Integrating Flood Susceptibility and Deforestation Mapping for Climate Vulnerability Assessment: A Geospatial and AI-Based Approach

Serah Akojenu
Data Science Nigeria
Lagos, Nigeria
serah@datasciencenigeria.ai

Chinazo Anebelundu
Data Science Nigeria
Lagos, Nigeria
chinazo@datasciencenigeria.ai

Godwin Adegbeingbe
Data Science Nigeria
Lagos, Nigeria
godwin@datasciencenigeria.ai

Blessing Gideon
Data Science Nigeria
Lagos, Nigeria
blessing@datasciencenigeria.ai

Olamide Shogbamu
Data Science Nigeria
Lagos, Nigeria
shogbamu@datasciencenigeria.ai

Tochuckwu Abia
Data Science Nigeria
Lagos, Nigeria
tochuckwu@datasciencenigeria.ai

Anthony Soronadi
Data Science Nigeria
Lagos, Nigeria
anthony@datasciencenigeria.ai

Olubayo Adekanmbi
Data Science Nigeria
Lagos, Nigeria
olubayo@datasciencenigeria.ai

Abstract

Climate change has intensified global natural disasters, with floods and droughts posing the greatest threats to human settlements and economic systems. In Nigeria alone, flooding causes 80% of climate-induced deaths and inflicts catastrophic economic losses of 60 billion dollars annually, while persistent drought threatens the food security of millions in northern regions, which contribute significantly to the food consumed across the country. Despite these devastating impacts, existing machine learning disaster prediction studies focus on single hazards with limited scope and insufficient risk indicator integration, leaving communities vulnerable and unprepared. This study addresses this critical research gap by developing a comprehensive Climate Vulnerability Index (CVI) that integrates flood susceptibility and drought risk mapping across Nigeria using advanced Random Forest, XGBoost, and Light GBM algorithms with multi-criteria decision analysis. This framework represents a shift from fragmented, single-hazard approaches to a unified, multi-hazard assessment system that incorporates diverse risk indicators, creating a scalable tool essential for protecting lives and economic stability at national, state, and local levels.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

1 Introduction

The global climate system has undergone unprecedented changes, fundamentally altering natural disaster patterns and creating cascading impacts on societal stability, economic security, and ecological integrity [1]. Among climate-induced disasters, floods and droughts represent the most devastating threats, with floods accounting for 47% of weather-related disasters, affecting over 250 million people annually and causing hundreds of billions in economic losses [2]. Nigeria exemplifies acute climate vulnerability as Africa's most populous country with over 220 million inhabitants experiencing severe climate-related disasters. Flooding represents 80% of all natural disasters in Nigeria [3], displacing millions, destroying infrastructure, contaminating water resources, and creating disease outbreak conditions [4, 5]. Simultaneously, northern regions face increasing drought stress threatening agricultural productivity and food security, with high temperatures and irregular rainfall disrupting crop cycles and creating interconnected risks for rural communities [6]. This dual flood-drought threat represents a critical vulnerability requiring comprehensive assessment and prediction capabilities [7]. This study addresses these gaps by developing an automated preprocessing framework using Google Earth Engine and integrating comprehensive flood risk indicators with drought indices to create a unified Climate Vulnerability Index (CVI) that maps combined flood and drought risks across Nigeria.

2 Literature Review

Machine learning (ML) classifiers like Random Forest (RF), Support Vector Machine (SVM), and XGBoost have been increasingly used for flood susceptibility mapping due to their ability to handle high-dimensional, nonlinear data, with RF showing particular robustness through interpretability [8]. Nigerian studies have employed various approaches: Ghile et al., 2022 used ANN and logistic regression with fifteen indicators (elevation, slope, rainfall, soil type, land cover, etc.) demonstrating ANN's effectiveness for flood prediction [9], while North Central Nigeria studies found GBM reliable for flood predictions and ensemble models efficient for the region [10]. For drought prediction, [11] used SPEI with meteorological parameters to train RF, XGB, CNN, and LSTM models over four decades, with XGB showing early warning potential.A study also combined AHP with RF using four predictors (elevation, slope, rainfall, distance to river) to create a Flood Susceptibility Index, though this approach was limited by excluding variables like soil type and infrastructure proximity. Despite ML advances, the "black-box" problem has led to hybrid approaches integrating ML with expert frameworks like Analytic Hierarchy Process (AHP), which allows systematic weight assignment based on domain knowledge [8]. However, most Nigerian flood susceptibility studies remain limited to specific regions and rely solely on ML or deterministic modeling. The reviewed studies reveal gaps in indicator comprehensiveness and multi-criteria analysis application, highlighting the need for approaches that incorporate diverse risk indicators and integrate flood-drought indices into unified vulnerability assessments.

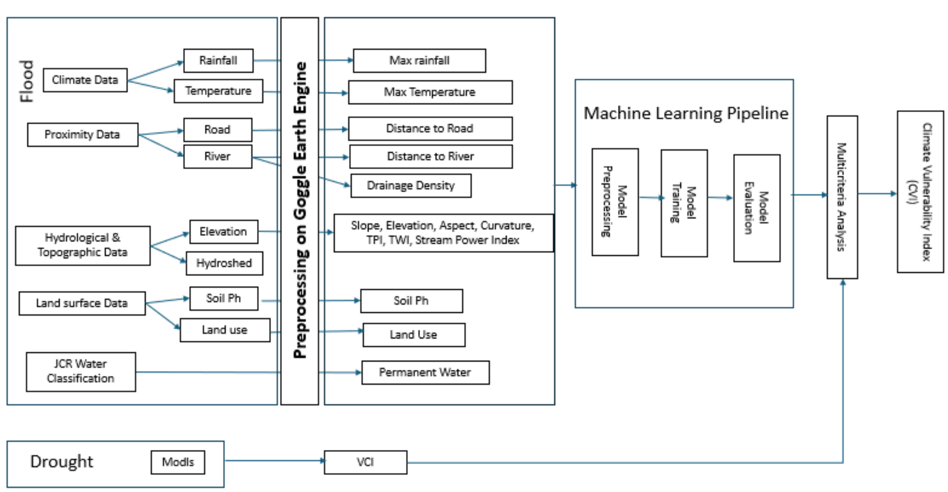

Figure 1: Methodology flowchart

3 Methodology

3.1 Data Collection

Group	Dataset	Source	Observation period / Last Updated	Format	Spatial resolution
Climate data	Rainfall	CHIRPS daily raster	2019 to 2024	Raster	5566 meters
	Temperature	TerraClimate	2019 to 2024	Raster	4638.3 meters
Proximity data	Roads	DivaGIS	April 2025	Shapefile	–
	Rivers	DivaGIS	April 2025	Shapefile	–
Hydrologic & Topographic data	Elevation	Shuttle Radar Topography Mission (SRTM)	2020	Raster	30 meters
	HydroSHEDS	WWF HydroSHEDS Flow Accumulation	2020	Raster	463.83 meters
Land surface data	NDVI	Modis Data	2018 to present	Raster	201 meters
	Soil pH	OpenLandMap	2018	Raster	250 meters
	Land Use	ESA WorldCover 10m v100	2021	Raster	10 meters
Administrative boundary data	Admin boundary (country, state, local)	DivaGIS	April 2025	Shapefile	–
Surface water data	Surface water	JRC Yearly Water Classification History	2022	Raster	30 meters

Table 1: Summary of geospatial datasets used in the study

3.1.1 Preprocessing in Google Earth Engine

The preprocessing done in Google Earth Engine can be divided into 3 parts namely;

Indices Calculation: The MODIS dataset was used to calculate the Vegetation Condition Index (VCI), which served as the indicator used for the mapping of drought.

Feature Derivation and Spatial Analysis: Several spatial features were derived for this analysis. The JRC Surface Water dataset was masked to extract permanent water bodies, which were used to assess flood history by applying a threshold of 50% water occurrence to identify areas consistently inundated. From the elevation dataset, multiple topographic parameters were derived, including elevation, slope, aspect, curvature, topographic wetness index (TWI), topographic position index (TPI), and stream power index (SPI). The flow accumulation layer was further processed to calculate drainage density. Additionally, proximity analyses were carried out using the Euclidean distance method to generate distance-to-road and distance-to-river maps, with a 100 km maximum search radius applied to ensure adequate visibility at the national scale.

Furthermore, rainfall and temperature datasets were used to produce maps of maximum temperature and rainfall distribution. The land use dataset was filtered to retain five key classes of interest: built-up areas, water bodies, vegetation, farmlands, and croplands. While the soil pH dataset was incorporated without modification. Upon completion of the preprocessing phase, all derived maps were exported as GeoTIFF files to serve as spatial predictor variables in the flood modeling pipeline.

3.2 Machine Learning Pipeline for Flood Prediction

Following the preprocessing stage, a machine learning pipeline was developed to predict flood vulnerability across Nigeria.

3.2.1 Data Preprocessing for Model Training

All predictors and the target variable (permanent water) were exported as GeoTIFF raster layers. These layers were ingested into Python and read using rasterio, which converted them into matrices. To ensure consistency in spatial resolution and alignment, each raster was resampled and reshaped to match a common reference grid (based on the elevation layer). Missing or undefined pixel values were replaced by interpolation to preserve more pixel values and a stratified 70/30 split for training and testing, preserving the proportion of flood vs. non-flood observations.

3.2.2 Model Training

Three machine learning models - Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) were implemented to classify flood susceptibility

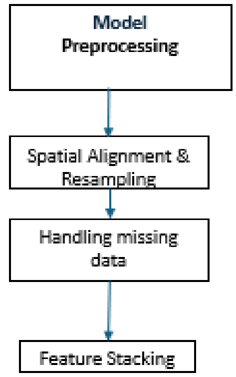

Figure 2: Flood dataset preprocessing workflow for model training.

based on the input features. All models were trained using 70% for training and the remaining 30% for evaluation. Prior to training, the target variable was rounded and converted to binary integers, 0s and 1s. Each model was configured with 100 estimators and trained using tuned hyperparameters specific to each algorithm, with the appropriate objective parameter defined for the LightGBM model to handle binary classification. The probabilistic output for the RandomForest model was retained because it gave the best performance, as seen in Table 2.

3.2.3 Evaluation

The performance of the Random Forest, XGBoost, and LightGBM classifiers was evaluated using the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (TPR) against the false positive rate (FPR) at various classification thresholds. The ROC curves were generated using the predicted probabilities (yprob) from each model, providing a visual and quantitative measure of their ability to distinguish between flood-prone and non-flood-prone areas.

3.3 Multicriteria Analysis

A multicriteria analysis was conducted in QGIS for the two disasters using a weighted overlay of flood and drought vulnerability layers, assigned a 60:40 ratio to emphasize flooding due to its higher prevalence across Nigeria. This process produced the Climate Vulnerability Index (CVI) map for the entire country. The final output was subsequently reclassified into five categories (15), with 5 indicating the highest level of vulnerability and 1 representing the lowest.

4 Results and Discussion

The evaluation of the three classification models as seen in (table2) showed that Random Forest (RF) achieved the highest performance, yielding an Area Under the Curve (AUC) of 0.85, which confirmed its robustness in predicting flood susceptibility across Nigeria, outperforming both XGBoost (0.79) and LightGBM (0.76). The feature importance as seen in (figure 4) revealed that elevation, rainfall, temperature, and proximity to roads and rivers were the most influential geospatial predictors of flooding. The resulting climate vulnerability index(CVI) map, as seen in (figure 3) indicated that states in the northeastern and central regions exhibit the highest climate vulnerability, consistent with areas frequently impacted by both extreme weather events and persistent drought stress.

Table 2: Model Performance Comparison based on AUC

Model	**Area Under Curve (AUC)**
Random Forest	0.85
XGBoost	0.79
LightGBM	0.76

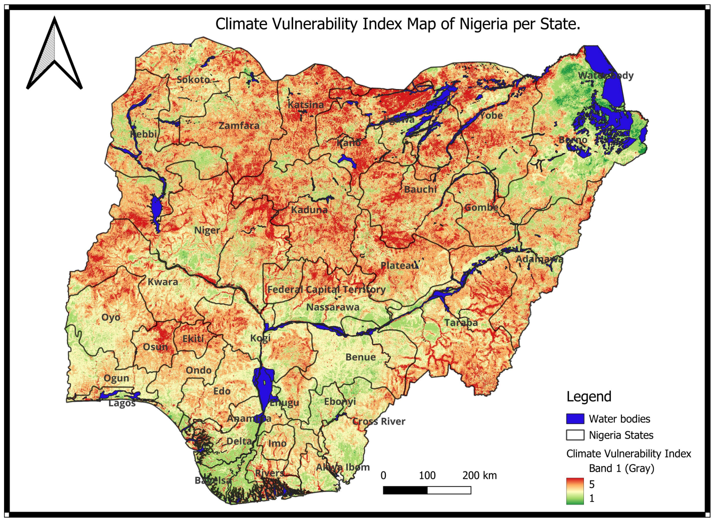

Figure 3: Climate Vulnerability Index Map of Nigeria

5 Conclusion

This study demonstrated the effectiveness of integrating geospatial datasets and machine learning, with Random Forest achieving the highest accuracy in mapping climate vulnerability across Nigeria and identifying hotspot regions most at risk. The findings support SDG 13 (Climate Action), SDG 11 (Sustainable Cities and Communities), and SDG 2 (Zero Hunger) by providing evidence-based insights to guide adaptation strategies, risk-informed planning, and climate-resilient interventions.

References

[1] Kristie L Ebi, Jennifer Vanos, John W Baldwin, Jennifer E Bell, David M Hondula, Nicole A Errett, Katherine Hayes, Colleen E Reid, Shubhayu Saha, Jason Spector, and Peter Berry. Extreme weather and climate change: Population health and health system implications. *Annual Review of Public Health*, 42:293–315, 2021.

[2] David Eckstein, Vera Künzel, and Laura Schäfer. Global climate risk index 2021: Who suffers most from extreme weather events? weather-related loss events in 2019 and 2000 to 2019, 2021. Accessed: 2025-08-18.

[3] National Emergency Management Agency (NEMA). National disaster risk management policy 2018. PDF report, 2018. Accessed: 2025-08-18.

[4] A. T. Aborode, O. J. Otorkpa, A. O. Abdullateef, et al. Impact of climate change-induced flooding water related diseases and malnutrition in borno state, nigeria: A public health crisis. *Environmental Health Insights*, 19, 2025.

[5] R. S. Chandel, P. K. Rai, and R. K. Misra. Advancing sustainability: a comprehensive review of geospatial technology in monitoring climate change-driven landslide hazards. *GNSS Applications in Earth and Space Observations*, pages 80–103, 2025.

[6] C. H. Ojeilua, A. E. Folorunso, F. N. Eze, E. L. Okaah, and O. R. Alonge. Evolutionary adaptations in drought-resilient crops and their impact on agricultural sustainability in nigeria. *African Journal of Agricultural Science and Food Research*, 19(1):44–64, 2025.

[7] A. Tsatsaris, K. Kalogeropoulos, N. Stathopoulos, P. Louka, K. Tsanakas, D. E. Tsesmelis, V. Krassanakis, G. P. Petropoulos, V. Pappas, and C. Chalkias. Geoinformation technologies in support of environmental hazards monitoring under climate change: An extensive review. *ISPRS International Journal of Geo-Information*, 10(2):94, 2021.

[8] A. J. Ademusire and J. A. Eyinade. Geospatial machine learning for flood risk assessment in contrasting physiographic environments. *World Journal of Advanced Engineering Technology and Sciences*, 16(1):436–446, 2025.

[9] E. H. Ghile, H. Shirakawa, and H. Tanikawa. Application of gis and machine learning to predict flood areas in nigeria. *Sustainability*, 14(9):5039, 2022.

[10] A. S. Sikiru, O. S. Aisha, H. Zaid, and E. E. Timothy. Artificial neural network and ensemble models for flood prediction in north-central region of nigeria. *Journal Of Engineering Research And Reviews*, 1(3):80–87, 2024.

[11] P. Eguagie-Suyi, B. Dada, and E. C. Okogbue. Machine learning based drought prediction using the standardized precipitation evapotranspiration index (spei) in kebbi state, nigeria. *Journal of Atmospheric Science Research*, 8(2):1–21, 2025.

6 Apendix

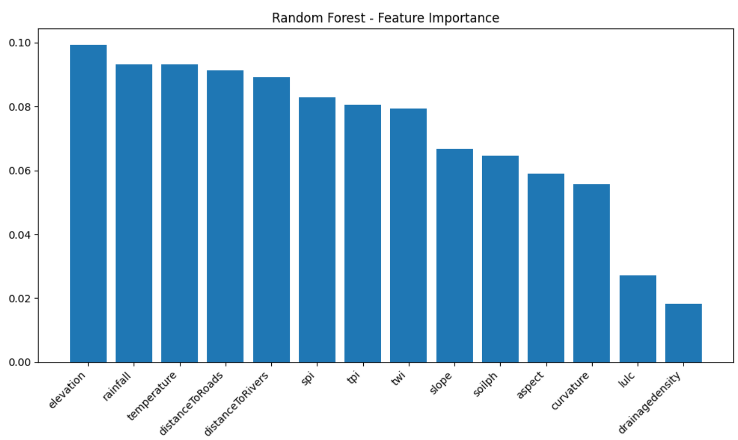

Figure 4: Feature importance of the Random Forest model