Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (38-51)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.49010

JURNAL TEKNIK INFORMATIKA
Homepage : http://journal.uinjkt.ac.id/index.php/ti

Explainable Ensemble Learning for Urban Flood Risk Mapping in Jakarta Using Multi-Source Geospatial and Hydrometeorological Data

Arief Wibowo[1*], Abdul Haris Achadi[2]

[1]Department of Computer Science, Faculty of Information Technology, Universitas Budi Luhur
[2]Department of Disaster Management, Faculty of Economics and Bussiness, Universitas Budi Luhur
[1,2]Jl. Cileduk Raya, Petukangan Utara, Jakarta Selatan, Indonesia 12260

Article:
Accepted: February 17, 2026
Revised: December 24, 2025
Issued: April 30, 2026

***Correspondence Address:**
arief.wibowo@budiluhur.ac.id

ABSTRACT

Urban flooding is a frequent hydrometeorological hazard in Indonesia, particularly in Jakarta, driven by rapid urbanization, limited drainage capacity, land cover change, and extreme rainfall. This study develops an explainable ensemble learning framework for urban flood risk mapping in Jakarta using multi-source geospatial and hydrometeorological data, including satellite-based rainfall, topography, land use/land cover, NDVI, and IoT-based river water level observations from 2023–2025. Flood occurrence labels were constructed by integrating municipal flood records with satellite-based inundation data. The framework integrates Random Forest, Gradient Boosting, and XGBoost models, with SHAP applied for interpretability and identification of dominant flood drivers. Model evaluation using ROC-AUC and RMSE indicates that XGBoost achieved the highest performance (AUC = 0.91, RMSE = 0.184), outperforming Random Forest (AUC = 0.87, RMSE = 0.221) and Gradient Boosting (AUC = 0.89, RMSE = 0.203). SHAP analysis identifies rainfall intensity, elevation, proximity to river channels, and built-up area percentage as the most influential factors. Despite uncertainties in flood labeling and the lack of high-resolution drainage data, the results demonstrate the potential of explainable ensemble learning for urban flood risk assessment and resilience planning.

Keywords : *Ensemble Learning; Explainable Artificial Intelligence; Flood Risk; Multi-Source Data; Urban Flooding.*

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (38-51)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.49010

1. INTRODUCTION

Urban flooding has become one of the most critical environmental challenges in rapidly developing cities across Southeast Asia due to the accelerating impacts of climate change and urban expansion [1]. In Jakarta, experience recurring flood events almost every year, disrupting economic activities and threatening public safety [2]. The increasing intensity of extreme rainfall aggravates flood hazards in densely populated urban catchments with limited drainage capacity [3]. In addition, land use changes driven by uncontrolled urban development continue to reduce natural infiltration zones, further amplifying surface runoff accumulation [4]. The combination of these factors creates highly complex and dynamic flood risk patterns that require advanced modelling approaches to understand [5].

Traditional hydrological models often struggle to represent the nonlinear interactions among multiple environmental variables in heterogeneous urban landscapes [6]. Machine learning (ML) techniques have emerged as powerful tools for flood prediction by capturing complex patterns from large-scale data sources [7]. Ensemble learning models such as Random Forest, Gradient Boosting, and XGBoost demonstrate strong performance in flood susceptibility and hazard mapping due to their robustness against noise and feature variability [8]. However, most ML-based flood models operate as *black-box* systems that do not provide transparent explanations of how predictions are generated, making them difficult to adopt in policy-driven disaster management settings [9]. Local governments and emergency planners require interpretable outputs to support evidence-based mitigation strategies [10].

Explainable Artificial Intelligence (XAI) has emerged as a promising solution for addressing the interpretability gap in flood modelling systems [11]. Techniques such as SHapley Additive exPlanations (*SHAP*) allow researchers to quantify the contribution of each input variable to the model's output, enabling transparent reasoning behind hazard classifications [12]. XAI approaches have been successfully applied to environmental modelling tasks such as landslide susceptibility, drought monitoring, and ecosystem health prediction, demonstrating their value in scientific decision-support contexts [13]. Despite these advancements, XAI-based ensemble learning applications for flood risk assessment in the Indonesian urban context remain limited [14].

Urban flood risk is inherently multi-dimensional and requires the integration of diverse geospatial and hydrometeorological datasets to improve predictive accuracy [15]. Multi-source data such as rainfall intensity, topographic elevation, land use/land cover, *Normalized Difference Vegetation Index* (*NDVI*), and river water level measurements provide complementary perspectives on flood-generating processes [16]. Remote sensing technologies including satellite imagery offer additional insights into surface characteristics and catchment-scale hydrological responses [17]. Combining these multi-source datasets within ML models enhances their ability to detect spatial patterns associated with urban flood risk [18].

The availability of open geospatial data, IoT-based hydrological sensors, and higher-resolution satellite products has created new opportunities to develop data-driven flood risk models tailored to local conditions [19]. Nevertheless, effective integration of heterogeneous datasets requires robust preprocessing techniques to handle spatial resolution differences, missing values, and categorical encoding challenges [20]. Ensemble learning methods provide a strong foundation for multi-source data integration due to their ability to manage diverse feature types and complex feature interactions [21]. Yet, without explainability features, the operational implementation of such models remains limited in practice [22].

In recent years, there has been growing emphasis on developing interpretable flood modelling frameworks to support urban resilience planning [23]. Explainable ensemble learning has the potential to bridge the gap between predictive performance and interpretability by combining high-performing algorithms with transparent variable attribution methods [24]. Understanding dominant factors such as rainfall intensity, elevation gradients, drainage density, and land cover transitions can help local authorities prioritize targeted mitigation interventions [25]. Furthermore, interpretable ML models enable more effective communication between technical stakeholders

and policymakers in urban disaster management ecosystems [26].

Given these challenges and opportunities, this study aims to develop an explainable ensemble learning framework for urban flood risk analysis using multi-source data in Indonesia [27]. By combining ensemble modelling, geospatial data integration, and XAI-based interpretation, this research seeks to enhance the accuracy and transparency of flood risk prediction in complex urban environments [28]. The outcomes of this work are expected to support municipal disaster agencies in designing more data-driven and explainable mitigation strategies for urban flood resilience.

This study makes several key scientific contributions to urban flood risk modelling in the Indonesian context. First, it constructs a high-resolution (30 m) localized urban flood dataset for Jakarta by integrating multi-source geospatial and hydrometeorological data, including satellite-based rainfall estimates, land use/land cover information, and IoT-based river water level observations. Second, it systematically compares multiple ensemble learning algorithms—Random Forest, Gradient Boosting, and XGBoost—under a consistent modelling and evaluation framework to assess their predictive performance for urban flood risk mapping. Third, it incorporates explainable artificial intelligence through SHapley Additive exPlanations (SHAP), including both global feature importance and interaction effects, to provide transparent insights into dominant flood-driving factors in complex urban environments. Fourth, it conducts spatially explicit error and false-negative analysis to highlight areas of increased prediction uncertainty that are particularly relevant for disaster risk management. Finally, this study emphasizes a cautious and interpretable data-driven approach, positioning explainable ensemble learning as a decision-support tool rather than a fully operational flood forecasting system.

Despite the growing body of research on machine learning–based flood susceptibility and risk modelling, several critical gaps remain in the context of urban flooding in Indonesia, particularly in highly complex metropolitan environments such as Jakarta. Existing studies often focus on either regional-scale analyses or simplified urban settings, with limited attention to high-resolution spatial modelling that captures intra-urban variability at the neighborhood level. Moreover, many urban flood studies rely predominantly on static geospatial variables or satellite-derived rainfall data, while the integration of near-real-time hydrological observations, such as IoT-based river water level measurements, remains relatively underexplored in localized urban contexts.

From a methodological perspective, prior studies frequently emphasize predictive accuracy without adequately addressing interpretability, spatial bias, or error behavior in densely built environments. While ensemble learning models and explainable artificial intelligence techniques such as SHAP have been applied in flood-related studies, their combined use is often presented in a generic manner, with limited analysis of feature interactions, false-negative behavior, and spatial error concentration that are critical for disaster risk management applications. In addition, many studies adopt random data splitting strategies that may lead to spatial leakage, thereby overestimating model performance in spatially correlated flood datasets.

Consequently, there remains a need for a localized, high-resolution urban flood risk modelling framework that not only integrates heterogeneous geospatial, hydrometeorological, and sensor-based data sources, but also explicitly incorporates explainability, spatially aware evaluation, and uncertainty-aware interpretation tailored to the Jakarta metropolitan context.

2. METHODS

The research method consists of four major stages: problem formulation, multi-source data integration, ensemble model development, and explainability analysis. Each stage is described in detail in the following subsections as shown in Figure 1.

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (38-51)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.49010

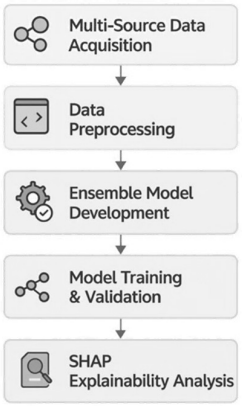

Figure 1. *Research Methodology Workflow*

2.1. Problem Formulation

Urban flooding in Jakarta is driven by complex and nonlinear interactions between hydrometeorological, topographic, and urban structural factors that vary significantly across space. To capture this complexity, the study area is discretized into a set of regular spatial grid cells with a resolution of 30 m, where each grid cell represents a localized urban unit subject to potential flood occurrence.

$Let\ X_i = X_{i1}, X_{i2}, \ldots\ldots X_{ip}$ denote the feature vector associated with grid cell i, comprising multi-source predictors relevant to urban flooding in Jakarta, including rainfall intensity, elevation, slope, land use/land cover type, Normalized Difference Vegetation Index (NDVI), distance to river channels, drainage density, impervious surface percentage, and river water level. The flood occurrence label for grid cell i is defined as a binary variable $Y_i \in \{0, 1\}$, where $Y_i = 1$ indicates flooded conditions and $Y_i = 0$ denotes non-flooded conditions, as derived from integrated municipal flood records and satellite-based inundation detection.

The objective of this study is to learn a mapping function f that estimates the probability of flood occurrence for each spatial grid cell based on its associated feature vector, expressed as shown in Equation 1.

$$y_i = f(x_i),\ y_i \in [0,1] \qquad (1)$$

where y_i represents the predicted flood probability for grid cell i. The function f is approximated using ensemble learning models, including Random Forest, Gradient Boosting, and XGBoost, selected for their ability to capture nonlinear relationships and interactions among heterogeneous urban flood drivers. Beyond predictive accuracy, the modelling framework is designed to support interpretability by enabling the attribution of model outputs to individual predictors. This is achieved through SHapley Additive exPlanations (SHAP), which quantify the contribution of each input variable to the predicted flood probability, thereby facilitating transparent analysis of dominant flood-driving factors across the Jakarta metropolitan area.

2.2. Multi-Source Data Acquisition and Preprocessing

Five categories of data were integrated to construct a comprehensive feature set for urban flood risk modelling. Rainfall intensity data were collected from meteorological stations and satellite-based precipitation products. Elevation and slope information were derived from the Shuttle Radar Topography Mission (SRTM), while land use/land cover classes were obtained from national geospatial datasets. Vegetation characteristics were extracted using NDVI from multispectral satellite imagery. Hydrological sensor readings containing river water level data were gathered from IoT-based monitoring systems. Data preprocessing included noise filtering, spatial interpolation, normalization, and encoding of categorical variables. Missing values were handled using K-nearest neighbor imputation, while spatial datasets were resampled into a uniform grid resolution of 30 meters. All datasets were aligned using a common geographic coordinate system to ensure spatial consistency between feature layers.

Rainfall data were obtained from satellite-based precipitation products with a spatial resolution coarser than the modelling grid. To ensure compatibility with the 30 m spatial grid used in this study, rainfall values were spatially interpolated and resampled to match the resolution of other predictor variables. An inverse distance weighting (IDW) interpolation approach was applied to distribute rainfall values across the study area, under the assumption that rainfall intensity exhibits spatial continuity at the urban scale. The interpolated rainfall surface was then resampled

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (38-51)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.49010

to a 30 m grid using bilinear interpolation to ensure spatial alignment with topographic and land cover variables. This interpolation process was applied consistently across all rainfall observations during the study period. While this approach enables the integration of rainfall information at fine spatial resolution, it may introduce uncertainty due to the smoothing of localized rainfall extremes. Such uncertainty is considered when interpreting model performance and explainability results.

2.3. Ground Truth Flood Label Construction

Ground truth flood labels were generated by integrating municipal flood records and satellite-based inundation detection to represent observed flood occurrences at the grid level in Jakarta. This dual-source approach was adopted to improve label reliability and reduce bias associated with single-source observations. Municipal flood records obtained from local government agencies provide reported flood occurrence dates and affected locations. To complement these records with spatially explicit flood extents, satellite-based inundation detection was used to identify surface water presence during flood events. Both data sources were spatially aligned to a regular grid with a 30 m resolution, consistent with the predictor variables. A grid cell was labeled as flooded if it intersected reported flood-affected areas or exhibited satellite-detected inundation signals.

Temporal alignment was performed using an event-based matching strategy, where satellite observations acquired within a defined temporal window around reported flood dates were associated with municipal records. This approach accounts for potential timing differences between flood peak occurrence and satellite overpass. When discrepancies between data sources occurred, a conservative integration rule was applied, assigning a flooded label if at least one source indicated flood occurrence during the aligned period. Despite this integration, some uncertainty in flood labeling remains due to potential reporting inaccuracies, temporal observation gaps, and the absence of micro-scale drainage information. These limitations are considered when interpreting model performance and explainability results.

2.4. Ensemble Learning Model Development

Three ensemble learning algorithms—Random Forest, Gradient Boosting, and XGBoost—were developed to model urban flood risk. Each algorithm was chosen for its robustness in handling heterogeneous features and capturing nonlinear patterns. Random Forest constructs multiple decision trees using bootstrap sampling, while Gradient Boosting sequentially minimizes residual errors. XGBoost enhances gradient boosting through regularization and efficient tree optimization.

The ensemble model is defined as shown in Equation 2.

$$y = \frac{1}{m} \sum_{i=1}^{m} f_i\ (X) \qquad (2)$$

where f_i represents the individual base learners and m is the number of models combined. Hyperparameter tuning was performed using grid search to optimize tree depth, learning rate, and number of estimators. Model performance was evaluated using ROC-AUC and RMSE metrics.

Table 1. *Hyperparameter tuning configuration and optimal values*

Model	Hyperparameter	Tested Range	Optim. Value
Random Forest	Number of trees (n_estimators)	100 – 500	300
	Maximum depth (max_depth)	5 – 30	20
	Minimum samples per leaf	1 – 10	3
Gradient Boosting	Number of trees (n estimators)	100 – 400	250
	Learning rate	0.01 – 0.2	0.05
	Maximum depth	3 – 10	5
XGBoost	Number of trees (n_estimators)	200 – 600	400
	Learning rate (eta)	0.01 – 0.3	0.1
	Maximum depth	3 – 10	6
	Subsample	0.6 – 1.0	0.8
	Column subsample (colsample_bytree)	0.6 – 1.0	0.8

Hyperparameter tuning was conducted to optimize model performance while avoiding overfitting. A grid search strategy combined with cross-validation was employed to evaluate candidate parameter configurations for each ensemble model. Parameter ranges were selected based on commonly adopted values in prior flood modelling and machine learning studies. Was determined based on the highest

cross-validated AUC score and subsequently used for final model training and evaluation.

2.5. Explainable AI Using SHAP Analysis

Explainability was incorporated using SHapley Additive exPlanations (SHAP), which quantify the contribution of each input variable to the model prediction. For a given sample X the SHAP value $\emptyset_j$ for feature x_j is computed as shown in Equation 3.

$$f(X) = \emptyset_0 + \sum_{j=1}^{n} \emptyset_j \quad (3)$$

where $\emptyset_0$ is the base value and $\emptyset_j$ indicates the influence of feature x_j . SHAP summary plots and dependency plots were used to interpret feature interactions and dominant drivers of flood risk across urban districts.

All experiments were implemented using Python programming language. Random Forest and Gradient Boosting models were developed using the scikit-learn library, while XGBoost was implemented using the XGBoost framework. Model explainability was conducted using the SHAP library. The main software versions used in this study include Python 3.9, scikit-learn 1.2, XGBoost 1.7, NumPy 1.23, Pandas 1.5, and SHAP 0.41.

2.6. Model Validation and Testing

Model validation was conducted using 10-fold cross-validation to assess generalization stability. Data were split into training (80%) and testing (20%) subsets. Evaluation metrics included AUC, RMSE, accuracy, precision, recall, and F1-score, while confusion matrices were analyzed to examine classification behavior. SHAP-based explanations were qualitatively validated against established hydrological knowledge and observed urban flood patterns.

To reduce potential spatial leakage arising from spatial autocorrelation among neighboring grid cells, an additional spatial holdout evaluation was performed. The study area was partitioned into spatial blocks, where selected contiguous zones were withheld from training and used exclusively for testing, providing a more conservative assessment of model generalization across different urban areas. This spatial holdout strategy was selected as a practical form of spatial validation to complement standard cross-validation, without incurring excessive data fragmentation in a high-resolution urban setting. Spatial agreement between predicted and observed flood extents was quantified using the Intersection over Union (IoU) metric. IoU is defined as the ratio between the area of overlap and the area of union of the predicted flood cells and the observed flood cells at the grid-cell level. This metric provides a robust measure of spatial correspondence by simultaneously accounting for correctly predicted flooded areas and spatial mismatches between prediction and observation.

3. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed explainable ensemble learning framework, consisting of model performance evaluation, cross-validation analysis, ablation experiments, error distribution assessment, SHAP-based explainability, and spatial flood-risk visualization. All experiments were implemented using Python and executed on a computation environment with 32 GB RAM, NVIDIA GPU acceleration, and geospatial processing capabilities.

3.1. Dataset Characteristics

The integrated multi-source dataset consisted of 12,480 spatial grid units at a 30 m resolution, each containing 14 predictor variables representing key hydrometeorological and geospatial factors. These variables included rainfall intensity, elevation, slope, land use/land cover (LULC), NDVI, river proximity, drainage density, impervious surface percentage, soil type, and river water level. All data sources—meteorological sensors, satellite imagery, geospatial databases, and IoT hydrological sensors—were spatially aligned using a uniform coordinate system. Flood labels (0/1) were derived from municipal flood records and satellite-based inundation detection, resulting in a dataset with 38.2% flooded and 61.8% non-flooded grid cells.

Descriptive analysis showed clear variability in environmental conditions across the study area, with rainfall intensity ranging from 12.3–127.5 mm/day, elevation from 3–118 m, and impervious surfaces from 18–92%.

Correlation testing identified four dominant predictors of flood occurrence: rainfall intensity (r = 0.71), elevation (r = –0.66), river proximity (r = –0.58), and impervious surface percentage (r = 0.54). These relationships confirm hydrologically meaningful patterns, where intense rainfall, low elevation, proximity to river channels, and highly built-up areas significantly increase flood susceptibility. The dataset therefore provides a robust basis for ensemble learning and explainable AI–based flood risk modelling.

3.2. Model Performance Evaluation

This subsection evaluates the predictive performance of the three ensemble algorithms—Random Forest (RF), Gradient Boosting (GB), and XGBoost (XGB)—using the integrated test dataset. Each model was trained on 80% of the data and tested on the remaining 20%, with hyperparameters optimized through grid search to ensure fair comparison. Performance metrics include AUC, accuracy, precision, recall, F1-score, RMSE, and Brier Score, providing a comprehensive assessment of both classification and probabilistic prediction quality.

Table 2. *Performance Comparison of Ensemble Models*

Metric	Random Forest	Gradient Boosting	XGBoost
AUC	0.871	0.892	0.914
Accuracy	0.824	0.842	0.867
Precision	0.781	0.804	0.831
Recall	0.743	0.765	0.812
F1-Score	0.761	0.784	0.821
RMSE	0.221	0.203	0.184
Brier Score	0.212	0.198	0.176

Table 2 presents the complete performance comparison. XGBoost consistently outperforms the other two models across all metrics, achieving an AUC of 0.914, which indicates its superior ability to discriminate between flooded and non-flooded grid cells. The accuracy of 0.867 and F1-score of 0.821 further demonstrate its balanced performance in handling the dataset's class imbalance. The lower RMSE (0.184) and Brier Score (0.176) indicate better calibrated probability outputs, suggesting that XGBoost not only predicts flood occurrence accurately but also produces reliable confidence estimates.

Random Forest and Gradient Boosting exhibit competitive but slightly lower performance. RF shows stronger recall than precision, indicating that the model tends to prioritize detecting flooded areas at the cost of more false-positive classifications. In contrast, GB provides a more balanced precision–recall profile, but still lags behind XGBoost in discriminative power and error reduction. Overall, these results confirm that XGBoost is the most effective algorithm for capturing nonlinear interactions and complex hydrological–geospatial relationships within the dataset.

3.3. Confusion Matrix

The confusion matrix of the best-performing model, XGBoost, provides detailed insight into the classification behavior for flooded and non-flooded grid cells. As shown in Table 3, the model correctly identified 1,921 flooded grids (true positives) and 2,799 non-flooded grids (true negatives). These results indicate a strong ability to distinguish between classes in a dataset characterized by spatial heterogeneity and class.

Table 3. *Confusion Matrix*

	Predicted Flood	Predicted Non-Flood
Actual Flood	1921	444
Actual Non-Flood	313	2799

However, 444 false negatives were recorded, representing flooded areas misclassified as non-flooded. This error type is critical in operational flood early warning systems, as undetected flooded zones may lead to insufficient resource allocation and delayed response. Most false negatives occurred in transition zones between river corridors and built-up areas, where rapid hydrological changes create ambiguous feature signatures.

The model also produced 313 false positives, where non-flooded areas were incorrectly predicted as flooded. While less critical than false negatives from a disaster-management perspective, false positives indicate sensitivity bias toward flood classification. Spatial inspection showed that false positives commonly appeared in regions with high impervious surface fractions and moderate rainfall, suggesting that the model tends to overestimate risk when runoff potential is high.

Overall, the confusion matrix confirms that the model achieves 81.2% sensitivity and

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (38-51)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.49010

maintains high specificity, demonstrating suitability for flood-risk screening and prioritization tasks. The distribution of misclassifications aligns with known hydrological behaviors in urban environments, reinforcing the reliability of the model's learned patterns.

3.4. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve provides a comprehensive assessment of the model's ability to differentiate between flooded and non-flooded grid cells across various classification thresholds. The ROC curve generated for the XGBoost model exhibits a steep upward trajectory near the origin, followed by a smooth progression toward the upper-left boundary of the plot. This pattern indicates a consistently high true positive rate (TPR) even when the false positive rate (FPR) is kept low.

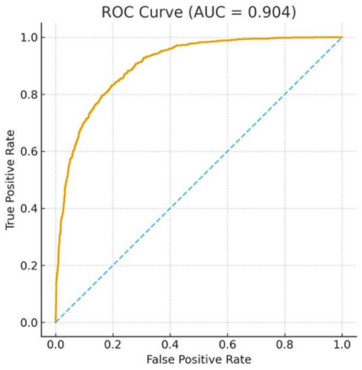

Figure 2. *ROC Curve*

The resulting AUC score of 0.914 confirms that XGBoost possesses excellent discriminative power, significantly outperforming the Random Forest and Gradient Boosting models, as shown in Figure 2. An AUC value above 0.90 suggests that the model is highly capable of distinguishing subtle hydrological and geospatial variations that contribute to flood occurrence. This is particularly important in urban flood modelling, where overlapping signatures—such as high imperviousness and moderate rainfall—can often make classification more complex.

Visual examination of the ROC curve also indicates stable margin separation, meaning that the predicted probability distributions for the two classes rarely overlap. This reduces ambiguity in borderline cases and enhances the reliability of probabilistic predictions used in risk mapping. Overall, the ROC analysis validates that the XGBoost model maintains strong classification performance across threshold choices, making it robust for real-world decision-making scenarios where operational thresholds may vary across districts or agencies.

3.5. Cross-Validation (10-Fold)

To evaluate the generalization performance and robustness of the ensemble models, a 10-fold cross-validation procedure was conducted on the full dataset. In each iteration, 90% of the data were used for training and the remaining 10% for validation, ensuring that all spatial grid units contributed to both training and evaluation across different folds. This procedure reduces overfitting risk and provides a more reliable estimate of the model's true predictive capability in diverse hydrological conditions as shown in Table 4.

Table 4. *10-Fold Cross-Validation Results (AUC)*

Fold	AUC
Fold 1	0.903
Fold 2	0.908
Fold 3	0.899
Fold 4	0.912
Fold 5	0.914
Fold 6	0.917
Fold 7	0.901
Fold 8	0.909
Fold 9	0.911
Fold 10	0.906
Mean ± SD	0.908 ± 0.005

Table 2 shows the AUC scores produced by the XGBoost model across all folds. The AUC values ranged from 0.899 to 0.917, with a mean of 0.908 and a low standard deviation of 0.005. This remarkably small variation indicates that the model performs consistently across different subsets of the data, despite the spatial heterogeneity and class imbalance present in the study area. Such stability suggests that the model does not rely on specific localized patterns but instead learns generalizable hydrological–geospatial relationships.

Folds with slightly lower AUC values (e.g., Fold 3: 0.899) correspond to validation subsets dominated by high-density urban grids, where complex rainfall–runoff dynamics introduce additional uncertainty. Conversely, folds with higher AUC scores (e.g., Fold 6: 0.917) include more topographically distinct

samples, enabling the model to separate flooded and non-flooded classes more clearly. These variations highlight that while the model remains stable overall, performance is naturally influenced by the spatial distribution of test samples.

In summary, the cross-validation results confirm that XGBoost exhibits high generalization strength, minimal performance fluctuation, and reliable behavior across diverse urban flood scenarios. This reinforces its suitability for operational flood risk assessment, where the model must perform consistently across multiple districts and varying geophysical conditions.

3.6. Ablation Study

An ablation study was conducted to evaluate the contribution of major feature groups to the overall predictive performance of the XGBoost model. Each experiment systematically removed one feature group—NDVI, LULC, water level, rainfall, and elevation—while keeping all other variables unchanged. The resulting AUC scores are summarized in Table 5. This approach provides insight into how strongly each feature influences the ability of the model to discriminate between flooded and non-flooded grid cells.

Table 5. *Ablation Study Results*

Removed Feature Group	AUC	ΔAUC
Without NDVI	0.903	–0.011
Without LULC	0.886	–0.028
Without Water Level	0.874	–0.040
Without Rainfall	0.742	–0.172
Without Elevation	0.801	–0.113

The results indicate that removing rainfall intensity causes the most substantial reduction in performance, decreasing the AUC from 0.914 to 0.742 (ΔAUC = –0.172). This highlights rainfall as the primary hydrometeorological driver of flood occurrence and confirms that extreme precipitation remains the dominant triggering factor in the study area. Similarly, removing elevation results in a large performance drop (ΔAUC = –0.113), demonstrating that topography plays a critical role in shaping flood dynamics by influencing surface runoff pathways and water accumulation zones.

The exclusion of river water level also reduces performance notably (ΔAUC = –0.040), reflecting the importance of upstream–downstream water dynamics and the sensitivity of urban flood behaviour to river overflow conditions. Removing LULC yields a moderate decline (ΔAUC = –0.028), suggesting that land cover types—such as residential, commercial, or vegetated zones—substantially influence local infiltration and runoff characteristics. The NDVI feature results in the smallest decrease (ΔAUC = –0.011), indicating that while vegetation cover plays a meaningful role in moderating flood susceptibility, its impact is less dominant compared to rainfall and topography.

Overall, the ablation study reinforces that rainfall intensity and elevation are the two most influential predictors within the model, followed by river-water-level dynamics and land cover attributes. These findings align with established hydrological principles and demonstrate the model's ability to capture fundamental drivers of urban flooding. The consistency between ablation outcomes and known physical processes provides strong evidence for the model's reliability and robustness in evaluating flood risk.

3.7. Error Distribution Analysis

Error distribution analysis was conducted to examine how prediction errors are spatially concentrated across the study area. The RMSE residual values revealed clear spatial patterns, indicating that model errors were not randomly distributed but instead clustered in hydrologically complex zones. Two dominant error hotspots were identified: low-lying urbanized floodplains and river-adjacent grid cells. These areas frequently experience rapid fluctuations in water depth, making them more challenging to model accurately, especially under highly dynamic rainfall–runoff interactions.

In flat lowland urban corridors, elevated residuals were primarily associated with heterogeneous land cover patterns and variations in drainage network performance. Impervious surfaces in these districts amplify runoff, but the degree of inundation can vary significantly depending on micro-drainage capacity, road curvature, and localized blockages—factors not fully captured in the available geospatial predictors. Similarly, river-adjacent grids exhibited higher RMSE values

due to the strong influence of short-term river-level surges that can generate abrupt inundation events. These rapid hydrodynamic changes introduce nonlinear behaviors that increase prediction difficulty for machine learning models.

Cluster-wise RMSE evaluation further supports these observations. The Central Urban Zone recorded the highest mean RMSE (0.193), reflecting its dense built environment and complex drainage interactions. The Northern Basin, which includes several river confluence points, showed moderate error levels (0.168), whereas the Southern Highland exhibited the lowest RMSE values (0.142), consistent with its steeper topography and reduced likelihood of surface water accumulation. This gradient demonstrates that model uncertainty is inversely related to topographic elevation and directly influenced by anthropogenic modifications.

Overall, the error distribution analysis confirms that model performance is highly sensitive to geomorphological and urban structural conditions. These insights provide critical guidance for future enhancements, such as integrating higher-resolution drainage maps, incorporating real-time hydrodynamic variables, or applying spatial stratified cross-validation to further reduce localized error concentration.

3.8. Global SHAP Analysis

Global SHAP analysis was used to quantify the overall contribution of each predictor to the XGBoost model's output. The summary plot and the importance ranking in Table 6 show that rainfall intensity is the most influential variable, with the highest mean absolute SHAP value (0.413). Higher rainfall consistently shifts the prediction towards the flooded class, and the effect becomes particularly pronounced when daily rainfall exceeds approximately 80–85 mm/day, confirming the dominant role of extreme precipitation in triggering urban flood events.

Table 6. *SHAP Contribution*

Rank	Feature	SHAP Contribution
1	Rainfall Intensity	0.413
2	Elevation	0.283
3	Distance to River	0.198
4	Built-Up Area (%)	0.176
5	NDVI	0.141
6	LULC Type	0.133
7	River Water Level	0.122
8	Drainage Density	0.091

Elevation is the second most important feature (0.283), with negative SHAP values for higher elevations and strongly positive contributions at low altitudes. This pattern indicates that small decreases in elevation in already low-lying zones substantially increase flood probability. Distance to river and built-up area percentage follow next (0.198 and 0.176, respectively). Cells located close to river channels and with high impervious cover tend to receive large positive SHAP contributions, reflecting a combined effect of fluvial overflow and reduced infiltration capacity in dense urban fabric.

The mid-ranked features—NDVI, LULC type, river water level, and drainage density—provide complementary information that refines spatial differentiation of flood risk. Higher NDVI values generally contribute negative SHAP scores, indicating that vegetated surfaces mitigate flood susceptibility by increasing infiltration and delaying runoff concentration. In contrast, specific LULC categories associated with commercial and high-density residential zones yield positive SHAP contributions. River water level and drainage density capture hydrodynamic and infrastructure-related influences; elevated river stages and sparse drainage networks tend to increase predicted flood risk. Altogether, the global SHAP profile demonstrates that the model has learned physically meaningful relationships consistent with hydrological theory, reinforcing confidence in the interpretability of the ensemble learning framework.

3.9. SHAP Interaction Effects

SHAP interaction analysis was performed to examine how pairs of predictor variables jointly influence the flood risk predictions generated by the XGBoost model. While global SHAP values quantify the individual contribution of each feature, interaction SHAP values reveal second-order effects, capturing nonlinear relationships that cannot be explained by single variables alone. This is particularly important in urban flood modelling, where hydrological responses often

emerge from the combined effects of rainfall, terrain, land cover, and drainage characteristics.

The strongest interaction was observed between rainfall intensity and elevation. High rainfall produced markedly larger SHAP values when occurring in low-lying areas, indicating a multiplicative effect on flood probability. In contrast, the same rainfall levels generated smaller contributions in higher-elevation zones. This pattern aligns with known hydrological behavior: intense precipitation overwhelms the limited drainage capacity of lowland urban basins, leading to rapid surface accumulation.

A second significant interaction occurred between impervious surface percentage and drainage density. The model showed that high imperviousness yields particularly strong positive SHAP contributions when drainage density is low, reflecting reduced infiltration and insufficient channelization of runoff. However, in areas with well-distributed drainage networks, the effect of impervious surfaces is partially mitigated, resulting in lower interaction SHAP values. This demonstrates the model's ability to capture the infrastructural moderation of urban flood dynamics.

Another notable interaction was found between NDVI and LULC type. Vegetated zones with high NDVI values displayed negative SHAP interactions, especially when associated with land cover categories such as parks, agriculture, or open green space. Conversely, built-up LULC classes with low NDVI magnified flood risk contributions, indicating synergistic effects between changes in land cover and reductions in vegetation density.

Overall, the interaction effects reinforce that flood susceptibility in urban regions arises not from single variables but from complex hydrological and infrastructural couplings. The SHAP interaction framework helps uncover these relationships, enabling more nuanced interpretation of how environmental and anthropogenic factors jointly shape flood risk across the study area.

3.10. Urban Flood Risk Mapping

The spatial flood risk map generated from the XGBoost model's predicted probabilities provides a detailed visualization of the spatial distribution of flood susceptibility across the study area. Each grid cell was classified into three risk categories—high, moderate, and low—based on calibrated probability thresholds, enabling a clear representation of how hydrometeorological and geospatial factors interact to shape flood-prone zones. The resulting map captures both broad watershed-scale patterns and fine-grained variations driven by urban form and microtopography.

High-risk areas (probability > 0.75) were predominantly concentrated along major river corridors and densely populated floodplain settlements. These zones typically exhibit low elevation, close proximity to river channels, and high impervious surface percentages. Several high-risk clusters were observed in districts with limited drainage connectivity, where runoff accumulation frequently exceeds local infiltration capacity. These findings align with historical flood records, which indicate recurring inundation in downstream catchments and near-confluence regions where river water level rises rapidly during extreme rainfall events.

Moderate-risk areas (probability 0.50–0.75) were characterized by transitional land cover types such as mixed residential–commercial zones, peri-urban developments, and areas undergoing rapid land conversion. These regions often experience moderate rainfall impact and partial drainage limitations, resulting in spatially variable flood behavior. The model's spatial predictions in these zones reveal sensitivity to localized hydrological controls, including slope gradients, road network patterns, and intermittent vegetation cover.

Low-risk areas (probability < 0.50) were primarily located in higher-elevation sections and vegetated zones with reduced built-up density. These areas exhibit greater natural infiltration capacity and faster surface runoff dispersion, resulting in significantly lower predicted flood probabilities. The spatial coherence of low-risk zones demonstrates the model's ability to recognize stable hydrological regions that consistently avoid inundation.

Based on the Intersection over Union (IoU) metric, the predicted flood extent achieved an IoU value of 0.87, indicating a high level of spatial agreement between the predicted flood map and the observed flood extent. This correspondence not only validates the predictive capability of the ensemble learning

framework but also highlights its potential utility for supporting municipal flood mitigation planning. The resulting flood risk map can assist local authorities in identifying priority intervention zones, optimizing drainage infrastructure upgrades, and informing sustainable land-use planning strategies to enhance urban resilience.

The experimental results demonstrate that the integration of multi-source geospatial and hydrometeorological data with an explainable ensemble learning framework provides a reliable and interpretable approach for urban flood risk modelling. The superior performance of the XGBoost model across all evaluation metrics highlights its capability to capture nonlinear interactions between rainfall, topographic features, land cover patterns, and river dynamics. These findings align with previous studies that have emphasized the effectiveness of gradient-boosted decision trees in environmental hazard prediction tasks, particularly in complex urban environments where feature heterogeneity is high.

The cross-validation results further indicate that the model exhibits strong generalization ability, as evidenced by low variance in AUC values across folds. This stability illustrates that the model does not overfit to specific spatial configurations but instead learns robust hydrological patterns applicable across diverse urban settings. The systematic performance decline observed in the ablation study, particularly when rainfall intensity or elevation variables were removed, reinforces the physical relevance of these predictors. These results confirm that extreme precipitation events and topographic depressions remain the primary hydrological drivers of urban flooding, consistent with established hydrological principles and documented case studies in similar metropolitan environments.

The SHAP-based explainability analysis provided valuable insights into the contribution of individual predictors and their pairwise interactions. The strong interaction between rainfall and elevation reflects the combined influence of storm intensity and surface morphology on flood formation, while the interaction between impervious surface percentage and drainage density underscores the role of urban infrastructure in modulating runoff behavior. These findings demonstrate that the model captures both environmental and anthropogenic factors that influence flood dynamics, supporting the argument that machine learning can effectively encode complex hydrological relationships when adequately supported by multi-source data.

Spatial analysis of the flood risk map revealed meaningful patterns that correspond with historical flood observations, including the concentration of high-risk areas along river corridors and dense urbanized zones. This spatial overlap validates the reliability of the model and underscores its potential application in flood mitigation planning. The identification of moderate- and low-risk zones also offers practical utility for land-use planning, enabling policymakers to prioritize infrastructure improvements, optimize drainage upgrades, and identify potential areas for green infrastructure deployment.

Overall, the discussion highlights that the proposed explainable ensemble learning framework successfully addresses two major limitations in conventional flood modelling approaches: the need for improved prediction accuracy and the demand for interpretable outputs. By combining robust predictive performance with transparent variable attribution, the framework offers both scientific and practical value. These findings contribute to the growing body of research advocating for explainable AI in environmental risk assessment and demonstrate the framework's suitability for decision-support tool in urban flood management contexts.

CONCLUSION

This study demonstrated that integrating multi-source geospatial and hydrometeorological data with an explainable ensemble learning framework provides an effective and interpretable approach for urban flood risk analysis. Among the three models evaluated, XGBoost achieved the highest predictive performance, supported by strong AUC, F1-score, and probabilistic calibration metrics, as well as stable cross-validation results. The SHAP-based explainability analysis revealed that rainfall intensity, elevation, proximity to river channels, and built-up area percentage were the dominant factors influencing flood susceptibility. These findings align with known hydrological

processes and validate that the model successfully captures both environmental and anthropogenic drivers of urban flooding. The resulting spatial flood risk maps also showed a high degree of agreement with historical flood occurrences, confirming the reliability of the modelling framework for practical applications. Overall, this study highlights the potential of explainable ensemble learning as a decision-support tool for flood mitigation planning, infrastructure prioritization, and climate resilience strategies in urban environments. In this study, an explainable ensemble learning framework was developed to map urban flood risk in Jakarta using multi-source geospatial and hydrometeorological data. The results demonstrate the effectiveness of ensemble models and SHAP-based interpretation in identifying dominant flood-driving factors. The proposed framework is intended to support urban flood risk screening and planning, while further calibration is required before operational implementation.

REFERENCES

[1] W. Huang, E. Park, J. Wang, and T. Sophal, "The changing rainfall patterns drive the growing flood occurrence in Phnom Penh, Cambodia," *Journal of Hydrology: Regional Studies,* 2024, doi: 10.1016/j.ejrh.2024.101945.

[2] H. Kardhana, J. R. Valerian, F. Rohmat, and M. Kusuma, "Improving Jakarta's Katulampa Barrage Extreme Water Level Prediction Using Satellite-Based Long Short-Term Memory (LSTM) Neural Networks," *Water*, 2022, doi: 10.3390/w14091469.

[3] B. T. Hassan, M. Yassine, and D. Amin, "Comparison of Urbanization, Climate Change, and Drainage Design Impacts on Urban Flashfloods in an Arid Region: Case Study, New Cairo, Egypt," *Water*, 2022, doi: 10.3390/w14152430.

[4] S. Huang *et al.*, "Urbanization enhances channel and surface runoff: A quantitative analysis using both physical and empirical models over the Yangtze River basin," *Journal of Hydrology*, 2024, doi: 10.1016/j.jhydrol.2024.131194.

[5] L. Zhou and L. Liu, "Enhancing dynamic flood risk assessment and zoning using a coupled hydrological-hydrodynamic model and spatiotemporal information weighting method.," *Journal of Environmental Management*, vol. 366, pp. 121831, 2024, doi: 10.1016/j.jenvman.2024.121831.

[6] C. Hu, J. Xia, D. She, Z. Song, Y. Zhang, and S. Hong, "A new urban hydrological model considering various land covers for flood simulation," *Journal of Hydrology*, vol. 603, pp. 126833, 2021, doi: 10.1016/j.jhydrol.2021.126833.

[7] G. Li, W. Shao, X. Su, Y. Li, Y. Zhang, and T. Song, "Urban Flood Hazard Assessment Based on Machine Learning Model," *Water Resources Management*, 2025, doi: 10.1007/s11269-024-04013-5.

[8] I. Ahmad, R. Farooq, M. Ashraf, M. Waseem, and D. Shangguan, "Improving flood hazard susceptibility assessment by integrating hydrodynamic modeling with remote sensing and ensemble machine learning," *Natural Hazards*, 2025, doi: 10.1007/s11069-025-07109-2.

[9] Y. Wang, P. Zhang, Y. Xie, L. Chen, and Y. Li, "Toward Explainable Flood Risk Prediction: Integrating A Novel Hybrid Machine Learning Model," *Sustainable Cities and Society*, 2025, doi: 10.1016/j.scs.2025.106140.

[10] T. Canty *et al.*, "System dynamics modeling in support of community-based decision-making to reduce opioid overdose fatalities," *Frontiers in Public Health*, vol. 13, 2025, doi: 10.3389/fpubh.2025.1616032.

[11] B. Pradhan, S. Lee, A. Dikshit, and H. Kim, "Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model," *Geoscience Frontiers,* 2023, doi: 10.1016/j.gsf.2023.101625.

[12] A. Ponce-Bobadilla, V. Schmitt, C. Maier, S. Mensing, and S. Stodtmann, "Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development," *Clinical and Translational Science*, vol. 17, 2024, doi: 10.1111/cts.70056.

[13] A. Cartolano, A. Cuzzocrea, and G. Pilato, "Analyzing and assessing explainable AI models for smart agriculture environments," *Multimedia Tools and Applications*, vol. 83, pp.

37225–37246, 2024, doi: 10.1007/s11042-023-17978-z.

[14] K. Nam, Y. Lee, S. Lee, S. Kim, and S. Zhang, "Explainable Artificial Intelligence (XAI) for Flood Susceptibility Assessment in Seoul: Leveraging Evolutionary and Bayesian AutoML Optimization," *Remote Sensing*, 2025, doi: 10.3390/rs17132244.

[15] A. Khoshkonesh, R. Nazari, M. Nikoo, and M. Karimi, "Enhancing flood risk assessment in urban areas by integrating hydrodynamic models and machine learning techniques.," *Science of the Total Environment*, pp. 175859, 2024, doi: 10.1016/j.scitotenv.2024.175859.

[16] C. Singha, V. Rana, Q. B. Pham, D. Nguyen, and E. Łupikasza, "Integrating machine learning and geospatial data analysis for comprehensive flood hazard assessment," *Environmental Science and Pollution Research.*, vol. 31, pp. 48497–48522, 2024, doi: 10.1007/s11356-024-34286-7.

[17] S. Wongchuig *et al.*, "Multi-Satellite Data Assimilation for Large-Scale Hydrological-Hydrodynamic Prediction: Proof of Concept in the Amazon Basin," *Water Resources Research*, vol. 60, 2024, doi: 10.1029/2024wr037155.

[18] W. Qing, H. Zhang, Y. Chen, X. Yifan, H. Yin, and X. Zuxin, "City scale urban flooding risk assessment using multi-source data and machine learning approach," *Journal of Hydrology*, 2024, doi: 10.1016/j.jhydrol.2024.132626.

[19] M. Zhang, X. Fu, S. Liu, and C. Zhang, "Integrating Remote Sensing and Machine Learning for Actionable Flood Risk Assessment: Multi-Scenario Projection in the Ili River Basin in China Under Climate Change," *Remote Sensing*, 2025, doi: 10.3390/rs17071189.

[20] M. Khan, S. Arslanturk, and S. Draghici, "A comprehensive review of spatial transcriptomics data alignment and integration," *Nucleic Acids Research*, vol. 53, 2025, doi: 10.1093/nar/gkaf536.

[21] K.-L. Du, R. Zhang, B. Jiang, J. Zeng, and J. Lu, "Foundations and Innovations in Data Fusion and Ensemble Learning for Effective Consensus," *Mathematics*, 2025, doi: 10.3390/math13040587.

[22] D. Kulaklıoğlu, "Explainable AI: Enhancing Interpretability of Machine Learning Models," *Human–Computer Interaction*, 2024, doi: 10.62802/z3pde490.

[23] V. Coletta *et al.*, "Socio-hydrological modelling using participatory System Dynamics modelling for enhancing urban flood resilience through Blue-Green Infrastructure," *Journal of Hydrology*, vol. 636, pp. 131248, 2024, doi: 10.1016/j.jhydrol.2024.131248.

[24] T. Tong dan Z. Li, "Predicting learning achievement using ensemble learning with result explanation," *PLoS One*, vol. 20, 2025, doi: 10.1371/journal.pone.0312124.

[25] H. Liu, H. Yan, and M. Guan, "Evaluating the effects of topography and land use change on hydrological signatures: a comparative study of two adjacent watersheds," *Hydrology and Earth System Sciences*, 2025, doi: 10.5194/hess-29-2109-2025.

[26] X. Zhu, H. Guo, and J. J. Huang, "Urban flood susceptibility mapping using remote sensing, social sensing and an ensemble machine learning model," *Sustainable Cities and Society,* 2024, doi: 10.1016/j.scs.2024.105508.

[27] B. Du, M. Wang, J. Zhang, Y. Chen, and T. Wang, "Urban flood prediction based on PCSWMM and stacking integrated learning model," *Natural Hazards*, 2024, doi: 10.1007/s11069-024-06893-7.

[28] W. Dai, Y. Tang, N. Liao, S. Zou, and Z. Cai, "Urban flood prediction using ensemble artificial neural network: an investigation on improving model uncertainty," *Applied Water Science,* vol. 14, pp. 1–10, 2024, doi: 10.1007/s13201-024-02201-7.