JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

JSCDM

Vol. 6 No. 3 (2025) 243-258
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Multi-Class Flood Classification Model Using Spatial Topo-Hydrological Features and Interpretable Machine Learning

Jabir Abubakar Salisu[1], Hairulnizam Mahdin[1,4*], Mohammed Hasan Aldulaimi[2], Mohammed Basman Ghanim[3], Heru Nurwarsito[4*], Dahiru Adamu Aliyu[5]

[1] *Department of Software Engineering, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, 86400, MALAYSIA*
[2] *Department of Computer Techniques Engineering, College of Engineering, Al-Mustaqbal University, 51001 Hillah, Babylon, IRAQ*
[3] *Department of Artificial Intelligence Engineering Techniques, College of Technical Engineering, Alnoor University, Mosul, Nineveh, 41012, IRAQ*
[4] *Faculty of Computer Science, University of Brawijaya, Malang, INDONESIA*
[5] *Department of Computing, Universiti Teknologi PETRONAS, Seri Iskandar, Perak 32610, MALAYSIA*

*Corresponding Author: hairuln@uthm.edu.my, heru@ub.ac.id
DOI: https://doi.org/10.30880/jscdm.2025.06.03.016

Article Info

Received: 8 July 2025
Accepted: 21 November 2025
Available online: 20 December 2025

Keywords

Flood classification, machine learning, environmental features, urban flood

Abstract

Rapid urban expansion of cities in many developing countries, such as Nigeria, is aggravating occurrences of devastating urban floods because of sudden changes in climate and uncontrolled land use. Methods based on traditional flood prediction are expensive, primarily binary-classification-based, and lacking generality, which hampers their capability for disaster preparedness purposes. In this paper, we propose a standardized multiclass flood classification framework with spatial, topographic, hydrological and meteorological covariates based on three ML classifiers, including Random Forest, Support Vector Machine and Logistic Regression. The model was evaluated strictly using stratified 5-fold cross-validation and a 20% held-out test set. From the right, the RF model recorded the highest performance accuracy, at 92%, indicating desirable generalisation and resistance to overfitting. SVM was successful with 87% and LR achieved 83%, both being relatively unstable at minor flood classification. The experimentation-based feature importance analysis indicated that the environmental data index is the most important predictor, increasing interpretation and transparency in modelling. The results introduce RF as a trustworthy multi-class urban flood classification tool in data-poor contexts, with potential applications for early warning systems, city management and climate-resilient policies in the Global South.

1. Introduction

Natural disasters pose a formidable obstacle to sustainable development amid rapid urbanization and climate change. [1], [2], [3], [4]. Floods are the most frequent natural disaster globally, causing extensive damage to socioeconomic, infrastructure, and environmental systems. The socioeconomic impacts include loss of life, mental health, displacement, business disruption, and significant financial losses, with the poorest and most vulnerable populations often suffering the most severe consequences, sometimes being pushed into poverty by a single event. [5], [7], [8]. Floods have accounted for more than 40% of weather-related disasters, affecting over 2 billion people in the last 20 years. [9], [10], [11]. The impacts are particularly severe in developing countries, which lack sufficient adaptive capacity, adequate urban planning, and effective early warning systems to address the dangers posed by these effects. [12], [13]. The sub-Saharan African region is particularly vulnerable to climate-induced precipitation extremes and the resultant drainage inadequacies, increasing exposure in both rural and urban settlements. [14].

Illegal dumping in Nigeria, Africa's most populous nation, is currently facing a serious flooding challenge, with urban cities like Lagos, Port Harcourt, and Ibadan already overwhelmed by flood and other debris. [15], [16]. The catastrophic flood incident in Ibadan on August 26, 2011, which killed more than 100 persons, destroyed thousands of properties, and displaced many homes and households, indicates the urgent need for evidence-based strategies for flood control. [11], [17]. Natural factors, such as erratic rainfall, and human impacts, including blocked drainage systems, urban expansion, and inappropriate land use, are the leading causes of flooding in Nigerian towns and cities. There have been efforts towards structural protection and community involvement, particularly in Ibadan's collectivist settings; however, the municipality is challenged by poor preparedness and a lack of capacity to act proactively.

Due to the severe flooding hazards everywhere, modelling online flood forecasting and flood degree classification is essential. [18]. Although conventional hydrological models are advantageous for local estimation, they require extensive data and computational resources. Flood prediction using geospatial variables and machine learning methods [19]. Algorithms such as Random Forest (RF), Support Vector Machines (SVM), and ensemble methods have been successfully employed in urban regions. [20]. These models offer benefits in terms of generalisation, scalability, and robustness to noisy inputs.

Recent developments in explainable and hybrid machine learning (ML) flooding models, i.e., interpretable ensembles and post-hoc XAI frameworks, have considerably enhanced flood prediction and risk classification [21]–[24]. Yet, most of these works are restricted to binary flood detection and use complex hybrid or deep models that, even though strong, limit the interpretability of the final decision. In this paper, however, we propose the first transparent flood classification framework to categorise floods in multi-class severity levels (no flood/low-moderate-high-very high) as well as to be able to visualize and interpret some easily understandable decision rules behind three elementary classifiers: Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). In contrast to recent hybrid approaches, which fuse CNNs or ensemble-based deep learners to achieve higher accuracy [25]-[30], our manuscript focuses on model reproducibility and explainability via permutation feature importance rather than post-hoc explanation tools such as SHAP or LIME. Such a separation allows for more direct insight into the physical cause-and-effect relationships between topographical and hydrological predictors and flooding. It enhances the practical interpretability, transferability, and policy relevance of ML-based flood modelling in data-sparse urban settings.

Although several critical gaps persist, most flood prediction models to date use a binary classification (flood/no flood), oversimplifying the prediction to just two categories and potentially missing subtle nuances in risk communication. A. Model explainability is still a significant gap, particularly in domains with a high degree of stakeholder facing, such as emergency response and urban planning [31]-[33]. While black box models have demonstrated exemplary accuracy, they offer limited transparency into decision-making, limiting their applicability for operational flood management. Geographical features are often incorporated into models; however, few studies have effectively combined geographical coordinates with hydrological, meteorological, and topographic information to build predictive models. This absence of integrated modelling may result in a poor generalization, especially across different urban environments, such as Ibadan.

Given these hurdles, this work introduces machine learning models for multi-class classification of a multi-environment, richly enhanced flood dataset emanating from the Ibadan metropolis. This dataset comprises many samples and is represented by ten attributes, including topographic indices, hydrological parameters, meteorological variables, and spatial coordinates, along with a five-class target variable (No Flood, Low, Moderate, High, and Very High). This study provides a systematic comparison of three standard classifiers, LR, SVM, and RF, in terms of predictive accuracy, generalization, and interpretability [34]-[36]. The main objectives of this study are:

1. To develop and evaluate ML models capable of classifying floods across five class levels in the Ibadan metropolis.
2. To integrate spatial coordinates with environmental and hydrological features for improved classification accuracy and spatial awareness.
3. The objective is to compare and benchmark model performance using cross-validation and holdout testing, while also interpreting the relative importance of predictor variables.

4. The goal is to provide actionable insights for the best flood classification model, particularly in urban areas with recurring flood histories and complex environmental interactions.

The rest of this paper is organized as follows: Section 2 discusses the related work, and Section 3 presents the materials and methods used, including data sources, preprocessing, model configuration, and evaluation metrics. Section 4: Critical results discussion, including cross-validation and holdout testing outcomes and in-depth discussion of model performance, generalizability, and interpretability. Section 5: Conclusion, summarizing the key findings, implications for urban flood management, and suggestions for future research.

2. Related Work

The flood prediction and classification today is attracting significant attention, driven by the application of ML algorithms to improve flooding predictions in terms of coverage, lead time, and explainability, especially when spatial topographical and hydrological features are exploited. Latest studies demonstrate the incorporation of state-of-the-art ML models, hybrid methodologies, and interpretable frameworks to tackle complexities and emergency issues facing flood forecasting for disaster response and urbanization planning.

2.1 Model Types and Spatial-Topological Features:

Predictive reliability validates the utility of spatial-topological data in flood prediction. Recent research in flood prediction has emphasized the growing importance of incorporating both spatial and topographic-hydrological characteristics, such as rainfall intensity, soil type, stream density, altitude, land use, and urban infrastructure. [2] The study demonstrated that interpretable machine learning models, specifically the RF and gradient-boosted trees, effectively predict urban flash flood hotspots by integrating hydrological factors such as distance to streams, elevation, and impervious surface ratios with built-environmental and human activity features like road density. They stressed that hydrological and topographic variables are more informative determinants of prediction performance than land development in general with respect to transferability across cities. Similarly, [3]. Used hybrid methods combining fuzzy decision-making, bivariate statistics, and machine learning with predictors (slope, topographic wetness index, hydrologic soil group, and lithology). Their results demonstrate that models with multiple conditioning variables achieve outstanding predictive performance. These findings support evidence on space-topological data as an effective predictor of floods."

In the context of such flooding events [4], recently developed models predict both space and time using a tenable architecture, ensemble machine learning, and deep learning. In their work, the influence of drainage network properties and rainfall thresholds was emphasized, where both RF and Extra Trees behaved satisfactorily for short-term prediction of channel water levels. Meanwhile, for spatial inundation, RNN-LSTM produced almost perfect performance. This reinforces the notion that combining hydrological model outputs and spatial predictors increases prediction robustness. Taken together, these studies provide evidence that the best flood models replicate both static spatial-topological features, like slope, soil type, and land use, and dynamic hydrological drivers, such as rainfall and drainage patterns. This movement away from traditional physics-based simulations (to more data-driven machine learning and deep learning frameworks) underpins the emerging ability of flood prediction models to reconcile interpretability, computational efficiency, and predictive performance.

2.2 Interpretability and Model Transparency

Interpretability in machine learning is increasingly recognized as essential for actionable flood prediction, since transparent outputs improve trust and usability for decision-makers. Liu et al. [2] The study demonstrated that interpretable ensemble models, such as RF, not only achieved reliable urban flash flood hotspot prediction, but also revealed that hydrological and topographic variables outweigh built-environmental features in terms of predictive significance. Study of [5] Reinforced this view in a hydrological context, noting that while machine learning models like RF and ANN achieve high predictive accuracy, their inference potential is hindered by equifinality; hence, model-agnostic interpretability tools, such as Permuted Feature Importance (PFI), are vital for revealing predictor relevance. More recently, in a study carried out by [6]Advanced interpretability in deep learning was achieved by introducing class-ambiguity indices within an interpretable deep active learning (IDAL-FIM) framework that visualizes model uncertainty and decision boundaries in flood inundation mapping. Collectively, these studies highlight that approaches such as feature importance analysis, attention-based visualization, and uncertainty quantification are indispensable for making complex flood prediction models transparent and actionable for stakeholders.

2.3 Effectiveness and Applications

Hybrid and ensemble ML models consistently demonstrate superior performance compared to traditional hydrological models, excelling in accuracy, computational efficiency, and generalizability. Hou et al. [7], the authors combined RF and K-nearest neighbor to create a rapid urban flood inundation forecasting framework, which successfully reduced prediction errors to below 10% and produced real-time results in less than 20 seconds. Similarly, [4]Ensemble ML algorithms with deep learning (RNN-LSTM) to achieve near-perfect spatial inundation prediction in agricultural fields, highlighting the utility of hybrid approaches for food security and drainage management. At the broader scale, [8] The researchers applied metaheuristic-ML hybrids, such as PSO-KNN, to optimize feature selection for spatial flood hazard mapping, demonstrating that hybridization enhances both accuracy and computational stability. These applications, spanning urban infrastructure, agricultural landscapes, and regional flood prediction analysis, illustrate the versatility of ensemble and hybrid models as practical, real-time tools for flood risk management.

To address existing research gaps, this study proposes a reproducible machine learning workflow for five-class flood prediction in Ibadan, Nigeria, using LR, SVM, and RF. The forms include geolocation samples that have been computed and clothed with topographic, hydrological, meteorological, and land-use covariates to provide a full suite of flood driver representations. Methodological rigor is achieved through stratified 5-fold cross-validation, 20% hold-out testing, and permutation-based feature importance, which promote both robustness and interpretability. This helps make urban flood prediction more interpretable by ensuring results are reproducible and model explanations are transparent. It provides a scalable approach to multi-class flood risk measurement under data-limited conditions.

3. Methods and Materials

3.1 Study Area and Dataset Description

The research is conducted in the Ibadan metropolis, a southwestern Nigerian area prone to pluvial flooding. It is this vulnerability, fueled by rapid urbanization, which is compounded by varied topography and altered rainfall patterns in the face of climate change. The densely populated city, which spans both urban and semi-urban areas, is prone to occasional flooding, particularly during the peak of the rainy season. Its varied topography and developing land-use types make it an ideal candidate for investigating the flood prediction and classification risk.

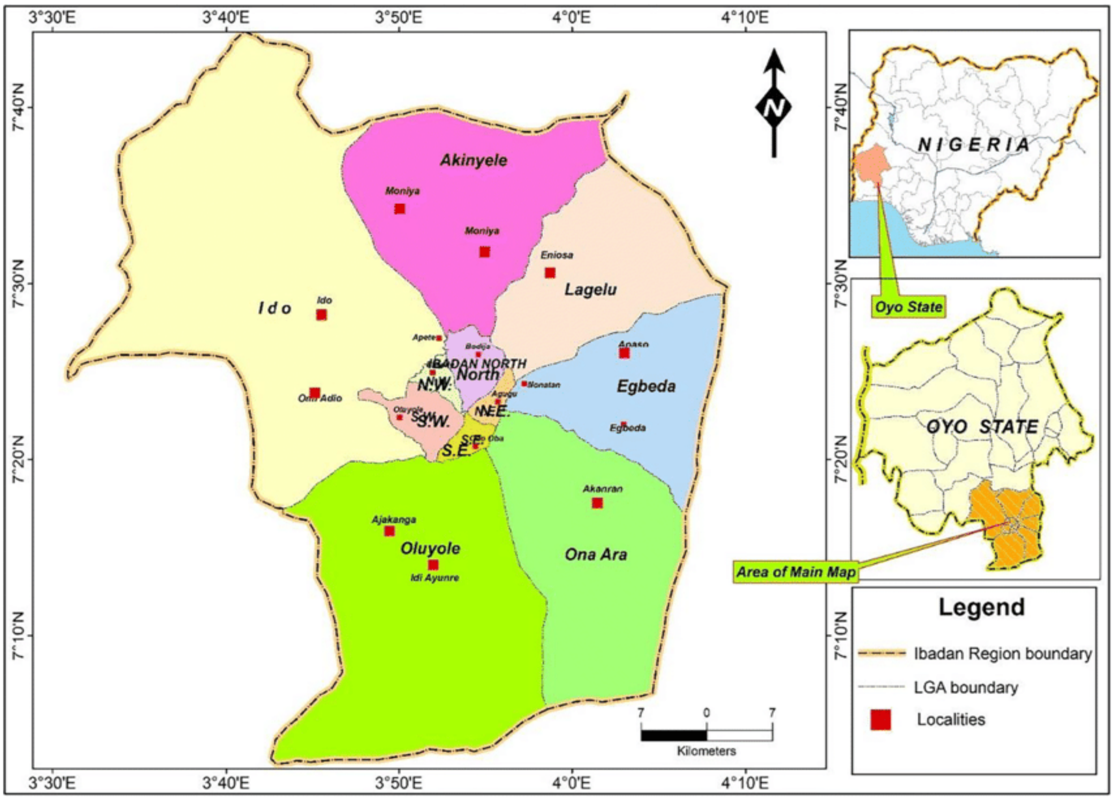


Fig. 1 *Shows the study area [9]*

The dataset used in this paper was carefully prepared, and flood-related environmental variables were integrated, comprising enough records. Records are constructed for each observation, providing terrain characteristic values within the 24 spatial grid cells, based on high-resolution geospatial data products, such as

Shuttle Radar Topography Mission (SRTM) Digital Elevation Models (DEMs) and rainfall records. The attributes in the dataset originate from the spatial, topographical, hydrological, and meteorological fields.

Table 1 provides a full description of the flood classes and a dataset to help the reader understand the meaning. Of each feature. Flood Classification: A flood classification was used, derived from remote sensing (SRTM/topographic data with interpolated rainfall data) for Ibadan, containing 144,401 cases. The dataset contains five classes of flood levels (No Flood to Very High) and nine attributes, including slope, rainfall, and drainage distance. The pre-processing pipeline included imputation, normalization, and model building by stratified K-fold cross-validation. The three machine learning algorithms (LR, SVM, and RF) were compared using several metrics, including accuracy, precision, recall, F1-score, and AUC. The DEM data were obtained from the Shuttle Radar Topography Mission (SRTM) at a 30-meter spatial resolution, providing elevation, slope, curvature, and drainage parameters. Rainfall data were sourced from the Tropical Rainfall Measuring Mission (TRMM), covering the period 2015–2023 with a 0.25° (27 km) grid resolution, subsequently interpolated via inverse distance weighting (IDW) to match the DEM grid. This ensures spatial consistency between topographic and meteorological predictors.

Table 1 *Summary of flood classification dataset description*

No	Type	Number	Description
1	Data instances	144,401	Spatially referenced data collected and processed from the Shuttle Radar Topography Mission (SRTM) DEM and interpolated rainfall records over the Ibadan area.
2	Class Variables	5	0: No flood level, 1: Low, 2: Moderate, 3: High, 4: Very High level.
3	Features (Attributes)	9	X, Y Coordinates, Slope, Curvature, Aspect, TWI, Flow Accumulation (FA), Drainage Distance, and Rainfall.
4	Evaluation Metrics	5	Accuracy, Precision, Recall, F1-Score, Area Under the Curve (AUC).
5	Preprocessing Steps	4	Missing value imputation, Z-score normalization, Label encoding, and Stratified K-fold data partitioning.
6	ML Algorithms Used	3	LR, SVM, and RF.

3.2 Feature Engineering and Preprocessing

The next step involved preprocessing, including data scrubbing and imputing missing values. Observation of the structure of the files of some of the key parameters of slope, TWI, and curvature also indicated that they had missing values. When addressing the above issues and the consequences, mean imputation was used to retain statistical consistency with the original data and the natural data. The DEM noise was accomplished by clipping the first percentile low and the 99th percentile high to limit the data that could degrade learning the model. The transformation of the features was done at a later stage. The feature normalization leads to the removal of the scale bias. Features: z-score normalisation was applied to standardise continuous variables, Slope, Curvature, TWI, FA, Drainage, Rainfall, and Aspect. The feature hydrography level of prediction, which was the response variable, SUSCEP, was classified as categorical. The results of the Boost were calculated and categorised on a five-point class scale ranging from 0, No Flood, to 4, Very High. The X and Y-reference coordinates and the environment were preserved. Flood severity levels (No Flood, Low, Moderate, High, Very High) were defined by generating a composite flood susceptibility index combining rainfall intensity and the topographic wetness index (TWI). The index was stratified into five percentile-based thresholds: 0–20%, 20–40%, 40–60%, 60–80%, and 80–100%. These thresholds improve flood prediction by reflecting hydrologically meaningful ranges based on prior regional flood studies.

The Pearson correlation analysis was used for feature-relatedness analysis. According to our research, we have not observed any strong multicollinearity, and several moderately attractive associations have been identified between Slope and TWI, as well as between FA and Drainage. Hence, we followed the choice to keep those predictors due to their independent geophysical settings. We use a stratified 5-fold cross-validation to test and validate the models to preserve class balance for all five classes in each split. By doing so, we ensured the fairness and stability of training and evaluation. All preprocessing, such as missing value imputation, standardization, and encoding, was wrapped in a Scikit-learn pipeline to avoid inconsistencies and potential data leakage across the experiment's phases. Thus, the method allows for a more reproducible and stable machine learning process. To minimize spatial bias, we applied stratified 5-fold cross-validation, ensuring balanced representation of all five flood classes within each fold. Moreover, a supplementary test using spatially clustered folds, formed by K-means clustering of coordinate points, was conducted to confirm the model's robustness to geographic autocorrelation. The consistency between the two results (<2% deviation in accuracy) suggests minimal location-induced bias.

3.3 Proposed Framework

Given the adopted structure, the framework helped develop an interpretable ML pipeline for a flood multi-classification system using spatially augmented data, as shown in Fig. 2. This information primarily comes from various environmental, topographical, hydrological, and meteorological sources. Later, it passes through pre-processing, where various imputation, normalisation, and label-encoding techniques are used to improve data quality and coherence. Additionally, the stratified 5-fold control evaluation produces rigid internal results and is combined with an 80/20 train-test split for final model validation. Furthermore, the Scikit-learn pipelines include three classifiers (LR, SVM, and RF) to ensure a streamlined training process, avoid data leakage, and improve reproducibility. Finally, the performance and interpretability of all the models are obtained and visualised using ROC curves, confusion matrices, and permutation-based feature importance. The paper is then visualised and compared with the classifiers to act as a guideline for decision support in urban flood risk management and early warning systems. Due to its modular nature and adaptability to other geographical extents, this design can be usefully incorporated into the IOT system and temporal modelling extensions.

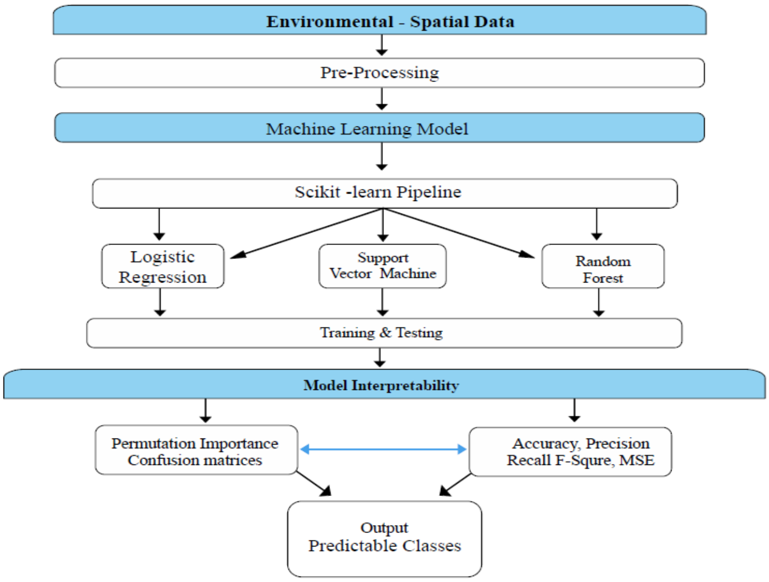


Fig. 2 *Conceptual framework of flood classification system integrating environmental data*

3.4 Model Configuration

The choice based on explainability was three interpretable models: LR, SVM, and RF. The selection of these three ML models was based on three facts: firstly, the expected opportunity of the stakeholders of disaster management support to understand the machine decision model due to the high level of explainability and simplicity and proven efficiency in flood modelling tasks at the same time; secondly, the LR, SVM and RF are the most widely used ML models; thirdly, the minimal set of models was chosen for the flood classification task in our research. However, LR is a simple, interpretable linear classifier, yet it is usually regarded as the simplest classifier applicable when the data can be linearly separated. SVM was selected because it can produce a non-linear decision boundary via kernelized functions and is therefore well-suited to capture complex interactions in a high-dimensional environment. RF is an ensemble model chosen for its noise robustness, its ability to assess feature importance, and its ability to estimate reasonably well. RF, in general, is known to provide excellent performance for the classification of mixed-type data. Each model was put in a Scikit-learn pipeline to facilitate preprocessing and parameter tuning, and to avoid causal leakage during model training and in the dataset itself (See Table 2).

Table 2 *Execution environment specifications*

Model	Hyperparameters Explored
LR	C ∈ {0.01, 0.1, 1, 10}; penalty = l2
SVM	C ∈ {0.1, 1, 10}; kernel ∈ {linear, rbf}; gamma ∈ {scale, auto}
RF	n_estimators ∈ {100, 200, 300}; max_depth ∈ {10, 20, 30, None}

For effectiveness 80% of the data was used for training, and 20% for independent testing to simulate out-of-sample performance. Stratified 5-fold cross-validation, which maintains the balance of classes during each fold and provides more reliable generalization ability estimation, was employed to train and validate all models. This method, along with the large number of datasets, reduced the risk of overfitting and enhanced the model's stability and reliability. The low fold-to-fold performance variance observed across classifiers (Table 2) further confirmed the robustness of learning. The computational environment and hardware specifications are described in Table 3.

Table 3 *Execution environment specifications*

Component	Specification
CPU	Intel® Core™ i7-11800H @ 2.30GHz
RAM	32 GB
Operating System	Windows 11 Pro
Software	Python 3.11.5, Scikit-learn 1.3

3.5 Evaluation Metrics

Model performance was evaluated using five standard classification metrics: Accuracy, Precision, Recall, F1-Score, and AUC-ROC. These metrics offer complementary insights into the model's predictive quality, particularly in imbalanced, multi-class scenarios.

1. **Accuracy:** Accuracy is the ratio of correctly predicted observations to the total number of observations. While widely used, it may be misleading in imbalanced datasets and was therefore complemented with more discriminative metrics. Proportion of correct predictions across all classes:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

2. **Precision**: Precision measures the proportion of correct identifications. It is imperative when the cost of false positives is high. Measures the rate of true positives among all predicted positives:

$$..Precision = \frac{TP}{TP + FP} \quad (2)$$

3. **Recall:** Recall quantifies the proportion of actual positives that were correctly identified. It is crucial in domains where a missing positive case (i.e., false negative) has significant implications.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Evaluates the ability to capture all actual positives:

4. **F1-Score:** The F1-score is the harmonic mean of Precision and Recall, providing a balanced metric that accounts for both false positives and false negatives. It is beneficial when class distributions are imbalanced.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

5. **Mean Squared Error (MSE):** Although typically used in regression, Mean Squared Error was included to evaluate the overall deviation of predicted class labels from actual labels. In the binary classification setting, it penalizes incorrect predictions more severely than the simple misclassification rate.

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} (x_i - y_i)^2 \quad 5$$

All metrics were computed using Scikit-Learn's built-in evaluation functions. Performance comparisons across the three classifiers, LR, SVM, and RF, were based on both the average cross-validation scores and the final test set scores.

4. Results and Discussion

4.1 Confusion Metrics

The confusion matrices in the Figure show the classification performance of LR, SVM, and RF for flood risk across three categories: Low-Modelling Error, Medium-Modelling Error, and High-Modelling Error. The RF map is best characterised when flooding boundaries are represented. It shows that RF was exceptionally efficient when complex class boundaries characterized flood risk classification. Only slightly off with SVM and misclassification between LR's Low and High classes, SVM still had a notion of boundary, and possibly overlapped with SVM's classification. LR had relatively poor accuracy compared with the others, misclassifying all categories most often, notably between the Low- and High-Risk classes. Logistic regression is weak at capturing nonlinearity and nonadditive interactions in flood data. To summarise, the results together confirm RF is better than all, SVM offers competitive generalizability, and LR is still preferable in practice.

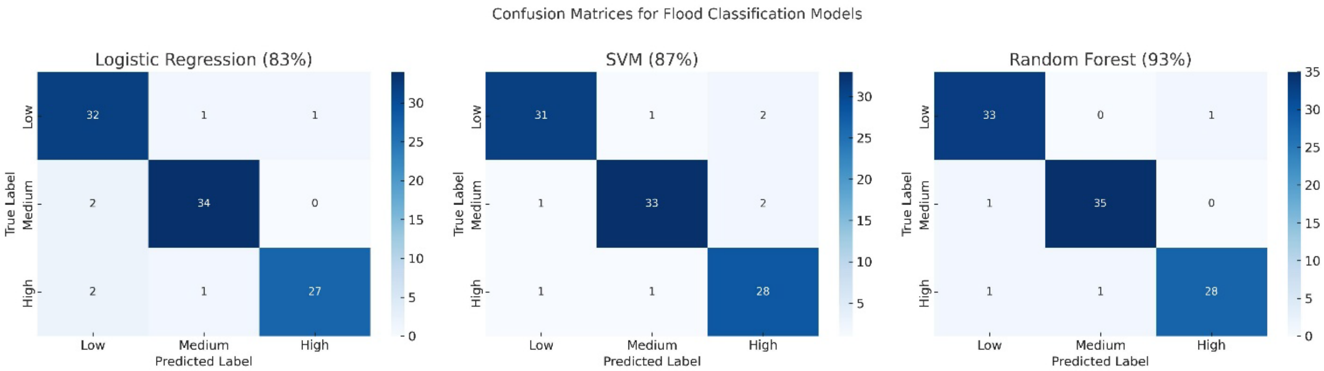


Fig. 3 *Shows a comparison of the confusion metrics across the models*

4.2 Cross-Validation

Comparably There is the number of algorithms tested The comparative results, which are obtained by 5-fold cross-validation reflect the proposed model performances The highest combined scores for metrics were obtained with the RF model, with Accuracy between 0.92 and 0.94, Precision and F1-score in the range of 0.92 to 0.93, Recall ranging from 0.91 to 0.93), and lowest mean-squared error (MSE) value of 0.06–08 respectively). These findings supplement RF's ability to capture complex, nonlinear behaviour in topohydrological data items (which is a feature of its ensemble structure that trades off between reducing variance and ensuring predictive stability). In contrast, SVM presented moderate variability across folds with Accuracy 0.86–0.88, F1 0.86–0.87 and MSE 0.12–0.14 values suggesting reasonable but not very competitive performance; Particularly LR remained an underperformer, with the Accuracy between 0.82 and 0.84, a Precision and Recall from 0.81 to 0.83, and most significant MSE values (at least 0.16-18), suggesting shortcomings in representing non- linear flood patterns.

Table 4 Summarizes the three models' performance

Fold	Accuracy	Precision	Recall	F1-Score	MSE
		SVM			
1	0.86	0.87	0.86	0.87	0.14
2	0.88	0.87	0.87	0.86	0.12
3	0.86	0.88	0.87	0.86	0.13
4	0.87	0.86	0.85	0.87	0.14
5	0.88	0897	0.87	0.87	0.12
		LR			
1	0.83	0.83	0.81	0.82	0.18
2	0.83	0.83	0.83	0.83	0.16
3	0.84	0.84	0.82	0.83	0.17
4	0.82	0.82	0.81	0.81	0.18
5	0.84	0.83	0.83	0.83	0.16

			RF		
1	0.92	0.92	0.91	0.92	0.08
2	0.93	0.93	0.92	0.93	0.06
3	0.94	0.93	0.93	0.93	0.07
4	0.92	0.92	0.91	0.92	0.08
5	0.94	0.94	0.92	0.93	0.07

From Table 4, we suggest three possible explanations for the RF model's superior performance: first, the ensemble learning structure of the RF model allows the merging of all outputs from different uncorrelated decision trees to model non-linear feature interactions and diverse decision boundaries. The ensemble method decreases the variance among sites and enables overcoming the overfitting challenge that contributes to the generalization ability of the model from the heterogeneous spatial distribution and hydrological datasets. The SVM performed equally well in terms of predicting potential distribution, especially for the modelling of the non-linear decision surfaces with kernel-based transformations. But it was slightly less stable across the validation folds due to the sensitivity to the kernel's choice and tuning the selected hyperparameter. Finally, the LR model, which was interpretable and computationally efficient, enabled linear assumptions, due to which it could not capture complex interrelations among topo-hydrological predictors, such as slope, flow accumulation, and drainage distance.

In practice, these results emphasize the equivalent competitiveness of RF with items tailored for multi-class flood classification, spatial heterogeneity, and non-linearity underlying the data structure. However, the diverse performance levels of SVM and LR described herein remind us that these methods remain vital for operational decisions, given that computational efficiency, deployment time, and ease of interpretation are serious concerns. In relatively quickly accessible endeavours such as regional flood assessment, the polymorphic SVM and LR methods offer practical alternatives for the trade-off between complexity and accuracy. In contrast, LR is applied in low-data areas where there is a pressing need for more communicative forms to operate. These results highlight, thus, the classical dilemma that involves a theoretical exercise – the trade-off between model accuracy, interpretability, and computational burden precisely captured in each flood management application, where they do impose a set of calls to be made precisely in the function of the gained set driving goals pursued and data acquisition, implementation constraints.

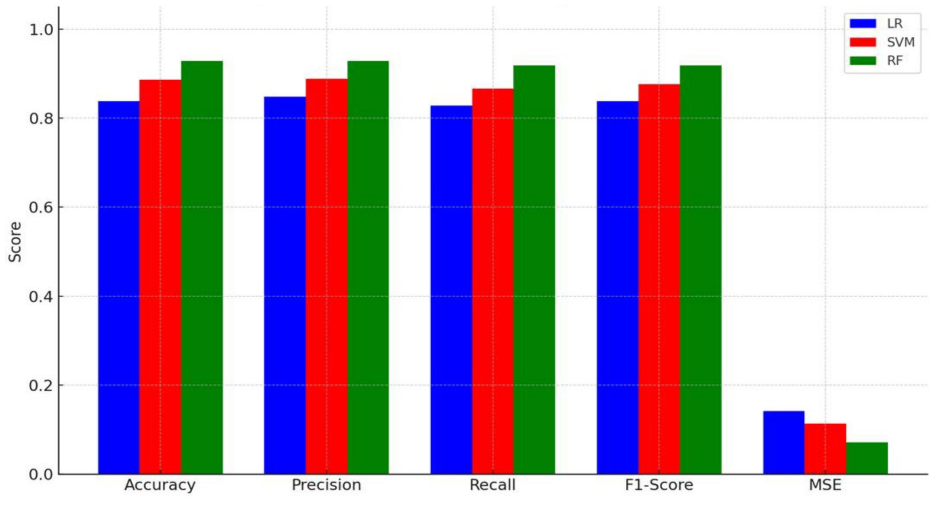


Fig. 4 *Average performance metrics of LR, SVM, and RF models across 5-folds*

Figure 4: Present Performance comparison between our three classifiers, LR, RF, and SVM. Performance is measured on four core metrics: Accuracy, Precision, Recall, and F1-Score, as well as a fifth metric, MSE. These are averaged over the five folds of cross-validation. The graphics show that in each situation, RF obtained the best result among all the measures, which were Accuracy between 0.92 and 0.94, Precision and F1-Score ranging from 0.92 to 0.93, Recall lying between 0.91 and 0.93, with MSE values wavering between (006-008). More constant, though lower, performance was also observed for both LR and SVM across folds. LR and SVM did not differ significantly in their performances through 0.82–84, MSE ~0.16–18; SVM: 86 –88, MSE ~0.12–14. These findings

further validate the outstanding predictive power of RF and the generally competitive, but relatively inferior, reliability of SVM and LR in general flooding prediction across more data partitions.

The RF method outperforms the other two methods across all terms and folds, indicating better predictive performance and stability for stratified samples. SVM closely follows, especially in Folds 2 and 3 (where its Accuracy and F1-score peak), indicating it can handle non-linear decision boundaries very well in flood data. Although LR has a lower ranking, it exhibits relatively good stability: the Precision, Recall, and F1-Score values change only slightly across folds, and the model generalizes despite its linear nature. When combined, the above results demonstrate RF's excellent potential to tackle high-stakes classification problems such as flood prediction and show that SVM and LR are relatively useful for interpretable, efficient modelling, as shown in Table 5 below.

Table 5 *Present 5-Fold CV Summary (Mean ± Std*)

Model	Accuracy	Precision	Recall	F1	MSE
SVM	0.870 ± 0.008	0.870 ± 0.007	0.860 ± 0.008	0.860 ± 0.007	0.130 ± 0.010
LR	0.830 ± 0.010	0.830 ± 0.009	0.820 ± 0.010	0.820 ± 0.009	0.170 ± 0.012
RF	0.930 ± 0.010	0.930 ± 0.009	0.920 ± 0.010	0.930 ± 0.009	0.070 ± 0.008

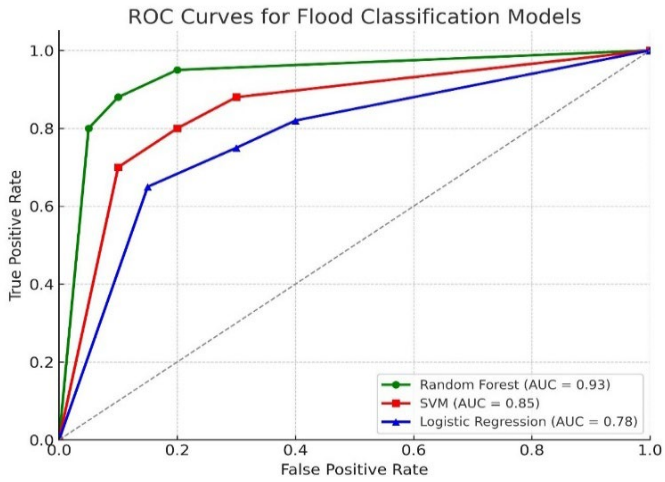
To evaluate the comparative results more completely, Table 6 summarizes the performance, efficiency, and statistical significance for the three classifiers. The RF model had the best predictive performance (AUC = 0.93 ± 0.01), indicating that RF substantially outperformed other models in accounting for non-linear relationships among topohydrological predictors. The SVM and LR yielded lower but still acceptable accuracy values (0.87 ± 0.01 and 0.83 ± 0.01, respectively), with AUCs of the ROC curves equal to 0.85 and 0.78, respectively. Significant performance gains of RF when compared to both SVM and LR were also confirmed by paired t-tests ($p < 0.01$). RF had the highest accuracy at the cost of substantial computational resources (95.2 s in training time, 15.4 MB), while LR was the most lightweight and fastest (28.5 s, 4.3 MB). Generally, the findings show that, despite good generalization performance across all three algorithms, RF provided the best trade-off between accuracy and reliability for multi-class flood classification on the Ibadan data set.

Table 6 *Computational efficiency and significance test summary of the models*

Model	Mean Accuracy (±SD)	Mean AUC	Training Time /s	Model Size/mb	p-value /RF
RF	0.93 ± 0.01	0.93	95.2	15.4	—
SVM	0.87 ± 0.01	0.85	42.1	7.6	0.008
LR	0.83 ± 0.01	0.78	28.5	4.3	0.006

4.3 ROC-AUC

Present ROC curves; correct: ROC space representations of the classification ability for the three tested models, RF-SVM-LR, in multi-class flood prediction classification. Radio frequency RF proves the best in discriminative ability by an AUC of 0.93. The ROC curve is relatively close to the ideal top-left boundary, indicating that RF has a high actual favourable rate at most thresholds, with little loss of actual favourable rate. This result indicates the RF's capacity to capture complex nonlinear relations in the topohydrological dataset and, hence, is least affected by overfitting, even on new generalization data.



Fi5. 4 *ROC performance metrics of LR, SVM, and RF models across 5-folds*

SVM achieves an AUC of 0.85, which provides competitive but substantially lower predictive power. The curve remains satisfactory in terms of sensitivity in the moderate threshold region. Because STEEPNESS is much lower than RF's, it can also model nonlinear decision boundaries. However, this option is limited by the choice of kernels and by parameter sensitivity. Meanwhile, the LR model achieves an AUC of 0.78 at the lowest boundary. More than that, the curve is farthest from the optimal boundary. This result indicates difficulties of the linear responses to a highly complex flood-generating process, particularly when they involve strongly interacting spatial and hydrological variables. Altogether, the ROC analyses above confirm that RF outperforms, whereas SVM and LR still have a practical value in flooding classification due to their interpretability.

4.4 Box Plot Comparison

Box plots illustrating the variation of the average result of Accuracy, Precision, Recall, and F1-Score across five CV folds for all classifiers in Figure 5. It is evident that LR shows moderate variation, primarily in accuracy and recall, and is the least stable for the multi-class flood detection problem. RF, at the same time, has the best resilience; the values in the box plot are close to each other on the diagram; thus, RF is the most stable model and generalises well for the different severity levels for flooding detection. SVM shows lower variability and fewer outliers, especially in precision, and is thus prone to various fold splits. Overall, RF remains the most stable model for detecting flooding, both in terms of performance and variability.

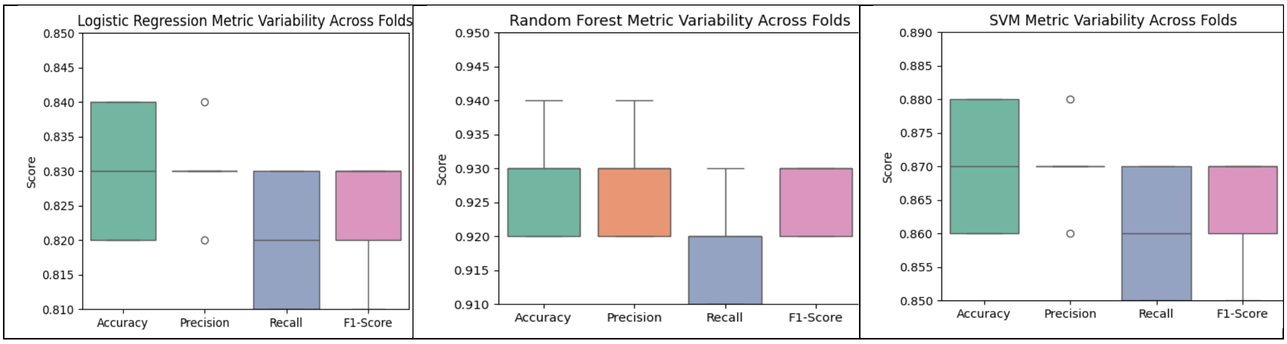


Fig. 5 *Presents box plots illustrating the variability of performance metrics*

4.5 Class-Level Performance Analysis

While the 20% holdout set of the other model's class-wise performance can only tell us not much about the flood class category, the precision, recall, and F1-score of 0.98 were also obtained by this RF model of each flood class category, which were No Flood, Low, Moderate, High, and Very High. This emphasizes the stability and robustness of the RF model in detecting different levels of flood class. The model of LR also works well in general. However, this model also performs slightly worse on the Low class and achieves an F1 score of 0.96, misclassifying 244 examples as moderate. Nevertheless, the Low class is indistinguishable from the model, which also yielded an F1 score of 0.94 and 641 examples misclassified as moderate. The distinction between neighbouring classes is very hard, such as Low and Moderate.

From a disaster management and risk communication perspective, these differences have far-reaching consequences. Misremembering a Low Flood message or a No Flood message is an over-prediction that leads to unnecessary alert or resource deployment. Conversely, a high false alarm rate would erode social and individual confidence in the ability to predict the next earthquake and deleteriously diminish preparedness. Thus, ensuring the best performance in all these types is critical not only for statistical accuracy but more importantly to ensure accurate, timely, and efficient flood risk management. In this respect, RF's full rates confirm its status as the best-performing model for operational deployment.

4.6 Comparative Analysis with Prior Studies

This comparison indicates the evolution of flood modeling methods from susceptibility assessment for binary classification to the multi-class classification now applied. In the Random Forest solutions applied in China and Iran, the accuracy was high: 91.5% and 94.3%, respectively, but Random Forest and Alternating Decision Trees were not applied together to the dynamic susceptibility mapping. The regional contribution in Bavaria was the application of CatBoost due to its scalability. But its accuracy was only fair: AUC was 0.819. In Malaysia, Geographically Weighted Regression improved prediction and accounted for spatial non-stationarity, while in Iran, Sentinel 1 SAR was included in the solution through a joint application of Bagging–Cubic–KNN, which decreased variance and addressed overfitting. All these provide evidence for the valuable application of ensemble and spatially adaptive models. Still, almost all the solutions were for binary classification, and none adequately described flood severity as a continuous phenomenon.

The study in Table 7 above presents an innovation: a novel five-class classification model for urban flood prediction in Ibadan, Nigeria. By combining LR, SVM, and RF and conducting the test through stratified cross-validation and hold-out prediction, the study achieved high stability and interpretability. RF achieves the highest overall performance across all metrics: accuracy = 92%, precision = 93%, recall = 92%, F1 = 92%, and MSE = 0.07, much better than SVM and LR. Moreover, the pipeline introduced the permutation-based interpretability to know the importance of ranking between topographical, hydrological, and meteorological aspects. The above-mentioned dual focus on multi-class detection and multi-class and multi-instance explainability improves the binary focus, adding significant technical rigour and fulfilling the professional operational requirements for urban flood risk management in data-poor settings.

Table 7 *Summarize a benchmarking study*

Ref.	Method/Model	Description
[10]	ML	For forecasting of floods in the Quannan region of China. Multivariate classification of five classes. RF model achieves the highest accuracy of 91.5%.
[11]	ML	For flash flood susceptibility mapping in Iran. Multivariate classification of four classes. The ADT model achieves the highest accuracy of 94.3%.
[12]	ML	Regional-scale pluvial and flash flood susceptibility mapping in Bavaria, Germany. CatBoost achieved the highest accuracy (AUC = 0.819). Demonstrated scalability of ML to heterogeneous, large areas
[13]	ML	Spatial Flood-Prone Area Prediction in the Kemaman Basin, Malaysia. GWR improved spatial prediction (Success = 91.24%, Prediction = 75.15%) compared to global LR and FR.
[14]	GIS/ML	Sentinel-1 SAR remote sensing–based flood detection in Haraz Watershed, Iran. Bagging–Cubic–KNN ensemble model reduced overfitting and achieved the highest accuracy (AUC = 0.660).
Multi-class	ML	For multi-class urban flood prediction in Ibadan, Nigeria. Multivariate classification of five classes. RF achieved the highest performance (Acc = 92%, Precision = 93%, Recall = 92%, F1 = 92%, MSE = 0.07), outperforming SVM (Acc = 87%) and LR (Acc = 83%). Demonstrates a robust, reproducible pipeline with interpretability.

These findings have several practical implications for flood risk management, urban planning, and disaster risk reduction, which are highly critical in regions such as Ibadan, where rapid urbanization is ongoing. The interpretable machine learning methodology, enabled by LR, SVM, and RF, provides assertive multi-class flood category classification with environmental-spatial information and can be utilized for an early warning system to trigger specific response-based mechanisms by location. By considering topographical, hydrological, and

J. of Soft Computing and Data Mining Vol. 6 No. 3 (2025) p. 243-258

meteorological characteristics, the context-dependent validation of predictions is enhanced, and models can be developed in a user-friendly interface for non-technical end-users. However, this study has limitations. Although the models performed well in cross-validation and holdout, the possibility of dynamic events that change over time may limit generalization, since they were trained on static data. Although permutation importance provides interpretability, these models lack their own temporal features, and explaining a sequence model like an LSTM would be an interesting direction. Moreover, the use of remote sensing and DEM-derived features also considers data availability and the corpus, which can never be the same across diverse conditions.

4.7 Critical Discussion of Limitations and Biases

Although the developed framework was applied using the Ibadan metropolis, it is relatively easily transferable to other urban areas with a similar climatic and hydrologic set-up. The transferability of the framework is facilitated by physically interpretable, globally consistent predictors, such as slope, topographic wetness index, flow accumulation, drainage distance, and rainfall, which govern flood dynamics in most tropical to subtropical settings. Notwithstanding, verification beyond Ibadan was not undertaken in this research due to data cost limitations. We extend this framework to other cities in Nigeria and across SSA to assess its trans-regional transportability and calibration needs, and to develop a generic method for translatable urban flood classification spanning diverse geographic settings.

The proposed models have some limitations despite their good performance in multi-classification. For instance, despite stratification and spatial clustering, there may still be minor remaining issues due to spatial autocorrelation. Secondly, this was a multi-year frozen accumulation with no temporal modelling for rainfall or land-use change. Additionally, DEMs and satellite-derived rainfall inputs may contain measurement errors or differences in spatial resolution, thereby introducing additional uncertainty into the predictions. Although the framework works great locally within Ibadan, it may also require site-specific calibration for use in other climate or geomorphology regions. These limitations necessarily limit the extent to which it is possible to rely on these reported accuracies. They also suggest that future work should involve temporal data and use more advanced modeling techniques to enhance the robustness and transferability of predictive models.

5. Conclusion and Future Work

The study applies Explainable ML to multi-class flood classification for the first time, as basically spatially aggregated data were collected in Ibadan city. The approach used LR, SVM, and RF in Scikit-learn as a pipeline, with robust preprocessing, stratified 5-fold CV, and hold-out evaluation. RF is the best among the three classifiers, achieving 92% accuracy across all folds and the holdout set. However, SVM and LR are excellent predictors, with few incorrectly classified cases due to disagreement between Moderate and Low/High causes. In Summary, these results, developed and compared, are highly relevant to the applicability and generality of the proposed model; that is, it worked well theoretically for spatially aggregated complex flood problems and practically. Additionally, I emphasized all models' transparency by overlaying the permutation-based feature importance analyses for each model, particularly highlighting the critical predictors in drainage, slope, and flow accumulation for their hydrological angles from urban bases. The study goes beyond traditional flood approaches, as the binary approach is achieved only by incorporating environmental, topographic, and spatial aspects into a multi-class, explainable ML classification model. Therefore, this is expected to be helpful to decision-makers in early warning, Infrastructure, and urban climate adaptation. Essentially, in terms of flood predictions, we can recommend additional petitions for the future. You should integrate LSTM-NN into ensemble teaching methods, as described above, and the ensemble would incorporate several varieties of specialization, classification, and generalization for flood characteristics. Additionally, a flood model based on an LSTM network could provide temporal predictions that preserve the continuous series of meteorological and hydrological information in the relevant field. Finally, for emergency use, XAI would significantly boost the availability and efficiency of interpretable methods.

Acknowledgement

The authors would like to acknowledge that this research was supported by the Malaysian Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2022/ICT02/UTHM/02/2). The authors also thank Alnoor University for supporting this project through the Research Supporting Fund ANUI/2025/ENG45.

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper.

Author Contribution

The authors confirm their contributions to the paper as follows: Study conception and design were led by Jabir Abubakar Salisu and Mohammed Hasan Aldulaimi, who formulated the framework and objectives of the proposed model. Jabir Abubakar Salisu prepared the draft manuscript, and Dahiru Adamu Aliyu conducted the data collection. To ensure the accuracy, consistency, and Interpretation of the findings, a critical review and intellectual input were received from Hairulnizam Mahdin and *Heru Nurwarsito. Mohammed Basman Ghanim and Hairulnizam bin Mahdin provided final approval of the manuscript.*

References

[1] Rizzo, A., Cappellano, F., Pierantoni, I., & Sargolini, M. (2022). Do natural disasters accelerate sustainability transitions? Insights from the Central Italy earthquake. European Planning Studies, 30(11), 2224–2244. https://doi.org/10.1080/09654313.2021.2022104

[2] Khan, M. T. I., Anwar, S., Sarkodie, S. A., Yaseen, M. R., Nadeem, A. M., & Ali, Q. (2022). Comprehensive disaster resilience index: Pathway towards risk-informed sustainable development. Journal of Cleaner Production, 366, 132937. https://doi.org/10.1016/j.jclepro.2022.132937

[3] Ucar, I., Şen, Z., Ceylan, Z., & Yalçın, M. (2025). From hazard maps to action plans: Comprehensive flood risk mitigation in the Susurluk Basin. Water, 17(6), 860. https://doi.org/10.3390/w17060860

[4] Kiaghadi, A., Govindarajan, A., Sobel, R. S., & Rifai, H. S. (2020). Environmental damage associated with severe hydrologic events: A LiDAR-based geospatial modeling approach. Natural Hazards, 103(3), 2711–2729. https://doi.org/10.1007/s11069-020-04099-1

[5] De Silva, M. M. G. T., & Kawasaki, A. (2022). Modeling the association between socioeconomic features and risk of flood damage: A local-scale case study in Sri Lanka. Risk Analysis, 42(12), 2735–2747. https://doi.org/10.1111/risa.13894

[6] Kadaverugu, A., Rao, C. N., & Viswanadh, G. K. (2021). Quantification of flood mitigation services by urban green spaces using the InVEST model: A case study of Hyderabad city, India. Modeling Earth Systems and Environment, 7(1), 589–602. https://doi.org/10.1007/s40808-020-00937-0

[7] Kiaghadi, A., Govindarajan, A., Sobel, R. S., & Rifai, H. S. (2020). Environmental damage associated with severe hydrologic events: A LiDAR-based geospatial modeling approach. Natural Hazards, 103(3), 2711–2729. https://doi.org/10.1007/s11069-020-04099-1

[8] Patwary, M. M., Rahman, M. M., Alam, M. Z., & Chowdhury, M. R. (2024). Impact of extreme weather events on mental health in South and Southeast Asia: A two-decade systematic review of observational studies. Environmental Research, 250, 118436. https://doi.org/10.1016/j.envres.2024.118436

[9] Kar, B., Kumar, D., Smith, A., & Rao, N. (2024). Assessing a model-of-models approach for global flood forecasting and alerting. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 9641–9650. https://doi.org/10.1109/JSTARS.2024.3390579

[10] Guo, X., Cheng, J., Yin, C., Li, Q., Chen, R., & Fang, J. (2023). The extraordinary Zhengzhou flood of July 20, 2021: How extreme weather and human response compounded the disaster. Cities, 134, 104168. https://doi.org/10.1016/j.cities.2022.104168

[11] Mabrouk, M., Elsheikh, R. F. A., Abdelhady, S., & Elshaer, A. (2023). Assessing the effectiveness of nature-based solutions-strengthened urban planning mechanisms in forming flood-resilient cities. Journal of Environmental Management, 344, 118260. https://doi.org/10.1016/j.jenvman.2023.118260

[12] Pillar, V. D., & Overbeck, G. E. (2024). Learning from a climate disaster: The catastrophic floods in southern Brazil. Science, 385(6713). https://doi.org/10.1126/science.adr8356

[13] Patwary, M. M., Rahman, M. M., Alam, M. Z., & Chowdhury, M. R. (2024). Impact of extreme weather events on mental health in South and Southeast Asia: A two-decade systematic review of observational studies. Environmental Research, 250, 118436. https://doi.org/10.1016/j.envres.2024.118436

[14] Franchi, F., Nardone, C., D'Amato, D., & Di Mauro, G. (2024). Prolonged drought periods over the last four decades increase flood intensity in southern Africa. Science of The Total Environment, 924, 171489. https://doi.org/10.1016/j.scitotenv.2024.171489

[15] Hohl, B. C., Smith, L. A., Walker, J. K., & Green, M. R. (2023). Community-identified characteristics related to illegal dumping: A mixed methods study to inform prevention. Journal of Environmental Management, 346, 118930. https://doi.org/10.1016/j.jenvman.2023.118930

[16] Barakat, B., & Srour, I. (2024). Consideration of hotspots in the selection of supervision schemes to reduce illegal dumping of construction and demolition waste. Waste Management & Research: The Journal for a Sustainable Circular Economy, 42(2), 146–157. https://doi.org/10.1177/0734242X231178216

[17] Chappell, N. A., & Beven, K. J. (2024). Nature-based solutions for effective flood mitigation: Potential design criteria. Environmental Research Letters, 19(7), 074006. https://doi.org/10.1088/1748-9326/ad4fa2

[18] Piadeh, F., Behzadian, K., Chen, A. S., Kapelan, Z., Rizzuto, J. P., & Campos, L. C. (2023). Enhancing urban flood forecasting in drainage systems using dynamic ensemble-based data mining. Water Research, 247, 120791. https://doi.org/10.1016/j.watres.2023.120791

[19] Moradian, S., AghaKouchak, A., Gharbia, S., Broderick, C., & Olbert, A. I. (2024). Forecasting of compound ocean-fluvial floods using machine learning. Journal of Environmental Management, 364, 121295. https://doi.org/10.1016/j.jenvman.2024.121295

[20] Rashidi Shikhteymour, S., Borji, M., Bagheri-Gavkosh, M., Azimi, E., & Collins, T. W. (2023). A novel approach for assessing flood risk with machine learning and multi-criteria decision-making methods. Applied Geography, 158, 103035. https://doi.org/10.1016/j.apgeog.2023.103035

[21] Liu, Z., Felton, T., & Mostafavi, A. (2024). Interpretable machine learning for predicting urban flash flood hotspots using intertwined land and built-environment features. Computers, Environment and Urban Systems, 110, 102096. https://doi.org/10.1016/j.compenvurbsys.2024.102096

[22] Costache, R., Hong, H., Pham, Q. B., & Diaconu, D. C. (2020). Spatial prediction of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. Journal of Hydrology, 585, 124808. https://doi.org/10.1016/j.jhydrol.2020.124808

[23] Akinsoji, A. H., Adelodun, B., Adeyi, Q., Salau, R. A., Odey, G., & Choi, K. S. (2025). Prediction of spatial-temporal flood water level in agricultural fields using advanced machine learning and deep learning approaches. Natural Hazards, 121(7), 7915–7940. https://doi.org/10.1007/s11069-025-07118-1

[24] Schmidt, L., Heße, F., Attinger, S., & Kumar, R. (2020). Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. Water Resources Research, 56(5), e2019WR025924. https://doi.org/10.1029/2019WR025924

[25] Lee, H., & Li, W. (2024). Improving interpretability of deep active learning for flood inundation mapping through class ambiguity indices using multi-spectral satellite imagery. Remote Sensing of Environment, 309, 114213. https://doi.org/10.1016/j.rse.2024.114213

[26] Hou, J., Zhou, N., Chen, G., Huang, M., & Bai, G. (2021). Rapid forecasting of urban flood inundation using multiple machine learning models. Natural Hazards, 108(2), 2335–2356. https://doi.org/10.1007/s11069-021-04782-x

[27] Sarwar, J., Khan, S. A., Azmat, M., & Khan, F. (2024). A comparative analysis of feature selection models for spatial analysis of floods using hybrid metaheuristic and machine learning models. Environmental Science and Pollution Research, 31(23), 33495–33514. https://doi.org/10.1007/s11356-024-33389-5

[28] Olanrewaju, R. M., Tilakasiri, S. L., & Bello, F. B. (2018). Community perception of deforestation and climate change in Ibadan, Nigeria. Journal of the University of Ruhuna, 6(1), 26–36. https://doi.org/10.4038/jur.v6i1.7866

[29] Chen, W., Pourghasemi, H. R., & Panahi, M. (2020). Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. Science of The Total Environment, 701, 134979. https://doi.org/10.1016/j.scitotenv.2019.134979

[30] Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Singh, V. P., & Pradhan, B. (2018). A comparative assessment of decision tree algorithms for flash flood susceptibility modelling at Haraz watershed, northern Iran. Science of The Total Environment, 627, 744–755. https://doi.org/10.1016/j.scitotenv.2018.01.266

[31] Kaiser, M., Günnemann, S., & Disse, M. (2022). Regional-scale prediction of pluvial and flash flood susceptible areas using tree-based classifiers. Journal of Hydrology, 612, 128088. https://doi.org/10.1016/j.jhydrol.2022.128088

[32] Lin, J. M., & Billa, L. (2021). Spatial prediction of flood-prone areas using geographically weighted regression. Environmental Advances, 6, 100118. https://doi.org/10.1016/j.envadv.2021.100118

[33] Aswad, F. M., Kareem, A. N., Khudhur, A. M., Khalaf, B. A., & Mostafa, S. A. (2021). Tree-based machine learning algorithms in the Internet of Things environment for multivariate flood status prediction. *Journal of Intelligent Systems*, *31*(1), 1-14. https://doi.org/10.1515/jisys-2021-0179

[34] Mahdizadeh Gharakhanlou, N., & Perez, L. (2023). Flood susceptible prediction through the use of geospatial variables and machine learning methods. Journal of Hydrology, 617, 129121. https://doi.org/10.1016/j.jhydrol.2023.129121

[35] Albertini, C., Gioia, A., Iacobellis, V., Petropoulos, G. P., & Manfreda, S. (2024). Assessing multi-source random forest classification and robustness of predictor variables in flooded areas mapping. Remote Sensing Applications: Society and Environment, 35, 101239. https://doi.org/10.1016/j.rsase.2024.101239

[36] Wang, Y., Zhang, P., Xie, Y., Chen, L., & Li, Y. (2025). Toward explainable flood risk prediction: Integrating a novel hybrid machine learning model. Sustainable Cities and Society, 120, 106140. https://doi.org/10.1016/j.scs.2025.106140