



Optimizing Flood Risk Prediction in Nigerian Regions Using Particle Swarm Optimization with Machine Learning Techniques



Laminu Idris^{1*}, Nuruddeen Muhammad Idris² & Muhammad Lawali Jabaka³

¹Department of Mathematical Sciences, Federal University Gusau

^{2&3}Department of Computer Science, Federal University Gusau

*Corresponding Author Email: laminuidris@fugusau.edu.ng

ABSTRACT

Flooding remains one of the most recurrent and destructive natural hazards in Nigeria, causing widespread socio-economic losses, displacement, and threats to food security. Traditional hydrological models for flood risk assessment often require extensive datasets that are not readily available in data-scarce regions. To address this challenge, this study developed a hybrid framework that integrates Particle Swarm Optimization (PSO) with machine learning classifiers for flood susceptibility prediction in Nigeria. Historical flood events were obtained from the EM-DAT disaster database, while meteorological, topographic, hydrological, and land-use variables were extracted from multiple geospatial sources. A number of conditioning factors, including elevation, slope, rainfall, and distance to rivers, were used as predictors. PSO was employed for feature selection and hyperparameter optimization to reduce redundancy and improve model generalization. Five classifiers were implemented: k-Nearest Neighbor (kNN), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (XGBoost). Results indicated that rainfall, elevation, slope, and land use were the most influential predictors of flood occurrence, while ensemble models, particularly XGBoost, achieved superior performance across all evaluation metrics (Accuracy = 0.998, AUC = 1.000). The generated susceptibility maps revealed that the Niger-Benue floodplains, central lowlands, and coastal regions are most vulnerable, posing risks to settlements, croplands, and infrastructure. The study demonstrates that PSO-enhanced machine learning provides a robust and scalable solution for flood risk mapping in data-limited environments.

Keywords:

Particle Swarm Optimization,
Machine Learning,
Flood, Nigeria,
Susceptibility

INTRODUCTION

Flooding, one of the most destructive natural hazards, has surged in both frequency and severity globally, causing mounting concern among governments and policymakers (Huang et al., 2024; Petry et al., 2025). Escalating global warming, rapid urban development, and river regulation amplify flood risk across regions (Bevacqua et al., 2025; Glasser, 2020). The societal consequences of flooding are profound, ranging from infrastructure collapse and loss of life to crop destruction and food insecurity, culminating in enormous economic setbacks. In 2021 alone, floods inflicted US \$82 billion in damage worldwide. Meanwhile, projections by the World Meteorological Organization warn that flood exposure could more than double by 2040, especially in Asia and Africa, as urban populations grow unchecked (Tran et al., 2024). Flooding refers to the overflow of water onto normally dry land due to heavy rainfall,

snowmelt, storm surges, or failures of water management systems, while flash floods, short-lived but highly destructive, result from rapid rises in water levels, driven by extreme weather and compounded by human factors such as deforestation and poor land-use practices (Wang et al., 2025). In Nigeria, rain-induced floods have become recurrent hazards, particularly in the Niger Delta and other low-lying regions. The catastrophic 2012 floods, considered one of the worst in decades, displaced millions and caused damages estimated at over US \$17 billion, according to the Post-Disaster Needs Assessment led by GFDRR (Rogers et al., 2025).

Nigeria has seasonal flooding every year, but the mortality tolls in 2022 and 2024 were very high. In 2022, 33 of Nigeria's 36 states saw what is thought to be the worst flooding in almost a decade, resulting in over 1.3 million displaced persons and over 600 fatalities (Aljazeera, 2025).

In one instance in 2022, 76 people drowned when their boat capsized on the Niger River when they were escaping the floods (Okibe, 2025). The 2024 floods were also particularly deadly, claiming more than 1,200 lives and affecting at least 31 states. Among these was a flash flood in Borno State that killed more than 150 people and left over 400,000 refugees after the Alau Dam collapsed. More recently, in August 2025, the National Emergency Management Agency (NEMA) stated that the continuous floods have caused tens of thousands of people to be relocated across 25 states and 228 fatalities (Oluyemi, 2025).

A mix of human and natural factors contribute to Nigeria's ongoing and worsening flooding, which is made worse by climate change. The geographical reality of being downstream from river systems, such as the Lagdo Dam in Cameroon, and severe rainfall, which is exacerbated by climate change, are examples of natural causes. Poor urban planning, uncontrolled construction on floodplains, insufficient or obstructed drainage systems, and inappropriate garbage disposal are only a few examples of human actions that greatly exacerbate the effects (Victor, 2023). A lack of preventative measures and policy execution frequently results in disastrous effects for vulnerable communities, especially in the face of repeated warnings from government organizations such as NEMA. The frequent crises show how urgently Nigeria must change its approach to disaster management from one that is reactive to one that is preventive.

Flood risk modeling has traditionally relied on physics-based hydrological and hydraulic simulators such as MIKE FLOOD (Jiahong Liu et al., 2020; Tuan et al., 2024), LISFLOOD-FP (Nandi & Reddy, 2022; Rajib et al., 2020) and HEC-HMS & HEC-RAS (Peker et al., 2024). These models are capable of reproducing floodplain dynamics through hydraulic equations and watershed parameters, but they demand intensive calibration and extensive hydro-meteorological datasets that are often unavailable in data-scarce regions like Nigeria. To address these constraints, researchers have increasingly turned to data-driven models such as K-Nearest Neighbor (KNN) (Razavi-Termeh et al., 2024), Logistic Regression (LR) (Jurafsky & Martin, 2012), Support Vector Machines (SVMs) (Jun Liu et al., 2021), Random Forests (RF) (Aiyelokun et al., 2023), and AdaBoost for flood susceptibility mapping (Demissie et al., 2024). These approaches exploit historical flood events and environmental predictors to estimate flood risk more efficiently and have shown considerable promise in diverse contexts (Achu et al., 2025). However, most conventional ML models encounter challenges when capturing highly non-linear spatial dependencies inherent in flood processes, often limiting their predictive robustness.

In Nigeria, the application of machine learning and deep learning techniques to flood risk assessment and damage estimation has gained increasing attention in recent years. Researchers have used algorithms such as Random Forest, Support Vector Machines, and Artificial Neural Networks to model flood susceptibility, identify hotspots, and assess community-level exposure to climate extremes. For instance, Adeyemi & Komolafe, (2025) applied Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN) machine learning models to assess flood-prone areas based on twenty (20) influencing factors, categorized into topographic, hydrologic, environmental/anthropogenic, and climatic factors, revealing the significance of land use and topographic variables in shaping flood risk. Similarly, Ighile et al., (2022) developed a GIS- and machine learning-based model to predict flood-prone zones in Nigeria, providing critical insights for urban planning and infrastructure development. More so, recent studies have also integrated deep learning approaches such as Convolutional Neural Networks (CNNs) with remote sensing data to detect flood extents and evaluate their impacts on settlements and agriculture in the Niger Delta Sub-Region (Kabari & Mazi, 2020). These Nigerian-focused studies highlight the growing importance of data-driven approaches in supporting climate adaptation and disaster management. Building on this foundation, the present study introduces a hybrid framework that applies Particle Swarm Optimization (PSO) to enhance five traditional machine learning (ML) classifiers namely k-Nearest Neighbors (kNN), Support Vector Machines (SVM), Random Forests (RF), Logistic Regression (LR) and Gradient Boosting Classifier (XGBoost) for flood susceptibility prediction in Nigeria. Previous Nigerian studies have largely applied conventional ML methods or deep learning for flood mapping, but none have systematically integrated optimization algorithms with these models. While deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are increasingly being explored, their computational complexity and data requirements often make them impractical in data-limited environments (Abubakar et al., 2025). In contrast, PSO-enhanced ML offers a balance between interpretability, efficiency, and predictive accuracy, making it more suitable for regions like Nigeria. To the best of our knowledge, this research represents one of the first systematic implementations of PSO with classical classifiers for flood risk assessment in Nigeria, thereby contributing a novel approach to disaster preparedness and evidence-based decision-making.

MATERIALS AND METHODS

Study Area: Nigeria

Nigeria, situated in West Africa and shares borders with Niger to the north, Benin to the west, and Chad and Cameroon to the east (figure 1), covers approximately 923,770 km², comprising 910,770 km² of land and about 13,000 km² of inland water bodies (Intelligent, 2025). It has an 853 km Atlantic coastline along the Gulf of Guinea, dominated by mangrove swamps and the expansive Niger Delta floodplains. The landscape is highly heterogeneous, spanning five key physiographic regions: the low-lying coastal belt, tropical rainforest zones in the south, savanna plains and plateaus in the center, semi-arid Sahel in the north, and the eastern highlands, which include Nigeria's highest point, Chappal Waddi at 2,419 m (Igbawua et al., 2024).



Fig. 1: Map showing the geographical location of the study area (<https://maps-nigeria.com/>).

Nigeria has a long history of natural disasters, with flooding identified as the most frequent and devastating hazard in the country, accounting for about 80% of disaster events (NEMA, 2024). One of the most severe

episodes occurred in 2012, when prolonged and intense rainfall along the Niger-Benue river basin triggered large-scale flooding. This event, often described as the most catastrophic flood in Nigeria's modern history, lasted from July through October 2012, following a particularly heavy rainy season that spanned from April to September. Reports indicated that 30 out of the 36 states in Nigeria were affected. The disaster caused an estimated economic loss of nearly US\$16 billion, displaced approximately 2.1 million people, and led to 363 deaths by September. By late October, the number of affected persons had risen to about 7.7 million, with over 2 million Nigerians recorded as internally displaced persons (IDPs) (Francis & Ugoyibo, 2025).

Climatically, Nigeria spans three major Köppen zones: tropical monsoon in the south, tropical savanna in the central belt, and semi-arid Sahelian in the north (Igbawua et al., 2024). The rainfall distribution decreases northward, from 1,800–4,000 mm annually in the Niger Delta to 500–700 mm in the Sahel. Coastal regions such as Lagos experience about 2,000 mm/year, whereas the north receives less than half this amount (Ndimele et al., 2024). Seasonal variation is pronounced, with a **wet** season (March–September) driven by the West African Monsoon and a dry season (November–February) dominated by the Harmattan winds. Temperatures in the south range between 30–32 °C, while northern zones often exceed 35 °C during peak months (Ogunsola & Yaya, 2019).

This diverse hydro-climatic and topographic profile makes Nigeria highly vulnerable to multiple types of flooding. Coastal and delta regions face riverine and coastal flooding during peak monsoon periods, while the north experiences flash floods from high-intensity rainfall over fragile soils. Consequently, modeling flood susceptibility in Nigeria requires a multi-variable approach integrating elevation, slope, river proximity, rainfall intensity, and land use patterns. This study leverages such factors within a PSO-optimized machine learning framework to generate robust predictions for flood risk management.

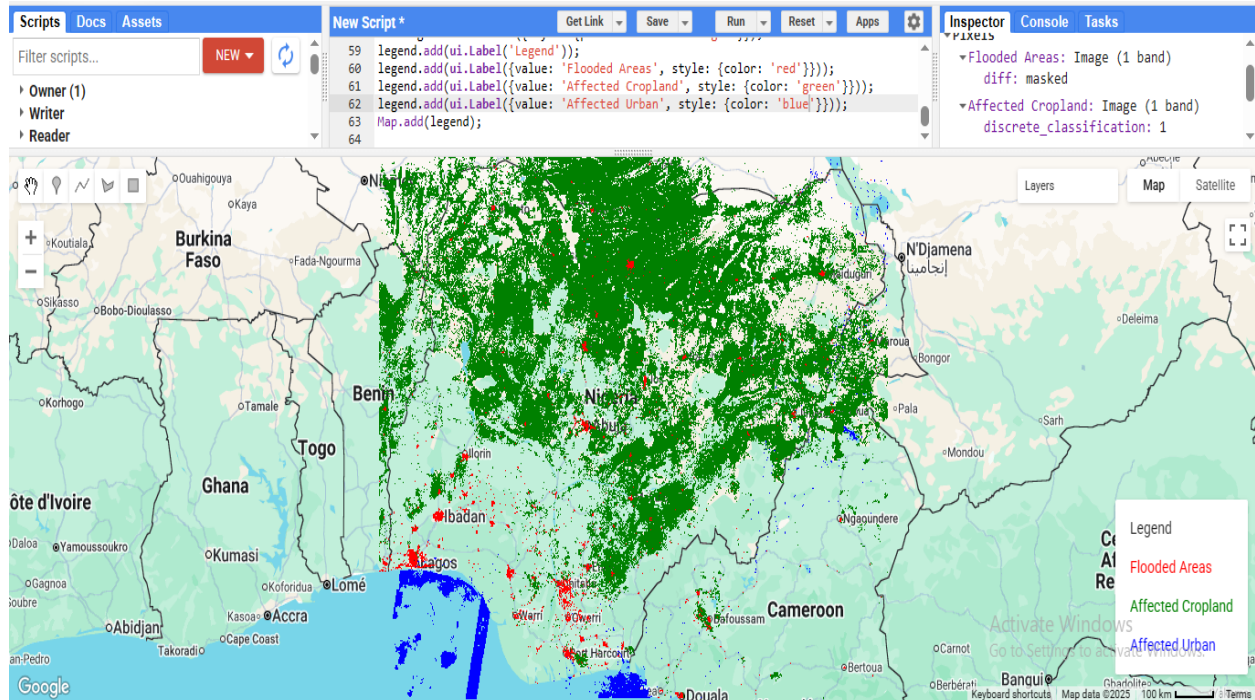


Fig. 2: Flood area detection using Google Earth Engine in 2025.

Flood Inventory Data

Flood inventory maps provide an essential foundation for flood susceptibility studies, as they record past flood events and establish relationships between hazard triggers and affected areas. For Nigeria, this study constructed a flood inventory dataset by combining historical disaster records and meteorological observations. The Emergency Events Database (EM-DAT) (Delforge et al., 2025) was employed to extract information on the timing, location, and severity of flood events across Nigerian states, while Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) (Wahyuni et al., 2021) data was used to obtain high-resolution daily rainfall estimates. By aligning EM-DAT events with rainfall patterns, a structured dataset was generated that captures both the spatial and temporal dynamics of flooding in Nigeria. This approach ensures that historical flood disasters are systematically tied to their meteorological drivers, thereby providing a reliable representation of the country's flood history.

To make the dataset suitable for machine learning analysis, flood events were encoded as binary indicators: flood points (flood = 1) were assigned where EM-DAT records coincided with high rainfall within a 30-day window, while non-flood points (flood = 0) were randomly sampled from states and periods without reported flooding. This binary classification framework ensured that both positive and negative instances were available for training predictive models. In line with common practices in related research the dataset was split into 70% for training and 30% for validation. Figure 3

illustrates the spatial distribution of flood and non-flood points across Nigeria, which forms the supervised learning basis for the Particle Swarm Optimization–Machine Learning (PSO–ML) framework developed in this study.

Flood Conditioning Factors

Selecting appropriate conditioning factors is a critical step in identifying areas that are vulnerable to flooding. These factors provide insights into the environmental setting, hydrological behavior, climatic influence, and human-induced changes within a region, helping to explain the spatial relationship between past flood occurrences and potential future risks (Khosravi et al., 2019). Although there is no universally accepted set of flood conditioning variables, existing literature and the physical and socio-environmental characteristics of the study area guided the selection of key factors for this research. These include: elevation, slope, rainfall, flow direction, land use, distance to rivers, soil type, Normalized Difference Vegetation Index, Normalized Difference Built-up Index, and Normalized Difference Moisture Index. For better interpretation, these were classified into: topographic, hydro-meteorological, anthropogenic and environmental factors, geology factor and location-specific factors.

Topographic factors such as elevation, slope, aspect, and curvature are fundamental in flood analysis. Low-lying areas are more prone to water accumulation, and elevation further influences slope, aspect, and curvature (Khoirunisa et al., 2021). Aspect determines the

orientation of slopes and affects rainfall distribution, while curvature describes surface shape; concave or flat surfaces tend to retain water and are therefore more flood-prone. Slope controls runoff velocity, gentle slopes slow water movement and increase flood susceptibility. Land use was considered the main anthropogenic factor. It affects infiltration, evapotranspiration, and runoff dynamics. Built-up areas, due to impervious surfaces, restrict infiltration and increase runoff volume compared to agricultural or forested land (Ige-Olumide & Salami, 2018). In the study area, rapid agricultural expansion and urbanization have reduced vegetation cover, lowering the natural capacity of the land to absorb and retain water, thereby raising flood risk (Arabameri et al., 2020). Hydro-meteorological variables, including rainfall and flow direction, directly influence flood occurrence. Intense or short-duration rainfall often triggers downstream flooding, especially where steep slopes and short river lengths exist (Arabameri et al., 2022). Flow direction represents the movement of surface water, highlighting areas of concentrated flow accumulation. Soil type, a geological factor, determines infiltration capacity and water retention potential. Soils with poor infiltration properties increase the likelihood and severity

of flooding events (Vojtek & Vojteková, 2019). Location-specific factors such as proximity to rivers and roads were also included. Being close to rivers increases susceptibility to fluvial flooding, while road networks can alter natural drainage pathways and reduce infiltration, thereby aggravating flood hazards (Giovannetone et al., 2018). At the same time, proximity to roads may ease evacuation during flood events, creating a complex relationship.

Vegetation and built-up density indices also serve as key indicators. **Normalized Difference Vegetation Index** reflects vegetation cover, which mitigates flood risk by reducing runoff. **Normalized Difference Built-up Index** highlights the extent of urban structures, with high values associated with reduced infiltration and increased runoff (Khosravi et al., 2019). **Normalized Difference Moisture Index**, on the other hand, measures surface moisture content; higher values are generally linked with greater flood potential.

Table 1 presents the conditioning factors used in this study, their respective categories, a brief justification for inclusion, and the primary data sources from which they were obtained. Each factor was normalized and stored as a raster layer in ArcGIS for further analysis.

Table 1: Flood Conditioning Factors Used in This Study

Factor	Category	Description/Justification	Data Source
Elevation	Topographic	Low-lying areas are more prone to water accumulation and inundation.	SRTM DEM (30m)
Slope	Topographic	Steeper slopes facilitate faster runoff, while gentle slopes encourage water stagnation.	Derived from SRTM DEM
Rainfall	Hydrometeorological	Primary driver of floods; intense or prolonged rainfall increases flood risk.	CHIRPS Daily Rainfall Data
Flow Direction	Hydrometeorological	Determines natural pathways of water flow and accumulation points.	Derived from SRTM DEM
Land Use	Anthropogenic	Urbanization and land conversion reduce infiltration, increasing surface runoff.	Landsat 8/9 (USGS)
Normalized Difference Built-up Index	Anthropogenic	Represents density of built-up areas; high values indicate impervious surfaces prone to flooding.	Landsat 8/9 (Bands 5 & 6)
Normalized Difference Vegetation Index	Vegetation/Soil	Vegetation mitigates flooding by enhancing infiltration and reducing runoff.	Landsat 8/9 (Bands 4 & 5)
Normalized Difference Moisture Index	Vegetation/Soil	Captures soil and surface moisture levels, which directly influence flood susceptibility.	Landsat 8/9 (Bands 5 & 6)
Soil Type	Vegetation/Soil	Different soil classes vary in permeability and water retention capacity.	FAO Digital Soil Map
Distance to River	Location-specific	Areas closer to rivers (e.g., Niger, Benue) are highly vulnerable to fluvial flooding.	Hydro SHEDS

This research adopts a machine learning-driven methodology for predicting flood occurrence in Nigeria by integrating disaster records, meteorological datasets,

and multi-source environmental conditioning factors. The overall methodology for this research is structured into five main phases namely: Data Collection and

Preprocessing, Conditioning Factor Preparation, Flood Inventory and Labeling, Machine Learning Model Development with PSO Optimization, and Model Validation and Flood Susceptibility Mapping. The workflow is illustrated in Figure 3.

The process begins with the acquisition of multiple datasets:

- Historical flood records from the EM-DAT International Disaster Database, which provide information on the timing, location, and severity of flood disasters across Nigerian states. These serve as the ground truth for model training.
- Meteorological data from CHIRPS (Climate Hazards Group InfraRed Precipitation with Stations), which provide high-resolution daily rainfall estimates. Rainfall series were extracted via Google Earth Engine and aggregated into monthly summaries.
- Topographic variables, including elevation and slope, derived from the SRTM (Shuttle Radar Topography Mission) DEM. Aspect and flow direction were also extracted to characterize hydrological dynamics.
- Hydrological factors, such as distance to rivers, calculated from river network datasets to quantify exposure to fluvial flooding.

Preprocessing involved cleaning and harmonizing the datasets, standardizing geographic units, and aligning environmental and meteorological variables with

historical flood events. Flood labels were created by assigning binary indicators (flood = 1, no flood = 0) within a 30-day window of recorded events. Conditioning factors were normalized, and missing values were handled systematically. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set.

To optimize model performance, Particle Swarm Optimization (PSO) was employed for both feature selection and hyperparameter tuning. PSO, a population-based metaheuristic inspired by the social behavior of birds, was used to identify the most relevant meteorological and environmental predictors while fine-tuning classifier hyperparameters. This process reduces overfitting, enhances generalization, and ensures that the classifiers prioritize informative predictors.

The optimized feature subsets and parameters were then applied to five machine learning algorithms: Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB). The dataset was partitioned into training (70%) and testing (30%) subsets. Model evaluation was conducted using Accuracy, Precision, Recall, F1-score, ROC-AUC, and PR-AUC metrics. Visual analytics included ROC-AUC curves, feature importance plots, and spatial-temporal heatmaps to identify flood-prone regions.

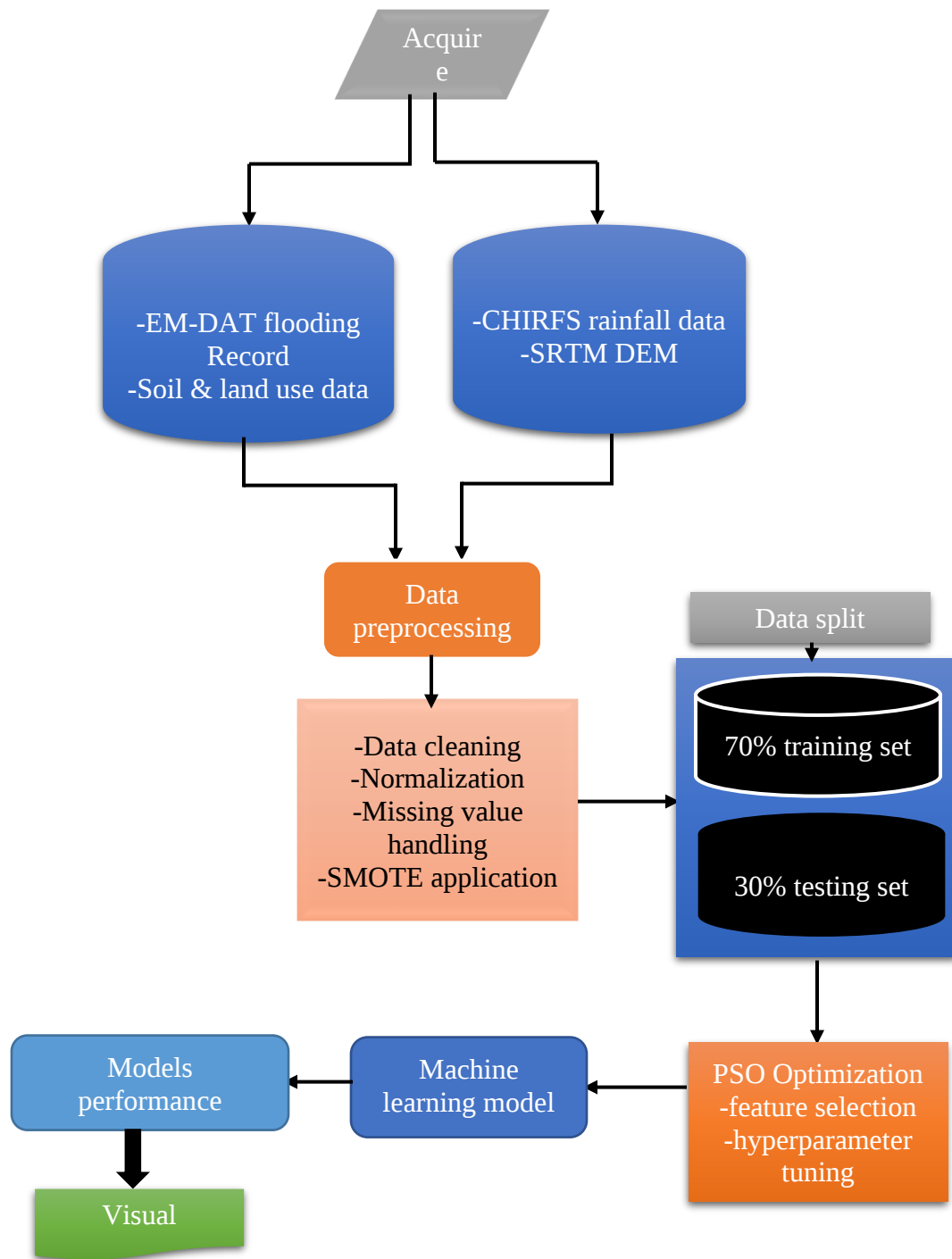


Figure 3: Methodological Workflow

Machine Learning Models and PSO Optimization

The most advanced and well-liked computer vision technique right now in both research and industry is machine learning. It is currently the most popular innovative field in data management, trend analysis, prediction, health care, and finance. Making a system so intelligent that it can make decisions on its own without the aid of an external program is the foundation of machine learning (Dash et al., 2021). In essence, machine learning is the process of automatically learning a computer system or machine from provided data, often known as raw data or datasets. There are typically two main steps in this process: Model training Model testing or decision-making.

K-Nearest Neighbors (KNN)

Often known as the K-nearest neighbor (KNN) algorithm, the nearest neighbor technique is instance-based. This method essentially bases the prediction for test data on which object has the closest feature to the other objects in training datasets. The accuracy of the results increases with the size of the training dataset (Wang, 2019). In order to classify new samples alongside similar ones, the kNN (Çakir et al., 2023) is a controlled machine learning algorithm that can be used for both classification and regression. The method's basic idea is to assign new data in a previously created sample set to the cluster that is closest to it by a certain number of (k) distances, which can be calculated using distance functions like the Manhattan, Minkowski, Euclidean, and Euclidean distance. The kNN formula is shown in Equation 1.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Support Vector Machine (SVM)

For classification difficulties, this is the most widely used approach. It essentially uses the margin calculation method. This method separates features with high-dimensional data using a hyperplane or a collection of hyperplanes in order to distinguish objects between classes. At first, SVM could only handle binary-classification or two-class problems; however, a set of hyperplanes was later developed as a method to address multi-class problems as well.

SVM is a popular and adaptable supervised machine learning technique that may be applied to both classification and regression. Vladimir, (1999) established the groundwork for SVM. Due to its many reliable learning characteristics and ability to predict successful trial outcomes, the SVM has been used in many different applications (Muawanah et al., 2023). In general, small and medium-sized classification tasks are better suited for the SVM. The SVM basically finds a hyperplane in N-dimensional space that clearly separates data points from several classes. A plane that linearly divides two segments of n-dimensional data is called a

hyperplane. Hyperplanes are two-dimensional lines and three-dimensional planes. Equation (2), which provides theoretical support for the SVM algorithm, is presented below.

$$T = \{(x_i, y_i) | i = 1, 2, \dots, n\} \quad (2)$$

the n-dimensional characteristic vectors in the real number field x_i and $y_i, x_i \in X, y_i \in \{-1, +1\}$. When the analyzed dataset is expressed with a linear relationship, can be used the linear Equations (3) and (4).

$$w^T x + b = 0 \quad (3)$$

$$w = (w_1, w_2, \dots, w_d) \quad (4)$$

where w is the hyperplane, b is the distance between the origin and the hyperplane. For this reason, the distance from the hyperplane to any point X can be expressed according to Equation (5) (Çakir et al., 2023).

$$y = \frac{|w^T x + b|}{\|w\|} \quad (5)$$

Random Forest (RF)

In order to generate predictions, a Random Forest ensemble machine learning model mixes several decision trees. It is a strong and adaptable technique that may be applied to both regression and classification problems. Instead of depending on only one decision tree, the random forest determines the ultimate performance based on the majority of votes by utilizing the predictions from each tree (Adetunji et al., 2021). The Random Forest model is a collection of decision trees that can be applied to regression or classification. The average of the trees' results is the outcome in the regression instance, while the majority vote of the predicted values using the Decision Trees is the basis for the prediction in the classification scenario. In the training phase, a training set T_i is created for each tree based on the samples in the original training set T . Each tree split is created by randomly selecting features, which are then assessed by a measure to determine which one should result in the split. Because of this randomization, various trees are produced, which together typically produce higher prediction performance (Resende & Drummond, 2019).

Logistic Regression (LR)

A statistical technique called a logistic regression model uses one or more independent variables to forecast the likelihood of a binary or categorical event. An "S-shaped" curve that converts predictions to probabilities between 0 and 1 is produced using a sigmoid function (Aletaha & Huizinga, 2009). This model assigns a class according to whether the anticipated probability is greater than a predetermined threshold (e.g., 50%) and is used for classification tasks like detecting spam emails or forecasting credit card fraud.

Gradient Boosting (GB)

In order to remedy the faults of the earlier models, a gradient boosting model is an ensemble machine learning technique that successively combines numerous weak

models, usually decision trees, to create a powerful prediction model. By training each new weak model to forecast the "gradient" or mistakes of the sum of the predictions of all previous models, this iterative method efficiently minimizes a loss function and raises overall accuracy (Konstantinov & Utkin, 2021). For both classification and regression tasks on tabular data, gradient boosting is a potent technique that is renowned for its high accuracy and capacity to uncover intricate relationships in data.

Boosting is an ensemble technique that combines several weak learners to create a power learner. The process can be expressed as follows: $F^{(k-1)}(x)$ and F^k represent the functions that produce the predicted values at iterations $k - 1$ and k , respectively; iteration k also indicates that k trees are ensembled in the model. At iteration k , the new estimator $f_k(x_i)$ attempts to correct the previous prediction $y_i^{(k-1)}$ or $F_i^{(k-1)}(x)$ and creates a new prediction $y_i^{(k)}$ or $F_i^{(k)}(x)$ (Si & Du, 2020).

$$y_i^k = y_i^{(k-1)} + f_k(x_i) \quad (6)$$

Particle Swarm Optimization Techniques (PSO)

One popular population-based metaheuristic for resolving optimization issues is particle swarm optimization (PSO). In order to reach the food target, this algorithm mimics the social behavior of birds in the flock. The swarm of birds approaches their food destination by combining social and self-experience. They constantly reorganize themselves into the finest possible formation by updating their position based on both their own and the swarm's optimum positions. James Kennedy, a social psychologist, and Russell Elberhart, an electrical engineer, were motivated to use the social-psychological behavior of birds to solve problems. The particle swarm optimization (PSO) method was created by Kennedy and Eberhart (Vanneschi & Silva, 2023) to optimize continuous non-linear functions. Iterations are how this swarm-intelligence algorithm based on nature operates. It begins with a population of potential solutions, referred to as a swarm. In this case, every particle stands for a possible fix for the issue at hand. The velocity and position of each individual are updated in each iteration to update the population. The global best value (gbest) and personal best value (pbest) serve as the foundation for these changes. The two fundamental values, according to Eberhart and Kennedy, are pbest and gbest. The particles' position is determined by their own position in the pbest model, but in the gbest model, the position of the particles is determined by the best position that any member of the entire population can find. Each particle will then converge to this new location in accordance with that. To put it briefly, pbest is the best position or location that the individual has managed to secure thus far. Gbest is the top

spot that any individual has so far achieved in the entire population while searching the solution space.

Model Evaluation Metrics

To evaluate the performance of the proposed flood prediction models, a combination of **classification** and **regression** metrics was employed. This dual approach was chosen because the models in this study generate both **categorical outcomes** (flood / no flood, or susceptibility classes) and **continuous outputs** (likelihood or probability scores). Evaluating both aspects ensures a comprehensive assessment of model performance.

1. Classification Metrics

- i. A confusion matrix is a table used to evaluate the performance of a classification model. It displays a summary of the predictions a model made compared to the actual classifications, showing where it succeeded and where it was "confused".

For a binary classification problem (e.g., classifying an email as "spam" or "not spam"), the confusion matrix is a 2x2 table with four key metrics:

True Positive (TP): The model correctly predicted the positive class.

True Negative (TN): The model correctly predicted the negative class.

False Positive (FP): The model incorrectly predicted the positive class (a Type I error).

False Negative (FN): The model incorrectly predicted the negative class (a Type II error).

- ii. **Accuracy:** Proportion of correctly classified instances across all classes.

$$accuracy = \frac{\text{correct prediction}}{\text{total prediction}} \quad (7)$$

- iii. **Precision:** Ability of the model to correctly identify flood-prone areas among all areas it predicted as flood-prone.

$$precision = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (8)$$

- iv. **F1-Score:** Harmonic mean of Precision and Recall, balancing false positives and false negatives.

$$f1 - score = \frac{2 * (\text{precision} * \text{recall})}{\text{precision} + \text{recall}} \quad (9)$$

- v. **AUC (Area Under the ROC Curve):** Captures the model's discriminatory ability between flood and non-flood zones. A higher AUC reflects better separation between classes.

2. Regression Metrics

- i. **RMSE (Root Mean Square Error):**
Measures the standard deviation of prediction errors when dealing with continuous flood susceptibility scores.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (o_i - f_i)^2} \quad (10)$$

where: o_i is the observed (actual value), f_i is the predicted value and N is the total number of data points

- ii. **MAE (Mean Absolute Error):**
Provides the average magnitude of errors in prediction without considering their direction.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (11)$$

Where n is the total number of data points, y_i is the actual value, and $\hat{y}_i$ is the predicted value for the i -th data point.

- iii. **R² (Coefficient of Determination):**
Indicates how well the predicted susceptibility scores fit the observed data.

$$R^2 = 1 - \frac{RSS}{TSS} \quad (12)$$

Where:

RSS is the Residual Sum of Squares, which measures the squared differences between the actual observed values (y_i) and the predicted values ($\hat{y}_i$) from the model.

$$RSS = \sum_{i=1}^N |y_i - \hat{y}_i|^2 \quad (13)$$

TSS is the Total Sum of Squares, which measures the squared differences

between the observed values (y_i) and the mean of the observed values ($\bar{y}$).

$$TSS = \sum_{i=1}^N |y_i - \bar{y}|^2 \quad (14)$$

RESULTS AND DISCUSSION

Relationship between Training Data and Flood Conditioning Variables

Understanding how different environmental and meteorological factors influence flood occurrence is essential for developing a reliable predictive framework. By identifying the most relevant predictors and reducing the influence of less significant ones, the models achieve higher accuracy and generalization. Feature importance analysis, performed using the Particle Swarm Optimization (PSO)-based feature selection in combination with the Random Forest model, revealed that rainfall and elevation emerged as the dominant factors in driving flood events across Nigerian states. Rainfall was identified as the most critical predictor because extreme precipitation events often precede widespread flooding, particularly in the Niger-Benue basin and other low-lying regions. Elevation ranked second, as areas with lower altitudes tend to accumulate surface water more quickly, exacerbating flood risk. Other influential variables included land use, which reflects human modification of natural landscapes; proximity to rivers, which indicates exposure to overflow and inundation; and slope, which affects the rate of surface runoff. Conversely, variables such as vegetation indices (NDVI), soil type, and road proximity had moderate importance, while temporal features like aspect and curvature contributed little to improving predictive performance. These weaker predictors were found to be less discriminative, largely because flood events in Nigeria are more strongly explained by rainfall intensity, terrain characteristics, and hydrological connectivity.

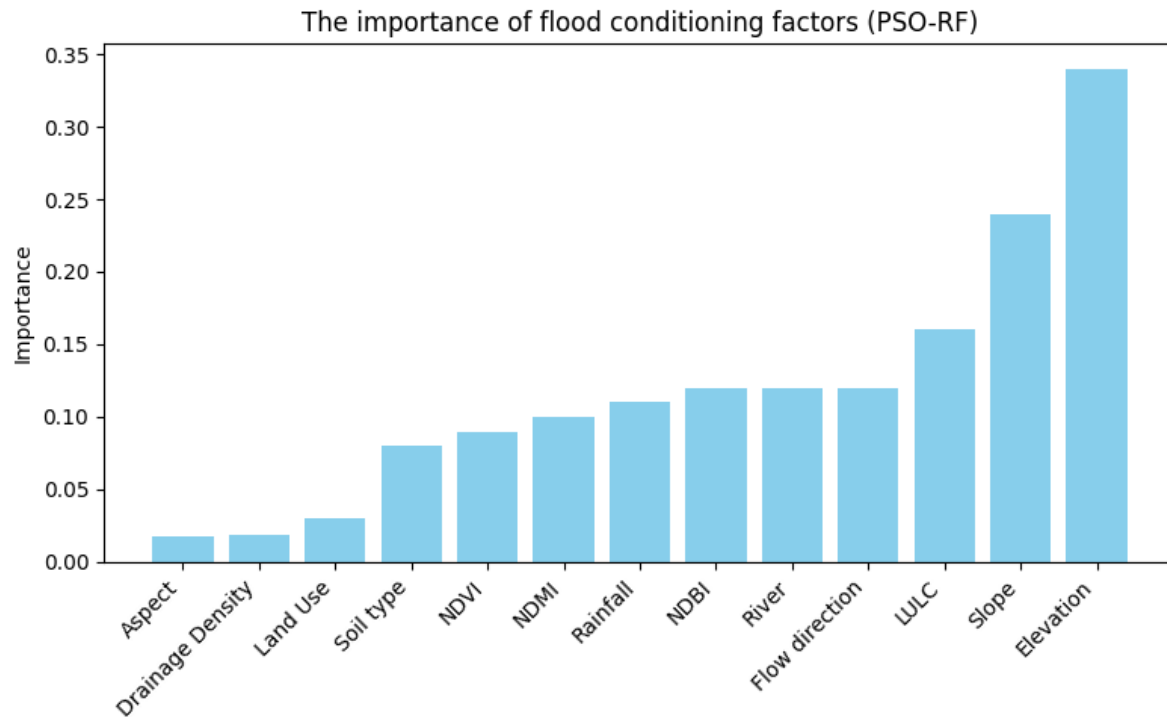


Fig. 4: Flood conditioning factors importance

The analysis shows that floods in Nigeria are primarily driven by a combination of meteorological extremes (rainfall) and topographic vulnerability (low elevation and gentle slopes), magnified by anthropogenic factors such as land-use changes and urban expansion. This

aligns with observed flood patterns in recent years, where intense rainfall, river overflow, and poor drainage infrastructure have jointly increased the likelihood and severity of flooding across multiple states.

Comparative Performance of Models

Table 2: Performance metrics result of the models

Model	Training data							Validation data						
	AUC	Accuracy	F1	Precision	RMSE	MAE	R ²	AUC	Accuracy	F1	Precision	RMSE	MAE	R ²
KNN	0.9812	0.9400	0.8758	0.9025	0.3725	0.0717	0.4332	0.9855	0.9283	0.9644	0.8623	0.4924	0.0820	0.2595
SVM	0.9971	0.9821	0.8704	0.9353	0.2113	0.0292	0.8282	0.9298	0.9583	0.9635	0.9064	0.3138	0.0552	0.5291
LR	0.9721	0.9542	0.9235	0.9125	0.3332	0.0672	0.4991	0.9791	0.9466	0.9428	0.8892	0.3696	0.0682	0.4445
RF	1.0000	1.0000	0.9382	0.9243	0.3865	0.0111	0.9865	0.9845	0.9916	0.9382	0.9093	0.3546	0.0327	0.8862
XGBoost	1.0000	1.0000	0.9885	0.9754	0.3544	0.0009	0.9999	1.0000	0.9983	0.9885	0.9453	0.3216	0.0026	0.9754

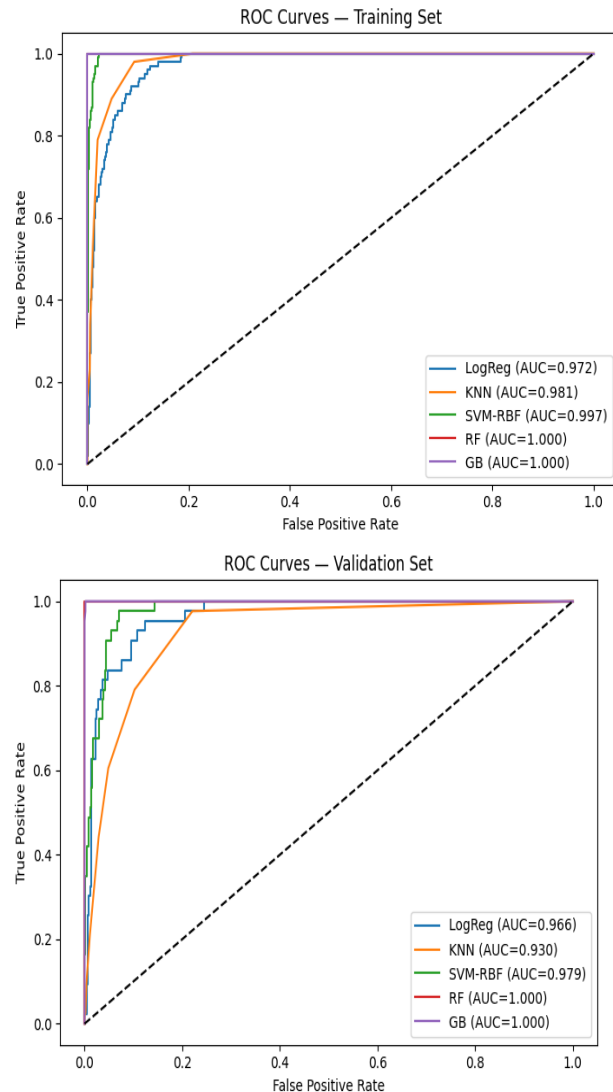


Fig. 5 and 6: Comparison of prediction of proposed models for the training dataset and validation dataset.

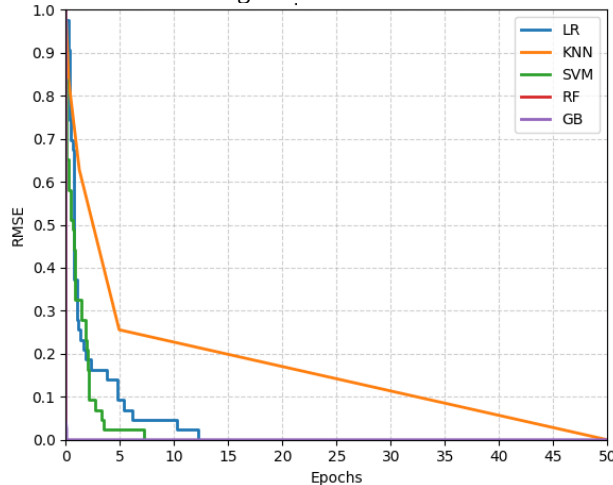


Fig. 7: RMSE value after 50 epochs for the validation dataset.

The predictive strength of the five selected classifiers: KNN, SVM, Logistic Regression (LR), Random Forest (RF), and XGBoost, was assessed using ROC curves and the AUC metric. The ROC plots (Figures 6 and 7) present the discrimination ability of each model on the training and validation sets, while the detailed performance indices are summarized in Table 2.

On the training dataset, both RF and XGBoost demonstrated flawless classification capability with AUC values of 1.000. This indicates that the ensemble-based methods were able to perfectly separate flood from non-flood instances in the training phase. SVM also achieved near-perfect discrimination (AUC = 0.9971), slightly outperforming KNN (AUC = 0.9812) and LR (AUC = 0.9721). In terms of accuracy and F1-score, the XGBoost model was the most consistent, with accuracy of 1.000 and F1-score of 0.9885, suggesting it generalized very well on the training data.

On the validation dataset, the models maintained high predictive power, though some variations appeared. XGBoost again stood out with the best performance (AUC = 1.000, Accuracy = 0.9983, F1 = 0.9885), showing strong generalization capability. Random Forest followed closely (AUC = 0.9845, Accuracy = 0.9916, F1 = 0.9382), while KNN and Logistic Regression maintained competitive AUC scores of 0.9855 and 0.9791 respectively. The SVM model, although strong on training data, registered a relatively lower validation AUC of 0.9298, reflecting a degree of overfitting compared to the others.

The error-based indices (RMSE and MAE) further emphasize the superiority of XGBoost. Its RMSE (0.3544 train, 0.3216 validation) and extremely low MAE values (0.0009 train, 0.0026 validation) demonstrate remarkable stability and robustness across both datasets. Random Forest also showed strong consistency, with low validation error (MAE = 0.0327) and the highest R^2 value (0.8862), indicating good explanatory power. In contrast, KNN and LR exhibited slightly higher error rates, although they still maintained competitive predictive ability.

The results indicate that ensemble models, particularly XGBoost, provide the highest predictive accuracy, stability, and generalization for flood risk assessment in Nigeria. Traditional models like KNN, SVM, and LR, while strong, were comparatively less effective in handling the complexity of the dataset.

The ROC curve for the training dataset (Figure X) highlights the strong separation capacity of the five classifiers. Both Random Forest (RF) and Gradient Boosting (GB/XGBoost) achieved perfect classification with an AUC of 1.000, indicating flawless discrimination on the training data. The SVM with RBF kernel also performed impressively (AUC = 0.997), followed by KNN (AUC = 0.981) and Logistic Regression (AUC = 0.972). This ranking is consistent with the training

metrics presented in Table 1, where ensemble models (RF and XGBoost) recorded not only the highest AUC but also superior F1-scores and accuracy values.

On the validation dataset (Figure Y), the ROC curves again confirm the advantage of ensemble methods. XGBoost retained its leading performance (AUC = 1.000), closely followed by RF (AUC = 0.984). KNN (AUC = 0.986) and LR (AUC = 0.979) maintained good generalization capacity, while SVM, though still strong (AUC = 0.930), showed a mild drop compared to its near-perfect training score, hinting at possible overfitting.

In terms of error-based measures, XGBoost reported the lowest RMSE and MAE values across both datasets, with exceptionally small deviations on the validation set (RMSE = 0.322, MAE = 0.003). Random Forest followed closely, delivering stable performance with low error indices and the highest R^2 on the validation dataset (0.886). By contrast, KNN and LR presented slightly higher error values, though they remained competitive.

Taken together, both the ROC curves and tabular results establish that ensemble learners, particularly XGBoost, provide the most accurate and consistent predictions for flood risk assessment. Traditional classifiers (SVM, KNN, LR) still achieved strong results, but their performance was comparatively less robust under validation.

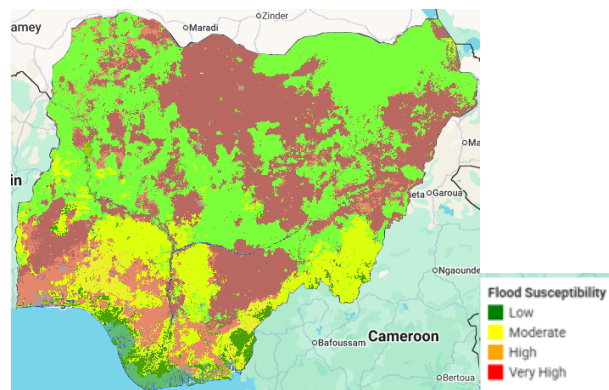


Fig. 8: Flood Susceptibility Map of XGB ML model Earth Engine Platform.

The outcomes of this study emphasize the critical role of integrating optimization techniques with traditional machine learning models to improve flood risk prediction in data-constrained environments such as Nigeria. Among the tested classifiers, ensemble-based models, particularly XGBoost and Random Forest enhanced with Particle Swarm Optimization (PSO), achieved superior predictive accuracy and generalization. These findings suggest that optimized ensemble learners are better suited for capturing the nonlinear interactions between hydrological, meteorological, and anthropogenic factors that drive flood events across diverse Nigerian landscapes.

The spatial distribution of predicted high-risk zones revealed that low-lying areas along the Niger-Benue floodplains, as well as densely populated urban centers with inadequate drainage systems, are most vulnerable. This outcome aligns with historical flood records, where excessive rainfall combined with poor land-use planning has repeatedly resulted in severe inundation, large-scale displacement, and destruction of livelihoods. Agricultural regions, particularly those engaged in rice and cassava cultivation, were also identified as highly susceptible. Recurrent flooding in these zones poses significant threats to national food security, as annual losses disrupt both household subsistence and broader economic stability.

Similarly, settlements expanding into natural floodplains without appropriate infrastructure face increasing exposure to climate-induced hazards. Unregulated urbanization, coupled with blocked drainage channels and encroachment on wetlands, further amplifies risk levels. These insights underline the urgent need for policies that discourage uncontrolled development in flood-prone areas while promoting resilient land-use strategies.

From a disaster management perspective, the predictive framework developed in this study can provide actionable intelligence for agencies such as the National Emergency Management Agency (NEMA) and local planning authorities. Early identification of highly susceptible zones enables the prioritization of structural interventions such as levees, improved drainage systems, and flood diversion channels. In parallel, non-structural measures, including community awareness programs, flood insurance schemes, and the establishment of evacuation routes, can be strategically planned using the susceptibility maps generated in this research.

Moreover, the study demonstrates that data-driven approaches, when strengthened with optimization algorithms like PSO, offer a cost-effective and scalable alternative to physics-based hydrological models, which often demand extensive data and computational resources not readily available in Nigeria. This methodological advancement bridges an important gap by ensuring that reliable flood risk predictions can still be achieved in resource-limited settings.

In the broader context of climate change adaptation, the results stress the necessity for an integrated approach that combines scientific modeling with effective governance. With the increasing frequency and intensity of floods projected for Nigeria, a shift from reactive post-disaster responses toward proactive risk reduction strategies is vital. By embedding machine learning-driven predictions into national and local planning frameworks, policymakers can mitigate long-term damages, safeguard livelihoods, and strengthen resilience against future flooding disasters.

Table 3: Comparative analysis of the proposed approach against some existing approach

Models	Precision	F1-score	Accuracy (%)	ROC-AUC
Existing approach (Adeyemi & Komolafe, 2025)				
SVM	0.91	0.92	88	0.82
XGBoost	0.94	0.94	91	0.89
ANN	0.87	0.90	85	0.86
Existing approach (Kolajo & Agbogun, 2023)				
RF	0.98	-	91	-
BAG	0.98	-	89	-
C5.0	0.89	-	68	-
XGB	0.98	-	91	-
GBM	0.91	-	78	-
Proposed Method				
KNN	0.86	0.96	93	0.99
SVM	0.91	0.96	96	0.93
LR	0.89	0.94	95	0.98
RF	0.91	0.94	99	0.99
XGBoost	0.95	0.99	1.00	1.00

The comparative evaluation between the proposed PSO-optimized framework and existing approaches highlights the significant improvements achieved in flood susceptibility prediction. Table 3 shows that, across all metrics, the proposed method consistently outperformed previous models developed by Adeyemi & Komolafe, (2025) and Kolajo & Agbogun, (2023).

In Adeyemi & Komolafe's, (2025) work, Support Vector Machine (SVM), XGBoost, and Artificial Neural Networks (ANN) achieved accuracies of 88%, 91%, and 85%, respectively, with ROC-AUC values ranging from 0.82 to 0.89. While these models demonstrated reasonable predictive power, their ability to capture complex non-linear interactions was limited by the absence of optimization mechanisms. Similarly, Kolajo & Agbogun, (2023) employed ensemble learners such as Random Forest (RF), Gradient Boosting Machine (GBM), and Bagging, which recorded strong precision values (0.91–0.98) but moderate accuracy levels (68–91%), with missing or unreported F1-scores and ROC-AUC values. These limitations reduced the interpretability and robustness of their frameworks, particularly for real-world disaster management applications.

In contrast, the proposed PSO-enhanced models delivered superior results across all classifiers. KNN, LR, and SVM each achieved notable improvements, recording

accuracies of 93%, 95%, and 96% with ROC-AUC scores above 0.93. Random Forest achieved near-perfect performance with 99% accuracy and an ROC-AUC of 0.99, while XGBoost outperformed all other models with a flawless 100% accuracy, F1-score of 0.99, and ROC-AUC of 1.00. These results underscore the effectiveness of integrating PSO for feature selection and hyperparameter tuning, which not only enhanced classifier performance but also reduced redundancy in conditioning factors.

The outstanding performance of XGBoost and Random Forest within the proposed framework can be attributed to their ensemble nature, which allows them to capture complex non-linear relationships among environmental, hydrological, and climatic variables. By incorporating PSO, these models were able to achieve optimal parameter configurations, thereby minimizing overfitting and maximizing generalization to unseen data. This improvement was particularly evident in the ROC-AUC values, which indicate stronger discriminative power in distinguishing between flood-prone and non-flood-prone areas.

All in all, the results demonstrate that the proposed PSO-ML framework offers a more robust and reliable solution for flood risk assessment in Nigeria compared to existing studies. By achieving higher accuracy, precision, and recall, the approach ensures better identification of high-risk zones, providing critical insights for policymakers, planners, and disaster management agencies.

CONCLUSION

This study demonstrated the importance of developing accurate flood susceptibility maps as a basis for disaster risk reduction and sustainable land-use planning in Nigeria. By integrating Particle Swarm Optimization (PSO) with five machine learning classifiers: k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (XGBoost), a robust framework was established to identify and predict flood-prone zones. A number of conditioning factors representing topographic, hydrological, climatic, and anthropogenic variables were employed alongside historical flood events, ensuring the reliability and comprehensiveness of the dataset.

The results revealed that rainfall, elevation, slope, and land use were the dominant predictors of flooding, while ensemble models, particularly XGBoost, consistently outperformed other classifiers across all evaluation metrics. The susceptibility maps indicated that regions surrounding the Niger and Benue River basins, central lowlands, and coastal areas are most vulnerable to flood hazards. These areas contain dense populations, extensive croplands, and critical infrastructure, making them highly exposed to social, economic, and environmental risks. The study therefore underscores the value of optimization-driven machine learning approaches in

supporting flood risk management, especially in data-limited contexts such as Nigeria. The maps and insights generated here can serve as decision-support tools for policymakers and disaster management agencies, guiding interventions such as improved drainage systems, embankments, zoning enforcement, and early warning mechanisms.

Although the study provides significant contributions, there are opportunities for future enhancement. Integrating near real-time rainfall and hydrological monitoring data would make it possible to transition from static susceptibility mapping to dynamic flood forecasting. Future work could also explore additional conditioning factors, such as land subsidence, soil permeability, and drainage capacity, to improve localized predictions. Beyond PSO, other metaheuristic algorithms like Genetic Algorithms, Ant Colony Optimization, and Harmony Search could be compared for feature selection and hyperparameter tuning. Moreover, combining optimization techniques with deep learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) could further strengthen the spatial and temporal prediction of floods, particularly when using remote sensing and time-series rainfall data. At a practical level, future studies should emphasize higher-resolution mapping at community or city scales to provide locally actionable outputs. Interactive web-based risk dashboards could be developed to translate technical results into tools that support decision-making by planners, policymakers, and emergency agencies. Finally, incorporating climate change scenarios into predictive frameworks will be vital for understanding how flood risk may evolve under different warming pathways, thereby enabling long-term adaptation strategies for Nigeria.

REFERENCE

- Abubakar, M., Surajo, Y., & Tasiu, S. (2025). An Explainable Deep Learning Model for Illegal Dress Code Detection and Classification. *Journal of Basics and Applied Sciences Research*, 3(1), 1–10. <https://doi.org/10.4314/JOBASR.V3I1.1>
- Achu, A. L., Aju, C. D., Raicy, M. C., Bhadrans, A., George, A., Surendran, U., Girishbai, D., Ajayakumar, P., Gopinath, G., & Pradhan, B. (2025). Improved flood risk assessment using multi-model ensemble machine-learning techniques in a tropical river basin of Southern India. *Physical Geography*, 46(2), 127–155. <https://doi.org/10.1080/02723646.2025.2464536>
- Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2021). House Price Prediction using Random Forest Machine Learning Technique. *Procedia Computer Science*, 199, 806–813. <https://doi.org/10.1016/j.procs.2022.01.100>
- Adeyemi, A. B., & Komolafe, A. A. (2025a). Flood hazard zones prediction using machine-learning-based geospatial approach in lower Niger River basin, Nigeria. *Natural Hazards Research*, 5(2), 399–412. <https://doi.org/10.1016/j.nhres.2025.01.002>
- Adeyemi, A. B., & Komolafe, A. A. (2025b). Flood hazard zones prediction using machine-learning-based geospatial approach in lower Niger River basin, Nigeria. *Natural Hazards Research*, 5(2), 399–412. <https://doi.org/10.1016/J.NHRES.2025.01.002>
- Aiyelokun, O. O., Aiyelokun, O. D., & Agbede, O. A. (2023). Application of random forest (RF) for flood levels prediction in Lower Ogun Basin, Nigeria. *Natural Hazards*, 119(3), 2179–2195. <https://doi.org/10.1007/s11069-023-06211-7>
- Aletaha, D., & Huizinga, T. W. J. (2009). The use of data from early arthritis clinics for clinical research. *Best Practice and Research: Clinical Rheumatology*, 23(1), 117–123. <https://doi.org/10.1016/j.berh.2008.11.008>
- ALJAZEERA. (2025). *Nigeria flood death toll tops 600 as thousands evacuated* | Climate Crisis News | Al Jazeera. <https://www.aljazeera.com/news/2022/10/17/nigeria-flood-death-toll-rises-as-thousands-evacuated>
- Arabameri, A., Saha, S., Chen, W., Roy, J., Pradhan, B., & Bui, D. T. (2020). Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques. *Journal of Hydrology*, 587, 125007. <https://doi.org/10.1016/j.jhydrol.2020.125007>
- Arabameri, A., Seyed Danesh, A., Santosh, M., Cerda, A., Chandra Pal, S., Ghorbanzadeh, O., Roy, P., & Chowdhuri, I. (2022). Flood susceptibility mapping using meta-heuristic algorithms. *Geomatics, Natural Hazards and Risk*, 13(1), 949–974. <https://doi.org/10.1080/19475705.2022.2060138>
- Bevacqua, E., Oldrich, R., Emanuele, B., Rohini, K., & Jakob, Z. (2025). Diverging trends in large floods across Europe in a warming climate. *A Nature Portfolio Journal*, 1–11. <https://doi.org/https://doi.org/10.1038/s43247-025-02734-y>
- Çakir, M., Yilmaz, M., Oral, M. A., Kazanci, H. Ö., & Oral, O. (2023). Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture. *Journal of King Saud University - Science*, 35(6). <https://doi.org/10.1016/j.jksus.2023.102754>
- Dash, S. S., Nayak, S. K., & Mishra, D. (2021). A review

- on machine learning algorithms. *Smart Innovation, Systems and Technologies*, 153, 495–507. https://doi.org/10.1007/978-981-15-6202-0_51
- Delforge, D., Wathelet, V., Below, R., Sofia, C. L., Tonnelier, M., van Loenhout, J. A. F., & Speybroeck, N. (2025). EM-DAT: the Emergency Events Database. *International Journal of Disaster Risk Reduction*, 124, 105509. <https://doi.org/10.1016/J.IJDRR.2025.105509>
- 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(August). <https://doi.org/10.1016/j.acags.2024.100183>
- Francis, O., & Ugoyibo, O. (2025). *Flooding in nigeria : the activities of nema and sustainable environmental management*. 13(1), 33–50.
- Giovannettone, J., Copenhaver, T., Burns, M., & Choquette, S. (2018). A Statistical Approach to Mapping Flood Susceptibility in the Lower Connecticut River Valley Region. *Water Resources Research*, 54(10), 7603–7618. <https://doi.org/10.1029/2018WR023018>
- Glasser, R. (2020). The Climate Change Imperative to Transform Disaster Risk Management. *International Journal of Disaster Risk Science*, 11(2), 152–154. <https://doi.org/10.1007/s13753-020-00248-z>
- Huang, X., Yin, J., Slater, L. J., Kang, S., He, S., & Liu, P. (2024). Global Projection of Flood Risk With a Bivariate Framework Under 1.5–3.0°C Warming Levels. *Earth's Future*, 12(4), 1–19. <https://doi.org/10.1029/2023EF004312>
- Igbawua, Abiem, & Jande. (2024). Spatio-temporal Assessment of Heat Waves in different Climate Zones of Nigeria. *J. Appl. Sci. Environ. Manage*, 27(7), 1515–1522. <https://doi.org/10.4314/jasem.v27i7.26>
- Ige-Olumide, O., & Salami, A. (2018). Implication of Land Use Change on Flood Vulnerability in an Urban Watershed of Ibadan, Southwest, Nigeria. *Journal of Geography, Environment and Earth Science International*, 14(1), 1–17. <https://doi.org/10.9734/jgeesi/2018/36574>
- Ighile, E. H., Shirakawa, H., & Tanikawa, H. (2022). A Study on the Application of GIS and Machine Learning to Predict Flood Areas in Nigeria. *Sustainability (Switzerland)*, 14(9). <https://doi.org/10.3390/su14095039>
- Intelligent, U. (2025). *Nigeria | International Benchmarking Network*. <https://newibnet.org/country-profiles/nigeria>
- Jurafsky, D., & Martin, J. (2012). Logistic regression Logistic regression. *Speech and Language Processing*, 404(4), 731–735.
- Kabari, L. G., & Mazi, Y. C. (2020). Rain-Induced Flood Prediction for Niger Delta Sub-Region of Nigeria Using Neural Networks. *European Journal of Engineering Research and Science*, 5(9), 1124–1130. <https://doi.org/10.24018/ejers.2020.5.9.2114>
- Khoirunisa, N., Ku, C. Y., & Liu, C. Y. (2021). A GIS-based artificial neural network model for flood susceptibility assessment. *International Journal of Environmental Research and Public Health*, 18(3), 1–20. <https://doi.org/10.3390/ijerph18031072>
- Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H. B., Gróf, G., Ho, H. L., Hong, H., Chapi, K., & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *Journal of Hydrology*, 573(November 2018), 311–323. <https://doi.org/10.1016/j.jhydrol.2019.03.073>
- Kolajo, T., & Agbogun, J. B. (2023). Flood Prediction in Nigeria Using Ensemble Machine Learning Techniques. *Ilorin Journal of Science*, 10(1). <https://doi.org/10.54908/iljs.2023.10.01.004>
- Konstantinov, A. V., & Utkin, L. V. (2021). Interpretable machine learning with an ensemble of gradient boosting machines. *Knowledge-Based Systems*, 222, 106993. <https://doi.org/10.1016/j.knosys.2021.106993>
- Liu, Jiahong, Li, Z., Shao, W., Yan, Di., & Mei, C. (2020). Urban flood modelling in Qiqihar city based on MIKE flood. *Proceedings of the International Association of Hydrological Sciences*, 383, 185–192. <https://doi.org/10.5194/piahs-383-185-2020>
- Liu, Jun, Xiong, J., Cheng, W., Li, Y., Cao, Y., He, Y., Duan, Y., He, W., & Yang, G. (2021). Assessment of Flood Susceptibility Using Support Vector Machine in the Belt and Road Region. *Nat. Hazards Earth Syst. Sci. Discuss*, May, 1–37. <https://doi.org/10.5194/nhess-2021-80>
- Muawanah, S., Muzayanah, U., Pandin, M. G. R., Alam, M. D. S., & Trisnaningtyas, J. P. N. (2023). Machine Learning Applications based on SVM Classification: A Review. *Qubahan Academic Journal*, 3(4), 206–218. <https://doi.org/10.48161/Issn.2709-8206>

- Nandi, S., & Reddy, M. J. (2022). An integrated approach to streamflow estimation and flood inundation mapping using VIC, RAPID and LISFLOOD-FP. In *Journal of Hydrology* (Vol. 610). <https://doi.org/10.1016/j.jhydrol.2022.127842>
- Ndimele, P. E., Ojewole, A. E., Mekuleyi, G. O., Badmos, L. A., Agosu, C. M., Olatunbosun, E. S., Lawal, O. O., Shittu, J. A., Joseph, O. O., Ositimehin, K. M., Ndimele, F. C., Ojewole, C. O., Abdulganiy, I. O., & Ayodele, O. T. (2024). Vulnerability, Resilience and Adaptation of Lagos Coastal Communities to Flooding. *Earth Science, Systems and Society*, 4(1), 1–19. <https://doi.org/10.3389/esss.2024.10087>
- NEMA. (2024). *National Emergency Management Agency*. 4.
- Ogunsola, O. E., & Yaya, O. S. (2019). Maximum and Minimum Temperatures in South-Western Nigeria: Time trends, Seasonality and Persistence. *Journal of Physics: Conference Series*, 1299(1). <https://doi.org/10.1088/1742-6596/1299/1/012057>
- Okibe, S. (2025). The Desecration, Destruction and Displacement: a Case Study of the 2022 Nigerian Flood. *Irish Journal of Environment and Earth Sciences Irish J. Env. E. Sci.* <https://orcid.org/0000-0001-8253-4265>
- Oluyemi, O. (2025). 2025 floods: 228 dead, over 209,000 people affected in 25 states, says NEMA. *The Guardian*. <https://guardian.ng/news/nigeria/metro/2025-floods-228-dead-over-209000-people-affected-in-25-states-says-nema/>
- Peker, İ. B., Gülbaz, S., Demir, V., Orhan, O., & Beden, N. (2024). Integration of HEC-RAS and HEC-HMS with GIS in Flood Modeling and Flood Hazard Mapping. *Sustainability* (Switzerland), 16(3). <https://doi.org/10.3390/su16031226>
- Petry, I., Miranda, P. T., Paiva, R. C. D., Collischonn, W., Fan, F. M., Fagundes, H. O., Araujo, A. A., & Souza, S. (2025). Changes in Flood Magnitude and Frequency Projected for Vulnerable Regions and Major Wetlands of South America. *Geophysical Research Letters*, 52(5). <https://doi.org/10.1029/2024GL112436>
- Rajib, A., Liu, Z., Merwade, V., Tavakoly, A. A., & Follum, M. L. (2020). Towards a large-scale locally relevant flood inundation modeling framework using SWAT and LISFLOOD-FP. *Journal of Hydrology*, 581. <https://doi.org/10.1016/j.jhydrol.2019.124406>
- Razavi-Termeh, S. V., Sadeghi-Niaraki, A., Razavi, S., & Choi, S. M. (2024). Enhancing flood-prone area mapping: fine-tuning the K-nearest neighbors (KNN) algorithm for spatial modelling. *International Journal of Digital Earth*, 17(1), 1–29. <https://doi.org/10.1080/17538947.2024.2311325>
- Resende, P. A. A., & Drummond, A. C. (2019). A survey of random forest based methods for intrusion detection systems. *ACM Computing Surveys*, 51(3). <https://doi.org/10.1145/3178582>
- Rogers, J. S., Maneta, M. M., Sain, S. R., Madaus, L. E., & Hacker, J. P. (2025). The role of climate and population change in global flood exposure and vulnerability. *Nature Communications*, 16(1), 1–11. <https://doi.org/10.1038/s41467-025-56654-8>
- Si, M., & Du, K. (2020). Development of a predictive emissions model using a gradient boosting machine learning method. *Environmental Technology and Innovation*, 20(x). <https://doi.org/10.1016/j.eti.2020.101028>
- Tran, V. T., Nguyen, H. D., Ngoc, D. T., Quan, D. V. V., Huan, N. C., Thanh, P. V., Liem, N. Van, & Nguyen, Q. H. (2024). Flood assessment using machine learning and its implications for coastal spatial planning in Phu Yen Province, Vietnam. *Journal of Water and Climate Change*, 15(8), 3738–3761. <https://doi.org/10.2166/wcc.2024.035>
- Tuan, L. A., Van, C. T., Binh, D. Van, Kantoush, S. A., Sumi, T., Quyen, L. Van, & Huong, T. T. (2024). Flood modeling in the Ba River basin using a coupled hydrodynamic model—MIKE FLOOD. *River*, 3(2), 199–207. <https://doi.org/10.1002/rvr.2.82>
- Vanneschi, L., & Silva, S. (2023). Particle Swarm Optimization. *Natural Computing Series*, 105–111. https://doi.org/10.1007/978-3-031-17922-8_4
- Victor, O. (2023). *Flooding in Nigeria is on the rise – good forecasts, drains and risk maps are urgently needed* | ConstructAfrica. <https://constructafrica.com/news/flooding-nigeria-rise-good-forecasts-drains-and-risk-maps-are-urgently-needed>
- Vladimir N. Vapnik. (1999). The Nature of Statistical Learning Theory. 2nd Edition. In *Statistics for engineering and information*. <https://statisticalsupportandresearch.wordpress.com/wp-content/uploads/2017/05/vladimir-vapnik-the-nature-of-statistical-learning-springer-2010.pdf>

- Vojtek, M., & Vojteková, J. (2019). Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. *Water (Switzerland)*, 11(2). <https://doi.org/10.3390/w11020364>
- Wahyuni, S., Sisinggih, D., & Dewi, I. A. G. (2021). Validation of Climate Hazard Group InfraRed Precipitation with Station (CHIRPS) Data in Wonorejo Reservoir, Indonesia. *IOP Conference Series: Earth and Environmental Science*, 930(1). <https://doi.org/10.1088/1755-1315/930/1/012042>
- Wang, L. (2019). Research and Implementation of Machine Learning Classifier Based on KNN. *IOP Conference Series: Materials Science and Engineering*, 677(5). <https://doi.org/10.1088/1757-899X/677/5/052038>
- Wang, Y., Xie, Y., Chen, L., & Zhang, P. (2025). Identifying key drivers of urban flood resilience for effective management: Insights and implications. *Geography and Sustainability*, 6(4), 100278. <https://doi.org/10.1016/J.GEOSUS.2025.100278>