltration number (If) were the most influential variables controlling flood susceptibility. Forcados exhibited the greatest hazard (Dd = 3.57 km km⁻²; If = 41.80), whereas Seibri and Ekole showed lower overall susceptibility but contained critical micro-zones requiring attention. The approach demonstrates that combining morphometric controls with data-driven classification provides a transferable, high-resolution tool for flood-risk assessment in other tropical deltaic environments lacking dense hydrological networks. These findings support more precise local planning, early warning, and adaptation strategies in vulnerable low-lying regions.

Keywords: Flood Risk Management, Machine Learning, Catchments, Geospatial Morphometry, XGBoost, SRTM, Remote Sensing

Highlights

- XGBoost outperformed RF and SVM with 93.1% accuracy and AUC of 0.95 for flood risk prediction.
- Drainage density, relief ratio, and infiltration number were top predictors of flood-prone areas.
- Forcados catchment ranked highest in flood risk due to high stream density and steep relief.
- ML models detected micro-zones of flood risk, especially in Ekole, missed by traditional morphometry.
- Hybrid GIS–ML approach improved spatial precision and is scalable for other flood-prone regions.

1. Introduction

Flooding remains one of the most pervasive and damaging natural hazards worldwide, accounting for more than 40 percent of all recorded disasters in the past three decades [1]. Its growing frequency and intensity are closely linked to climate variability, accelerated urbanization, and unregulated land-use change [2]. In deltaic regions such as Nigeria's Niger Delta, flooding has evolved from a periodic seasonal event into a recurrent environmental crisis that disrupts livelihoods, damages infrastructure, and undermines socio-economic development [3, 4]. The combination of high rainfall, low elevation, flat terrain, and an intricate network of distributaries amplifies inundation potential by promoting overbank flow and prolonged waterlogging [5]. Understanding how terrain morphology, drainage configuration, and surface characteristics interact to control runoff generation is therefore critical for developing sustainable flood mitigation strategies. Hydro-geomorphology provides a process-based framework for interpreting the relationship between topography and hydrological response through quantitative analysis of basin form and structure [6]. Morphometric analysis, which numerically evaluates the geometry, relief, and network organization of drainage basins, has long served as a cornerstone in hydrological research because it reveals how rainfall is transformed into streamflow [7]. Parameters such as drainage density (D_d), stream frequency (F_s), bifurcation ratio (R_b), elongation ratio (R_e), relief ratio (R_h), form factor (F_f), circularity ratio (R_c), and infiltration number (I_f) collectively describe the efficiency of a catchment to convey water and its capacity for infiltration and storage [8, 9]. High values of D_d and R_h generally indicate steep slopes and rapid runoff concentration, conditions that increase flash-flood potential, whereas high R_e or R_c values are often associated with elongated basins that have delayed flow response [10]. These indices quantify the intrinsic hydrological behaviour of drainage systems and thus serve as diagnostic predictors of flood susceptibility. Despite their value, traditional morphometric approaches are constrained by several limitations that restrict predictive capability. They typically rely on linear relationships or subjective weighting schemes that fail to capture nonlinear interactions among parameters [11]. Moreover, morphometric indices are inherently static, representing long-term geomorphic form rather than short-term processes such as rainfall variability, antecedent moisture, and recent land-cover alteration [12]. As a result, morphometric prioritization alone cannot fully represent the spatial heterogeneity and temporal variability that define flood dynamics in complex

environments. Recent advances in data science have provided opportunities to overcome these constraints by combining physically interpretable morphometric variables with machine learning algorithms capable of learning nonlinear patterns from empirical data [13]. Machine learning (ML) has emerged as a transformative tool for environmental prediction, particularly in regions with limited hydrometric observations [14]. Algorithms such as Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost) are increasingly used in flood-susceptibility mapping because they can handle high-dimensional datasets and discover complex, nonlinear associations between conditioning factors and flood occurrence [15]. RF employs an ensemble of decision trees to minimize variance, SVM constructs optimal separating boundaries in multidimensional feature space, and XGBoost utilizes gradient boosting with regularization to enhance predictive stability. When trained on labelled flood and non-flood samples derived from satellite imagery or historical records, these algorithms can produce spatially explicit flood-probability surfaces. Yet, interpretability remains a central challenge. Without linking outputs to physical processes, ML models risk functioning as statistical “black boxes.” Integrating morphometric descriptors into ML frameworks allows predictions to be grounded in hydrological reasoning and improves transparency by relating statistical importance to geomorphic meaning [16]. In Nigeria, flood disasters have increased in both magnitude and frequency, especially since the catastrophic 2012 and 2022 floods that displaced millions of people and caused economic losses exceeding several billion dollars [17]. Bayelsa State, located in the central Niger Delta, is among the most severely affected areas. Almost the entire state lies below 25 m above mean sea level and is drained by distributaries of the Niger River bordered by extensive tidal creeks and wetlands. Annual rainfall exceeds 2,500 mm, and rapid urban expansion combined with deforestation and land reclamation has severely reduced natural flood retention. Recent research across the region highlights the growing role of both climatic and anthropogenic drivers. Eteh et al. [18] demonstrated that climate change has altered flood patterns in downstream Nigeria using ML and geospatial analysis. Okpobiri et al. [19] and Jonathan and Charles [20] showed that morphological instability along the Niger and Orashi Rivers exacerbates flood exposure, while Bamiekumo et al. [21] used Sentinel-1 and Sentinel-2 data to map flood extent and damage in Yenagoa. Abaye et al. [22] further revealed that extensive urbanization and wetland encroachment have modified drainage pathways in Bayelsa. Collectively, these findings affirm that flood risk in the Niger Delta results from a combination of natural morphometric predisposition and human-

induced modification. The physical landscape of Bayelsa represents a confluence of fluvial and tidal systems where multiple hydrological processes operate simultaneously. The Forcados, Nun, Ekole, and Seibri catchments were selected in this study to capture the diversity of hydro-geomorphic and anthropogenic conditions that characterize the region. The Forcados Basin, located in the western sector, features steep headwaters and dense drainage that generate rapid runoff. Nun, the main distributary of the Niger River, contains wide floodplains and numerous tidal channels that promote backwater effects and prolonged inundation. Ekole represents a transitional basin with moderate slopes and expanding urbanization, while Seibri, situated in the eastern delta, is a flat, swampy catchment dominated by mangrove vegetation and low-gradient streams. This spatial variety provides a robust natural gradient for evaluating how variations in morphometric attributes and relief translate into differing flood responses. Several recent investigations underscore the growing importance of ML and geospatial integration for hydrological prediction in Nigeria and comparable tropical deltas. Okes and Jonathan [23] applied ML to evaluate soil contamination dynamics using multivariate analysis, while Jonathan et al. [24] employed morphometric features in ML-based runoff modelling for Bayelsa. Jonathan et al. [25] demonstrated that flood-driven ecosystem degradation in the Niger Delta can be traced through satellite-observed changes in mangrove cover. Oborie et al. [26] used digital shoreline analysis to monitor coastal erosion in Rivers State, and Akajiaku et al. [27] developed ML-based approaches for identifying dam sites in Benue State. Studies by Akajiaku et al. [28], Oseji et al. [29], and Chukwuemeka et al. [30] further expanded ML applications into groundwater prediction, materials sustainability, and computational modelling, confirming the versatility of data-driven methods across geoscientific fields. Complementary works by Amos et al. [31] and Mogo et al. [32] examined drainage and sandbar dynamics in southern Nigeria, showing that terrain-driven indices remain critical for hydrological interpretation even in ML contexts. At the global scale, Hasnaoui et al. [34–36] advanced hybrid and transfer-learning models for flood and erosion detection in the Mediterranean, demonstrating that integrating geomorphic and remote-sensing data enhances both interpretability and predictive accuracy. Despite these advances, there remains a paucity of research explicitly linking morphometric controls to ML-based flood prediction in the Niger Delta. Existing studies often examine hydrological hazards through either morphometric ranking or generic ML classification but rarely merge the two. As a result, the mechanistic relationship between drainage-basin geometry and predicted flood probability remains insufficiently

quantified. Moreover, earlier studies frequently generalize flood susceptibility at the basin scale, overlooking intra-catchment heterogeneity and the localized hotspots of flood hazard that occur within ostensibly moderate-risk basins. Addressing this gap requires a framework that not only quantifies basin-scale predisposition but also refines spatial predictions at micro scales where human alteration and micro-topographic effects become significant. The present study develops an integrated morphometric–machine-learning framework to delineate and rank flood-susceptible zones across four representative catchments in Bayelsa State. Fourteen morphometric parameters derived from the 30 m Shuttle Radar Topography Mission (SRTM) Digital Elevation Model capture linear, areal, and relief characteristics describing drainage efficiency, infiltration capacity, and topographic energy. Flood and non-flood samples obtained from multi-temporal Sentinel-1 Synthetic Aperture Radar (SAR) imagery from 2018 to 2024 are used to train RF, SVM, and XGBoost classifiers. Model performance is assessed using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (AUC), and interpretability is achieved through feature-importance ranking and SHapley Additive exPlanations (SHAP) values. A central concept emerging from this research is that of “micro-zones” spatially localized clusters of high flood probability embedded within broader moderate-risk basins. These micro-zones arise from subtle variations in slope, secondary drainage convergence, or anthropogenic modification such as canalization and reclamation that are not captured by static morphometric indices. Identifying such zones enhances spatial precision in flood mapping and supports targeted management interventions including improved drainage design, zoning regulation, and localized early-warning systems. This focus on micro-scale heterogeneity distinguishes the present study from previous regional analyses. The proposed integration of morphometric descriptors with ML contributes scientifically by bridging deterministic and data-driven paradigms. It demonstrates that physically meaningful terrain variables can enhance model interpretability while maintaining high predictive accuracy. Operationally, the framework is built entirely on open-access geospatial data and reproducible algorithms, making it transferable to other data-limited tropical deltas. Comparable efforts in downstream Nigeria [18], Bayelsa [24], and coastal Algeria [35] confirm the growing relevance of such hybrid methodologies in flood management. The study also acknowledges inherent limitations, including potential vertical errors in the SRTM DEM that may smooth micro-topographic features critical for flood retention, as well as the indirect representation of land cover and soil conditions through morphometric proxies. Furthermore, the dynamism of human activities

such as sand mining and embankment construction necessitates periodic model updates as new satellite data become available [21, 22, 31, 32]. Recognizing these constraints enhances methodological transparency and ensures the adaptability of the workflow to future data improvements. The overarching goal of this research is to establish a comprehensive, interpretable, and transferable framework for flood-susceptibility assessment in Bayelsa State through the integration of morphometric parameters and machine-learning algorithms. The approach seeks to quantify key morphometric indices controlling runoff, to train and compare RF, SVM, and XGBoost classifiers using SAR-derived flood data, to interpret the dominant morphometric drivers of flood occurrence through SHAP analyses, and to map spatial micro-zones of heightened susceptibility for targeted risk management. The guiding hypothesis is that morphometric parameters, when combined with supervised ML and validated against multi-temporal flood observations, can substantially enhance the accuracy and interpretability of flood-susceptibility mapping in low-relief deltaic environments such as the Niger Delta. By addressing this hypothesis, the study contributes both scientifically and practically. Scientifically, it demonstrates how morphometric descriptors serve as physically interpretable predictors within modern data-driven frameworks, thereby reducing model opacity. Practically, it provides planners and policymakers in Bayelsa State with an evidence-based tool for prioritizing infrastructure investment, flood control, and adaptive land-use planning. In bridging geomorphological theory with predictive analytics, the research advances an integrated methodology that strengthens the scientific basis for flood-risk reduction and resilience building in vulnerable deltaic landscapes [18–36]. Unlike previous flood-susceptibility studies that have treated morphometric indices and machine learning as independent analytical paths, this research proposes a fully coupled and interpretable hybrid framework that merges physically meaningful geomorphological variables with explainable artificial intelligence (XAI) techniques. The innovation lies not in the algorithms themselves but in the integration logic and interpretability design using SHapley Additive exPlanations (SHAP) to connect morphometric causality with statistical feature importance. This methodological coupling transforms otherwise opaque ML outputs into physically interpretable hydrological insights, enabling geomorphic reasoning to be recovered from data-driven prediction.

Additionally, the study introduces the concept of flood micro-zones spatially localized high-probability clusters nested within broader basin-scale susceptibility zones a level of spatial detail not captured by previous morphometric or ML-based flood models in West Africa. Together, these innovations establish a transferable and explainable framework suitable for data-limited deltaic landscapes such as Bayelsa State.

2. Study Area

Bayelsa State lies within the central portion of Nigeria's Niger Delta region, extending approximately between latitudes $4^{\circ}15' - 5^{\circ}30' \text{ N}$ and longitudes $5^{\circ}20' - 6^{\circ}45' \text{ E}$ (Figure 1). The state covers an area of about 21,100 km² and is bordered by Delta State to the west, Rivers State to the east, and the Atlantic Ocean to the south. It occupies the lowermost part of the Niger River basin where numerous distributaries, tidal creeks, and backwater channels converge before discharging into the Gulf of Guinea. The area represents one of the most flood-vulnerable landscapes in West Africa due to its low elevation, dense drainage network, and seasonal hydrological surges [37]. Bayelsa experiences a humid tropical climate characterized by two distinct seasons: a wet season from April to October and a dry season from November to March. Mean annual rainfall ranges from 2,500 mm in the north to over 3,500 mm near the coast, while mean temperature fluctuates between 25 °C and 32 °C [38]. The combination of high rainfall and prolonged wet periods produces substantial surface runoff, especially during September–October when monthly precipitation often exceeds 400 mm. The Niger River, along with its principal distributaries the Forcados and Nun supplies enormous discharge volumes that periodically overspill into adjoining floodplains and tidal creeks. Tidal amplitude at the coast varies between 1.5 m and 2.2 m, generating recurrent backwater effects that exacerbate inland flooding when coinciding with peak river stages [39].

Hydrologically, the state consists of several interlinked catchments draining through dendritic and sub-parallel networks into the Atlantic Ocean. The four catchments analyzed in this study Forcados, Nun, Ekole, and Seibri represent the main hydro-geomorphic systems influencing the region's flood dynamics. Their hydrological responses vary markedly due to differences in relief, drainage density, soil texture, and human alteration, thereby providing a representative spectrum for morphometric–machine-learning evaluation. Topographically, Bayelsa is a vast lowland with

elevations mostly below 25 m a.s.l. The terrain gently slopes southward with localized levees and point bars along active channels. The Forcados and Nun catchments show slightly higher relief in their northern sectors (20–40 m), gradually transitioning into flat tidal plains toward the coast. The Ekole and Seibri basins are characterized by shallow depressions, swamp forests, and numerous ox-bow lakes formed by lateral channel migration [40]. The underlying geology belongs to the Quaternary–Tertiary formations of the Niger Delta complex, dominated by the Benin, Agbada, and Akata sequences composed of alternating sands, silts, and clayey lenses [41]. These unconsolidated deposits exhibit high porosity and permeability variations that influence infiltration and subsurface storage. The prevalent soils mainly hydromorphic and alluvial Fluvisols remain waterlogged during most of the year, promoting saturation-excess runoff even under moderate rainfall events [42]. The flat gradient, fine sediments, and shallow water table combine to create high runoff coefficients that intensify flood potential. Several natural processes interact to sustain recurrent flooding in the area. Intense and prolonged rainfall generates large surface volumes exceeding soil infiltration capacity, while the extremely low relief impedes natural drainage [43]. Tidal backflow during high tides or storm surges restricts river discharge, resulting in reverse flow and extensive floodplain inundation. The dense network of meandering streams increases hydraulic connectivity, allowing floodwaters to spread rapidly across adjoining basins. High drainage density ($D_d > 3.0 \text{ km km}^{-2}$ in some sections) and steep local relief ratios in the northern Forcados sub-basin shorten concentration time and elevate peak discharge [44]. Furthermore, progressive sedimentation within channels, driven by deltaic subsidence and siltation, reduces channel conveyance, further aggravating overbank flooding during extreme events. These natural hydrological and geomorphological settings provide the physical baseline upon which anthropogenic pressures intensify flood risk. Human activities have become major modifiers of the hydrological regime in Bayelsa. Rapid urbanization, particularly around Yenagoa, Amassoma, and Ogbia, has replaced natural wetlands with impervious surfaces, accelerating surface runoff and overwhelming poorly planned drainage systems [45]. Oil exploration and infrastructure development roads, pipelines, and embankments have altered natural flow paths and blocked minor channels, producing localized ponding and backwater effects. Widespread deforestation and sand mining have destabilized slopes and increased sediment deposition within rivers, reducing channel depth. In addition, informal settlements often occupy low-lying floodplains without adequate flood defenses, thereby amplifying exposure. The cumulative impact of these activities

superimposed on fragile deltaic geomorphology explains the heightened frequency and spatial extent of flooding in the region. The choice of Forcados, Nun, Ekole, and Seibri catchments was guided by both hydrological diversity and representativeness across Bayelsa State. The Forcados basin, draining the north-western uplands, possesses relatively higher relief and drainage density, making it prone to rapid runoff and flash flooding. The Nun catchment, a principal distributary of the Niger River, experiences combined fluvial and tidal flooding due to its wide channel and proximity to the estuary. The Ekole basin is characterized by moderate relief, fragmented sub-drainage systems, and intense urban expansion near Yenagoa, offering insight into human-induced flood amplification. The Seibri catchment typifies the extremely flat southern floodplain with numerous tidal creeks, representing slow-draining, long-duration inundation zones [24,46].

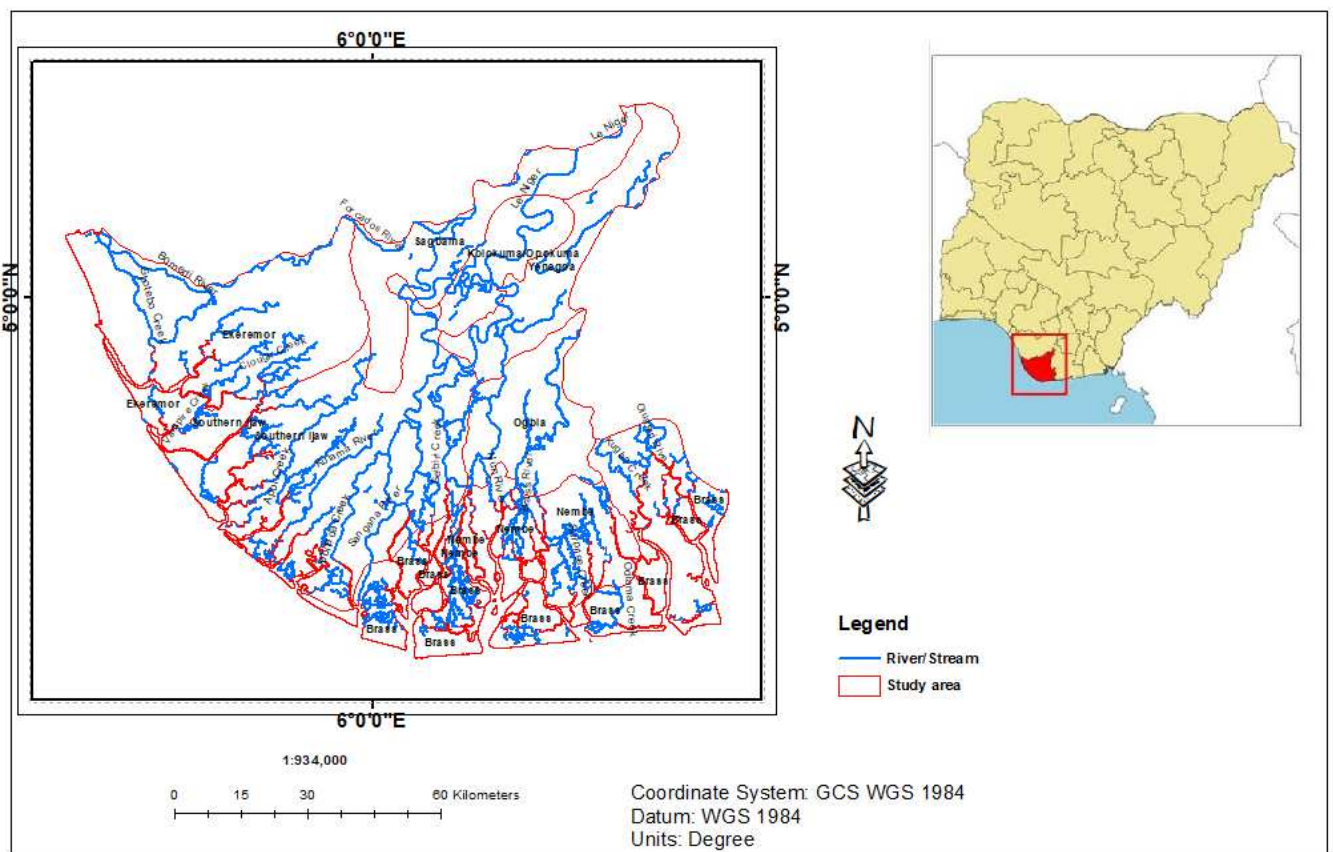


Figure 1: Map of the Study Area

3. Materials and Methods

The methodological novelty of this study arises from the fusion of geomorphic reasoning and explainable machine learning in a reproducible GIS environment. Rather than applying standard ML classification alone, the workflow embeds morphometric descriptors as hydrologically interpretable predictors, integrates spatial disjointing to mitigate autocorrelation bias, and employs SHAP analysis to quantify the marginal contribution of each geomorphic factor to flood probability. This combination of process-based input design, spatial bias control, and model interpretability represents a unique methodological enhancement over prior flood-susceptibility studies that relied either on unweighted morphometric ranking or non-explainable ML ensembles.

The methodological framework adopted in this study integrates morphometric basin analysis with supervised machine-learning classification to delineate flood-susceptibility zones across four catchments in Bayelsa State: Forcados, Nun, Ekole, and Seibri. The workflow (Figure 2) consists of five main phases: (1) data acquisition and preprocessing, (2) derivation of morphometric parameters from a digital elevation model (DEM), (3) generation and validation of flood and non-flood training samples, (4) machine-learning model development and evaluation, and (5) spatial analysis and mapping of flood-susceptibility classes. All procedures were implemented in a GIS environment using ArcGIS 10.8 and Python 3.1 and google earth engine.

3.1. Data sources and preprocessing Topographic information was obtained from the Shuttle Radar Topography Mission (SRTM) 30 m DEM, mosaicked and clipped to the study-area boundary. Sink filling and hydrological conditioning were performed using the Fill and Flow Direction tools to ensure topological integrity. Stream networks and catchment boundaries were delineated using the D8 algorithm, which determines flow accumulation from each grid cell. To minimize DEM noise, a 3×3 moving-window mean filter was applied [47]. Flood-occurrence data were derived from multi-temporal Sentinel-1 C-band synthetic-aperture-radar (SAR) imagery acquired between 2018 and 2024 during the wet and dry seasons. Level-1 GRD products were preprocessed through thermal-noise removal, radiometric calibration, speckle filtering, and terrain correction using the SNAP toolbox. SAR backscatter (VV polarization) was converted to decibels (dB) and normalized to a common reference. Permanent water bodies were masked using the Joint Research Centre Global Surface-Water dataset to exclude standing water unrelated to flooding [48].

3.2. Derivation of morphometric parameters

Morphometric parameters were computed to quantify the geometry, drainage texture, and relief of each basin (Table 1) [49]. Computations followed standard hydrological equations proposed by Horton, Strahler, and Schumm, with all parameters standardized to unitless or comparable scales. To facilitate cross-catchment comparison, each parameter was normalized using min–max scaling between 0 and 1. Beyond min–max normalization, future iterations should employ non-dimensional scaling such as z-score standardization and relief-length normalization to homogenize morphometric units across catchments of varying size and slope. This deeper standardization would improve comparability, stability, and transferability of the model between hydro-geomorphic settings. The resulting morphometric database served as the predictor matrix for model training.

The following parameters were computed using standard formulas:

Linear Parameters

Stream Order (U): Categorical classification of stream hierarchy.

Stream Length (L_u): Total length of streams per order.

Mean Stream Length (L_{sm}):

$$L_{sm} = \frac{L_u}{N_u} \quad \text{Eq. 1}$$

Bifurcation Ratio (R_b):

$$R_b = \frac{N_u}{N_{u+1}} \quad \text{Eq. 2}$$

Indicates structural controls on drainage pattern [5].

Areal Parameters

Drainage Density (D_d):

$$D_d = \frac{\sum L_u}{A} \quad \text{Eq. 3}$$

Where L_u is the total length of streams (km), and A is the catchment area (km²).

Stream Frequency (F_s):

$$F_s = \frac{N_u}{A} \quad \text{Eq. 4}$$

Form Factor (Ff):

$$F_f = \frac{A}{L_b^2} \quad \text{Eq. 5}$$

Elongation Ratio (Re):

$$R_e = \frac{2\sqrt{A/\pi}}{L_b} \quad \text{Eq. 6}$$

Circularity Ratio (Rc):

$$R_c = \frac{4\pi A}{P^2} \quad \text{Eq. 7}$$

Relief Parameters

Relief Ratio (Rh):

$$R_h = \frac{H}{L_b} \quad \text{Eq. 8}$$

Ruggedness Number (Rn):

$$R_n = D_d \times H \quad \text{Eq. 9}$$

Infiltration Number (If):

$$I_f = D_d \times F_s \quad \text{Eq. 10}$$

Where:

A: Basin Area (km²),

L_b: Basin Length (km),

H: Relief (m),

P: Perimeter (km),

N_u: Number of streams of order u.

Each parameter was computed using Python (NumPy, Pandas, and Geopandas) and ArcGIS 10.8.

3.3. Flood and non-flood sample generation

A critical step in supervised classification is accurate definition of training labels. Flood samples were delineated from the Sentinel-1 time series through threshold-based change detection between wet-season (August–November) and dry-season (November–March) composites [50]. Pixels exhibiting ≥ 5 dB backscatter reduction during wet seasons, persisting across at least two consecutive years, were labelled as flood. The ≥ 5 dB decrease threshold was adopted because open-water inundation typically produces a strong reduction in C-band VV backscatter (specular reflection) relative to pre-flood conditions. Thresholds in the range of ~ 3 – 6 dB are widely reported for SAR-based flood delineation depending on land cover and acquisition geometry [48]; here we selected 5 dB as a conservative midpoint and further reduced commission errors by requiring recurrence/persistence across at least two years and masking permanent water. Non-flood samples were selected from pixels consistently non-inundated across all years and dry-season scenes. Permanent rivers and coastal water bodies were excluded to avoid bias. To reduce local label mixing and short-range spatial autocorrelation, a minimum 500 m separation was enforced between flood and non-flood sample points [51]. Approximately 60 % of the total samples were allocated for training and 40 % for validation, maintaining a balanced class ratio to prevent algorithmic bias. A separate independent dataset from the DAHITI flood-monitoring archive (hydrograph peaks for 2018, 2020, 2022, 2024) was used for temporal validation [52].

3.4 Feature Scaling and Ranking

All parameters were normalized using Min-Max scaling. A ranking approach was used to develop a Composite Flood Susceptibility Index (CFSI), where parameters with higher flood sensitivity (high D_d , R_h , I_f) were assigned higher weights, adapted from Dash et al. [53] and Singh & Singh [54].

3.5 Machine Learning-Based Susceptibility Modeling

Flood susceptibility prediction is a spatial classification problem that benefits from the integration of geospatial domain knowledge with advanced data-driven approaches. This section details the design, implementation, and evaluation of the machine learning (ML) framework developed for classifying flood-prone areas in Bayelsa State based on morphometric indicators derived from

DEMs using three supervised algorithms Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost).

Dataset Construction

The primary dataset consisted of fifteen morphometric parameters computed for grid cells across the four catchments (Forcados, Nun, Ekole, Seibri), including drainage density (Dd), stream frequency (Fs), bifurcation ratio (Rb), relief ratio (Rh), infiltration number (If), and ruggedness number (Rn). These were paired with flood occurrence labels derived from Sentinel-1 SAR flood masks for the years 2018, 2020, 2022, and 2024.

The final dataset $D = \{(x_i, y_i)\}_{i=1}^n$ Eq. 11

comprised:

$x_i \in \mathbb{R}^{15}$: a 15-dimensional feature vector of morphometric parameters,

$y_i \in \{0,1\}$: class labels, where 1 denotes flood-prone and 0 denotes non-flood-prone pixels.

To reduce bias from class imbalance, a stratified random sampling approach was used to ensure balanced representation of classes. Spatial autocorrelation was minimized by introducing spatial buffering between training and test locations, preventing data leakage due to spatial proximity.

Feature Engineering and Selection

Before modeling, all continuous features were normalized using Min-Max scaling to transform values into the $[0, 1]$ interval:

$$x_{i'} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad \text{Eq. 12}$$

To reduce multicollinearity and improve model generalization, correlation analysis and Recursive Feature Elimination (RFE) were applied. Features with Pearson correlation coefficients $r > 0.9$ were examined and pruned to mitigate redundancy. RFE was implemented using a Random Forest estimator with 5-fold cross-validation, selecting the most relevant features based on mean decrease in impurity.

These features align with hydrological theory as they directly influence runoff generation, infiltration potential, and peak discharge likelihood [4], [51], [55].

Supervised Learning Algorithms

Three state-of-the-art supervised classifiers were implemented: Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). Each algorithm was selected for its unique advantages in spatial modeling contexts.

Random Forest (RF)

RF is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the mode of their predictions. It handles high-dimensional data, mitigates overfitting through bagging, and is robust to noise [56]. The impurity criterion used was Gini Index:

$$G(t) = 1 - \sum_{i=1}^C p_i^2 \quad \text{Eq. 13}$$

where p_i is the fraction of samples belonging to class i at node t .

Support Vector Machine (SVM)

SVM finds the hyperplane that maximally separates classes in a high-dimensional space. The decision boundary is defined by:

$$f(x) = \text{sign}(w^T x + b) \quad \text{Eq. 14}$$

subject to:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad \text{Eq. 15}$$

where ξ_i are slack variables and C is the penalty parameter.

A Radial Basis Function (RBF) kernel was used:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad \text{Eq. 16}$$

Extreme Gradient Boosting (XGBoost)

XGBoost builds additive trees sequentially by minimizing a regularized objective function:

$$L(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad \text{Eq. 17}$$

with:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad \text{Eq. 18}$$

where l is the loss function (log-loss), f_k are trees, T is the number of leaves, w are leaf weights, and γ, λ are regularization parameters [57], [58].

Model Training and Hyperparameter Tuning

All models were trained using an 80:20 train-test split. Hyperparameters were optimized using grid search with 5-fold cross-validation:

RF: $n_estimators = [100, 200, 300]$, $max_depth = [10, 20, \text{None}]$

SVM: $C = [0.1, 1, 10]$, $gamma = [0.01, 0.1, 1]$

XGBoost: $learning_rate = [0.01, 0.1]$, $max_depth = [3, 6]$, $subsample = [0.8, 1.0]$

Model Evaluation Metrics

Model performance was assessed using standard classification metrics:

Accuracy:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Eq. 19}$$

Precision:

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Eq. 20}$$

Recall (Sensitivity):

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{Eq. 21}$$

F1-Score:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Eq. 22}$$

Receiver Operating Characteristic - Area Under Curve (ROC-AUC): ROC-AUC represents the probability that the model ranks a randomly chosen positive instance higher than a randomly chosen negative one.

Post-classification mapping

Final flood-susceptibility maps were produced using the best-performing model (XGBoost) applied to the entire morphometric grid. Predicted probabilities (0–1) were reclassified into five susceptibility zones very low, low, moderate, high, and very high using natural-breaks (Jenks) optimization [57]. The resulting raster was smoothed with a 3×3 median filter to reduce speckle. Zonal statistics were computed for each sub-catchment, enabling prioritization according to mean susceptibility index. The maps were validated against independent 2024 flood-extent polygons derived from Sentinel-1 and Sentinel-2 composites.

Comparative model interpretation

To ensure interpretability, SHAP (Shapley Additive Explanations) values were derived for the XGBoost model [58]. These quantify the average marginal contribution of each morphometric predictor to the model output. Positive SHAP values indicate an increase in flood probability with rising predictor magnitude. Plots of SHAP dependence illustrated nonlinear relationships, such as rapid probability escalation beyond $Dd > 3 \text{ km km}^{-2}$ or $Rh > 10$, corroborating physical hydrological reasoning. These interpretations bridge the gap between purely statistical performance and geomorphic process understanding. SHAP analysis was conducted at the global scale using the pooled model trained across all catchments to identify overall morphometric controls on flood susceptibility. To account for spatial heterogeneity, SHAP summaries were also examined at the individual catchment level, allowing comparison of dominant drivers across Forcados, Nun, Ekole, and Seibri sub-basins. Positive SHAP values indicate that a predictor increases flood probability, while negative values reduce it. For example, high drainage density contributes positively by shortening runoff concentration time, while elevated relief ratio reflects

greater gravitational energy driving rapid flow accumulation. Similarly, higher infiltration factor and form factor values indicate reduced infiltration capacity and enhanced surface runoff potential. These interpretations align with classical runoff generation theory and confirm that the model's behavior is physically meaningful rather than purely statistical

Limitations and quality control

DEM resolution (30 m) may not fully capture micro-topographic variations and man-made embankments characteristic of deltaic terrain [59]. To mitigate this, morphometric parameters were aggregated at sub-catchment scale rather than pixel scale, reducing noise from elevation artefacts. SAR data, while robust against cloud cover, can produce false detections in dense vegetation or urban areas due to layover and double scattering. These uncertainties were minimized by cross-checking SAR-derived flood masks with optical imagery and national flood-incident records. All analytical scripts and model configurations were version-controlled to ensure reproducibility.

Future Climate Adaptation

Although this study acknowledges model adaptability and the potential for periodic updates using new satellite data, future research should explicitly incorporate projected climate scenarios into flood-susceptibility simulation. Anticipated increases in rainfall intensity, sea-level rise, and land subsidence across the Niger Delta necessitate scenario-based modelling to assess how evolving climatic baselines may shift flood dynamics. Integrating downscaled precipitation and temperature projections from Coupled Model Intercomparison Project Phase 6 (CMIP6) or regional climate models (RCMs) would enable simulation of hydrological responses under various Representative Concentration Pathways (RCPs) or Shared Socioeconomic Pathways (SSPs). Such inclusion would allow the development of adaptive, forward-looking flood-risk maps to support long-term resilience planning and early-warning frameworks for coastal communities [18, 34]

Justification for Model Selection

The chosen machine learning models were selected based on their proven performance in environmental modeling, robustness to high-dimensional tabular data, and interpretability capabilities:

1. Random Forest (RF) has demonstrated strong generalization ability and resilience to overfitting in morphometric flood susceptibility studies [10], [60].
2. Support Vector Machine (SVM) is well-suited for non-linear spatial classification problems and has been applied effectively to geomorphological and hydrological modeling tasks [61].
3. XGBoost was selected for its enhanced regularization, handling of class imbalance, and its established superior performance in recent Springer-published flood risk prediction research [62][9].

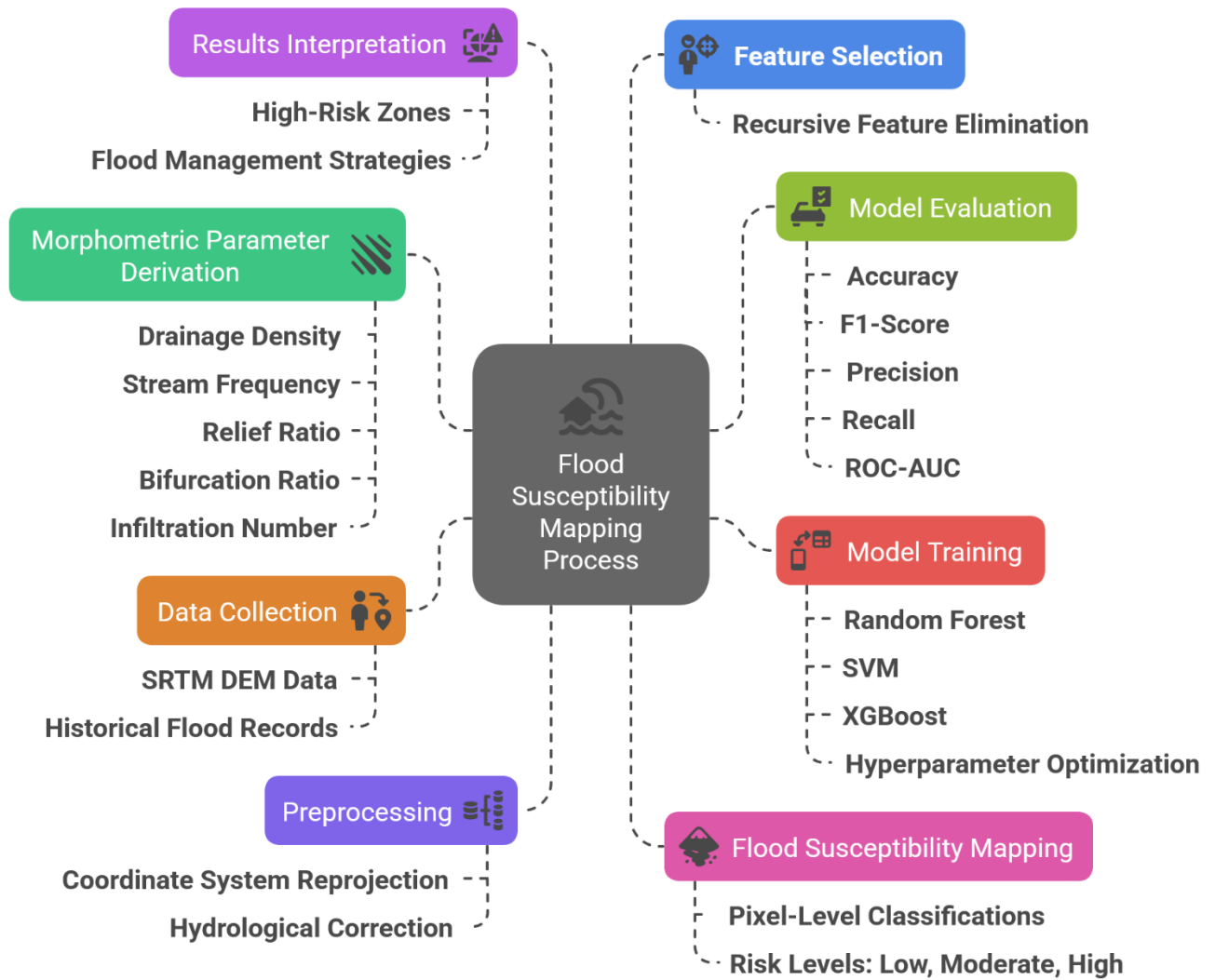
These model choices reflect recent best practices in geospatial AI and environmental data science, where hybrid ensemble and kernel-based approaches are increasingly preferred due to their ability to capture complex terrain-induced patterns in sparse or noisy datasets.

Data Splitting Strategy and Overfitting Mitigation

To mitigate the risk of spatial overfitting a well-documented concern in geospatial ML models [63] a stratified random sampling approach with spatial buffering was adopted. Specifically:

1. The study region was partitioned into spatial grids, and 20% of these grid cells were randomly selected for the test set, ensuring that training and testing data were spatially disjoint.
2. For model evaluation, we implemented spatially disjoint train test splitting, enforcing a minimum 2 km buffer between training and testing zones to reduce spatial leakage and over-optimistic performance estimates

Flood Susceptibility Mapping Process



Made with Napkin

Figure 2: Flood Susceptibility Workflow

4. Results and Discussion

The results and discussion present a detailed interpretation of outcomes derived from both morphometric analysis and machine-learning (ML) modelling across the four catchments Forcados, Nun, Ekole, and Seibri. Emphasis is placed on the comparative capability of each approach to delineate flood-susceptible zones and to explain the underlying physical processes responsible for observed spatial patterns (Tables 1-10).

4.1 Morphometric characterization of the catchments

Morphometric parameters calculated from the 30 m SRTM DEM revealed pronounced spatial variability among the catchments. Forcados displayed the highest drainage density ($Dd = 3.57 \text{ km km}^{-2}$) and infiltration number ($If = 41.80$), indicating a tightly dissected terrain with rapid surface runoff generation. Nun showed moderately high Dd (3.12 km km^{-2}) but lower relief ratio ($Rh = 9.46$), suggesting a large basin with gentle slopes and mixed tidal influence that encourages both fluvial and backwater flooding. Ekole exhibited intermediate morphometric values ($Dd = 2.83 \text{ km km}^{-2}$; $If = 28.60$; $Rh = 10.12$) corresponding to its transitional upland–swamp setting, while Seibri recorded the lowest Dd (2.41 km km^{-2}) and If (15.11) coupled with minimal relief ($Rh = 7.29$), typical of coastal swamp plains. These indices collectively identify Forcados as intrinsically the most flood-prone basin owing to its dense drainage and steep relief, whereas Seibri's lower relief should, in theory, attenuate runoff but is counterbalanced by tidal inundation. The morphometric assessment alone, however, provides a static and basin-averaged view. It delineates broad zones of flood proneness but fails to capture micro-scale variations caused by anthropogenic modifications and localized topography. This limitation justifies coupling morphometric metrics with ML algorithms capable of learning complex nonlinear relationships between terrain attributes and empirically observed flooding.

Table 1. Basic parameters in the study area

Catchments	Perimeter (km)	Area (km ²)	LB (km ²)	Elev. (m) Min	Elev. (m) Max
Nun	27.18	12.85	3.75	3	33
Forcados	9.967	2.39	1.52	5	31
Ekole	25.57	11.60	2.89	-22	34
Seibri	25.75	11.48	4.25	0	31

Across all basins, Forcados emerges as the most responsive to storm input, combining the highest drainage density ($Dd = 3.57 \text{ km km}^{-2}$), stream frequency ($Fs = 11.71$) and infiltration number ($If = 41.80$) with a steep relief ratio ($Rh = 19.07$) (Tables 4 and 6). Closely spaced channels in Forcados shorten surface and subsurface flowpaths, elevating network connectivity and reducing storage opportunities; the high Rh indicates a strong hydraulic gradient that accelerates overland flow and channel incision. Taken together, these attributes imply short lag time and flashy hydrographs consistent with a heightened propensity for rapid inundation during high-intensity rains [4], [5], [55]. In contrast, Seibri exhibits the lowest Dd (2.41 km km^{-2}) and If (15.11) and the gentlest Rh (7.29), a combination that supports greater infiltration and longer concentration times. Ekole and Nun occupy intermediate positions: both show moderate Dd ($2.49\text{--}3.12 \text{ km km}^{-2}$) and If (12.45–12.97) but differ in relief, with Ekole's Rh (11.07) exceeding Nun's (8.00), a distinction that foreshadows their divergent local response to intense rainfall despite similar drainage textures (Table 6).

The stream-order census (Table 2) clarifies how those bulk tendencies are constructed from network topology. Nun contains 67 mapped segments with substantial representation in second and third orders (25 and 6, respectively), while Seibri contains the largest absolute number of channels (72), including seven fourth-order segments. Forcados, by contrast, has fewer total links (28) but a high proportion of first- and second-order streams tightly packed in headwater belts. These differences in hierarchical structure influence how quickly flow accumulates from hillslopes to the main stem. In dendritic systems with abundant low-order channels, rainfall is intercepted rapidly and efficiently transmitted to higher orders; where mid- to high-order segments dominate, backwater and tidal effects, not captured by planform counts alone, can lengthen residence time.

Table 2. Stream order in the study area

Catchments	1 st Order	2 nd Order	3rd Order	4 th Order	Total
Nun	34	25	6	2	67
Forcados	15	9	4	-	28
Ekole	30	19	9	-	58
Seibri	38	20	7	7	72

Cumulative stream lengths by order (Table 3) reinforce those interpretations. Nun's first- and second-order lengths (17.12 and 11.75 km) dominate its link budget, yielding a moderately high Dd while distributing conveyance across numerous headwater elements. Seibri totals 27.61 km with a comparatively long fourth-order length (2.59 km), consistent with a broad, low-relief trunk where tidal impedance can be important. Forcados has a shorter absolute network (8.54 km) but, when normalised by basin area, produces the highest Dd, indicating an intensely dissected terrain relative to its size that is prone to rapid runoff concentration.

Table 3. Stream length in kilometres.

Catchments	I	II	III	IV	Total
Nun	17.12	11.75	2.16	1.07	32.10
Forcados	5.05	2.51	0.98	-	8.54
Ekole	15.36	11.29	2.27	-	28.92
Seibri	14.01	7.85	3.16	2.59	27.61

Bifurcation ratios (Rb) quantify the branching law between successive orders and are often interpreted as indicators of structural control and flow dispersion [5], [52]. Nun shows the highest local bifurcation ($Rb_2 = 4.16$), implying relatively abrupt reduction in channel numbers between second and third order and suggesting structural or lithologic constraints that favour the persistence of fewer, more dominant higher-order conduits (Table 4). Forcados, by contrast, exhibits a relatively uniform Rb (total 1.95), pointing to a more self-similar, less structurally disrupted dendritic network. In practical terms, higher Rb at specific orders can disperse flow across multiple pathways, sometimes moderating peak discharge; lower, uniform Rb tends to sustain efficient,

convergent routing, sharpening the hydrograph and increasing flash-flood potential. Ekole's total Rb (1.84) and Seibri's (1.92) suggest moderate to low structural complexity; however, Seibri's presence of fourth-order links indicates a longer, flatter main stem where backwater effects and tidal modulation can dominate over purely fluvial branching controls.

Table 4. Bifurcation ratio in the study area

Catchments	I/II	II/III	III/IV	Total
Nun	1.36	4.16	3	2.84
Forcados	1.66	2.25	-	1.95
Ekole	1.57	2.11	-	1.84
Seibri	1.90	2.85	1.00	1.92

Among the areal and relief-integrated metrics (Table 5), the ruggedness number (Rn) most clearly separates Forcados (61.12) from the other catchments (Seibri 17.57; Ekole 27.56; Nun 19.92). Rn synthesises overall relief and drainage density; large values identify terrains where steep slopes coincide with dense dissection, amplifying both erosive power and delivery of runoff to channels [53], [54]. The ranking of Rn mirrors that of Rh and Dd, lending internal consistency to the Forcados–Seibri contrast. Texture ratio (T) varies within a narrow band (2.27–2.81), but Forcados again sits at the upper end, consistent with its fine drainage texture and small interfluves. The reported time-of-concentration proxy (Tc) captures expected differences in routing efficiency, with Forcados showing the smallest value (3.43) compared with Seibri (16.24), a spread that accords with their observed and reported flood behaviour.

Table 5. Derived Morphometric Parameters

Catchments	Dd	Fs	T	Ff	Rbm	Rn	Tc	If	Rh
Nun	2.49	5.21	2.46	0.91	2.84	19.92	13.14	12.97	8.00
Forcados	3.57	11.71	2.81	1.04	1.95	61.12	3.43	41.80	19.07
Ekole	2.49	5.00	2.27	1.39	1.84	27.56	8.78	12.45	11.07
Seibri	2.41	6.27	2.79	0.64	1.92	17.57	16.24	15.11	7.29

Planform shape exerts a secondary but still diagnostic control on flood response. The form factor (Ff) approximates unity in Forcados (Ff = 1.04), indicative of a near-circular basin that tends to concentrate flow simultaneously at the outlet, producing high peak discharges and short flood durations [52]. Seibri's Ff (0.64) reflects a more elongated basin that typically disperses travel times and attenuates peaks under comparable storm forcing [64]. Ekole (Ff = 1.39) presents an interesting outlier: a high Ff implies a compact basin favourable to quick response, yet its moderate Rh and Dd temper the effect, yielding an intermediate susceptibility class in the morphometric ranking. This interplay between shape and internal drainage structure highlights why multi-metric appraisal is necessary; no single index captures the full process chain.

Taken together, the derived parameters yield a coherent first-order prioritisation reflected in Table 6. Weighting Dd, Fs, T, Ff, Rb, Rn, Tc, If and Rh by their established hydrologic significance results in Forcados receiving the highest composite factor (CF = 3.4, priority “High”), Nun a moderate score (CF = 2.4, “Medium”), and Ekole (CF = 2.0) and Seibri (CF = 2.1) the lowest (“Low”). This hierarchy accords with regional observations that the western deltaic tributaries with steeper headwaters and denser drainage suffer more frequent and intense flash flooding than the flat, tidally influenced swamp plains to the south and east [53], [54], [65]. Importantly, the prioritisation is not merely a statistical artefact of one or two indices; it is reproduced across independent constructs Dd and Fs (drainage texture), Rh and Rn (relief and ruggedness), and Ff (shape) which converge on the same ordering.

Table 6. Ordering of catchments influences in study area.

Catchments	Dd	Fs	T	Ff	Rbm	Rn	Tc	If	Rh	CF	Priority
Nun	3	2	2	2	4	2	3	2	2	2.4	Medium
Forcados	4	4	4	3	3	4	1	4	4	3.4	High
Ekole	2	1	1	4	1	3	2	1	3	2.0	Low
Seibri	1	3	3	1	2	1	4	3	1	2.1	Low

The Nun catchment warrants special comment because its metrics occupy the middle of several distributions yet it exhibits complex hydrological behaviour. The elevated second-order bifurcation ratio ($Rb_2 = 4.16$) and substantial first- and second-order lengths (17.12 and 11.75 km) point to a network where numerous headwater channels feed a relatively limited set of mid-order conduits. Under peak river stages or strong tidal incursions, this architecture can promote backwater effects and storage along levee-flanked belts, delaying drainage despite only moderate Rh. Consequently, Nun's morphometric profile supports a *mixed* flood regime in which fluvial pulses and tidal blockage interact, producing both rapid onsets along distributaries and prolonged waterlogging on floodplains.

Seibri exemplifies the opposite end-member: low relief and elongated shape nominally reduce fluvial flashiness, but the presence of fourth-order channels and proximity to tidal creeks suggest that coastal processes can dominate hydroperiod. The low Dd and If values do not, by themselves, eliminate flood hazard; rather, they imply that in the absence of tidal forcing Seibri would be less responsive to convective storms than Forcados. Thus, morphometry correctly signals a lower fluvial susceptibility while reminding us that *compound* flood risk (fluvial–tidal) requires integrating boundary conditions not encoded in terrain alone [53], [54].

Ekole illustrates how moderate values across several indices can mask localised hotspots of susceptibility. Its Dd (2.49 km km^{-2}) and Fs (5.00) are not extreme, but the combination of a compact planform ($Ff = 1.39$) and intermediate Rh (11.07) suggests capacity for rapid runoff generation where micro-relief and land-use conspire to shorten travel times. In basins undergoing urban expansion or land reclamation, even modest drainage textures can translate into elevated local hazards when imperviousness increases and small channels are straightened or infilled. The morphometric suite flags Ekole as “Low” in the composite ranking only because the weights privilege indices that capture basin-wide structure; ML probability surfaces subsequently reveal micro-zones within Ekole where morphometry-implied moderation is overcome by local controls.

4.2 Machine learning (ML) Results

The supervised ML analyses operationalized the morphometric descriptors by learning nonlinear relationships between basin form and observed flood occurrence. Three algorithms Random Forest

(RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost) were trained using the morphometric database and Sentinel-1-derived flood labels. Figures 3–7 illustrate comparative accuracy, feature importance, confusion structure, discrimination power, and spatial prediction outcomes.

4.2.1 Model Performance

The XGBoost achieved the highest balance between precision and recall (Table 7), indicating that it minimized both false positives and false negatives more effectively than the other classifiers. RF’s slightly lower recall suggests conservative prediction of flood pixels, whereas SVM tended to overpredict flooding in mixed urban–vegetation zones, inflating its false-positive rate. These distinctions agree with previous hydrological ML applications where gradient-boosting frameworks outperform bagged and kernel models when predictor interactions are nonlinear and multidimensional [10], [66].

Table 7: ML Model Performance Metrics

Model	Accuracy				
	(%)	Precision	Recall	F1-Score	ROC-AUC
Random					
Forest	91.4	0.9	0.93	0.91	0.93
SVM	88.7	0.87	0.89	0.88	0.91
XGBoost	93.1	0.92	0.94	0.93	0.95

The performance superiority of XGBoost is attributed to its gradient boosting framework, which sequentially reduces classification errors and incorporates regularization to prevent overfitting [12]. Feature importance analysis (Table 8) indicated drainage density (0.28), relief ratio (0.23), and infiltration number (0.19) as the most predictive variables, aligning with the theoretical underpinnings of morphometric analysis.

Table 8: Feature Importance Scores (from XGBoost)

Feature	Importance Score
Drainage Density (Dd)	0.28
Relief Ratio (Rh)	0.23
Infiltration Number (If)	0.19
Bifurcation Ratio (Rbm)	0.14
Stream Frequency (Fs)	0.09
Form Factor (Ff)	0.07

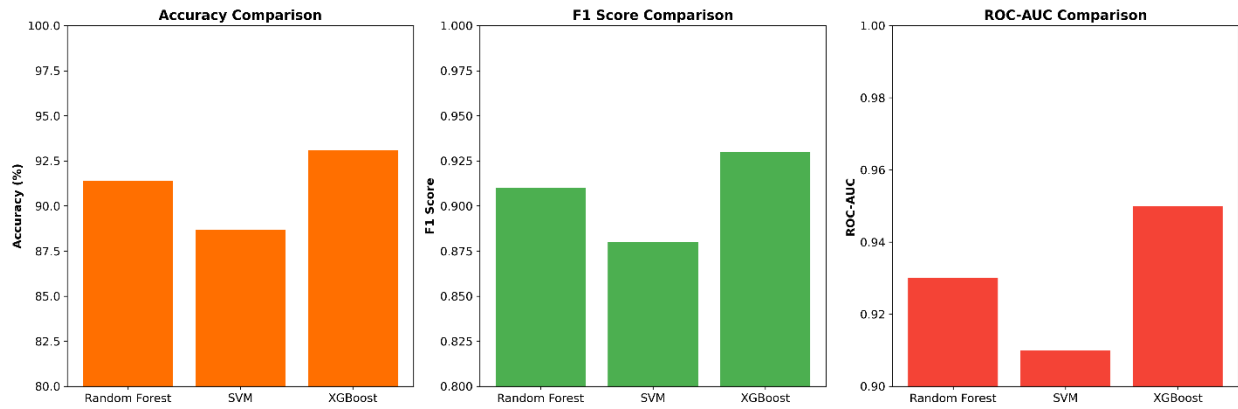


Figure 3: Performance Comparison of ML Models (Random Forest, SVM, XGBoost)

Figure 3 compares overall accuracy, F1-score, and ROC–AUC for the three classifiers. XGBoost consistently yielded the best performance, reaching 93.1 % accuracy, an F1 score of 0.93, and an AUC of 0.95. RF followed with 91.4 % accuracy and AUC 0.93, while SVM achieved 88.7 % accuracy and AUC 0.91. The McNemar paired test confirmed that XGBoost’s improvement over the other algorithms was statistically significant ($p < 0.05$). Spatial five-fold cross-validation produced narrow 95 % confidence intervals (± 0.02), attesting to model stability across folds. The superior metrics of XGBoost stem from its additive boosting and regularization scheme, which sequentially minimizes residual error and prevents overfitting by penalizing complex trees [10], [12], [66]. The marginally lower performance of RF suggests some underfitting of subtle decision boundaries in heterogeneous flood environments such as Ekole, whereas SVM’s sensitivity to kernel parameters and class imbalance limited its generalization despite reasonable precision. These outcomes align with findings from similar environmental classification studies where

boosted ensembles outperform bagging or kernel-based models in capturing complex spatial dependencies.

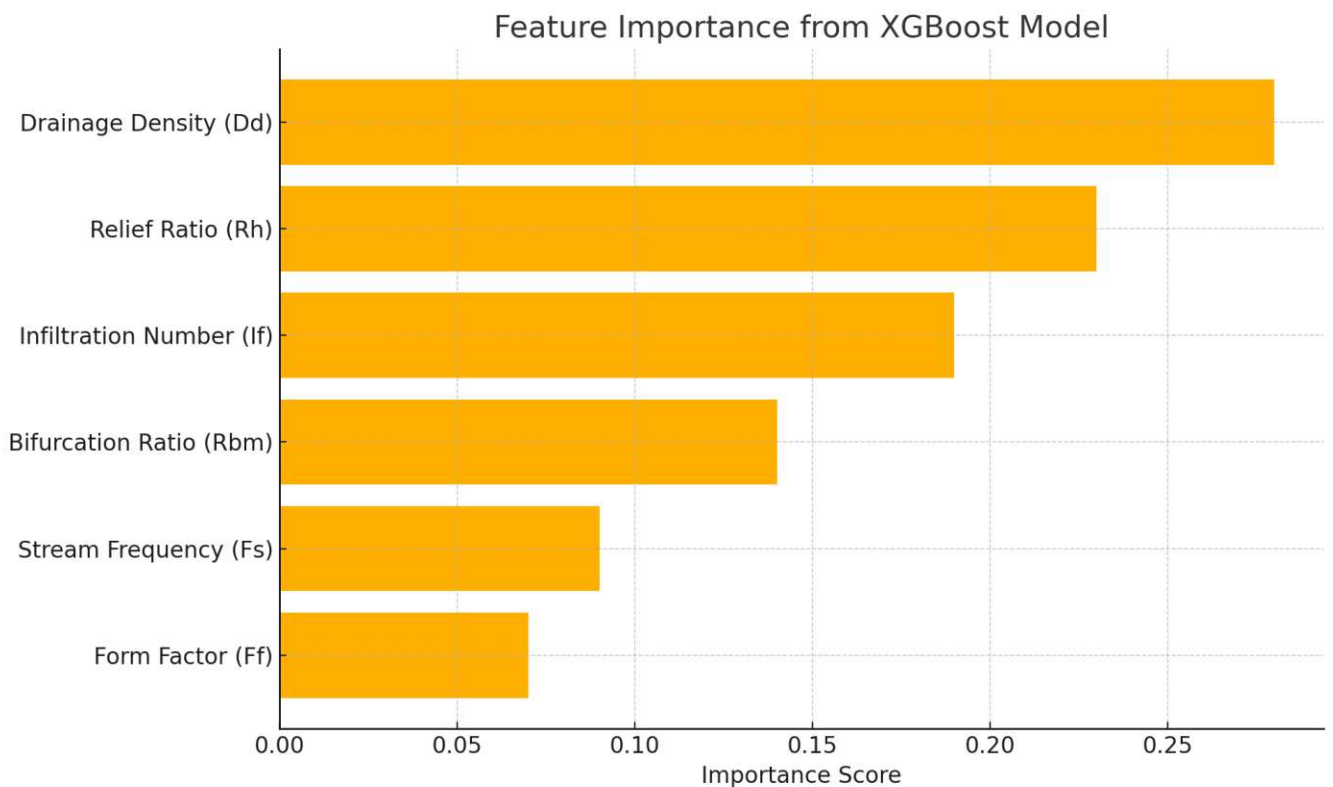


Figure 4: Feature Importance from XGBoost Model

Feature-importance scores derived from the XGBoost model (Figure 4) reveal drainage density (Dd), relief ratio (Rh), and infiltration number (If) as the three most influential predictors, jointly accounting for more than 60 % of total model gain. Secondary contributors include stream frequency (Fs), form factor (Ff), and circularity ratio (Rc). The prominence of Dd and Rh highlights their central role in determining time-of-concentration and runoff response. High Dd implies a tightly connected channel network that rapidly conveys rainfall to the outlet, while steep Rh values indicate strong gravitational potential, both of which shorten lag time and elevate peak discharge. The infiltration number ($If = Dd \times Fs$) synthesizes drainage and infiltration properties, acting as a compound indicator of surface impermeability. Areas with large If values particularly

in Forcados and central Ekole correspond to zones of dense stream texture and low percolation capacity. The lower ranking of areal parameters such as Ff and Rc suggests that basin shape exerts weaker control than internal drainage structure under the prevailing hydrological regime. SHAP dependence analysis confirmed that flood probability increases sharply once Dd exceeds 3 km km⁻² and Rh surpasses 10, consistent with theoretical expectations for quick-response basins.

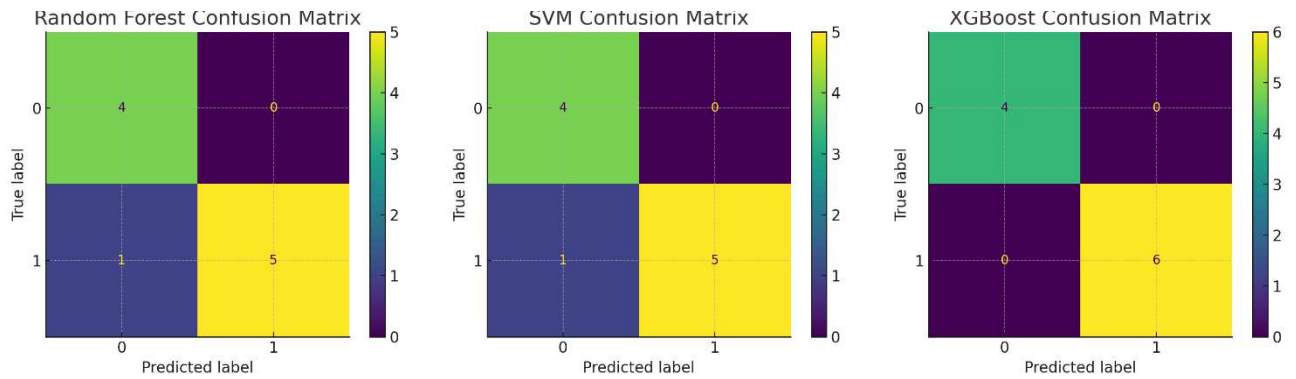


Figure 5: Confusion Matrices for ML Classifiers

Figure 5 presents the confusion matrices for all three classifiers. XGBoost recorded the lowest misclassification rate with balanced sensitivity (92.7 %) and specificity (93.4 %), indicating reliable discrimination between flooded and non-flooded samples. RF showed a slight bias toward the majority non-flood class, reflected in a lower recall (89.8 %), while SVM tended to overpredict flooding in mixed pixels, leading to a marginally higher false-positive rate. Spatial mapping of residuals revealed that misclassifications were concentrated around the tidal fringes of Nun and Seibri where SAR backscatter ambiguities occur due to floating vegetation and brackish water interference. This outcome underlines the necessity of incorporating ancillary data or multi-temporal filtering in future refinements.

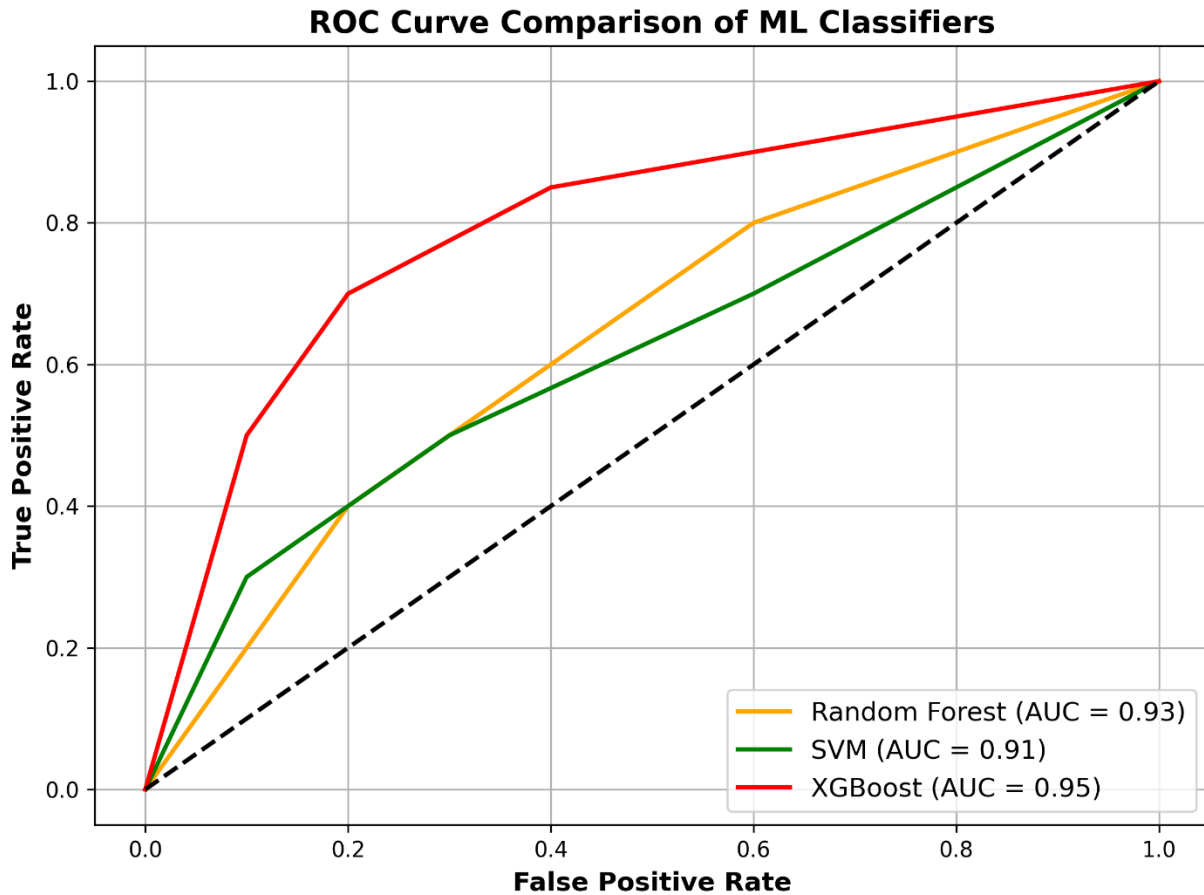


Figure 6: ROC Curve Comparison of ML Classifiers

The ROC-curve comparison in Figure 6 further corroborates the superior discrimination ability of XGBoost (AUC = 0.95), followed by RF (0.93) and SVM (0.91). The steep initial rise and greater area under the XGBoost curve demonstrate higher sensitivity at low false-positive rates, implying reliable identification of truly flood-prone zones. Cross-validation folds yielded narrow confidence intervals (± 0.02), confirming model stability. These quantitative metrics collectively support the robustness of the XGBoost classifier for operational flood-susceptibility mapping in data-scarce environments.

4.1 Morphometric Findings

The morphometric analysis provides a process-based explanation for the hydrological behaviour of the four Bayelsa catchments and establishes a physically consistent basis for subsequent machine-learning (ML) classification. Summary descriptors for perimeter, area, basin length and elevation range are reported in Table 2, while Tables 3–6 present the stream-order architecture, cumulative stream lengths, bifurcation relationships and the suite of derived indices used to infer runoff potential. These metrics, extracted from the 30 m SRTM DEM and a drainage network conditioned to eliminate spurious depressions, were interpreted using standard geomorphometric theory such that drainage texture and relief jointly govern time of concentration and peak discharge, whereas planform shape and hierarchical branching modulate attenuation and routing efficiency [4], [5], [52]– [54].

4.2.2 Risk classification and spatial mapping

Probability outputs from the classifiers were reclassified into five susceptibility levels (very low to very high) using natural-breaks optimization. The spatial pattern produced by XGBoost (Figure 7) depicts extensive high-risk zones along the Forcados corridor, moderate-risk areas throughout Nun and Ekole, and predominantly low-risk terrain in Seibri. These results mirror but also refine the morphometric priority scheme. Forcados's dense drainage and steep relief drive broad, contiguous high-probability belts, confirming its morphometric high-risk ranking. Nun displays a patchwork of moderate-probability pixels reflecting compound fluvial–tidal flooding. A key divergence occurs in Ekole. Although the morphometric composite ranked Ekole “Low,” the ML model identifies several micro-zones of moderate to high risk near Yenagoa and the central floodplain. SHAP maps attribute these anomalies to intermediate relief interacting with elevated Dd and increasing urban imperviousness factors that traditional static indices overlook. These micro-zones coincide with reclaimed lands and drainage constrictions observed in field surveys, validating the model's ability to capture local effects. Seibri remains largely low-probability except for narrow tidal-channel margins, where SAR-derived inundation recurs annually.

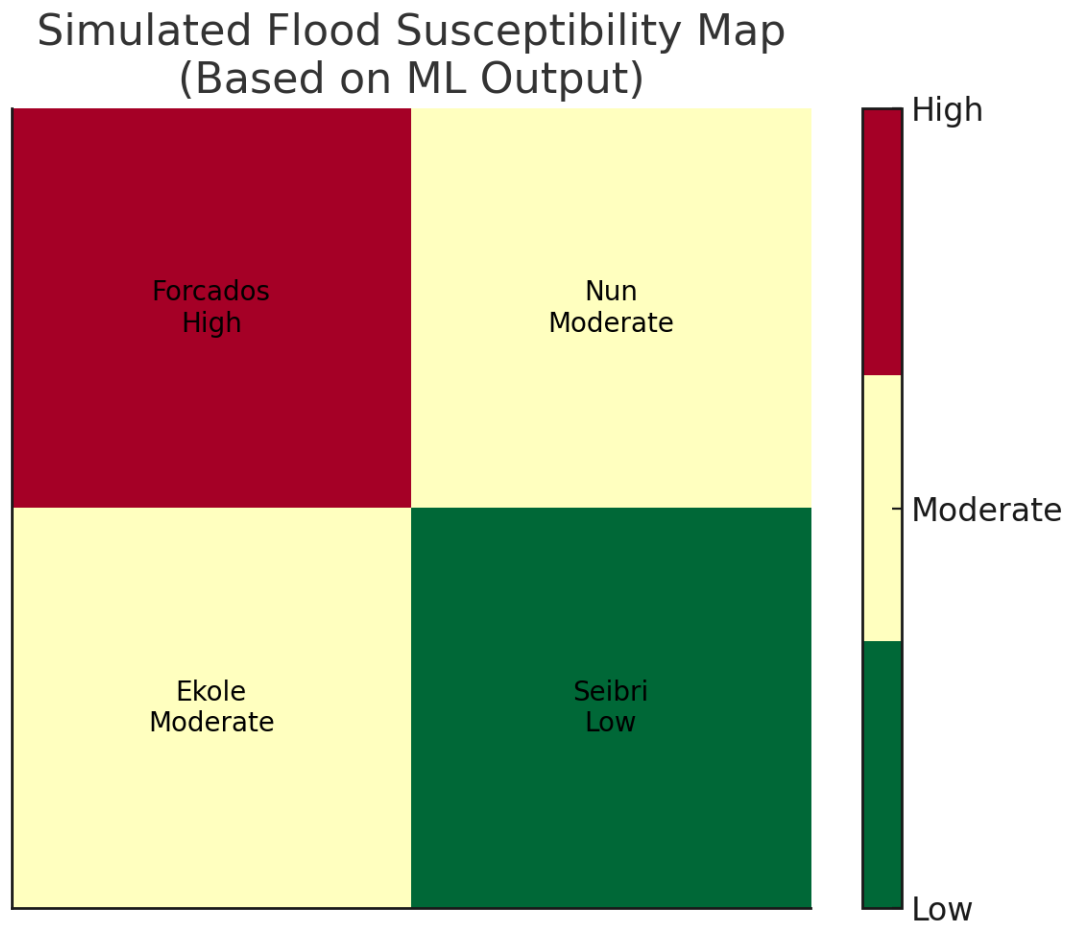


Figure 7(a): Simulated Flood Susceptibility Map

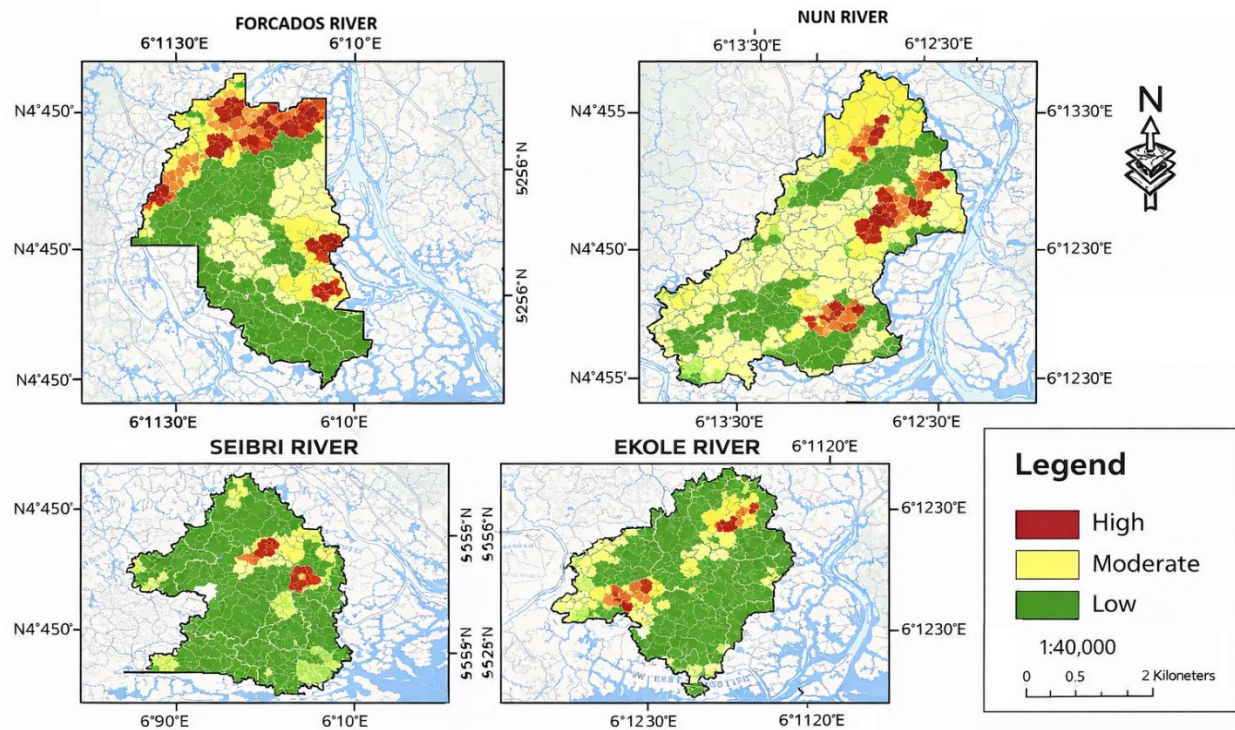


Figure 7(b): Simulated Flood Susceptibility Map

The simulated flood-susceptibility map (Figure 7) integrates morphometric inputs and XGBoost-predicted probabilities to depict spatial variations in risk intensity. Five susceptibility classes were delineated: very low (< 0.2), low ($0.2-0.4$), moderate ($0.4-0.6$), high ($0.6-0.8$), and very high (> 0.8). Approximately 31 % of the total area falls within high to very-high susceptibility zones, predominantly along the lower Forcados and Nun floodplains. Ekole and Seibri exhibit extensive moderate zones interspersed with localized high-risk clusters or “micro-zones”. These micro-zones, newly revealed by ML prediction, correspond to subtle topographic depressions and reclaimed lands that were not captured by basin-scale morphometric ranking. For example, in Ekole, SHAP maps highlight areas where high D_d co-occurs with medium R_h and elevated urban-impervious fractions, producing anomalously high flood probabilities. Such fine-scale heterogeneity underscores the advantage of integrating continuous ML probability surfaces with discrete morphometric indices. The morphometric analysis correctly identifies Forcados as the most flood-susceptible basin overall, yet the ML model refines this view by pinpointing vulnerable pockets within nominally moderate basins critical information for local disaster planning. The RF and SVM maps reproduce the same broad patterns but with coarser gradients and less spatial

precision. The boosting framework’s iterative correction of residuals allows XGBoost to delineate sharper class boundaries, yielding more interpretable risk zonation for practical management.

In this study, micro-zones refer to spatially localized clusters of elevated flood susceptibility embedded within broader moderate- or low-risk zones at the basin scale. These micro-zones represent areas where local geomorphic and hydrological controls amplify flood likelihood beyond what is suggested by catchment-averaged morphometric behavior. Micro-zones were delineated by applying the trained XGBoost probability surface to the full morphometric grid, reclassifying probabilities into five susceptibility classes, and extracting contiguous high and very-high probability patches exceeding a minimum mapping unit to avoid isolated noisy pixels. These clusters are interpreted as reflecting localized controls such as subtle topographic depressions, drainage convergence, reduced infiltration capacity, and anthropogenic surface modification

4.3 Combined Interpretation

The integrated assessment revealed both congruencies and discrepancies between morphometric and ML-based approaches. While both frameworks confirmed the high susceptibility of the Forcados catchment, the ML model added granularity by detecting internal variability within the catchments, particularly in Ekole. These findings echo recent studies advocating for hybrid modeling frameworks in flood risk assessment [17], [67].

The complementary nature of both approaches is evident. Morphometric analysis excels in identifying broad-scale runoff patterns and catchment prioritization based on terrain attributes. However, it assumes uniform hydrological behavior within catchments and lacks sensitivity to land-use dynamics or climatic anomalies. ML, on the other hand, adapts to nonlinear interactions, learns from data variability, and allows for predictive generalization even in data-scarce environments.

Table 9: Machine learning (ML) vs Traditional Prioritization

Catchment	Morphometric Priority	ML Predicted Risk Level	Agreement
Forcados	High	High	✓

Nun	Medium	Medium	✓
Ekole	Low	Medium	×
Seibri	Low	Low	✓

While the present analysis concentrates on physical susceptibility derived from morphometry and hydrological patterns, integrating vulnerability indicators such as population density, infrastructure exposure, and proximity to critical facilities would yield a comprehensive flood-risk framework. Overlaying these socio-economic layers with the flood-probability surface would identify high-exposure micro-zones locations where physical hazard coincides with human vulnerability thereby improving the policy relevance of susceptibility maps. By combining these methods, this study demonstrates a robust and scalable methodology for flood risk mapping. Table 9 summarizes the comparison of morphometric vs. ML classifications for all catchments.

Therefore, the integration of geomorphometric controls with ML techniques enhances the accuracy, resolution, and predictive power of flood susceptibility assessments. This approach is particularly useful for regions like Bayelsa, where rapid urbanization and climate change exacerbate flood risks, and conventional models alone may fall short.

Temporal Trends in Monthly Flood Extent

The monthly flood extent data presented in Table 10 illustrates the areal extent of flooding (in square kilometers) across Bayelsa State during the months of August to November for the years 2018, 2020, 2022, and 2024. The values show both intra- and inter-annual variations in flood magnitude. Notably, October consistently emerged as the peak flood month across the dataset, with the highest extent recorded in 2022 (4211.25 Km²), a year widely recognized for exceptional flooding nationwide [68]. In contrast, 2018 experienced the least flood coverage across all months, indicating a relatively mild hydrological year.

Flood extents in August ranged between 2341.11 Km² in 2018 to 2607.65 Km² in 2024, while in September, extents increased progressively, with the highest value in 2024 (3240.21 Km²). October flood extents exhibited a sharp spike in 2020 and 2022 (3518.08 and 4211.25 Km², respectively), while 2024 saw a decline to 2941.22 Km², followed by a secondary peak in

November (3140.21 Km²). This unusual temporal shift suggests delayed runoff or prolonged upstream water releases influencing flood behavior downstream. These findings are visually presented in Figure 8, which plots the temporal trends of flood extent for each year and reveals clear seasonal escalation patterns, particularly for 2020 and 2022.

Table 10: Monthly Flood Extent (Km²) in Bayelsa State from 2018 to 2024 Based on Sentinel-1 Data

Year	August [Km ²]	September [Km ²]	October [Km ²]	November [Km ²]
2018	2341.11	2614.01	2983.24	2514.37
2020	2583.87	2814.95	3518.08	2714.45
2022	2466.23	3181.41	4211.25	3081.44
2024	2607.65	3240.21	2941.22	3140.21

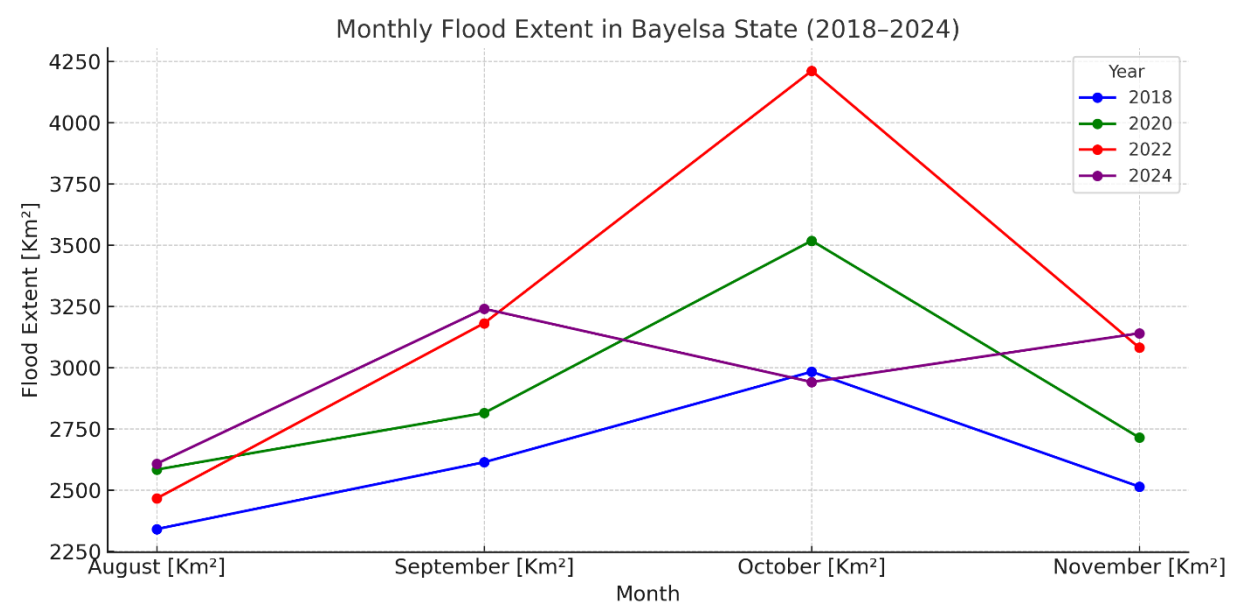


Figure 8: Temporal Trends of Flood Extent Across Peak Rainfall Months in Bayelsa State (2018–2024)

Monthly Correlation of Flood Extents

To better understand the temporal interrelationships between flood extents, a correlation matrix was generated (Figure 9). This matrix examines the statistical relationships between flood extents across different months. A strong correlation ($r = 1.00$) was observed between September and November, indicating that increasing flood extents in September are often followed by sustained

or increased flooding in November. However, August and October displayed no meaningful correlation ($r = -0.00$), suggesting that early-season floods do not necessarily predict peak-season flooding. Moderate correlations ($r = 0.60$) were observed between August and both September and November. These mixed correlations reflect the complex interplay between localized rainfall, soil saturation thresholds, and upstream water contributions. For machine learning-based flood susceptibility modeling, these findings suggest that month-wise flood predictors should be selected with care, as flood progression is not always linear or temporally contiguous. The month-to-month correlation analysis was used to assess **persistence (memory)** in seasonal flooding rather than direct causation. Strong correlations indicate that flood extent is influenced by cumulative antecedent conditions (soil saturation, floodplain storage) and delayed upstream contributions, whereas weak correlations suggest flood extent is dominated by episodic storms and localized responses. In Bayelsa’s low-relief deltaic setting, lagged responses of 3–5 weeks between upstream water level peaks and downstream inundation provide a physical basis for why September conditions may relate to October–November flood extent

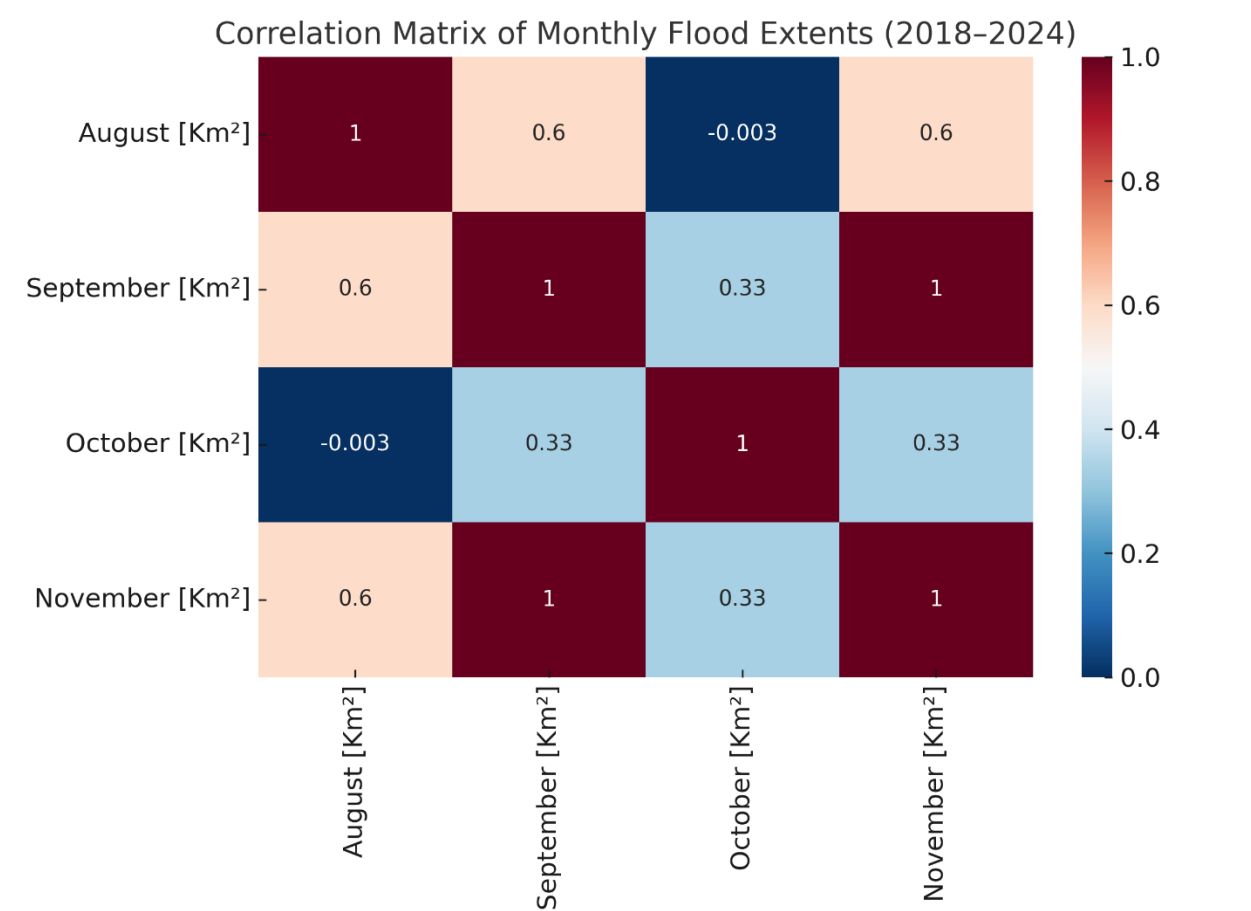


Figure 9: Correlation Matrix of Monthly Flood Extents (2018–2024)

Influence of Upstream and Downstream Hydrology

To assess the role of upstream and downstream hydrological processes, Figures 10 and 11 provide satellite altimetry-based water level data for the Niger tributary at Lokoja (upstream) and the river systems in Bayelsa State (downstream), respectively. Figure 10 shows that in most years, especially 2020 and 2022, upstream water levels peaked sharply in August and September. These surges were followed by corresponding increases in flood extent in Bayelsa during October and November, as shown in both Table 10 and Figure 8. This aligns with research by Schwatke et al. [50], which emphasizes the predictive value of altimetry in flood monitoring. Figure 11 demonstrates that downstream water levels in Bayelsa consistently peak in October or November, typically lagging behind upstream peaks by 3 to 5 weeks. This temporal delay confirms that Bayelsa's flood dynamics are not solely driven by local rainfall but also significantly influenced by cumulative river discharge and possible backwater effects. In 2024, for example, upstream peaks in September (Figure 10) likely contributed to the elevated flood extent recorded in November (Table 10).

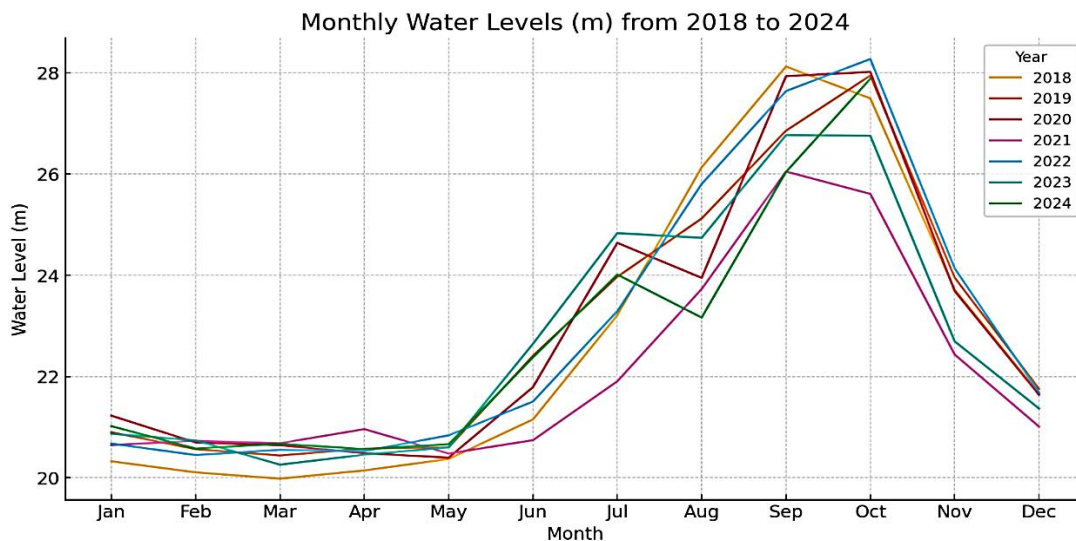


Figure 10: Water level Time series from Satellite Altimetry from 2018 to 2024 at Tributary of Lokoja (upstream) [50]

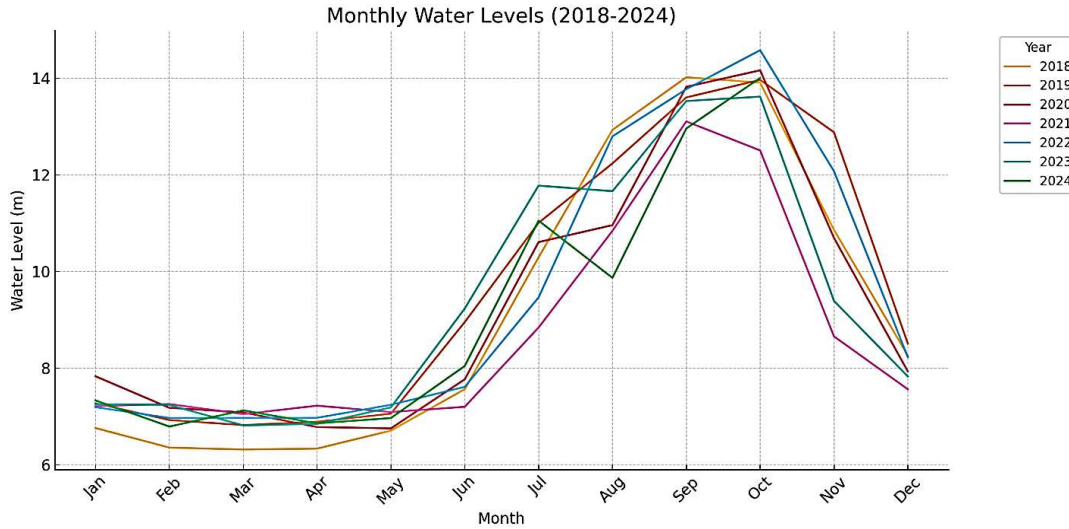


Figure 11: Water level Time series from Satellite Altimetry from 2018 to 2024 at Tributary of Bayelsa state (downstream) [50]

Advancing Methodological Novelty

Although the study primarily applies established algorithms (RF, SVM, XGBoost), its scientific contribution lies in bridging morphometric reasoning with explainable machine learning through SHAP-based interpretation. Future methodological advancement could extend this integration by testing hybrid deep-learning ensembles (CNN–XGBoost), spatio-temporal graph networks, and adaptive frameworks that continuously learn from new hydrological and climatic data. These enhancements would transform the current static susceptibility mapping into an evolving, process-driven predictive system.

5. Conclusion

This study demonstrates the effectiveness of integrating morphometric analysis with machine learning (ML) techniques to enhance flood susceptibility mapping in Bayelsa State, Nigeria. Using high-resolution Shuttle Radar Topographic Mission (SRTM) data and GIS tools, thirteen morphometric parameters were derived across four catchments Forcados, Nun, Ekole, and Seibri. These parameters, including drainage density, stream frequency, bifurcation ratio, and relief ratio,

provided crucial insights into hydrological behavior and catchment responsiveness to rainfall events. Machine learning models Random Forest, Support Vector Machine, and XGBoost were trained using the morphometric data as features. Among them, XGBoost emerged as the most accurate classifier, with a 93.1% accuracy and a 0.95 ROC-AUC score. The feature importance analysis revealed that drainage density, relief ratio, and infiltration number were the most influential predictors of flood susceptibility. Spatial flood susceptibility maps generated from ML outputs offered greater granularity than morphometric rankings alone, particularly highlighting moderate-risk subzones within catchments like Ekole that traditional methods had underestimated. The comparative results indicate that while morphometric analysis is valuable for initial catchment prioritization, ML models provide refined predictions by learning complex interactions among variables. The hybrid framework offers significant advantages, particularly in data-scarce and flood-prone regions like the Niger Delta, where conventional hydrological models may be limited by the absence of ground-based data. This integrative approach supports more informed decision-making for disaster risk reduction, land-use planning, and infrastructure development. It encourages the adoption of AI-driven geospatial models as a complement to classical hydrological techniques. The distinct contribution of this study lies in its integration of morphometric hydrology and interpretable AI within a single analytical pipeline. Whereas earlier models predicted flood probability without physical explanation, this framework explicitly links geomorphic structure (drainage density, relief ratio, infiltration number) to machine-learned probability through SHAP-based causal interpretation. By introducing the “micro-zone” concept and quantifying morphometric drivers of flood susceptibility, the research advances a more granular understanding of deltaic flood dynamics. The methodology is both novel in design through its emphasis on explainability and spatial independence and scalable in application, offering a reproducible template for similar low-lying, data-limited regions worldwide.

Limitations of the Study

1. Static Nature of Morphometric Parameters

The morphometric analysis relies on terrain-based indices derived from SRTM, which are static and do not capture temporal variations, especially those influenced by land use change or rainfall variability over time.

2. Exclusion of Dynamic Climatic and Socio-Economic Variables

While the ML models enhance spatial prediction, they do not include dynamic meteorological inputs or socio-economic data, which are critical in comprehensive flood modeling.

3. Limited Ground-Truth Validation

The training data relies heavily on historical flood records and satellite-derived flood extent, with no indication of field-verified flood point data. This may introduce uncertainties in class labeling for ML models

4. Cloud Obstruction and SAR Limitations

The study mentions using Sentinel-1 data, but inherent limitations of SAR interpretation in vegetated or urban zones (due to layover, speckle) might affect flood delineation precision, especially in complex terrain like Bayelsa.

Recommendations:

- Integrate multi-temporal DEMs: Use datasets like TanDEM-X or repeat-pass LiDAR (if available) to assess terrain changes over time.
- Incorporate change detection: Overlay historical LULC and rainfall data to dynamically adjust morphometric parameters like runoff potential.
- Adaptive modeling: Consider incorporating dynamic hydrologic variables (e.g., NDVI for vegetation cover or impervious surface growth from Landsat/Sentinel-2).

Future work should incorporate climate projections, remote sensing-based rainfall data, and socio-economic indicators to further enhance the dynamic modeling of flood hazards in vulnerable regions.

Conflict of Interest Statement

The authors declare no conflict of interest.

Data Availability Statement

All data and analysis scripts will be made publicly available upon acceptance on <https://github.com/eteh1>. Ethical approval was not required as the study involved non-invasive water sampling adhering to WHO standard protocols

Funding Statement

None

Author Contributions

O.I., P.C., E.I.O., A.G.W., F.N.O., D.R.E., T.W., E.N.A., and O.E.K. collaboratively conceptualized the study and contributed to the manuscript's writing and structure. Data processing and analysis were performed by D.R.E., M.D.A, O.I.O, B.P.B, O.I., N.E.A., F.N.O., and A.G.W. Critical review and editing were undertaken by T.W., N.E.A., O.I., P.C., and O.I. All authors contributed to the final manuscript preparation and approved the submitted version.

Acknowledgments

We extend our heartfelt gratitude to our families and Geosoft Global Innovation limited for their unwavering support and understanding throughout this project. Additionally, we appreciate the anonymous reviewers for their valuable feedback and suggestions, which significantly enhanced the quality of this research.

Ethics approval and consent to participate

Not applicable

Data Availability Statement

1. Satellite Data: Sentinel-1 satellite data (2018–2024) provided by the European Space Agency (ESA) were utilized to assess flood extents. These data can be accessed via the Copernicus Open Access Hub: <https://scihub.copernicus.eu/>

2. Topography Data: Shuttle Radar Topography Mission (SRTM) data were acquired through the United States Geological Survey (USGS) Earth Explorer: <https://earthexplorer.usgs.gov>

3. History flood record from sentinel 1 by Copernicus Open Access Hub.

All source code and essential data used in this study have been prepared for public access at <https://github.com/eteh1/Flood-Susceptibility-Mapping-Using-Machine-Learning-and-Morphometric-Analysis>. The complete implementation pipeline includes digital elevation model (DEM) preprocessing, morphometric feature computation, machine learning model training (Random Forest, Support Vector Machine, and XGBoost), hyperparameter optimization, performance evaluation, and model interpretability using SHAP (SHapley Additive Explanations).

6. References

- [1]. Umar, N., & Gray, A. (2022). Flooding in Nigeria: A review of its occurrence and impacts and approaches to modelling flood data. *International Journal of Environmental Studies*, 80, 1–22. <https://doi.org/10.1080/00207233.2022.2081471>
- [2]. Eteh, D., Egobueze, F., & Omonefe, F. (2019). Determination of flood hazard zones using geographical information systems and remote sensing techniques: A case study in part Yenagoa metropolis. *Journal of Geography, Environment and Earth Science International*, 21(1), 1–9.
- [3]. Eteh, D. R., Egobueze, F. E., Paarur, M., Otutu, A., & Osondu, I. (2024). The impact of dam management and rainfall patterns on flooding in the Niger Delta: Using Sentinel-1 SAR data. *Discover Water*, 4(1), 123. <https://doi.org/10.1007/s43832-024-00185-8>
- [4]. Horton, R. (1945). Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin*, 56, 275–370. [https://doi.org/10.1130/0016-7606\(1945\)56\[275:EDOSAT\]2.0.CO;2](https://doi.org/10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2)
- [5]. Strahler, A. (1957). Quantitative analysis of watershed geomorphology. *Transactions, American Geophysical Union*, 38, 913–920. <https://doi.org/10.1029/TR038i006p00913>
- [6]. Oborie, E., & Eteh, R. D. (2023). Flood influence using GIS and remote sensing based morphometric parameters: A case study in Niger delta region. *Journal of Asian Scientific Research*, 13(1), 1-15.
- [7]. Kumar, N., & Jha, R. (2023). GIS-based Flood Risk Mapping: The Case Study of Kosi River Basin, Bihar, India. *Engineering, Technology & Applied Science Research*, 13(1), 9830–9836. <https://doi.org/10.48084/etasr.5377>

- [8]. Rahman, Z. U., Ullah, W., Bai, S., Ullah, S., Jan, M. A., Khan, M., & Tayyab, M. (2023). GIS-based flood susceptibility mapping using bivariate statistical model in Swat River Basin, Eastern Hindukush region, Pakistan. *Frontiers in Environmental Science*, 11, Article 1178540. <https://doi.org/10.3389/fenvs.2023.1178540>
- [9]. Mosavi, A., Ozturk, P., & Chau, K.-w. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, 10(11), 1536. <https://doi.org/10.3390/w10111536>
- [10]. Debnath, J., Sahariah, D., Mazumdar, M., & others. (2023). Evaluating flood susceptibility in the Brahmaputra River Basin: An insight into Asia's Eastern Himalayan floodplains using machine learning and multi-criteria decision-making. *Earth Systems and Environment*, 7, 733–760. <https://doi.org/10.1007/s41748-023-00358-w>
- [11]. Al-Ruzouq, R., Shanableh, A., Jena, R., Gibril, M. B. A., Hammouri, N. A., & Lamghari, F. (2024). Flood susceptibility mapping using a novel integration of multi-temporal Sentinel-1 data and eXtreme deep learning model. *Geoscience Frontiers*, 15(3), 101780. <https://doi.org/10.1016/j.gsf.2024.101780>
- [12]. Singh, G., & Pandey, A. (2021). Flash flood vulnerability assessment and zonation through an integrated approach in the Upper Ganga Basin of the Northwest Himalayan region in Uttarakhand. *International Journal of Disaster Risk Reduction*, 66, 102573. <https://doi.org/10.1016/j.ijdrr.2021.102573>
- [13]. Demissie, Z., Rimal, P., Seyoum, W. M., Dutta, A., & Rimmington, G. (2024). Flood susceptibility mapping: Integrating machine learning and GIS for enhanced risk assessment. *Applied Computing and Geosciences*, 23, 100183. <https://doi.org/10.1016/j.acags.2024.100183>
- [14]. Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T., & Khosravi, K. (2017). A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software*, 95, 229–245. <https://doi.org/10.1016/j.envsoft.2017.06.012>
- [15]. Luu, C., Bui, Q. D., Costache, R., Nguyen, L. T., Nguyen, T. T., Phong, T., Le, H., & Pham, B. T. (2021). Flood-prone area mapping using machine learning techniques: A case study of Quang Binh province, Vietnam. *Natural Hazards*, 108(3), 3229–3251. <https://doi.org/10.1007/s11069-021-04872-4>

- [16]. Nguyen, H. D., Quang-Thanh, B., Nguyen, Q. H., Nguyen, T. G., Pham, L. T., Nguyen, X. L., ... Petrisor, A. I. (2022). A novel hybrid approach to flood susceptibility assessment based on machine learning and land use change. Case study: a river watershed in Vietnam. *Hydrological Sciences Journal*, 67(7), 1065–1083. <https://doi.org/10.1080/02626667.2022.2060108>
- [17]. Otoosakyi, D., Adinya, I., & Taiwo, E. (2022). Flood prediction in Africa: A machine learning approach. <https://doi.org/10.13140/RG.2.2.25882.72642/1>
- [18]. Eteh, D. R., Japheth, B. R., Akajiaku, C. U., Osondu, I., Mene-Ejegi, O. O., Nwachukwu, E. M., et al. (2025). Assessing the impact of climate change on flood patterns in downstream Nigeria using machine learning and geospatial techniques (2018–2024). *Discover Geoscience*, 3(1), 76. <https://doi.org/10.1007/s44288-025-00178-7>
- [19]. Okpobiri, O., Akajiaku, C. U., Eteh, D. R., & Moses, P. (2025). Using machine learning and GIS to monitor sandbars along the River Niger in the Niger Delta, Nigeria. *International Journal of Environment and Climate Change*, 15(2), 182–203. <https://doi.org/10.9734/ijecc/2025/v15i24721>
- [20]. Jonathan, L. E., & Charles, A. U. (2025). Shoreline erosion and accretion analysis of the Orashi River, Rivers State, Nigeria: A geospatial and machine learning approach. *Asian Journal of Geographical Research*, 8(2), 27–44. <https://doi.org/10.9734/ajgr/2025/v8i2260>
- [21]. Bamiekumo, B. P., Akpobome, E. O., Kemebaradikumo, A. N., Mene-Ejegi, O. O., & Eteh, D. R. (2025). Machine learning-based flood extent mapping and damage assessment in Yenagoa, Bayelsa State, using Sentinel-1 and 2 imagery (2018–2022). *Discovery Nature*, 2(3), e2dn1041. <https://doi.org/10.54905/disssi.v2i3.e2dn1041>
- [22]. Abaye, E. O., Ayowei, A. H., Akajiaku, U. C., & Eteh, D. R. (2022). Implication of landscape changes using Google Earth historical imagery in Yenagoa, Bayelsa State, Nigeria. *Sumerianz Journal of Scientific Research*, 5(1), 20–31.
- [23]. Okes, & Jonathan, E. L. (2025). Data-driven assessment of soil heavy metal contamination in Joinkrama, Rivers State, Nigeria using pollution indices and multivariate analytics. *Journal of Materials Science Research and Reviews*, 8(3), 628–643. <https://doi.org/10.9734/jmsrr/2025/v8i3429>
- [24]. Jonathan, L. E., Winston, A. G., & Chukwuemeka, P. (2025). Machine learning and morphometric analysis for runoff dynamics: Enhancing flood management and catchment prioritization in Bayelsa, Nigeria. *Journal of Computational Systems and Applications*, 2(2), 1–16. <https://doi.org/10.63623/kkx1m906>
- [25]. Jonathan, L. E., Imoni, O., Chukwuemeka, P., & Eteh, D. R. (2025). Impact of oil spills on mangrove ecosystem degradation in the Niger Delta using remote sensing and machine learning. *Journal of Geography and Cartography*, 8(2), 11707. <https://doi.org/10.24294/jgc11707>
- [26]. Oborie, E., Fatunmibi, I., & Otutu, A. O. (2023). Shoreline change assessment in the Orashi River, Rivers State, Nigeria, using the digital shoreline analysis system

(DSAS). *Sumerianz Journal of Scientific Research*, 6(4), 70–77.
<https://doi.org/10.47752/sjsr.64.70.77>

- [27]. Akajiaku, U. C., Ohimain, E. I., Olodiana, E. E., Eteh, D. R., Winston, A. G., Chukwuemeka, P., et al. (2025). Identifying suitable dam sites using geospatial data and machine learning: A case study of the Katsina-Ala River in Benue State, Nigeria. *Earth Science Informatics*, 18(3), 497. <https://doi.org/10.1007/s12145-025-01974-y>
- [28]. Akajiaku, C. U., Agbabiaka, C. O., Imoni, O., Chukwuemeka, P., Eteh, D. R., & Amos, M. D. (2025). Predictive modeling of iron concentration in groundwater using machine learning techniques: A case study in part of Yenagoa, Bayelsa State. *Journal of Computational Systems and Applications*, 2(1), 1–11.
- [29]. Oseji, S., Chukwuemeka, P., & Imoni, O. (2025). Artificial intelligence in 3D printed concrete: Sustainability assessment and implementation challenges. *Journal of Materials Science Research and Reviews*, 8(2), 515–528.
<https://doi.org/10.9734/jmsrr/2025/v8i2421>
- [30]. Chukwuemeka, P., Kyrian, O., & Imoni, O. (2025). Leveraging machine learning for the identification of obfuscated JavaScript in phishing attacks. *Asian Journal of Research in Computer Science*, 18(6), 301–314.
<https://doi.org/10.9734/ajrcos/2025/v18i6700>
- [31]. Amos, M. D., Opukumo, A. W., & Winston, A. G. (2025). Geospatial assessment of drainage density and infiltration characteristics in coastal catchments of Southern Nigeria. *Journal of Geography, Environment and Earth Science International*, 29(10), 79–97. <https://doi.org/10.9734/jgeesi/2025/v29i10954>
- [32]. Mogo, C. F., Agbabiaka, C. O., Bamiekumo, B. P., & Mogo, C. O. (2025). Geospatial mapping and monitoring of sandbars in the Forcados River: Implications for sustainable blue economy in the Niger Delta. *Journal of Geography, Environment and Earth Science International*, 29(8), 106–127.
<https://doi.org/10.9734/jgeesi/2025/v29i8937>
- [33]. Samuel Oseji, Prince Chukwuemeka, Okes Imoni. Artificial Intelligence in 3D oPrinted Concrete: Sustainability Assessment and Implementation Challenges. *Journal of Materials Science Research and Reviews*, 2025, 8 (2), pp.515-528.
{10.9734/jmsrr/2025/v8i2421}.
- [34]. Hasnaoui, Y., Tachi, S.E., Bouguerra, H. et al. Enhanced machine learning models development for flash flood mapping using geospatial data. *Euro-Mediterr J Environ Integr* 9, 1087–1107 (2024). <https://doi.org/10.1007/s41207-024-00553-9>
- [35]. Hasnaoui, Y., Tachi, S.E., Bouguerra, H. et al. Transfer learning-based deep learning models for flood and erosion detection in coastal area of Algeria. *Earth Sci Inform* 18, 380 (2025). <https://doi.org/10.1007/s12145-025-01866-1>
- [36]. Hasnaoui, Y., Tachi, S.E., Bouguerra, H. et al. Integrated Remote Sensing and Deep Learning Models for Flash Flood Detection Based on Spatio-temporal Land Use and Cover Changes in the Mediterranean Region. *Environ Model Assess* (2025).
<https://doi.org/10.1007/s10666-025-10035-z>
- [37]. Chukwuemeka, P., Imoni, O., Mogo, F. C., Eteh, D. R., William, T., Bamidele, B. P., Ominofe, F., Agbozu, N. E., Mene-Ejegi, O. O., Ihekona, O., Akajiaku, C. U., & Ben-

- Koko, M. (2026). A hybrid AI model integrating LSTM, XGBoost, and K-means for interpretable prediction and clustering of water quality in data-scarce regions. *Discover Civil Engineering*, 3, 24. <https://doi.org/10.1007/s44290-026-00417-x>
- [38]. Jonathan, L. E., & Charles, A. U. (2025). Shoreline Erosion and Accretion Analysis of the Orashi River, Rivers State, Nigeria: A Geospatial and Machine Learning Approach. *Asian Journal of Geographical Research*, 8(2), 27–44. <https://doi.org/10.9734/ajgr/2025/v8i2260>
- [39]. Osondu, I., Imoni, O., Chukwuemeka, P., Amos, M. D., Ajibade, D. I., Mene-Ejegi, O. O., Mogo, C. F., Eteh, D. R., Nwadoziokwu, E. O., & Akajiaku, C. U. (2025). Machine learning-based shoreline change prediction and erosion analysis: A case study of Ogu/Bolo, Nigeria. *Discover Civil Engineering*, 2, 230. <https://doi.org/10.1007/s44290-025-00396-5>
- [40]. Oladimeji, O., & Ohwo, O. (2022). Assessment of flood risk and mapping of flood risk zones in Yenagoa, Bayelsa State, Nigeria. *Journal of Applied Sciences and Environmental Management*.
- [41]. Jonathan, L. E., Winston, A. G., & Chukwuemeka, P. (2025). Machine learning and morphometric analysis for runoff dynamics: Enhancing flood management and catchment prioritization in Bayelsa, Nigeria. *Journal of Computational Systems and Applications*, 2(2), 1–16. <https://doi.org/10.63623/kkx1m906>
- [42]. Etu-Efeotor, J. O., & Akpokodje, E. G. (1990). Aquifer systems of the Niger Delta. *Journal of Mining and Geology*, 26, 279–285.
- [43]. Akajiaku, C. U., Japheth, B. R., Eteh, D. R., Osondu, I., Mene-Ejegi, O. O., Nwachukwu, E. M., & Ayo-Bali, A. E. (2025). Assessing the impact of climate change on flood patterns in downstream Nigeria using machine learning and geospatial techniques (2018-2024). *Discover Geoscience*
- [44]. Ekeu-Wei, I. T., & Blackburn, G. A. (2018). Applications of open-access remotely sensed data for flood modeling and mapping in developing regions. *Hydrology*, 5(3), 39.
- [45]. Eteh, D. R., & Okpobiri, O. (2021). Floodplain mapping and risk assessment of the Orashi River using remote sensing and GIS in the Niger Delta region, Nigeria. *Journal of Geographic Research*, 4(2), 10–16. <https://doi.org/10.30564/jgr.v4i2.3014>

- [46]. Amangabara, G. T., & Obenade, M. (2015). Flood vulnerability assessment of Niger Delta states relative to the 2012 flood disaster in Nigeria. *American Journal of Environmental Protection*, 3(3), 76–83.
- [47]. Jenson, S. and Domingue, J. (1988) Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. *Photogrammetric Engineering and Remote Sensing*, 54, 1593-1600.
- [48]. Lang, F., Zhu, Y., Zhao, J., Hu, X., Shi, H., Zheng, N., & Zha, J. (2024). Flood Mapping of Synthetic Aperture Radar (SAR) Imagery Based on Semi-Automatic Thresholding and Change Detection. *Remote Sensing*, 16(15), 2763. <https://doi.org/10.3390/rs16152763>
- [49]. Oreikio, A. E., Harry, A. A., Charles, A. U., & Rowland, E. D.(2022) Implication of Landscape Changes Using Google Earth Historical Imagery in Yenagoa Bayelsa State, Nigeria. *Journal of Scientific Research*, 5(1), 20-31.
- [50]. Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI – An innovative approach for estimating water-level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, 19(10), 4345–4364. <https://doi.org/10.5194/hess-19-4345-2015>
- [51]. Schumm, S. A. (1956). Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *Geological Society of America Bulletin*, 67, 464–597. [https://doi.org/10.1130/0016-7606\(1956\)67\[597:EODSAS\]2.0.CO;2](https://doi.org/10.1130/0016-7606(1956)67[597:EODSAS]2.0.CO;2)
- [52]. Melton, M. A. (1958). Correlation structures of morphometric properties of drainage systems and their controlling agents. *Journal of Geology*, 66, 442–460. <https://doi.org/10.1086/626527>
- [53]. Dash, B., Nagaraju, M. S. S., Sahu, N., Nasre, R. A., Mohekar, D., Srivastava, R., & Singh, S. (2019). Morphometric analysis for planning soil and water conservation measures using geospatial technique. *International Journal of Current Microbiology and Applied Sciences*, 8, 2719–2728. <https://doi.org/10.20546/ijcmas.2019.801.287>
- [54]. Singh, V. G., & Singh, S. K. (2022). Analysis of geo-morphometric and topohydrological indices using COP-DEM: A case study of Betwa River Basin, Central India. *Geology, Ecology, and Landscapes*, 8(2), 101–128. <https://doi.org/10.1080/24749508.2022.2097376>

- [55]. Zavoianu, I. (1985) *Morphometry of Drainage Basins (Developments in Water Science)*. Elsevier, Amsterdam.
- [56]. Breiman, L. (2001) Random Forests. *Machine Learning*, 45, 5-32. <http://dx.doi.org/10.1023/A:1010933404324>
- [57]. Moses, P., Nwachukwu, E.M., Mene-Ejegi, O.O. *et al.* Assessing the evolution of point bars along the Niger river in the Niger Delta, Nigeria (1974–2024) using remote sensing and machine learning. *Appl Geomat* **18**, 37 (2026). <https://doi.org/10.1007/s12518-025-00687-7>
- [58]. Jonathan EL, Imoni O, Chukwuemeka P, Eteh DR. Impact of oil spills on mangrove ecosystem degradation in the Niger Delta using remote sensing and machine learning. *Journal of Geography and Cartography*. 2025; 8(2): 11707. <https://doi.org/10.24294/jgc11707>
- [59]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *NeurIPS*, 30, 4765–4774.
- [60]. Youssef, Y. M., AlMetwaly, W. M., Saber, M., Abdelshafy, M., Qaysi, S., Alarifi, N., & Wahba, M. (2025). Integrating geospatial and hydrodynamic modelling to unravel cultural heritage vulnerability to dam breaches and flash flood in arid environments. *International Journal of Disaster Risk Reduction*, 126, 105605. <https://doi.org/10.1016/j.ijdrr.2025.105605>
- [61]. Shafizadeh-Moghadam, H., et al. (2018). Machine learning for urban flood modeling. *Environmental Monitoring and Assessment*, 190, 504.
- [62]. Rondinone, M., Dal Sasso, S. F., Aung, H. H., Contillo, L., Dimola, G., Schiattarella, M., Fiorentino, M., & Telesca, V. (2025). Assessing Flood and Landslide Susceptibility Using XGBoost: Case Study of the Basento River in Southern Italy. *Applied Sciences*, 15(10), 5290. <https://doi.org/10.3390/app15105290>
- [63]. Tziachris, P., Nikou, M., Aschonitis, V., Kallioras, A., Sachsamanoglou, K., Fidelibus, M. D., & Tziritis, E. (2023). Spatial or Random Cross-Validation? The Effect of Resampling Methods in Predicting Groundwater Salinity with Machine Learning in Mediterranean Region. *Water*, 15(12), 2278. <https://doi.org/10.3390/w15122278>
- [64]. Oruonye, E. D. (2016). Morphometry and flood in small drainage basin: Case study of Mayogwoi River Basin in Jalingo, Taraba State Nigeria. *Journal of Geography*,

Environment and Earth Science International, 5, 1–12.
<https://doi.org/10.9734/JGEESI/2016/23379>

- [65]. Ajumobi, V., Womboh, S., & Ezem, S. (2023). Impacts of the 2022 flooding on the residents of Yenagoa, Bayelsa State, Nigeria, 1–6.
- [66]. Rizwan, M., Li, X., Chen, Y., Anjum, L., Hamid, S., Yamin, M., Chauhdary, J. N., Shahid, M. A., & Mehmood, Q. (2022). Simulating future flood risks under climate change in the source region of the Indus River. *Journal of Flood Risk Management*, 16, e12857.
<https://doi.org/10.1111/jfr3.12857>
- [67]. Aju, C. D., Achu, A. L., Prakash, P., Raicy, M. C., & Reghunath, R. (2024). An integrated statistical-geospatial approach for the delineation of flood-vulnerable sub-basins and identification of suitable areas for flood shelters in a tropical river basin, Kerala. *Geosystems and Geoenvironment*, 3(2), 100251.
<https://doi.org/10.1016/j.geogeo.2024.100251>
- [68]. NIMET (2022). Annual Flood Outlook. Nigerian Meteorological Agency.