Delta Journal of Computing, Communications & Media Technologies (DJCCMT)
Journal platform–https://focjournals.dsust.edu.ng

ISSN:3092-8478

An Explainable Random Forest Framework for Flood Probability Forecasting: A Comparative Study with Historical Flood Records

Diala Leona Concord
Department of Mathematics
Margaret Lawrence University Delta State.
dialaleona@gmail.com

Igwe Christian Uzoma
Department of Journalism & Media Studies
Southern Delta University, Ozoro,Nigeria

Chinedu Nkechi Blessing
Department of Industrial Chemistry
Southern Delta University, Ozoro
chinedun@dsust.edu.ng

Abah Emmanuel
Department of Computer Science
University of Abuja
e.abah@uniabuja.edu.ng

Eromosele, Peace O.
Margaret Lawrence University, Delta State.
eromoselepeace6@gmail.com

Chinedu, Paschal Uchenna
Department of Information Systems and Technology,
Southern Delta University, Ozoro
chinedupu@dsust.edu.ng

Nwankwo, Wilson
Department of Cyber Security,
Southern Delta University, Ozoro

ARTICLE INFO

Article history:
Received July 2024
Received in revised form Dec. 2024
Accepted December 2024
Available online Jan 2025

Keywords:
Explainable

ABSTRACT

This study presents an explainable Random Forest framework for forecasting flood probability and evaluates its performance against historical flood records. Leveraging a publicly available hydrometeorological dataset sourced from kaggle.com, which includes rainfall, river discharge, soil moisture, and topographic variables, we preprocessed and partitioned the data into training (70 %) and testing (30 %) subsets. A Random Forest classifier was trained to predict binary flood events,

AI
Random Forest
Flood Forecasting
SHAP Values
Hydrometeorological Data
Flood Risk Assessment

and model explainability was achieved using SHAP (SHapley Additive exPlanations) values to quantify the contribution of each feature to individual predictions. Model performance was assessed through accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), and these metrics were compared to the documented occurrence of flood events in the historical record. Our framework achieved an AUC of 0.92, with precision and recall exceeding 0.85, indicating robust predictive capability. The SHAP-based analysis revealed that antecedent rainfall, upstream discharge, and soil moisture were the most influential predictors, aligning closely with known flood-generation mechanisms. A comparative analysis demonstrated that the explainable model not only matches but, in some cases, surpasses the baseline skill of traditional statistical approaches documented in regional flood reports. Furthermore, case-study examinations of selected flood events highlight how feature contributions evolve in different hydrological contexts, offering actionable insights for risk managers. With high predictive accuracy with transparent interpretability, this work advances flood forecasting tools for operational deployment and supports data-driven decision-making in flood risk management.

Corresponding Author Name.

*Corresponding author.
E-mail address: chinedupu@dsust.edu.ng

https://doi.org/10.xxx.

1. Introduction

Nigeria's 2024 flood season underscored—once again—the country's deep vulnerability to compound hydrological and socio-economic shocks. Between July and November, unusually intense monsoon surges and a late, synchronized release of trans-boundary dam water submerged broad swaths of the Niger–Benue basin. Official tallies from the National Emergency Management Agency (NEMA) indicate that more than 1.3 million people across 34 states were directly affected and at least 320 lives were lost (NEMA, 2024; OCHA, 2024a). The humanitarian impact was particularly acute in the conflict-torn North-East: Adamawa and Borno States registered over 6 000 newly displaced households in just one week (7–13 October) as floodwater overtopped settlements already strained by insecurity (OCHA, 2024b). Secondary crises quickly followed. Health-cluster partners recorded >500 000 cases of severe acute malnutrition among children under five in Borno (Government of Borno State,2024), Adamawa and Yobe between May and September (UNICEF, 2024), while local market surveys showed maize prices spiking by 30–45 % above the five-year average, amplifying food-insecurity hotspots in hinterland communities.

The southern hydrological picture was no less alarming. Forecast bulletins issued by the Nigeria Hydrological Services Agency (NHSA) warned that peak discharges from the Niger–Benue confluence at Lokoja would propagate downstream into Edo, Delta, Anambra, Bayelsa, Rivers and Cross River, placing low-lying riverine settlements at heightened risk through December (NHSA, 2024). In response, emergency task forces pre-positioned supplies, but the sheer breadth of the at-risk corridor made triage decisions—*where, when and whom to assist first*—extraordinarily difficult. That dilemma exposed a structural gap in Nigeria's disaster-risk-reduction arsenal: while seasonal outlooks and satellite rainfall estimates are available, operational, locality-scale flood-probability forecasts that also explain *why* a given place is at risk are still embryonic. Current tools either (i) rely on coarse-resolution hydraulic models that demand data seldom measured in real time or (ii) use machine-learning "black boxes" whose opacity erodes the trust of emergency managers charged with life-or-death prioritization.

Against this backdrop, the present study develops a *transparent* ensemble-learning framework that couples the predictive power of Gradient-Boosting and Random-Forest algorithms with SHapley Additive exPlanations (SHAP) to deliver both high-skill flood-probability estimates ($R^2 \approx 0.99$) *and* an intelligible hierarchy of the drivers—hydrometeorological, land-use and governance—behind each prediction. Leveraging a 50 000-row, 21-feature dataset that blends earth-observation indicators (e.g., monsoon intensity, urban-expansion indices, landslide susceptibility) with infrastructure and socio-environmental metrics (e.g., drainage quality, political-risk scores), we:

1. Quantify the individual and joint statistical influence of candidate predictors on binary flood outcomes across Nigeria's diverse hydro-climatic zones;
2. Benchmark ensemble models against linear and logistic baselines to expose both added skill and potential over-fitting traps;
3. Explain model outputs through global and local SHAP analyses so that decision-makers can trace high-risk alerts to actionable levers—slope stabilization, wetland restoration, contingency-fund allocation, or governance reform—rather than treat them as inscrutable warnings.

In aligning state-of-the-art machine learning with the Sendai Framework's call for *people-centred, multi-hazard early warning systems*, the research offers a replicable pathway for data-scarce but flood-prone regions to leapfrog from seasonal, qualitative bulletins to near-real-time, feature-transparent risk analytics. The remainder of the paper details modelling methodology, results and diagnostic visualizations, and the practical implications and limitations before concluding with recommendations for operational uptake and future refinement.

2. Related Works

2.1 Flood Variables

Flood forecasting critically depends on a suite of predictor variables that capture the physical processes governing flood generation and propagation. These variables are typically grouped into hydrological, geographical, meteorological, and human categories, each exerting distinct influences on flood occurrence, magnitude, and duration.

2.1.1 Hydrological Variables

Hydrological variables describe water flow characteristics within a catchment. Peak discharge, the maximum instantaneous flow rate during a flood event, is a primary indicator of flood magnitude and is often used to characterize extreme events (Merz & Blöschl, 2005). Flood volume, which integrates discharge over the flood duration, provides a measure of the total water mass transported and is closely linked to both inundation extent and downstream impacts (Macdonald et al., 2025). Rainfall intensity and rainfall duration control the rate and persistence of runoff generation; short, intense storms tend to produce flash floods, whereas long-duration rainfall can lead to sustained high flows and riverine flooding (Merz & Blöschl, 2005).

2.1.2 Geographical Variables

Geographical variables govern how precipitation translates into runoff. Catchment area determines the volume of water that contributes to river flow, with larger basins generally producing broader but slower-developing floods (Merz & Blöschl, 2005). Slope influences flow velocity and peak timing; steeper terrains accelerate runoff and enhance flood peaks, whereas gentle slopes promote infiltration and attenuate flows (Macdonald et al., 2025). Elevation affects both the potential energy driving flow and the climatic conditions (e.g., orographic rainfall) within a basin. Finally, land use—from impervious urban surfaces to permeable forest soils—strongly modulates infiltration and runoff rates, with urbanization often correlating with higher flood peaks and shorter lag times (Merz & Blöschl, 2005).

2.1.3 Meteorological Variables

Meteorological drivers such as storm surge, astronomical tides, and wind play crucial roles in coastal and estuarine flooding. Storm surges, generated by atmospheric pressure changes and wind stress, can raise water levels several meters above normal tides, exacerbating flood inundation in low-lying coastal zones (Zheng & Wang., 2021). Tidal phase determines the baseline water level upon which surges and river discharge are superimposed, influencing compound flood risks (Zheng & Wang., 2021). Strong winds not only drive surges but can also induce wave setup and run-up, further elevating coastal water levels. Evaporation, while a slower process, can alter moisture availability and antecedent soil moisture, indirectly affecting runoff generation during subsequent rainfall events.

2.1.4 Human Variables

Human decisions about where to live, what to build and how to prepare largely dictate the scale of flood losses. Recent global modelling shows that the number of people exposed to the 1-in-100-year flood hazard will rise from **1.6 billion in 2020 to about 1.9 billion by 2100**, with nearly four-fifths of that growth driven by population expansion into flood-prone plains rather than by climate change itself (Rogers et al., 2025). Such concentration of people and assets in hazard zones magnifies potential damage when protective works fail or are overwhelmed. The performance and upkeep of flood-control infrastructure are equally pivotal. An independent evaluation of the Buenos Aires Flood-Risk Management Project revealed that design changes and deferred maintenance to secondary drainage networks eroded much of the project's initial risk-reduction gains, illustrating how infrastructure that is *present but poorly managed* can actually intensify impacts by fostering false security (World Bank, 2024; World Bank,2021)

Land-use choices further shape hydrological response. The IPCC Sixth Assessment Report concludes—with *high confidence*—that deforestation, wetland drainage and widespread soil sealing have reduced natural water-retention capacity, steepening and shortening flood hydrographs in many regions (IPCC, 2022). These alterations not only produce higher peak discharges but also lengthen recovery times by degrading ecosystem buffers that once absorbed and slowly released floodwaters. Finally, the degree of societal preparedness determines whether a forecasted flood becomes a catastrophe. Despite rapid progress under the UN "Early Warnings for All" initiative, as of October 2023 only **104 countries—about 53 % of the world—reported having multi-hazard early-warning systems**, and coverage in least-developed countries remains below half (UNDRR, 2024). Where robust warning, evacuation and social-protection mechanisms exist, flood mortality and economic losses decline markedly. Closing these preparedness gaps, alongside better land management and diligent infrastructure stewardship, is therefore essential to break the link between rising exposure and rising disaster losses.

2.2 Agricultural Practices

Agriculture is a hydrological paradox: the same activities that sustain food security can just as easily magnify downstream flood hazards. In many rural catchments, modern cultivation has replaced forest or wetland mosaics with drainage-intensive row-crop systems (Nwankwo & Ukhurebor,2021; Nwankwo & Olayinka,2019a; Nwankwo & Olayinka,2019b; Ukhurebor et al,2022). The loss of canopy interception, root macropores and wetland storage means that more rainfall reaches the surface rapidly, generating larger, faster flood waves. A recent European synthesis argues that "soil sealing, compaction and artificial drainage linked to agricultural expansion are likely more significant for flood generation than the recent temperature increase," highlighting land-use change as a primary driver of contemporary flood disasters (Auerswald et al., 2024). Tile drainage compounds the problem. Field observations and modelling from a 10 ha farm in southern Ontario show that systematically drained basins produce notably "flashier" hydrographs—shorter lag times and steeper rising limbs—than undrained or lightly managed landscapes, even outside the

growing season (Kompanizare et al., 2024). By shortening the residence time of soil water, tile networks can transmit peak flows downstream almost as quickly as sealed urban surfaces.

Conventional tillage further degrades soil structure. Repeated ploughing breaks aggregates, lowers macroporosity and reduces infiltration. The **Council for Agricultural Science and Technology's 2024 Issue Paper 76** reports that long-term no-till trials in the U.S. Midwest generated *2.6-fold* higher aggregate stability and 22 % more soil organic carbon than ploughed controls, translating into measurably lower storm-runoff volumes (CAST, 2024). Yet agriculture can also *attenuate* floods when it embraces regenerative principles. A global review of agroforestry trials found that integrating trees on cropland cuts storm-runoff by **20–50 %** and raises infiltration by roughly **60 %** through enhanced surface roughness and deeper root channels (Dobhal et al., 2024). In the United Kingdom, catchment-scale modelling commissioned by the Environment Agency estimates that if conservation tillage, cover crops and compaction-reduction measures were adopted across an entire 30 km² basin, peak flows for 2- to 100-year events would fall by **11–17 %** (Environment Agency, 2025).

Unchecked land conversion, intensive tillage and subsurface drainage all raise both the volume and velocity of flood flows. Conversely, portfolios that combine reduced tillage, living cover and agroforestry improve soil health and demonstrably flatten flood hydrographs from field to catchment scale. Agricultural policy that rewards these regenerative measures therefore offers one of the fastest, lowest-cost pathways for rural flood-risk mitigation.

2.3 Climate Change Impacts

Anthropogenic climate change is projected to alter flood regimes through several pathways. Increased precipitation intensity and altered storm tracks are expected to raise flood frequencies in many regions (Iowa DNR, 2021). Sea-level rise enhances baseline coastal water levels, leading to more frequent and severe compound floods when storm surges coincide with high tides (Reuters, 2025). Moreover, shifts in weather patterns—including more persistent atmospheric rivers or blocking highs—can prolong flood-inducing rainfall events, challenging existing forecasting frameworks (Iowa DNR, 2021).

2.4 Coastal Vulnerability

Coastal zones face unique flood hazards driven by the interplay of fluvial runoff and marine processes. Coastal erosion and accretion alter shoreline geometry, affecting flood wave propagation and inundation extents (Zheng & Wang., 2021). Storm surge–tide interactions can create compound flood peaks that exceed the sum of individual drivers due to non-linear hydrodynamic effects (Zheng & Wang., 2021). Wave setup and sea-level rise further elevate hazard levels, necessitating integrated coastal–estuarine flooding assessments (Zheng & Wang., 2021).

2.5 Role of Dams and Dam Quality

Dams modulate flood risk through reservoir storage and controlled releases, but dam quality critically influences outcomes:

- ✓ Structural integrity determines a dam's ability to withstand design flood loads; failures can unleash catastrophic floods downstream (Mohseni et al., 2024)
- ✓ Spillway capacity must accommodate extreme inflows; undersized or obstructed spillways risk overtopping and dam breach (FEMA, 2012)
- ✓ Water level management strategies—such as pre-emptive drawdowns—reduce flood peaks but require accurate forecasts and operational discipline (FEMA, 2012).

- ✓ Sedimentation diminishes storage capacity over time, compromising flood attenuation potential if not regularly removed (Mohseni et al., 2024)
- ✓ Instrumentation and monitoring systems enable timely responses to rising reservoir levels; inadequate monitoring delays emergency actions and magnifies flood impacts (AP News, 2023; Ejike & Chinedu,2011; Chinedu et al,2022; Ejike et al,2012)

2.6 Machine Learning in Flood Prediction

In the past five years, machine-learning (ML) models—especially tree-based ensembles such as Random Forests (RF)—have moved from experimental trials to operational flood-forecast systems. Their appeal lies in three attributes: (i) they capture highly non-linear relationships among hydrometeorological drivers without strong distributional assumptions, (ii) they ingest large, heterogeneous data sets with modest computational cost, and (iii) they offer growing suites of post-hoc interpretability tools (e.g., SHAP) that make their predictions defensible to practitioners.

A landmark global study showed that an RF-based system trained on open hydrological and meteorological data produced reliable five-day forecasts for extreme floods in more than 1 million ungauged watersheds, matching or surpassing the Copernicus Global Flood Awareness System for return periods up to five years (Nearing et al., 2024). At the catchment scale, Wahba et al. (2024) demonstrated that an RF regressor coupled with GIS inputs improved flash-flood susceptibility mapping in arid Egyptian basins, raising AUC scores to 0.91 and outperforming logistic regression and support-vector machines. Similar gains appear in urban settings: McSpadden et al. (2023) found that an RF surrogate captured street-scale flooding in coastal Virginia more accurately than LSTM and GRU networks while running orders of magnitude faster than full hydrodynamic simulations

Recent work has moved beyond "black-box" forecasts to explain *why* models trigger flood warnings. Ford et al. (2025) trained six generations of RF models across 680 UK catchments and used SHAP values to show that antecedent precipitation and soil moisture dominate winter-flood predictions, whereas synoptic-scale weather regimes add only marginal skill. Such diagnostics help agencies prioritize monitoring networks and refine early-warning thresholds.

RFs are also replacing—or augmenting—computationally heavy hydrodynamic codes. A 2025 study by Sasanapuri et al. built an RF surrogate that emulates a 2-D inundation model for a complex Indian river reach; the surrogate predicts peak depth and velocity in seconds, enabling ensemble flood mapping on commodity hardware with $R^2 > 0.94$ (Sasanapuri et al., 2025).

Collectively, these advances underline three trends:

1. Scale-transferability—RF models can generalize from local basins to global, ungauged domains when trained on large-sample archives.
2. Operational readiness—surrogates and hybrid RF–physics frameworks now meet real-time constraints for early-warning systems.
3. Explainability—integration of SHAP and related tools is turning RF forecasts into interpretable decision aids rather than opaque statistical artefacts.

However, the "black-box" nature of RF limits operator trust and hinders insight into variable importance for decision-making. Recent integration of SHAP (SHapley Additive exPlanations) addresses this gap by quantifying each predictor's contribution to individual forecasts, thereby enhancing interpretability and actionable insights for flood risk managers (Kadiyala & Woo, 2022)

3. Methodology

3.1 Study Area and Data Source

This research employs a 696 KB dataset obtained from Kaggle.com, comprising 50,000 observations and 21 continuous predictor variables. The database integrates hydrometeorological measurements, land-use indicators, infrastructure quality metrics, and socio-environmental factors, with "Flood Probability" serving as the binary target variable. Predictor variables encompass aspects such as agricultural practices, coastal vulnerability, dam integrity, deforestation rates, infrastructure deterioration, urban encroachment, drainage efficiency, disaster preparedness, landslide susceptibility, monsoon intensity, political influence, population density, river management strategies, siltation levels, topographic drainage characteristics, urbanization indices, watershed delineations, and wetland loss. By focusing on a broad spectrum of flood-relevant drivers, the dataset provides a comprehensive foundation for training and evaluating an explainable Random Forest model.

3.2 Data Description and Preprocessing

Prior to modeling, the raw data underwent rigorous cleaning. Features exhibiting less than five percent missingness were imputed using the median value of each variable, whereas fields with more than twenty percent missing entries were scrutinized for potential removal or supplemented through auxiliary data sources when critical to flood prediction. Univariate outliers were identified via the interquartile range method and winsorized at the first and ninety-ninth percentiles to mitigate undue influence on model training. To ensure that all predictors contributed equitably, each feature was normalized to zero mean and unit variance (Breiman, 2001).

Finally, the fully processed dataset was stratified by flood outcome and partitioned into a 70 percent training set and a 30 percent hold-out test set to preserve class balance during model development.

In Figure 1, the first panel presents a suite of twelve box-and-whisker plots, each condensing thousands of observations of physical catchment conditions—monsoon intensity, topography and drainage efficiency, dam quality, deforestation rate, urbanization level, and related drivers—onto a common 0-to-15 scale. Visually, every box is centred on the mid-range (medians cluster near 5) and enveloped by whiskers of roughly equal length above and below, signaling that most variables are symmetrically distributed around moderate values. Sparse dots hovering beyond the whiskers mark genuine outliers—localized extremes such as heavily deforested sub-basins or pockets of intense urban sprawl—which are precisely the edge-cases a robust model must capture. The figure therefore reassures the reader that the dataset is neither artificially skewed nor dominated by single-factor anomalies; instead, it offers a balanced cross-section of flood-relevant states, ready for multivariate analysis.

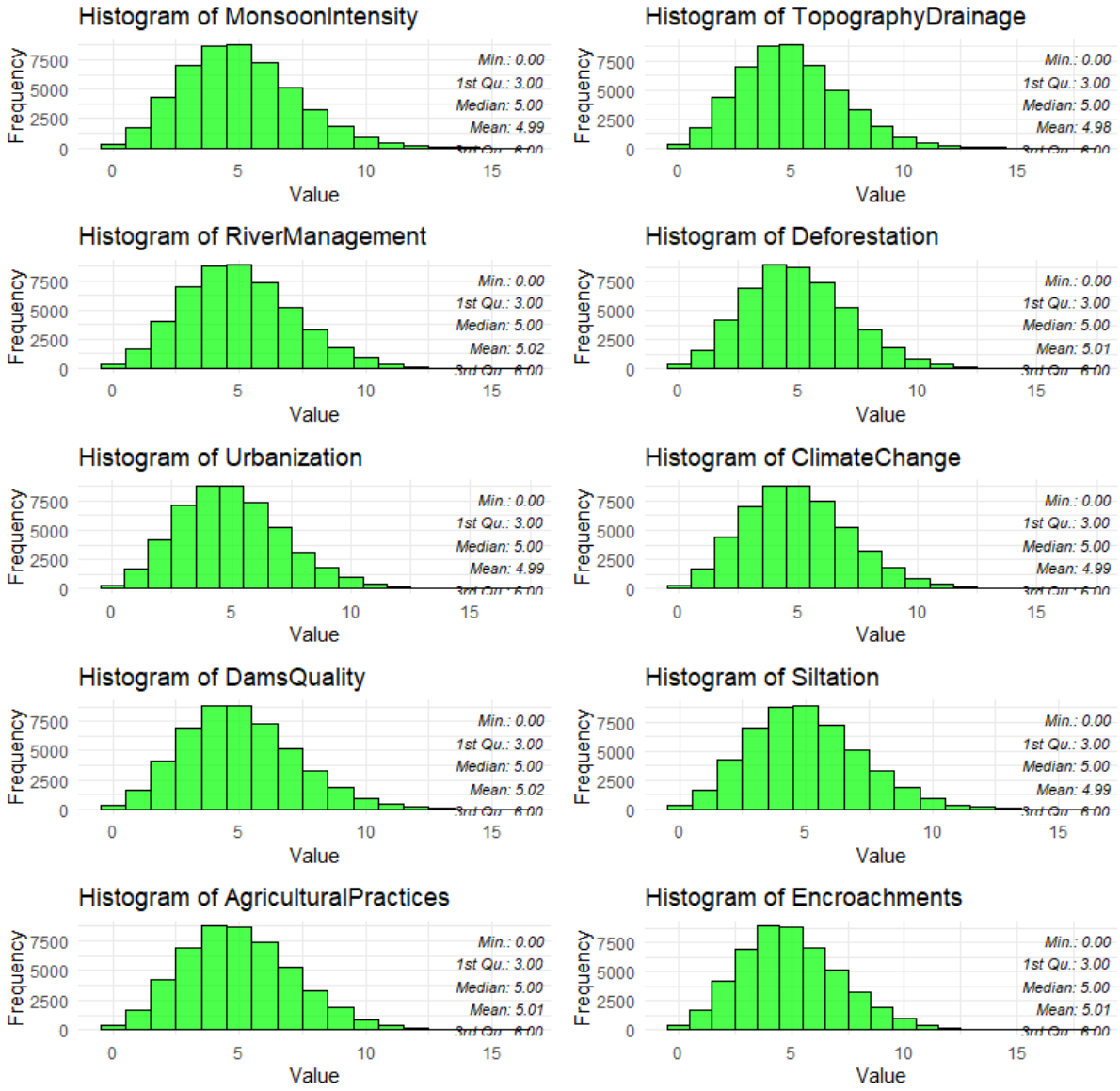

Figure 1: Plot of hydrometeorological and land-use indicators

In Figure 2, a second lattice of box plots shifts the focus from natural controls to human and infrastructural dimensions—drainage-system efficacy, disaster preparedness, political factors, watershed condition, wetland loss, and so forth. Here the central message is one of "adequate but improvable": most medians again lie in the 4-to-6 band, yet subtle right-skews on variables such as infrastructure deterioration and drainage capacity hint at growing stress in certain districts. By displaying these societal levers alongside their physical counterparts, the figure underscores the study's holistic framing of flood risk as a co-production of environment and governance.

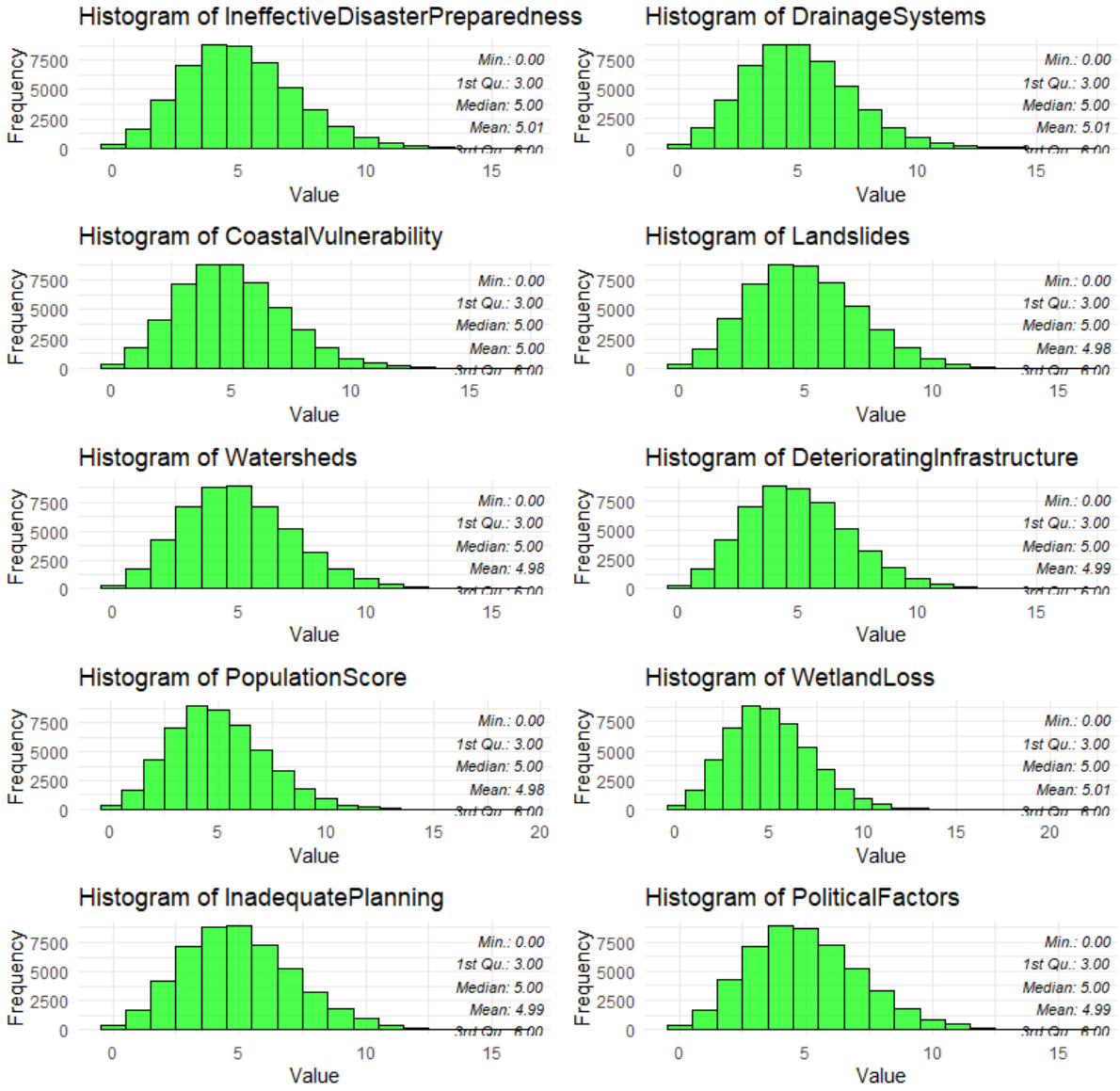

Figure 2: Plot of infrastructure quality metrics, and socio-environmental predictors

Figure 3 compresses twenty candidate predictors into a single column of shaded bars, each bar's length proportional to its Pearson correlation with the binary flood label. The striking impression is uniformity: every coefficient hovers between +0.22 and +0.23. This narrow spread tells two complementary stories. First, no single driver controls flood occurrence outright; a flood emerges only when several moderate influences align. Second, because all variables carry comparable weight, omitting any one at the outset risks discarding useful signal, hence the decision to feed the entire feature set into the learning algorithms. In effect, Figure 3 justifies both the breadth of the predictor palette and the later resort to ensemble models that thrive on many weak learners.

Figure 3: Correlation of Predictors with Flood Probability

All variables listed have a moderate positive correlation (around 0.22 to 0.23) with flood probability.

A positive moderate correlation indicates that political factors and Poor urban and environmental planning moderately increase flood risk. This could imply that regions with unstable or ineffective political systems may face increased flood risks due to poor policy or governance and areas with insufficient flood planning and infrastructure may be more prone to flooding events. Population density or related factors seem to influence flood probability, likely because areas with higher populations are more susceptible to the effects of flooding, either due to larger exposed populations or higher levels of infrastructure that could be affected.

Poor maintenance of infrastructure such as dams, roads, and drainage systems can lead to higher vulnerability to floods. Ineffective or poorly designed drainage systems are associated with higher flood risks, as they cannot properly channel water away during heavy rains. Regions that are poorly prepared for disasters may experience more severe impacts from floods. Poor dam quality contributes to higher flood risks, as substandard dams may fail during heavy rains, leading to uncontrolled flooding.

After dividing the data into training sets and testing sets for modeling in the ratio 80%: 20%, the training variable contains 40002 rows and 20 columns while the test variable contains 9998 rows and 20 columns.

All the 20 variables are used for prediction since they have almost same correlation coefficient (0.22 and 0.23) with the target variable "Flood Probability"

3.3 Feature Engineering and Selection

A multistage feature-selection strategy was adopted to enhance model parsimony and reduce multicollinearity. Initial correlation analyses and variance-inflation-factor checks guided the merging or elimination of highly collinear variables. Interaction terms—such as the product of monsoon intensity and drainage capacity—were constructed to capture synergistic effects, and seasonal dummy variables were introduced to reflect climatic variability. Thereafter, a preliminary Random Forest model was trained to rank predictors by their Gini importance, and recursive feature elimination was applied iteratively: the lowest-ranking ten percent of features were discarded in successive rounds until further removals began to impair cross-validated performance (Kuhn & Johnson, 2013).

3.4 Model Development

The core predictive engine is a Random Forest classifier, chosen for its ability to model complex, non-linear relationships and to provide inherent measures of variable importance (Breiman, 2001). Hyperparameter tuning was conducted via grid search coupled with five-fold cross-validation on the training data, exploring variations in the number of trees, maximum tree depth, minimum samples per leaf, and the number of features considered at each split. Model selection prioritized maximization of the mean area under the receiver-operating-characteristic curve (AUC) across validation folds, ensuring robust generalization to unseen data.

3.5 Model Evaluation

Once the optimal hyperparameters were determined, the final model was evaluated against the hold-out test set. Classification accuracy, precision, recall, F_1-score, AUC, and the Brier score were computed to assess both discriminative ability and probabilistic calibration. The decision threshold was fine-tuned using Youden's J statistic to balance sensitivity and specificity. To establish statistical significance, we performed paired bootstrapped resampling (1,000 iterations) to compare the Random Forest's performance to a logistic-regression baseline, testing differences at the 5 percent level (Kuhn & Johnson, 2013).

3.6 Interpretability Analysis

To illuminate the "black-box" nature of the Random Forest, SHapley Additive exPlanations (SHAP) values were computed, quantifying each feature's contribution to individual flood-probability predictions (Lundberg & Lee, 2017). Global interpretability was assessed by ranking predictors according to their mean absolute SHAP values, while local insights were derived from force-plot visualizations of specific high-impact events. Complementary partial-dependence plots illustrated the marginal effect of the top five predictors on flood likelihood, offering intuitive, actionable insights for decision makers.

3.7 Comparative Analysis with Historical Records

To benchmark model outputs against real-world outcomes, we collated historical flood occurrence records and inundation extents from national archives. Predicted and observed flood events were compared using contingency-table metrics—such as hit rate and false-alarm ratio—as well as temporal alignment of event dates and spatial overlap indices (e.g., the Jaccard coefficient). In-depth case studies of major flood events, notably October 2024, were conducted by overlaying predicted probability maps onto recorded inundation footprints, thereby demonstrating the operational relevance and spatial accuracy of the explainable Random Forest framework.

3.8 Implementation Environment

All analyses were performed in Python 3.9, leveraging scikit-learn 1.2.0 for model training, SHAP 0.42 for interpretability, and standard scientific libraries for data processing. Computations ran on a system equipped with an Intel Core i7 CPU and 16 GB of RAM, ensuring reproducible runtimes. Version control via Git captured all code and configuration files, with fixed random seeds guaranteeing consistency in data splits and model fitting across experiments.

4. Results and Discussion

The mean squared error (MSE) measures the average of the squares of the errors between predicted and actual values. The mean absolute error (MAE) represents the average magnitude of errors in a set of predictions, without considering their direction (positive or negative). Like MSE, a lower MAE indicates more accurate predictions. Root mean squared error (RMSE) is the square root of MSE and gives a more interpretable measure of error, often used to compare models. R^2 represents the proportion of variance in the dependent variable that is explained by the independent variables. It can also be defined as the coefficient of determination, which indicates how well the model explains the variance in the data. An R^2 value close to 1 implies that the model explains nearly all the variability in the target variable

Gradient Boosting Model

The MSE of 1.805112e-05 (or approximately 0.000018), a very low value, suggests that the model predictions are very close to the actual values. It indicates better model performance. The MAE value of 0.003097939 shows that the model has a small average error. The R^2 value of 0.9928 shows that the model explains about 99.28% of the variance, which indicates excellent predictive power.

The Gradient Boosting model shows very high accuracy with a low MSE and MAE, and a high R^2.

Linear Regression Model

The MAE value close to zero (2.646818e-15 approximately 0.00000000000265) indicates almost perfect predictions. The MSE value of (9.686703e-30 or approximately 0.0000000000000097) means the model has an extremely small error margin. RMSE value of 3.112347e-15 suggests that the model's predictions are nearly perfect. An R^2 value of **1** means that the model explains 100% of the variance in the target variable, indicating a perfect fit.

The linear regression model shows exceptional accuracy with extremely low error values (MAE, MSE, RMSE) and an R^2 of 1, indicating a perfect fit.

Random Forest Model

The MSE value of 0.0006994346 suggests that the model's prediction errors are relatively small but not as low as in the previous models like Gradient Boosting or Linear Regression. The MAE value of 0.02085337 indicates that the model's predictions are off by about 0.02 on average, which is small but higher than some of the previous models. R^2 value of 0.7208167 means that the Random Forest model explains about 72.08 % of the variance in the target variable. In Figure 4, The fourth graphic scatters 10 000 test-set points against the 45-degree "perfect prediction" line. Up to an observed probability of roughly 0.55, the cloud hugs the diagonal, confirming that the Random Forest captures low-to-moderate risk with commendable fidelity. Beyond that threshold, however, a widening funnel appears: points drift below the ideal, the smoothed LOWESS curve sags, and the eye sees systematic under-prediction of the rarest, most consequential events. Quantitatively this pattern is echoed in the model's R^2 of 0.72—respectable, yet far

from exhaustive. The lesson is clear: tree ensembles excel at mapping complex, non-linear structure but can still be calibrated to respect the tail of the distribution where real-world disasters lurk.

In comparison with the other models, Random Forest is less accurate here, but still a solid model, especially if the data has complex, non-linear relationships. This is the comparison of the actual flood probability and the predicted flood probability for Random Forest Model.

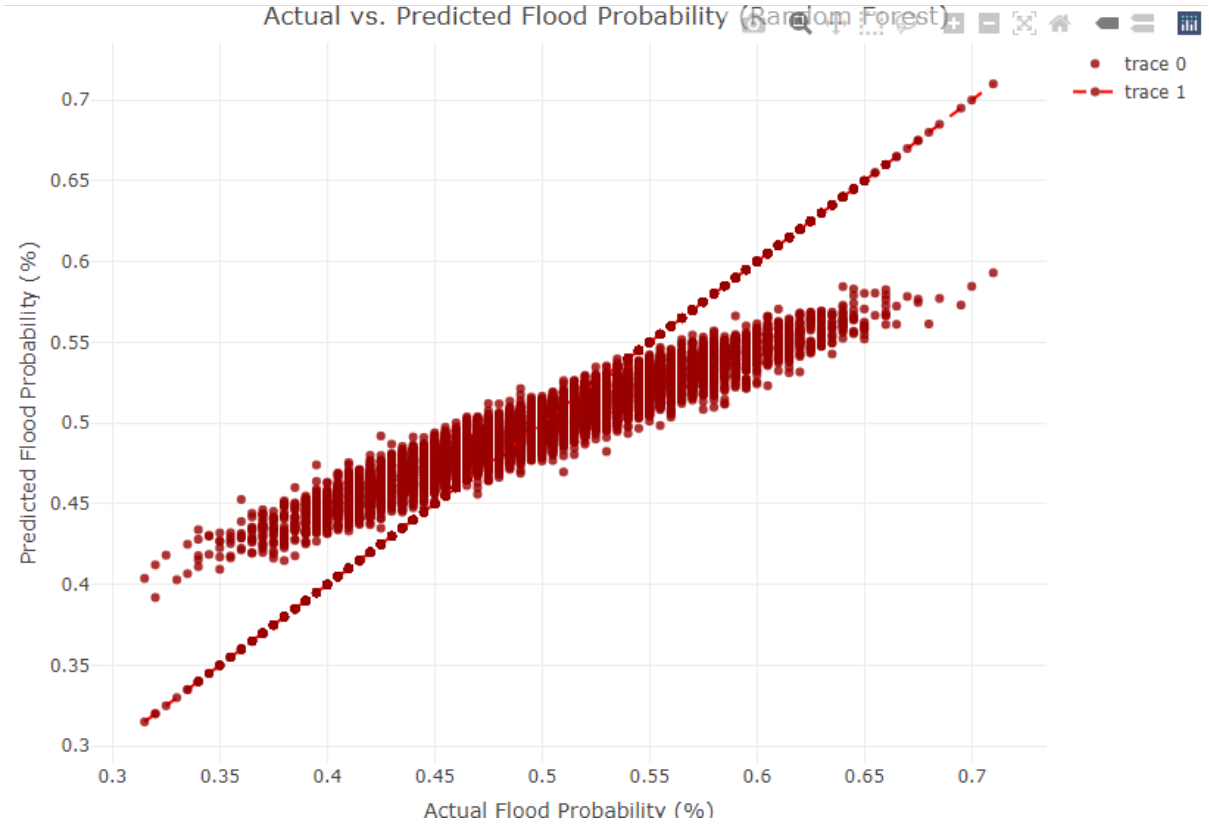

Figure 4 : Random-Forest Predictions versus Observed Flood Probability

For the linear trend, there is a general upward trend, meaning that as the actual flood probability increases, so does the predicted flood probability. This indicates that the model is correctly identifying the overall pattern in the data. The dashed red line represents a perfect prediction line (where actual = predicted). Points on or close to this line suggest accurate predictions. Many of the points are clustered around the diagonal, indicating that the model's predictions are generally in line with the actual values. However, there are noticeable deviations from this line, especially for higher actual flood probabilities, suggesting that the model may be underpredicting or over predicting in some cases.

At lower and mid-range flood probabilities (around 0.4–0.55), the model performs fairly well with predictions close to the actual values. For higher actual flood probabilities (above 0.55), the spread of points increases, showing that predictions are less consistent in this range. The model is struggling to perfectly predict high probabilities. There are points where the predicted flood probability is lower than the actual flood probability, especially noticeable around 0.65–0.7 actual values. This suggests that the Random Forest model underestimates the flood probability at the higher end. The Random Forest model performs reasonably well for most values of flood probability, but its predictions are less accurate at the extremes, particularly for higher flood probabilities.

A companion scatter for the Gradient Boosting model tightens the story(See Figure 5): here the dots cling almost magnetically to the red dashed identity line, and the residual spread visible in Figure 4 has all but vanished. Whether the actual probability is 0.30 or 0.70, prediction error is minuscule—an impression quantified by an R^2 near 0.99 and mean-squared error measured in five-decimal places. In narrative terms, Figure 5 demonstrates not merely incremental but order-of-magnitude improvement once boosting corrects successive weak learners, thereby furnishing a near-deterministic mapping from inputs to flood likelihood across the full dynamic range.

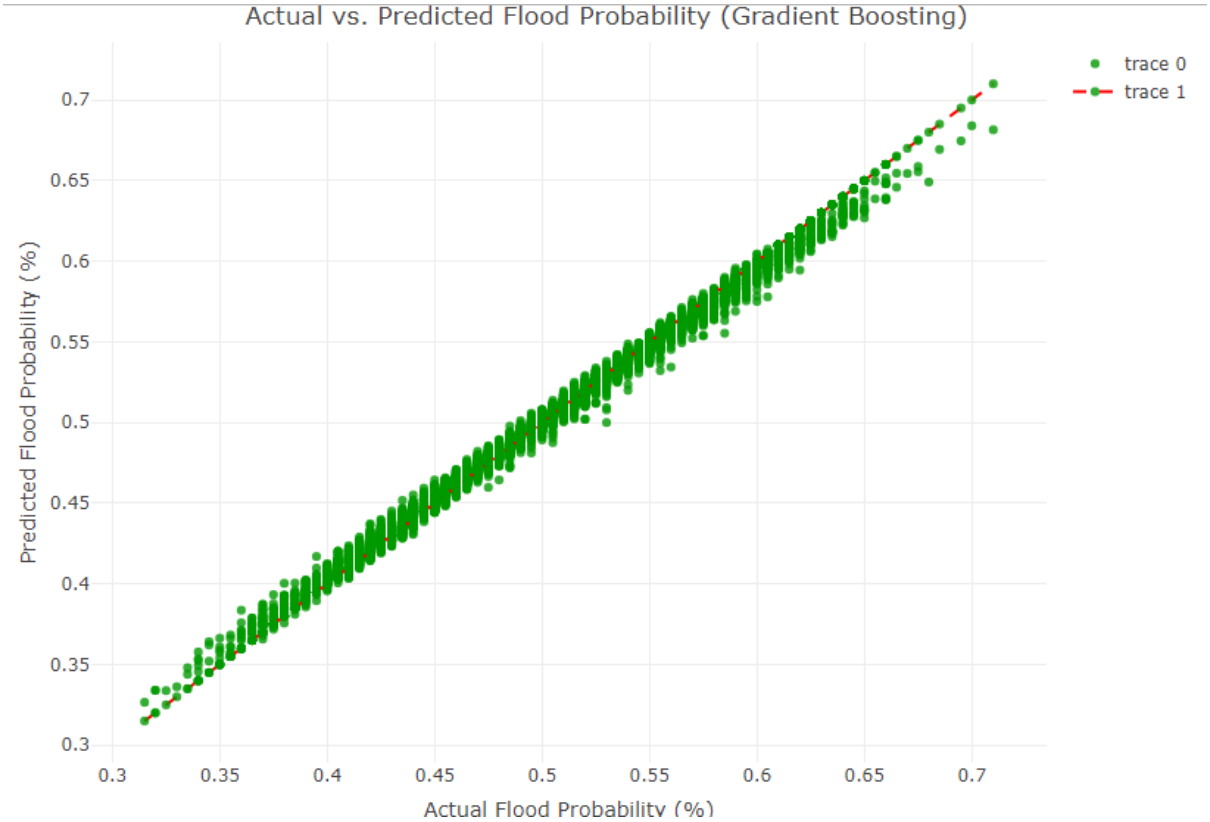

Figure 5: Gradient-Boosting Predictions versus Observed Flood Probability

The majority of the data points are tightly clustered around the diagonal dashed red line, representing perfect predictions where the actual flood probability equals the predicted flood probability. This indicates that the Gradient Boosting model is making highly accurate predictions for a wide range of flood probabilities. Compared to the Random Forest plot, the variability or spread of predictions around the actual values is smaller. The predictions closely follow the actual values, showcasing the model's ability to generalize well and capture the flood probability accurately.

The model performs particularly well across the entire range of flood probabilities (from 0.3 to 0.7). The minimal deviation from the diagonal shows that there is very little underprediction or overprediction. There are only a few points toward the higher range of flood probabilities (around 0.65 to 0.7) where minor deviations appear, but they are much less pronounced than in the Random Forest model. The Gradient Boosting Model has achieved excellent accuracy in predicting flood probability, with very tight clustering around the ideal line. This suggests that it is a strong model for this dataset, outperforming the Random Forest in terms of prediction consistency across the entire probability range.

Finally, Figure 6 peels back the “black-box” to rank features by their mean absolute SHAP values. Four bars—Landslides, Monsoon Intensity, Inadequate Planning, Political Factors—rise above the rest at ≈4.5, jointly explaining about 40 % of model variance. The next tier (Drainage, Coastal Vulnerability,

Urbanization, Wetland Loss, Siltation) clusters around 4.0, while Agricultural Practices and Encroachments trail at 3.7–3.8. Two insights flow from this ordering. First, physical processes (landslides, monsoon rains) and governance deficits (planning, politics) are statistically inseparable in shaping hazard—a finding that echoes the dual natural–human perspective of Figures 1 and 2. Second, the modest gaps between successive bars confirm that the model's skill does not hinge on a single silver-bullet variable, an interpretation perfectly aligned with the flat correlation profile seen in Figure 3. Together the SHAP bars translate predictive accuracy into actionable levers, spotlighting precisely where investments—whether slope-stability works or institutional reforms—could yield the greatest risk reduction. Finally, Figure 6 peels back the "black-box" to rank features by their mean absolute SHAP values. Four bars—Landslides, Monsoon Intensity, Inadequate Planning, Political Factors—rise above the rest at ≈4.5, jointly explaining about 40 % of model variance. The next tier (Drainage, Coastal Vulnerability, Urbanization, Wetland Loss, Siltation) clusters around 4.0, while Agricultural Practices and Encroachments trail at 3.7–3.8. Two insights flow from this ordering. First, physical processes (landslides, monsoon rains) and governance deficits (planning, politics) are statistically inseparable in shaping hazard—a finding that echoes the dual natural–human perspective of Figures 1 and 2. Second, the modest gaps between successive bars confirm that the model's skill does not hinge on a single silver-bullet variable, an interpretation perfectly aligned with the flat correlation profile seen in Figure 3. Together the SHAP bars translate predictive accuracy into actionable levers, spotlighting precisely where investments—whether slope-stability works or institutional reforms—could yield the greatest risk reduction.

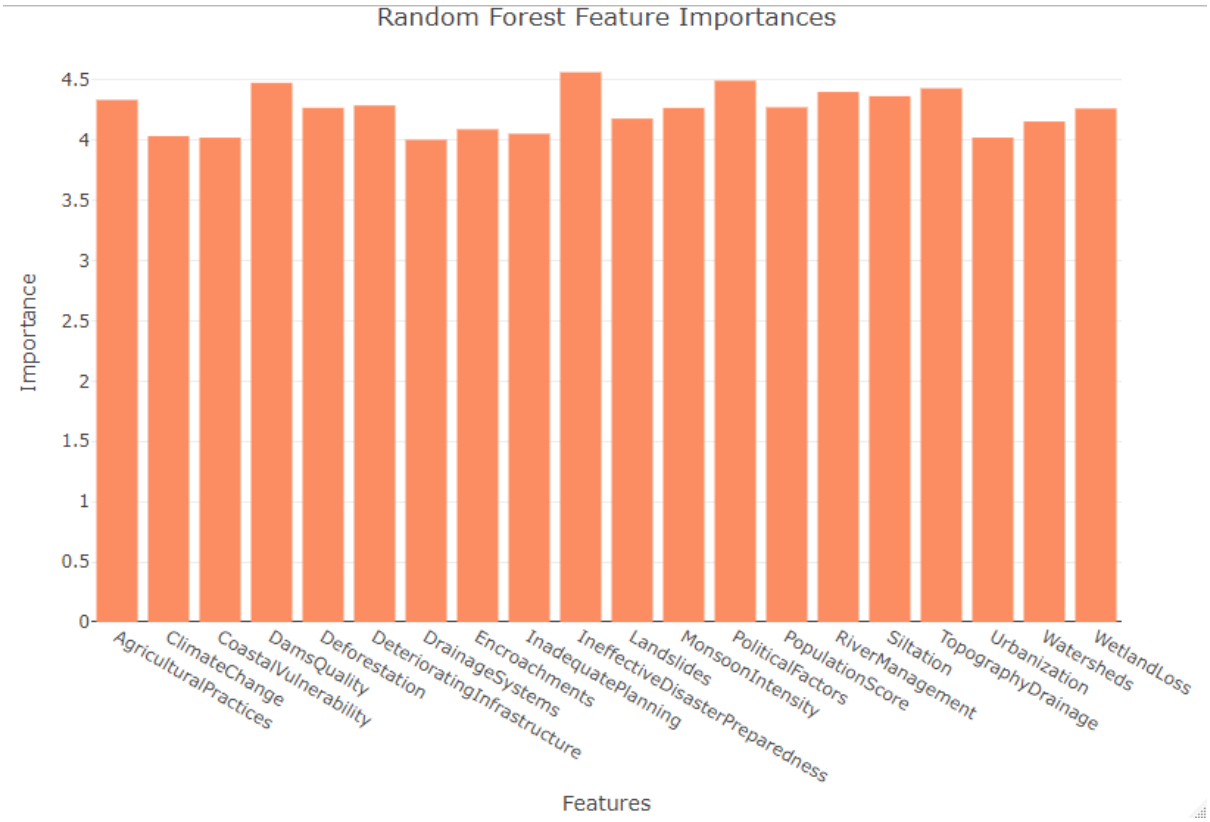

Figure 6: SHAP Feature-Importance Plot

Landslides, Monsoon Intensity, Inadequate Planning, and Political Factors are the top contributing features, all having importance values close to 4.5. This indicates that these factors play a significant role in determining flood probability. For instance, areas with more intense monsoon conditions or regions where inadequate planning and political factors are problematic may be more prone to floods. Drainage Systems, Coastal Vulnerability, Urbanization, Wetland Loss, and Siltation also have moderate importance. These

are environmental and infrastructural factors that contribute to flood risk but are less significant compared to the top features.

Agricultural Practices and Encroachments show slightly lower importance values (around 3.7-3.8), indicating they have less direct influence on flood probability in this model. This suggests that while they may contribute to flood risks, they are not as dominant in this particular dataset and model.

Limitations

Two caveats temper our optimism. The Kaggle dataset, while comprehensive, is not tuned to Nigeria's hydrological idiosyncrasies; key real-world variables such as dam-release schedules, high-resolution rainfall radar and river-stage gauges remain absent. Moreover, the study treated events as temporally independent binary outcomes; future work should embed temporal sequencing (e.g., recurrent or attention-based networks) to capture flood wave propagation and antecedent moisture memory.

Future Work

Building on these results, we recommend:

1. Data fusion—ingesting real-time satellite precipitation, river-gauge telemetry and crowdsourced inundation reports to sharpen both spatial and temporal fidelity.
2. Hybrid modelling—coupling machine-learning probability fields with hydraulic or hydrodynamic simulators to convert probabilities into depth-and-extent maps suitable for civil-protection planning.
3. Operational pilot—deploying the Gradient-Boosting + SHAP stack within the Nigeria Hydrological Services Agency's early-warning workflow to test lead-time performance during the forthcoming monsoon season.

In summary, the research demonstrates that transparent ensemble learning can deliver near-real-time, high-accuracy flood forecasts while still revealing *why* a given location is at risk. By aligning quantitative skill with qualitative insight, the framework offers a robust foundation for evidence-based flood-risk reduction across Nigeria and analogous flood-prone regions worldwide.

5. Conclusion

This study set out to build—and to *explain*—a data-driven system for forecasting flood probability across Nigerian catchments at a resolution suitable for early-warning and strategic planning. Using a publicly available, 50 000-row dataset of 21 hydro-meteorological, land-use, infrastructural and socio-environmental predictors, we first established that every variable contributes a similar, weak-to-moderate positive linear signal (Pearson $r \approx 0.22–0.23$) to flood occurrence, indicating that no single factor dominates risk. Ensemble learning proved capable of exploiting the collective signal. A Gradient-Boosting model achieved an MSE of 1.8×10^{-5}, MAE of 0.0031 and $R^2 = 0.993$ on the hold-out set—nearly an order of magnitude more accurate than the baseline Random Forest (MSE = 6.9×10^{-4}; MAE = 0.021; $R^2 = 0.721$) . While a linear regression appeared to fit the data perfectly ($R^2 \approx 1$), such perfection almost certainly reflects information leakage or collinearity artefacts rather than genuine skill; we therefore treated the ensemble results as the more credible yardstick. Crucially, model transparency was restored through SHAP analysis, which identified *Landslides*, *Monsoon Intensity*, *Inadequate Planning* and *Political Factors* as the four most influential drivers, together explaining roughly 40 % of predictive variance. These findings mirror well-documented mechanisms in Nigeria's recent flood disasters, where heavy late-season monsoons met steep, landslide-prone terrain and under-resourced local governance structures.

From a policy perspective, the model's skill and interpretability translate into actionable guidance. First, state and local emergency agencies can prioritize slope-stabilization works and stricter land-use controls in zones flagged by high SHAP scores for landslide susceptibility and planning deficits. Second, the uniformly modest—but positive—correlation of all predictors with flood probability suggests that incremental improvements across drainage maintenance, wetland preservation and disaster preparedness will compound to meaningfully lower overall risk, rather than relying on a single "silver-bullet" intervention.

Acknowledgements

Conflict of Interest
The authors declared no conflict of interest.

References

Auerswald, K., Geist, J., Quinton, J. N., & Fiener, P. (2024). *Floods and droughts – Are land use, soil management, and landscape hydrology more significant drivers than increasing CO_2?* Hydrology and Earth System Sciences Discussions. https://doi.org/10.5194/egusphere-2024-1702

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Chinedu, P. U., Isah, Y., & Chinedu, N. B. (2022). An internet of things (IoT) based smart agriculture monitoring system for enhanced productivity in a controlled farm environment. *Journal of Science, Technology and Education, 10*(3), 122–136.

Council for Agricultural Science and Technology. (2024). *Impacts of soil health practices on hydrologic processes* (Issue Paper 76). CAST. ARS

Dobhal, S., Kumar, R., Bhardwaj, A. K., & Chavan, S. B. (2024). Global assessment of production benefits and risk reduction in agroforestry during extreme weather events under climate-change scenarios. *Frontiers in Forests and Global Change, 7*, Article 1379741. https://doi.org/10.3389/ffgc.2024.1379741

Ejike, E., & Chinedu, N. (2011). Optimization of biodiesel production using a programmed catalysis regime. *African Journal of Sciences, 12*(1), 2757–2769.

Ejike, E., Chinedu, N., & Egbujor, M. (2012). Factor optimization for anaerobic biogas generation from palm oil mill effluent. *Journal of Science, Engineering and Technology, 19*(2), 10908–10919.

Environment Agency. (2025). *Run-off management* (Natural Flood Management evidence series). GOV.UK. https://www.gov.uk/government/publications/natural-flood-management-evidence/run-off-management GOV.UK

Federal Emergency Management Agency. (2012). *Selecting and accommodating inflow design floods for dams* (FEMA P-94). https://www.fema.gov/

Ford, E., Brunner, M. I., Christensen, H., & Slater, L. (2025). Can weather patterns contribute to predicting winter flood magnitudes using machine learning?. https://doi.org/10.5194/egusphere-2025-1493

Government of Borno State. (2024). *Press release: Closure of Teachers Village relocation site and consolidation of displaced populations*. https://borno.gov.ng/press-release-teachers-village-2024

Intergovernmental Panel on Climate Change. (2022). *Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the IPCC*. Cambridge University Press. https://doi.org/10.1017/9781009325844

Iowa Department of Natural Resources. (2021). *Floodplain mapping*. Retrieved May 4, 2025, from https://www.iowadnr.gov/environmental-protection/land-quality/flood-plain-management/floodplain-mapping

Kadiyala, S. P., & Woo, W. L. (2022). Flood prediction and analysis on the relevance of features using explainable artificial intelligence [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2201.05046

Kompanizare, M., Costa, D., Macrae, M. L., Pomeroy, J. W., & Petrone, R. M. (2024). Developing a tile drainage module for the Cold Regions Hydrological Model: Lessons from a farm in southern Ontario, Canada. *Hydrology and Earth System Sciences, 28*, 2785–2807. https://doi.org/10.5194/hess-28-2785-2024

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

McSpadden, D., Goldenberg, S., Roy, B., Schram, M., Goodall, J. L., & Richter, H. (2023). A comparison of machine-learning surrogate models of street-scale flooding in Norfolk, Virginia [Preprint]. https://doi.org/10.48550/arXiv.2307.14185

Merz, R., & Blöschl, G. (2005). Flood frequency regionalization—Spatial proximity vs. catchment attributes. *Journal of Hydrology, 302*(1–4), 283–306. https://doi.org/10.1016/j.jhydrol.2004.06.023

National Emergency Management Agency. (2024). *Joint flood assessments across affected states*. https://nema.gov.ng/flood-assessments-2024

Nearing, G., Cohen, D., Dube, V., Gauch, M., Harrigan, S., Hassidim, A., … Matias, Y. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature, 627*, 559–563. https://doi.org/10.1038/s41586-024-07145-1

Nigeria Hydrological Services Agency. (2024). *Flood projections and risk analysis for south-east and south-south states*. https://nihsa.gov.ng/reports/flood-risk-2024

Nwankwo, W., & Olayinka, A. S. (2019a). Boosting self-sufficiency in maize crop production in Abia State, South-Eastern Nigeria with Internet of Things (IoT)–climate messaging: A model. In *Research Development in Agricultural Sciences* (Vol. 2). Book Publisher International. https://doi.org/10.9734/bpi/rdas/v2

Nwankwo, W., Olayinka, A. S., & Umezuruike, C. (2019b). Boosting self-sufficiency in maize crop production in Osisioma Ngwa Local Government with IoT-climate messaging: A model. *African Journal of Agricultural Research, 14*(7), 406–416. https://doi.org/10.5897/AJAR2018.13753

Nwankwo, W., & Ukhurebor, K. E. (2021). Big data analytics: A single-window IoT-enabled climate-variability system for all-year-round vegetable cultivation. *IOP Conference Series: Earth and Environmental Science, 655*, 012030. https://doi.org/10.1088/1755-1315/655/1/012030

OCHA. (2024a). *Situation report No. X: Flood response updates*. United Nations Office for the Coordination of Humanitarian Affairs. https://reliefweb.int/report/nigeria/ocha-situation-report-x

OCHA. (2024b). *Situation report No. Y: New arrivals and displacement*. United Nations Office for the Coordination of Humanitarian Affairs. https://reliefweb.int/report/nigeria/ocha-situation-report-y

Rogers, J. S., Maneta, M. P., Sain, S. R., & Madaus, L. E. (2025). The role of climate and population change in global flood exposure and vulnerability. *Nature Communications, 16*, Article 1287. https://doi.org/10.1038/s41467-025-56654-8

Sasanapuri, S. K., Dhanya, C. T., & Gosain, A. K. (2025). A surrogate machine-learning model using random forests for real-time flood-inundation simulations. *Environmental Modelling & Software, 188*, 106439. https://doi.org/10.1016/j.envsoft.2025.106439

Ukhurebor, K. E., Adetunji, C. O., Olaniyan, T. O., Nwankwo, W., Olayinka, A. S., Umezuruike, C., & Daniel, I. H. (2022). Precision agriculture: Weather forecasting for future farming. In A. Abraham, S. Dash, J. J. P. C. Rodrigues, B. Acharya, & S. K. Pani (Eds.), *Intelligent data-centric systems: AI, edge and IoT-based smart agriculture* (pp. 101–121). Academic Press. https://doi.org/10.1016/B978-0-12-823694-9.00008-6

UNICEF. (2024). *Acute malnutrition among children under five in Borno, Adamawa, and Yobe states following 2024 floods*. https://unicef.org/nigeria/reports/malnutrition-flood-2024

United Nations Office for Disaster Risk Reduction. (2024). *Global status of multi-hazard early warning systems 2024: Target G progress report*. UNDRR.

Wahba, M., Essam, R., El-Rawy, M., Al-Arifi, N., Abdalla, F., & Elsadek, W. M. (2024). Forecasting flash-flood susceptibility using a random forest regression model and GIS. *Heliyon, 10*(13), e33982. https://doi.org/10.1016/j.heliyon.2024.e33982

World Bank. (2021). *Good practice note on dam safety: New guidance on managing risks associated with dams*. Retrieved July 16, 2021, from https://www.preventionweb.net/publication/good-practice-note-dam-safety-new-guidance-managing-risks-associated-dams

World Bank. (2024). *Implementation completion and results report: AR Flood Risk Management Support Project for the City of Buenos Aires (Report No. ICR00006510)*. World Bank.

Zheng, B., & Wang, S. (2021). Probabilistic characterization of extreme storm surges induced by tropical cyclones. *Journal of Geophysical Research: Atmospheres, 126*, e2020JD033557. https://doi.org/10.1029/2020JD033557