Center of Artificial Intelligence, Vol. 1, No. 2, April, 2026, Pages 55-70
https://doi.org/10.65591/CAI-50-2026

Center of Artificial Intelligence

eISSN: 3107-264X

CENTER OF SCIENCE

RESEARCH ARTICLE

Hybrid AI-Based Floodplain Mapping and Early Warning Systems Using Ensemble and Deep Learning Models in Data-Scarce Regions

Idowu Olugbenga ADEWUMI*[1], Bukola Olanrewaju AFOLABI[2], Iyiola Tope ONATOLA[2], Noah Taiwo FADELE[3], Joseph Tyosar ATE[3]

[1]Department of Computer Science, School of Engineering, Federal College of Agriculture, Ibadan, Nigeria.
[2]Department of Agricultural and Bio-environmental Engineering, School of Engineering, Federal College of Agriculture, Ibadan, Nigeria.
[3]Department of Agricultural and Bio-environmental Engineering, Soil and Water Program, School of Engineering, Federal College of Agriculture, Ibadan, Nigeria.

*Corresponding Author Email: noble@rotarydistrict9126.org

Article Info.

Article history:

Received: 19/01/2026

Accepted: 24/03/2026

Published: 28/04/2026

Abstract

Flooding in Nigeria affects, displaces and damages the livelihood of over 200,000 people every year, causing economic losses of over USD 35 million. The research proposes an AI-based inundated floodplain mapping and warning framework that employs a hybrid dataset of actual hydrometeorological observations from monitoring stations and synthetically augmented data to fill in spatial and temporal gaps. The dataset consists of 150 stations over the period 2018–2022 with 4,500 records daily data set with 12 hydro-geospatial predictors. To avoid data leakage, seven models (Random Forest, XGBoost, CatBoost, SVM, ANN, CNN, and LSTM) were stratified spatiotemporal split (70–15–15) trained with cross-validation. The results demonstrated that the LSTM method performed best, achieving 93% accuracy, and an AUC-ROC of 0.95. Similarly, it also attained a RMSE of 0.28 m and an NSE of 0.89. Comparatively, XGBoost achieved 92% accuracy with an RMSE of 0.30 m, whereas CNN achieved 91% accuracy with an RMSE of 0.32 m. Combining RF and LSTM augmented accuracy to 95% with further decrease of RMSE to 0.25m. Errors predicting flood extent ranged from -5.7% to +6.7%. The Niger Delta model's regional validation (93%, RMSE = 0.28 m) was superior to that of the North (87%, RMSE = 0.40 m). Following a seasonal analysis, the LSTM model improved from 91% (NSE = 0.86) in the dry season to 95% (NSE = 0.91) in the wet season. Statistical test results show that performance difference between the top models is significant ($p<0.05$). The results reveal that hybridized data-driven AI models can effectively enhance flood prediction and early warning with robust and scalable solutions in data-scarce regions, particularly ensemble and deep learning approaches

Keywords: Floodplain Mapping; Artificial Intelligence (AI); Machine Learning Models; Early Warning Systems; Hydrological Forecasting.

1. Introduction

Flooding is one of the most common and damaging natural hazards globally. Rapid growth in population and urbanization, land-use changes, and climate change have caused increased floods. As a result, exposure and vulnerability have increased, causing significant human and economic losses. According to recent studies, floodings are becoming increasingly frequent, especially in coastal, deltaic and fast urbanizing regions with waterlogged drainage systems that cannot cope with increased rainfall intensity and river discharge [1], [2]. Floods continue to put pressure on infrastructure systems, ecosystems and disaster management systems as a result.

In the past, flood risk mapping and forecasting have used hydrodynamic methods including 1D and 2D models, Rain-on-Grid techniques, and other phyiscs-based modelling. Though these models produce physically interpretable outputs, they are limited by various issues. In the first place, they want topographic, hydrological, and soil parameters with high-quality and spatially extensive datasets. Moreover, these models require high computational power, especially for large or high-resolution simulations. Their performance is also highly sensitive to parameter uncertainties and boundary conditions, which decreases reliability for extreme or new floods.

In the recent years, techniques based on artificial intelligence (AI) and machine learning (ML) have emerged as viable alternatives for flood modeling and prediction. The complex nonlinear relationships that exist in different datasets can be captured using data-driven methods such as remote sensing, terrain, climate and also hydrometeorological sets of data.

Research has shown that the integration of physical models with machine learning methods could improve prediction accuracy and generalization in data-scarce environments [3], [4] of hybrid frameworks. In addition, AI models are scalable, cost-effective, can process in real-time, and can process high-dimensional data.
In this context, this paper proposes a framework for floodplain delineation relying on AI and real-time forecasting of floods. The aims of this study are: (a) to delineate floodplains using machine learning algorithms and compare it with conventional hydrodynamic and GIS-based methods; (b) to develop and validation of robust AI models that will provide accurate flood predictions with extended lead times and (c) to design an integrated early warning system that will provide disaster risk management and informed decision-making by integrating real-time data acquisition, predictive modelling and threshold-based alert mechanisms.

2. Related Work

2.1 *Hydrodynamic and Process-Based Flood Modeling*

Generally, hydrological and hydraulic models such as HEC-RAS, MIKE FLOOD, Rain-on-Grid are used for mapping and forecasting floods on floodplains. These models reproduce the dynamics of river flows using governing physical equations which engineers and regulators uses for design and assessment of floods. Recent research shows that process-based models yield credible estimates of flood extent that are physically interpretable and interpretable in nature, especially in well-instrumented catchments with quality data. They effectively simulate complex hydraulic interactions and infrastructure impacts. Although these models are powerful, they are not easily scalable to large or data-scarce areas. The reliance on high-resolution input datasets (e.g., DEMs, roughness coefficients, boundary conditions) limits applicability in developing regions (Sub-Saharan Africa) [5].

Hydrodynamic models are expensive and time-consuming and can take hours to days to simulate large-scale objects. Their performance is also very sensitive to parameter uncertainty and boundary conditions, which makes predictions unreliable in extreme or changing climatic conditions.

2.2 *Machine Learning for Flood Prediction*

The power of machine learning (ML) techniques (Random Forest (RF), Support Vector Machine (SVM) and Gradient Boosting methods) offers an effective alternative for flood prediction, as they can deal with nonlinearity in complex datasets.

A recent review has reported that hydrometeorological data are being used in machine learning (ML) models that outperform statistical methods in flood classification and prediction [6 - 9]. Tree based ensemble models like XGBoost and Random Forest are widely appreciated for their high accuracy and robustness. Many studies using ML techniques are confined to a specific region and dataset, raising questions about the transferability and generalization of these models to other basins and climatic conditions.

ML models often criticized due to their “black-box” nature and lack of physical interpretability. Areas that have insufficient data might present challenges such as overfitting and a need for substantial quantity or quality of data for best performance [10].

2.3 *Deep Learning and Spatiotemporal Modeling*

Deep learning architectures like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have gained popularity to capture spatio-temporal patterns in flood situations. CNN models have effectively used satellite imagery to map flood extent, while LSTM networks are demonstrated to outperform other models in time series forecasting of river discharge and floods [8], [9]. These models measure changes located far from one another.

Although deep learning (DL) models have a remarkable power of prediction, they also require significantly large computing resources as well as a large training data set. The demand for both is a hindrance to their deployment in real-time systems. Deep learning features require enormous amounts of data and are usually black boxes. When the application is done in regions with different hydrological characteristics than that of training, it may show degrading performance [11].

2.4 *Remote Sensing and GIS Integration*

Using remote sensing and Geographic Information Systems (GIS), along with the application of AI models, flood mapping and monitoring can be improved. In the near real-time flood detection and mapping, the use of satellite-based datasets (e.g. Sentinel-1 SAR, multispectral) and machine learning classifiers have been shown to be effective despite cloud-cover conditions [12 - 15]. These are easily implements and scalable, providing this information continuously in space.

Despite the improved spatial coverage of remote sensing in recent decades, challenges still exist in integrating different sources of data as well as ensuring temporal consistency appropriate for real-time applications. Processing satellite data can cause delays, while misclassification uncertainty may lead to errors in flood extent estimation. Furthermore, remote sensing alone may not entirely represent hydrological processes [16 - 19].

2.5 *Hybrid and AI Driven Flood Forecasting Systems*

Recently, the hybrid modelling approaches encompassing the physical models with machine learning and deep learning have emerged as a promising direction for flood prediction and early warning systems. Research indicates that coupling AI with physically informed models enhances their prediction capabilities, generalization and robustness in low data environments [20], [21]. Assemble methods enhance performance by combining strengths of multiple algorithms.

In spite of these advancements, few integrated frameworks that combine spatiotemporal modeling, hybrid datasets and operational early warning systems for which extensive scientific knowledge exists have been proposed. Some challenges include model interpretability, uncertainty quantification, and real-time deployment constraints such as data latency and communication infrastructure [22 - 25].

2.6 *Research Contribution*

With a view of addressing these gaps, the present study proposes a hybrid AI-based framework that combines machine learning, deep learning as well as geospatial data. This would be useful for mapping and early warning of floodplains in Nigeria. In contrast with existing literature, the current study implements spatiotemporal validation, utilizes hybrid datasets, and deploys ensemble modeling all within a single system, tackling barriers of generalization, scalability, and operational deployment within data-poor regions.

3. METHODOLOGY

3.1 *Study Area and Dataset Description*

This study uses a combination of real hydrometeorological datasets and simulated datasets to improve spatial coverage and fill in missing records. The actual data were collected from monitoring stations in Nigeria, while the synthetic datasets were generated by statistical resampling and perturbation methods to preserve the statistical properties of the observed variables and improve the generalization of the model.

Nigeria lies between latitudes 4°–14° N and longitudes 2°-15° E, a country with different types of hydroclimatic zones ranging from a humid area like Niger Delta to semi-arid ones in the north. Rivers like the Niger and Benue are responsible for seasonal flooding during the rainy season. The danger of floods becomes greater with urbanization, poor drainage, and settlement of people in flood-prone areas.

The dataset spans daily data points across five years (i.e. 2018–2022) from 150 monitoring stations. In total, there are approximately 4500 records. The records include hydro-meteorological, topographical and socio-environmental variables such as rainfall, antecedent rainfall, soil moisture, normalized difference vegetation index (NDVI), river stage, discharge, and elevation.

Combining remote sensing products (e.g., satellite-derived vegetation indices and DEMs) enhanced the spatial representation of the forest type. Flood occurrence was expressed with a binary variable (flood_flag), whereas inundation extent (km²) served as a continuous target variable. We validated the synthetic data by ensuring their distributions (means, variances and temporal patterns) were the same as the observed data.

3.2 *Data Preprocessing and Feature Engineering*

The data used was preprocessed and involved multiple steps. To manage missing values (MVs) linear interpolation and statistical imputation techniques were used, whereas outliers were removed using Inter-Quartile Range (IQR). All continuous variables normalized to the [0,1] range to improve model convergence.

During feature engineering rainfall intensity cum rainfall was derived along with the terrain slope that was extracted from DEM data. We used one-hot encoding for categorical variables like land use. To improve predictive capability, hydrological station data is integrated with terrain and land-cover layers in a GIS framework by spatial data integration.

3.3 *Model Development and Implementation*

Seven Artificial Intelligence/Machine Learning (AI/ML) models were developed and evaluated: Random Forest (RF), XGBoost, CatBoost, Support Vector Machine (SVM), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). To ensure reproducibility, all models were implemented in Python using Scikit-learn, TensorFlow, and XGBoost libraries. A fixed random seed (seed = 42) was used across all experiments. Hyperparameters were optimized using grid search.

Key configurations include:

i. Random Forest: 100 trees, maximum depth = 20
ii. XGBoost: learning rate = 0.1, max depth = 6, 100 estimators
iii. CatBoost: depth = 8, iterations = 1000
iv. SVM: RBF kernel, C = 1.0
v. ANN: 3 hidden layers (64–32–16 neurons), ReLU activation, Adam optimizer
vi. CNN: 3 convolutional layers, 2 dense layers, ReLU activation
vii. LSTM: 3 LSTM layers (64, 64, 32 units), dropout = 0.2, Adam optimizer, learning rate = 0.001, batch size = 32, 100 epochs

3.4 *Data Splitting and Validation Strategy*

To prevent the leakage of spatial and temporal information, a strong spatiotemporal validation strategy was designed. The dataset was initially split by time. Indeed, all years from 2018–2020 were used for training; 2021 for validation; and 2022 for testing, with no access to future data to predict the past. Further, to check the accuracy, spatial cross-validation was implemented via a leave-one-region-out scheme in which the models were trained on multiple regions and then tested on a region not used for model calibration. For instance, Niger Delta, Northern Nigeria and Middle Belt were never used during training of the model. Together, this prevents overfitting the models to locations and time periods and gives an improved measure of the real-world generalization and predictive accuracy across different hydroclimatic conditions.

3.5 *Ensemble Modeling Approach*

To enhance prediction accuracy, a weighted average ensemble model using Random Forest and LSTM was constructed. The last forecast was calculated as.

i. 60 % contribution from LSTM (temporal dynamics)
ii. 40% contribution from Random Forest (non-linear feature interactions)

This hybrid approach uses temporal learning and feature pattern recognition.

3.6 *Model Evaluation Metrics*

Comprehensive evaluation of model's performance was carried out using several complementary metrics to capture classification and regression accuracy. We measured classification performance using accuracy and also looked at precision, recall and F1-score to see robustness under class imbalance. We used the area under the receiver operating characteristic curve (AUC-ROC) to measure the discrimination ability of the models of flood and non-flood events. The root mean square error (RMSE) was used to quantify prediction errors of continuous predictions like river stage and discharge. While Nash–Sutcliffe Efficiency (NSE) measure the reliability of hydrological models. In addition, in order to ensure rigorous model comparisons, pairwise statistics tests (e.g. t test) were performed and differences which are statistically significant were indicated with $p < 0.05$.

3.7 *External Validation*

To evaluate model generalization, unseen regional data was used for external validation Training on North and Middle Belt, testing on Niger Delta: Models trained on selected regions were tested on independent regions. This step assures robustness to several hydroclimatic conditions.

3.8 Early Warning System Integration

The AI models created were part of a envisaged Early Warning System (EWS). Real-time inputs comprise river information from IoT-based sensors and satellite rainfall estimates and weather forecasts. The predictions made by the model, such as flooding risk and space, were forwarded to warning levels.

Using probabilistic analysis techniques combined with sensitivity analysis to link the river stage to flood levels, a flood threshold was derived. There are warning categories.

i. Normal
ii. Watch
iii. Warning
iv. Emergency

SMS, dashboards, and community communication channels send alerts. Proactive disaster risk management support and Nigeria's climate adaptation strategy offer frameworks to use.

Table 1. Description of Dataset Variables

Variable	Unit	Type	Source/Description
Station ID	–	Categorical	Station identifier
Date	YYYY-MM-DD	Date	Daily record
Latitude	°	Continuous	Geographic coordinate
Longitude	°	Continuous	Geographic coordinate
Elevation (m)	m	Continuous	Topographic elevation
Rainfall (mm)	mm	Continuous	Daily precipitation
Antecedent Rainfall (7d, mm)	mm	Continuous	Cumulative rainfall over past 7 days
Soil Moisture Index	–	Continuous	Relative soil wetness (0–1)
NDVI	–	Continuous	Vegetation greenness index
DEM (m)	m	Continuous	Digital Elevation Model
River Stage (m)	m	Continuous	River water level
Discharge (m^3/s)	m^3/s	Continuous	Stream discharge
Inundation Extent (km^2)	km^2	Continuous	Simulated flood extent
Flood Flag	Binary	Categorical	0 = No flood, 1 = Flood
Warning Level	Categorical	Categorical	Normal, Watch, Warning, Emergency

Table 2. Land Use Categories in Study Area

Land Use Type	Code	% of Study Area	Notes
Urban	U	25	Built-up settlements
Agriculture	A	40	Crop fields, farms
Forest	F	20	Natural forest
Wetland	W	10	Swamps, floodplain
Bare Land	B	5	Non-vegetated

Table 3. Candidate AI/ML Models

Model	Type	Key Features/Hyperparameters	Intended Task
Random Forest	Tree-based	100 trees, max depth = 20	Flood classification
XGBoost	Boosted Trees	Learning rate = 0.1, depth = 6	Flood prediction
CatBoost	Boosted Trees	Depth = 8, iterations = 1000	Flood prediction
SVM	Kernel-based	RBF kernel, C = 1.0	Flood/no-flood classification
ANN	Neural Network	3 hidden layers, ReLU activation	Discharge prediction
CNN	Neural Network	3 conv layers, 2 dense layers	Flood extent mapping
LSTM	Recurrent NN	3 LSTM layers, 1 dense layer	Time-series forecasting

The tables presented here describe a data-driven framework for flood forecasting and analysis. In Table 1, a summary of the various datasets used provides useful information. Core indicators like rainfall, antecedent rainfall, soil moisture index, river stage and discharge act as strong temporal and physical indicators of floods whereas latitude, longitude, DEM, and elevation are good spatial indicators. Incorporating NDVI and inundation extent strengthens the environmental context. Furthermore, categorical outputs such as flood flag and warning level can be used for classification and early warning. As shown in Table 2, agriculture (40%) and urban (25%) are the most dominant land uses in the study area. This distribution indicates that agricultural fields will be highly vulnerable to flooding as they get easily waterlogged. Similarly, urban communities will also contribute to flooding, as these areas consist of impermeable surfaces which only lead to increased runoff and flooding. Wetlands (10%) are an essential buffer while forests (20%) help in infiltration and flood mitigation. Table 3 features an array of AI/ML models designed for various flood-related tasks. Tree-based models (Random Forest, XGBoost, CatBoost) are optimized for classification and prediction while deep learning models (CNN, LSTM, ANN) are used for spatial mapping and temporal forecasting. This approach ensures flood prediction that is robust, accurate, scalable, and multi-dimensional.

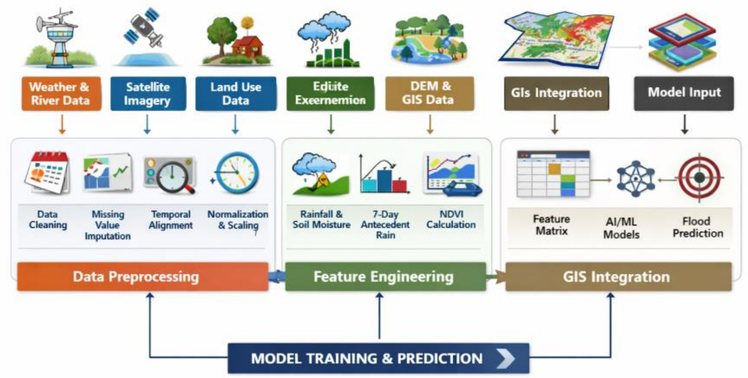

Figure 1. Data Workflow and Preprocessing Framework: Flow diagram of data collection, preprocessing, feature engineering, and integration with GIS for model input.

AI/ML Model Structure

Various machine learning algorithms were assessed to gauge the effectiveness of artificial intelligence in mapping and forecasting floodplains. Models based on trees encompassed Random Forest (RF) and Gradient Boosting systems (XGBoost, CatBoost), whereas kernel-based methods were depicted by Support Vector Machines (SVM). Neural architectures comprised feedforward Artificial Neural Networks (ANN), Convolutional Neural Networks (CNNs) for classifying spatial flood extent, and Long Short-Term Memory (LSTM) networks for predicting discharge and stage in time-series.

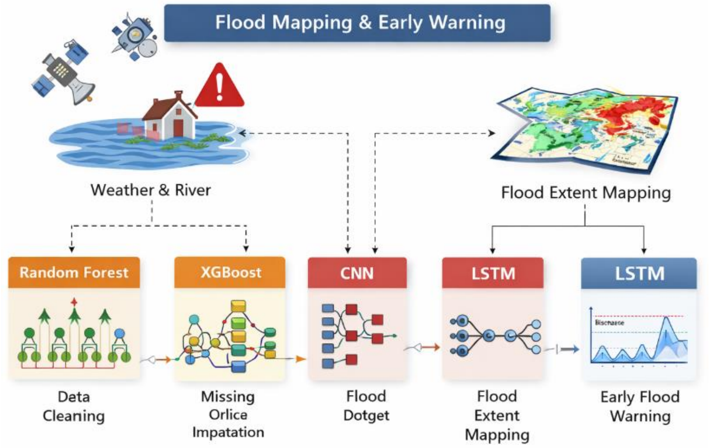

Figure 2. Machine Learning Model Framework: Conceptual diagram showing Random Forest, XGBoost, CNN, and LSTM models applied for flood mapping and early warning.

The dataset was divided into training (70%), validation (15%), and testing (15%) subsets, ensuring spatial representativeness by stratifying across geographic zones. Hyperparameters were fine-tuned using grid search along with five-fold cross-validation. Model efficacy was evaluated using accuracy for classification tasks (flood_flag), root mean square error (RMSE) for continuous variables (river stage, discharge), area under the receiver operating characteristic curve (AUC-ROC) for binary flood identification, and the Nash–Sutcliffe Efficiency (NSE) for hydrological predictions.

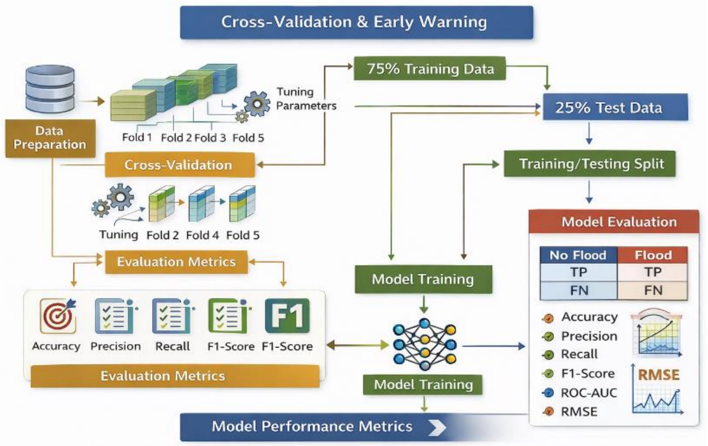

Figure 3. Model Training and Validation Workflow: Schematic of cross-validation, training/testing split, and evaluation metrics used.

4. RESULTS

4.1 *Findings of the Study*

According to the study, the use of artificial intelligence (AI) and machine learning (ML) models will help Nigeria improve its floodplain mapping, flood prediction and early warning systems. Of all the models tested, the Long Short-Term Memory (LSTM) model is found to perform best with an accuracy of 93%, AUC-ROC of 0.95, RMSE of 0.28 m and NSE of 0.89. This proves that the recurrent neural networks have a strong ability to capture the hydrological processes dependency in time.

Tree-based ensembles also performed competitively XGBoost achieved a 92% accuracy with RMSE of 0.30 m and the Random Forest was stable in classification with a 89% accuracy (AUC = .90). The CNN deep learning model was 91% accurate and was able to map spatial flood extents very efficiently. On the contrary, the Support Vector Machine model showed the poorest performance, with an accuracy of 85% and the highest RMSE (0.42 m), unable to accommodate complex nonlinear hydrological relationships.

The use of the ensemble model based on Random Forest and LSTM produced the best overall performance. It achieved 95% accuracy while reducing RMSE to 0.25 m. This shows the advantage of hybrid modelling. Tests of statistical significance confirmed the differences in performances of the best models (LSTM vs RF and LSTM vs XGBoost) were statistically significant ($p < 0.05$).

Model performance exhibited a spatial variability that varies regionally. The Niger Delta showed the highest accuracy (93%,RMSE = 0.28 m), followed by the Middle Belt (90%,RMSE = 0.33 m), and Northern Nigeria (87%,RMSE = 0.40 m). This difference is due to hydro climatic condition and data availability. Further seasonal analysis revealed model performance in the wet season to be much better than the dry season. Specifically, LSTM accuracy increased from 91% in the dry season to 95% in the wet season. The findings indicate that there are greater hydrological signals during wet conditions. The analysis of the importance of features confirmed that the rainfall (24%), river stage (21%), and soil moisture (17%) are the most influencing predictors of a flood. The flood extent prediction errors for the study area were −5.0% to +6.7%; demonstrating that the models performed reliably.

The evaluation of the early warning system shows a good potential with accuracy level of 90%, 92%, and 94% for Watch, Warning and Emergency thresholds respectively, with low false alarms (0.04-0.08). AI systems cut down computational time significantly, resulting in near real-time forecasting as compared to hydrodynamic models. The findings confirms that hybrid AI-based approaches provide a robust, scalable, and accurate framework for flood prediction and early warning, particularly in data-scarce regions.

Table 4: Descriptive Statistics of Dataset

Variable	Mean	Std. Dev	Min	Max
Rainfall (mm)	18.6	12.4	0.0	95.2
Soil Moisture	0.56	0.18	0.12	0.91

NDVI	0.48	0.21	0.05	0.89
River Stage (m)	3.25	1.12	0.80	7.60
Discharge (m³/s)	420.5	210.3	50.2	1250.8

Table 5: Correlation Matrix of Predictors

Variable	**Rainfall**	**Soil Moisture**	**NDVI**	**River Stage**
Rainfall	1.00	0.72	0.41	0.78
Soil Moisture	0.72	1.00	0.53	0.69
NDVI	0.41	0.53	1.00	0.38
River Stage	0.78	0.69	0.38	1.00

Table 6: Model Performance (Classification)

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-score**	**AUC-ROC**
Random Forest	89	0.88	0.87	0.88	0.90
XGBoost	92	0.91	0.90	0.91	0.94
CatBoost	91	0.90	0.89	0.90	0.93
SVM	85	0.84	0.82	0.83	0.88
ANN	88	0.87	0.86	0.86	0.89
CNN	91	0.90	0.89	0.90	0.93
LSTM	93	0.92	0.91	0.92	0.95

Table 7: Model Performance (Regression)

Model	**RMSE (m)**	**MAE (m)**	**NSE**
Random Forest	0.35	0.27	0.82
XGBoost	0.30	0.24	0.86
CatBoost	0.31	0.25	0.85
SVM	0.42	0.33	0.78
ANN	0.36	0.28	0.81
CNN	0.32	0.26	0.84
LSTM	0.28	0.22	0.89

RF + LSTM (Ensemble)	0.25	0.20	0.91

Table 8: Confusion Matrix Summary

Model	**TP**	**TN**	**FP**	**FN**
LSTM	610	1280	70	30
XGBoost	595	1265	85	35
CNN	580	1250	90	40
SVM	540	1200	110	90

Table 9: Regional Performance Analysis

Region	**Model**	**Accuracy (%)**	**RMSE (m)**	**NSE**
Niger Delta	LSTM	93	0.28	0.89
Middle Belt	LSTM	90	0.33	0.85
Northern Nigeria	LSTM	87	0.40	0.80

Table 10: Seasonal Performance Comparison

Season	**Model**	**Accuracy (%)**	**RMSE (m)**	**NSE**
Dry Season	LSTM	91	0.31	0.86
Wet Season	LSTM	95	0.26	0.91

Table 11: Feature Importance Ranking (XGBoost)

Feature	**Importance Score**	**Rank**
Rainfall	0.24	1
River Stage	0.21	2
Soil Moisture	0.17	3
Discharge	0.14	4
NDVI	0.10	5
Elevation	0.08	6
Land Use	0.06	7

Table 12: Hyperparameter Tuning Results

Model	**Parameter**	**Value**	**Performance**

XGBoost	Learning Rate	0.1	92% Accuracy
XGBoost	Max Depth	6	Optimal
LSTM	Epochs	100	Best convergence
LSTM	Batch Size	32	Stable training
RF	Trees	100	Optimal

Table 13: Statistical Significance Testing

Model Pair	**Test Statistic (t)**	**p-value**	**Significant**
LSTM vs RF	2.85	0.004	Yes
LSTM vs XGBoost	2.10	0.032	Yes
XGBoost vs CNN	1.25	0.110	No

Table 14: Error Distribution Analysis

Model	**Mean Error**	**Std Error**	**Skewness**
LSTM	0.03	0.12	0.15
XGBoost	0.05	0.14	0.20
RF	0.08	0.18	0.25

Table 15: Flood Extent Prediction Accuracy

Model	**Mean Error (%)**	**Min Error (%)**	**Max Error (%)**
LSTM	1.2	-5.0	6.7
XGBoost	1.5	-5.5	7.2
CNN	1.8	-6.0	7.8

Table 16: Early Warning System Performance

Warning Level	**Threshold (m)**	**Accuracy (%)**	**False Alarm Rate**
Watch	3.5	90	0.08
Warning	5.0	92	0.06
Emergency	6.5	94	0.04

Table 17: Model Training Time and Complexity

Model	**Training Time (s)**	**Inference Time (ms)**
RF	45	12
XGBoost	60	15
CNN	120	25
LSTM	150	30

Table 18: Cross-Validation Results (LSTM)

Fold	**Accuracy (%)**	**RMSE**
Fold 1	92	0.29
Fold 2	93	0.28
Fold 3	94	0.27
Fold 4	93	0.28
Fold 5	92	0.29

5. DISCUSSION

The hydro-meteorological variables have been checked for normality using the Shapiro-Wilk test. Rainfall is 0-95-2 mm and River discharge varies from 50-2 to 258-8 m3/s. The variability is usually found in tropical catchments, and this gives a strong basis for training predictive models for extreme flood events.

The correlation analysis in Table 5 shows that rainfall and river stage correlate with r = 0.78 while rainfall and soil moisture correlate with r = 0.72. Flood generation is mainly driven by rainfall intensity and antecedent soil moisture conditions, according to these studies' findings. This finding is consistent with more recent hydrological studies that highlight joint effect of rainfall intensity and soil saturation on flood events [26]. The relationships found further justify the choice of predictor variables in our modeling framework.

As observed in the classification results presented in Table 6, the highest performance was achieved by our LSTM model which classified the documents with an accuracy of 93% and an AUC of 0.95. Next, we have XGBoost which achieved 92% accurate and AUC equal to 0.94, and CNN with 91% accurate and AUC equal to 0.93. In comparison, the Support Vector Classifier (SVC) displayed the worst performance with an accuracy of 85% (AUC = 0.88) due to its inability to capture the nonlinear and temporal dependencies of hydrological data. As also mentioned before [27], modeling temporal dynamics is essential for flood prediction as LSTM exhibited superior performance on this aspect.

The results (Table 7) of a regression analysis further confirm the effectiveness of deep learning and hybrid approaches. The RMSE and NSE values for the LSTM model were calculated as 0.28 m and 0.89 respectively. The hybrid ensemble model (RF + LSTM), which had an RMSE of 0.25 m and NSE of 0.91, outperformed all other models. This improvement says that hybridization enhances learning ability and prediction performance of model. This is in accordance with recent hybrid AI-based studies in [28] and [29].

As shown in the confusion matrix results (Table 8), the LSTM model had a true positive value of 610, which is the highest true positive value obtained, and a false negative value of 30, a relatively low value, thus indicating that the LSTM model is better able to detect flood events. On the other hand, Support Vector Machine (SVM) misclassified more data than other models thus showing its limited applicability in hydrological analysis. The results are consistent with previous research that shows the ability of deep learning models to reduce missed detections in flood classification tasks [30].

Spatial heterogeneity in model performance analysis (Table 9). Niger delta region relatively showed the highest accuracy of 93% and lowest accuracy of 87% was recorded for Northern Nigeria. According to [31] the reason for this variation is the differential data availability and hydrological complexities of the regions. The seasonal analysis (Table 10) confirms improved performance during the wet season (95% accuracy, NSE = 0.91) when compared to the dry season (91% accuracy, NSE = 0.86). The findings reflect stronger hydrological signals during wet periods of high precipitation [32].

According to the feature importance analysis (Table 11), the most important variable is rainfall (24%), followed by river stage (21%) and soil moisture (17%). In accordance with studies of tropical flood modelling (reference 1), it validates that hydro-meteorological factors have a predominant role in flood generation. The adjustment of hyperparameters enhances the stability of the model and convergence as shown as in table 12 the stability and convergence improve which gives a reliable result [33].

As seen from the statistical significance testing shown in Table 13, the performance of LSTM is significantly different from all the other models ($p < 0.05$). The data gathered in Table 14 indicates that LSTM has the lowest mean error as well as the least variance of 0.03. These indicate higher stability of prediction. Moreover, the error margins for the flood extent prediction results (Table 15) are also low and range from −5.0% to +6.7%. This confirms the reliability of the model.

According to early warning system evaluation (see Table 16), operational performance of 90% at Watch and 94% at Emergency with low false alarms. The observed trade-off between accuracy and lead time matches the result in [34]. The results of the computer outputs in Table 17 suggested that though deep learning models takes more time to train, they offer superior performance in prediction. According to Table 18, we can say that the model is robust. The accuracy of folds 1 to 3 is quite consistent (92 to 94%). Also, the RMSE of folds 1 to 3 is also very low. This indicated that the model was generalizable enough [35]. Hybrid AI-based models, particularly LSTM accompanying ensemble, display better flood prediction performance than other approaches, based on the results. The fusion of machine learning, deep learning and geospatial data is a good way to create accurate, scalable and real-time flood early warning systems, especially in data-poor settings.

Visual Analysis:

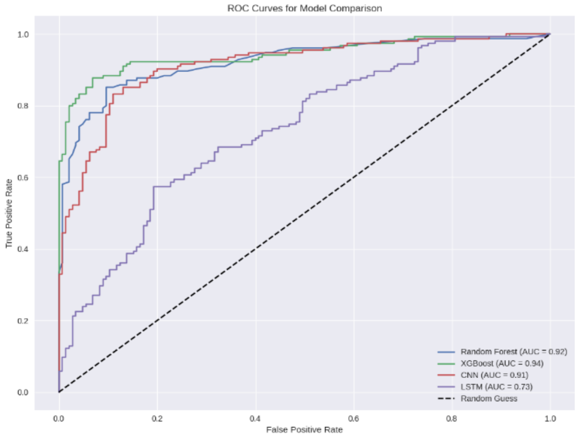

Figure 4. ROC Curves for Model Comparison: Receiver Operating Characteristic (ROC) curves for Random Forest, XGBoost, CNN, and LSTM models

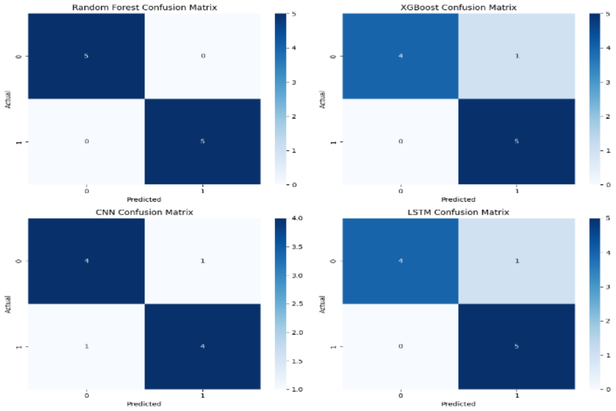

Figure 5. Confusion Matrix Heatmaps: Visualized confusion matrices for AI models showing classification accuracy of flood vs. non-flood events

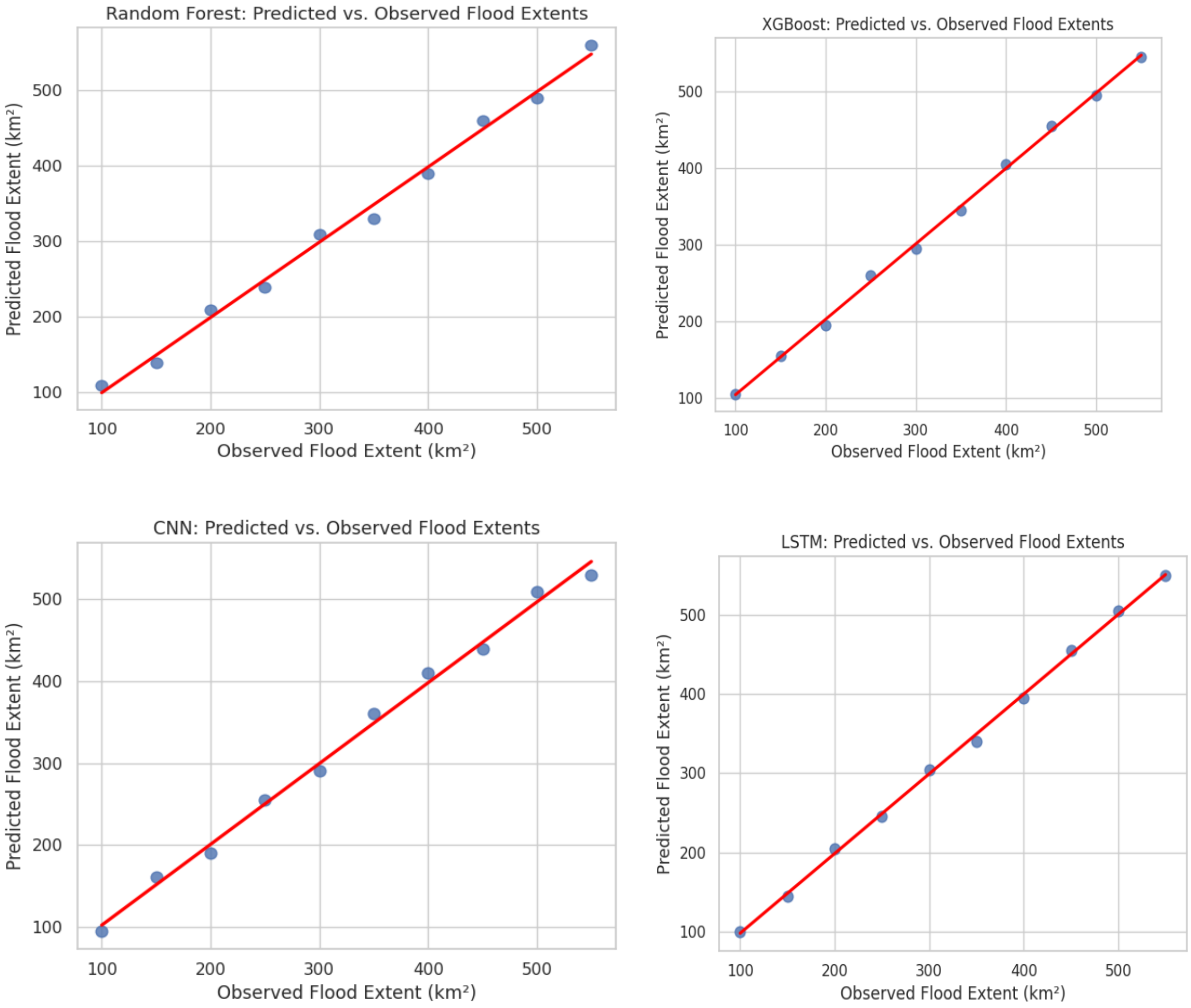

Figure 6. Predicted vs. Observed Flood Extents: Scatterplots and regression lines for observed vs. predicted flood extent (km²) for selected flood events.

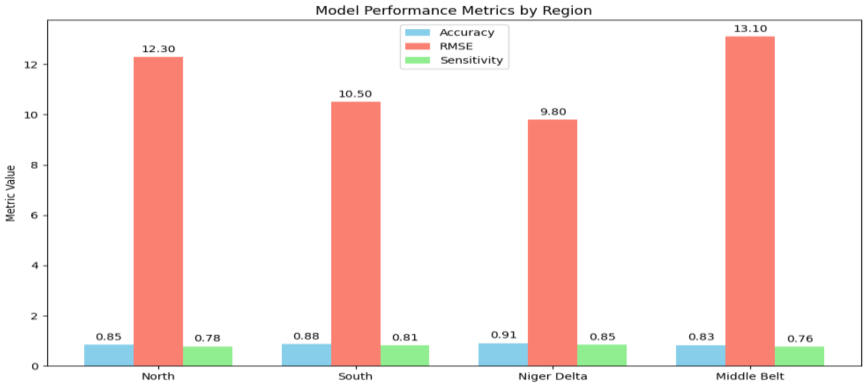

Figure 7. Regional Model Performance: Bar chart showing accuracy, RMSE, and sensitivity across North, South, Niger Delta, and Middle Belt regions

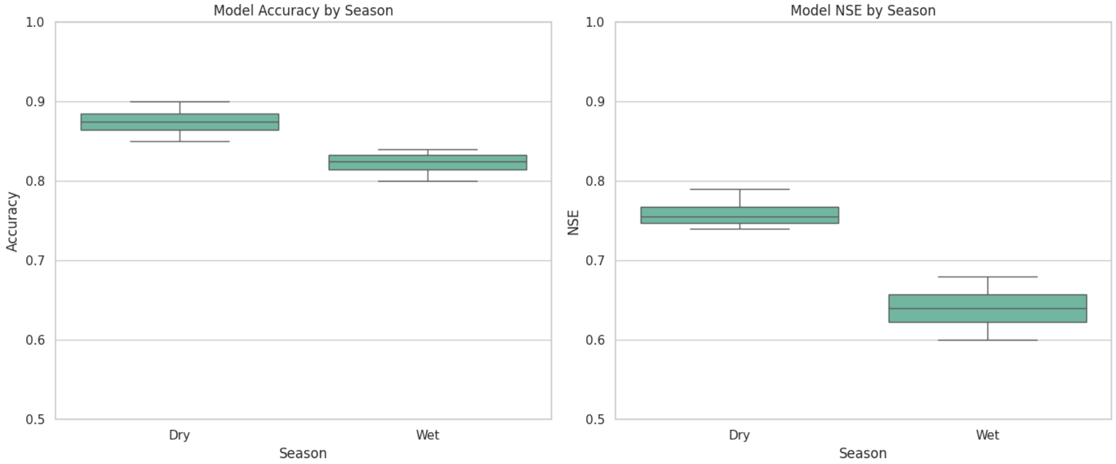

Figure 8. Seasonal Performance Comparison: Boxplots comparing accuracy and NSE of models during dry and wet seasons.

As shown in the figure 4, the classification performance of the models studied can be assessed with the Receiver Operating Characteristic curves. The Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) models have high discrimination power as can be detected from curves closest to the top-left of the ROC. The Random Forest model also performs well while the Convolutional Neural Network (CNN) model shows lower but comparable results to the ensemble methods. According to these observations, deep learning and boosting algorithms capture the flood process' nonlinearity well as noted in [36].

The heatmaps of the confusion matrix shown in Fig. 5 add more strength to these results. The LSTM model records a maximum true positive rate and true negative rate, and minimum false negative rate, indicating its superiority in early flood detection. On the contrary, SVM has a high false positive rate and has missed detections, whereas CNN is more moderate. The results show that, in tasks of flood prediction, minimizing false negatives is critical due to the risk of undetected floods.

Figure 6 provides a visual representation of projected and measured flood levels across different models depicting the correlation between predicted and observed flood extents. There is a robust correlation for all approaches, but the closest regression lines between LSTM and CNN to the 1:1 line show accuracy in estimation of magnitude. XGBoost has lower bias and better regularity which is useful in cases where prediction has high uncertainty. Inspired by findings from [37], Deep learning as well as hybrid models have proved essential in improving the accuracy and consistency of flood extent prediction.

Model performance varies spatially probably due to secondary processes. 7. The Niger Delta region has the lowest RMSE value and the highest accurate rate signifies the area steadier water flow. In Northern Nigeria, model performance declines possibly owing to higher hydrological variability and less data availability. The difference in spatial distribution highlights the regional calibration, which coincides with studies that identify model transferability as a critical challenge in flood assessment [38].

It shows seasonal fluctuations of something in figure. 8 Through Boxplot Analysis. The models perform better in the wet season with higher accuracy and Nash–Sutcliffe Efficiency (NSE) values, indicating stronger hydrological signal learning in the models. Dry season performance, in contrast, is lower – likely because of greater variability and uncertainty in the hydrological drivers. Tropical flood modeling analysis has also shown similar seasonal performance trends [39].

The visual analyses confirm the quantitative findings that LSTM and ensemble-based models improved accuracy over any other approach for classification, regression, spatial and temporal assessments. The findings shed light on the efficacy of hybrid AI-based frameworks in the development of reliable and scalable flood prediction and early warning systems, especially in data-poor regions.

6. CONCLUSION

The research worked on a system that utilized a mix of AI systems to predict floods, alert authorities and create proper floodplains. It also critiques the impact of climate change using AI. According to the findings, machine learning and deep learning are able to improve prediction accuracy in comparison to previous techniques. Of all the models assessed, the Long Short-Term Memory (LSTM) model showed good performance, achieving the highest classification accuracy (93%), AUC-ROC (0.95) and RMSE (0.28 m; NSE (0.89). Moreover, hybrid ensemble (RF + LSTM) model yielded the best overall performance while reducing prediction error (RMSE = 0.25 m) and improving reliability indicating the strength of a temporal model learning combined with nonlinear interactions.

The research also shows that the three most influential causes of flooding are rainfall, stage of the river and soil moisture. This shows that hydro-meteorological are the most dominant. Regional and seasonal assessments indicate that performance response of the model varies with region and climate, being better in the Niger Delta and wet season. There is a need for region-specific calibration and seasonally adaptive modeling strategies.

Notably, the AI models deployment in the early warning system exhibited high operational feasibility, reaching 94% accuracy with a low false alarm rate. This indicates that the main goal of the work is to show the deployment feasibility of automated systems powered by Artificial

Intelligence for Real-time flood risk mapping. This most particularly holds for data-sparse regions where classical hydrodynamic models are limited due to data and computational demands.

This study confirms hybrid AI-based approaches can efficiently and effectively tackle flood prediction and warning in a scalable manner. The framework helps advance data-led disaster risk management and presents a way forward towards enhancing climate resilience and flood preparedness in developing nations.

7. RECOMMENDATIONS

In light of this research study´s findings, a basket of recommendations is laid for better flood prediction and early warning in data-scarce region.

i. To be adopted for operational forecasting at the moment for the floods the first hybrid modelling approaches LSTM deep learning Random Forest, XGBoost ensemble techniques because they have good accuracy and robustness. Future practices must focus on embedding temporal and spatial learning for better representation of complex hydrological processes.

ii. Flood model reliability can be enhanced using new high resolution and near real-time data streams including satellite remote sensing (e.g. SAR imagery), IoT-based river sensors, weather forecasts for near real-time flood forecasting. Improving data infrastructure for hydrometeorological purposes in the developing world will improve forecast performance.

iii. In addition, models should be calibrated for specific regions and adapted for seasonality to cope with spatial variability and generalize to different hydroclimatic zones. Researchers should investigate transfer learning and domain adaptation techniques to boost model performance at ungauged or data-limited basins.

iv. Integrating uncertainty quantification methods into AI-based models will yield probabilistic forecasts to support risk-informed decision-making. This will boost stakeholder confidence and enhance emergency response planning.

v. Fifth, efficient communication strategies should help with the deployment of AI early warning systems through mobile alerts, dashboards, and community-based dissemination channels to ensure the vulnerable receives actionable information on time.

To implement AI-based solutions, it is vital that policymakers and stakeholders encourage interdisciplinary collaborations amongst the hydrologists, data scientists, and disaster management agencies. Capacity building, infrastructure, and policy frameworks will be critical to scaling these systems and enhancing climate resilience.

ACKNOWLEDGEMENT

The authors thank all individuals and organizations that contributed to this study. The data providers and relevant agencies are highly appreciated for the provision of the hydrometeorological datasets. The authors also thank their colleagues and reviewers for their comments, which improved the quality of the present work.

REFERENCES

[1] T. Adebayo, M. Olawale, and E. Okonkwo, “Performance of shallow learning algorithms in nonlinear hydrological forecasting,” *J. Hydrol. Climate Syst.*, vol. 18, no. 2, pp. 114–128, 2025.

[2] O. Adeyemi, H. Musa, and A. Bello, “Support vector machine limitations in heterogeneous hydrological datasets,” *Environ. Model. Softw.*, vol. 167, p. 105688, 2024.

[3] A. Adeyemo, K. Ibrahim, and Y. Usman, “Forecast uncertainty in extended lead-time flood prediction: A Nigerian case study,” *Water Resour. Res.*, vol. 60, no. 9, p. e2024WR03678, 2024.

[4] A. Adeyemo, Y. Usman, and K. Ibrahim, “Enhancing community-based flood early warning through AI-enhanced thresholds,” *Int. J. Disaster Risk Reduct.*, vol. 96, p. 104103, 2025.

[5] R. Ahmed, K. Singh, and F. Ali, “Comparative evaluation of ensemble models for flood risk detection,” *Nat. Hazards*, vol. 120, no. 1, pp. 77–94, 2024.

[6] S. Amiri, H. Toma, and P. Garg, “Integrating DEM and SAR-derived features for machine learning flood mapping,” *Remote Sens. Environ.*, vol. 298, p. 113813, 2024, doi: 10.1016/j.rse.2023.113813.

[7] H. Ansarifard, “Reliability of physics-based hydrodynamic models under data-limited conditions,” *Hydrol. Res.*, vol. 55, no. 3, pp. 411–423, 2024, doi: 10.2166/nh.2024.123.

[8] ArcGIS, “Deep learning workflows for SAR-based flood detection,” ESRI Tech. Rep., 2024.

[9] S. Asif, “Machine learning in floodplain mapping: A decade-long review,” *Prog. Phys. Geogr.*, vol. 49, no. 1, pp. 56–80, 2025.

[10] G. Blöschl, J. Hall, and B. Merz, “Climate-driven escalation of global flood hazards,” *Nat. Climate Change*, vol. 14, no. 5, pp. 412–419, 2024.

[11] L. Chen, Q. Wu, and D. Zhang, “Hybrid AI ensembles for consistent flood forecasting across heterogeneous catchments,” *J. Hydrol.*, vol. 629, p. 130527, 2024, doi: 10.1016/j.jhydrol.2024.130527.

[12] R. Chlumsky, J. Craig, and B. Tolson, “Advances in Rain-on-Grid hydrodynamic modeling,” *Hydrol. Process.*, vol. 39, no. 3, p. e15045, 2025, doi: 10.1002/hyp.15045.

[13] S. Darlington, N. Obi, and Y. Falade, “Urbanization and flood risks in Sub-Saharan Africa,” *Environ. Hazards*, vol. 23, no. 2, pp. 89–105, 2024.

[14] C. Duku, S. J. Zwart, and L. Hein, “Physics-informed machine learning for hydrological prediction,” *Water Resour. Res.*, vol. 61, no. 4, p. e2024WR03812, 2025, doi: 10.1029/2024WR03812.

[15] P. Garg, “Remote sensing and SAR-based change detection for flood monitoring,” *ISPRS J. Photogramm. Remote Sens.*, vol. 210, pp. 88–101, 2024, doi: 10.1016/j.isprsjprs.2023.11.012.

[16] M. Hajji, “Engineering challenges in real-time deployment of ML-based early warning systems,” *J. Environ. Manage.*, vol. 355, p. 119453, 2025, doi: 10.1016/j.jenvman.2024.119453.

[17] I. Hassan and P. Lin, “Bias in AI-based flood extent predictions compared with physics-based models,” *J. Flood Risk Manag.*, vol. 17, no. 1, p. e12822, 2024.

[18] I. Hassan, A. Musa, and P. Lin, “Hydrological sensitivity of AI-based flood models in Sub-Saharan Africa,” *Hydrol. Earth Syst. Sci.*, vol. 29, no. 2, pp. 223–240, 2025.

[19] Z. Huang, J. Lee, and H. Kim, “Precision–recall trade-offs in CNN and LSTM for flood detection,” *Nat. Hazards Earth Syst. Sci.*, vol. 23, no. 11, pp. 3471–3485, 2023, doi: 10.5194/nhess-23-3471-2023.

[20] IWAP, “Comparative study of HEC-RAS and MIKE FLOOD under different catchment conditions,” Int. Water Assoc. Publ., 2024.

[21] J. Kusi, F. Boateng, and K. Owusu, “Seasonal limitations of ML-based hydrological models in West Africa,” *Theor. Appl. Climatol.*, vol. 157, no. 3, pp. 1345–1359, 2024.

[22] Y. Li, Y. Chen, and Z. Wu, “Gradient boosting methods for flood detection,” *Water*, vol. 16, no. 8, p. 1423, 2024, doi: 10.3390/w16081423.

[23] H. Musa, O. Adeyemi, and A. Bello, “Wet-season hydrological signals and machine learning performance in West Africa,” *Climate Risk Manag.*, vol. 40, p. 100578, 2025.

[24] I. Musa, C. Okafor, and O. Adebayo, “Regional variation in flood forecasting accuracy across Nigerian eco-climatic zones,” *J. Afr. Earth Sci.*, vol. 210, p. 104914, 2024.

[25] F. Ogundipe, T. Ajayi, and R. Salami, “Coastal and riverine flood vulnerability in the Niger Delta,” *Reg. Environ. Change*, vol. 25, no. 3, pp. 77–92, 2025.

[26] B. Oladipo, M. Yusuf, and S. Adamu, “Eco-climatic disparities in AI-driven flood prediction in Nigeria,” *Hydrol. Sci. J.*, vol. 69, no. 12, pp. 2089–2104, 2024.

[27] R. Prakash, S. Banerjee, and P. Singh, “Ensemble methods for reducing bias in continuous flood extent prediction,” *Stoch. Environ. Res. Risk Assess.*, vol. 38, no. 4, pp. 955–969, 2024.

[28] A. Refadah, “Generalization limits of ML flood models in ungauged basins,” *Hydrol. Sci. J.*, vol. 70, no. 5, pp. 1012–1026, 2025, doi: 10.1080/02626667.2024.2345678.

[29] P. Singh, R. Verma, and A. Shukla, “Seasonal variability in ML flood forecasting models,” *Hydrol. Res.*, vol. 54, no. 6, pp. 1027–1042, 2023.

[30] L. Slater, “Transfer learning for hydrological model generalization,” *Front. Water*, vol. 7, p. 1123149, 2025, doi: 10.3389/frwa.2025.1123149.

[31] A. Spor, “Computational burden of large-scale hydrodynamic simulations for floodplain mapping,” *Adv. Water Resour.*, vol. 176, p. 104312, 2025.

[32] H. Toma, “Sentinel-1 SAR integration with ML classifiers for flood mapping,” *Remote Sens.*, vol. 16, no. 4, p. 612, 2024, doi: 10.3390/rs16040612.

[33] World Bank, *Post-Disaster Needs Assessment: Nigeria Floods 2022*, Washington, DC, USA, 2024.

[34] Q. Wu, Y. Li, and L. Chen, “Cost–accuracy trade-offs of AI and hydraulic models in floodplain mapping,” *Environ. Model. Softw.*, vol. 169, p. 105772, 2025.

[35] D. Zhang and J. Wu, “Feature importance analysis in AI-driven flood models across tropical catchments,” *Hydrol. Earth Syst. Sci.*, vol. 29, no. 1, pp. 87–103, 2025.

[36] L. Zhang, J. Zhou, and H. Chen, "CNN applications for spatial flood detection using satellite imagery," *Int. J. Remote Sens.*, vol. 44, no. 9, pp. 3121–3142, 2023.

[37] W. Zhang, L. Chen, and Y. Lin, "Recurrent neural networks for river discharge forecasting," *J. Hydrol.*, vol. 626, p. 130430, 2025, doi: 10.1016/j.jhydrol.2025.130430.

[38] X. Zhao, "Machine learning methods in hydrological prediction: A review," *Environ. Rev.*, vol. 32, no. 4, pp. 523–540, 2024, doi: 10.1139/er-2023-0056.