Wujud: Journal of Engineering and Technology
2026, 1 (1), 40-67

Predicting Urban Flood Susceptibility Using Machine Learning and Geospatial Indicators for Climate-Resilient Infrastructure Planning

Muhammad Farhan Akbar[1]*; Intan Permata Sari[2]; Arif Rahman Hakim[1]; Dewi Lestari Wulandari[2]

[1]Department of Civil and Environmental Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia

[2]Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia

Article History
Received: 2026-01-17
Revised: 2026-02-21
Accepted: 2026-04-15
Published: 2026-05-31

Keywords
urban flood susceptibility; machine learning; geospatial indicators; climate-resilient infrastructure; Semarang City

DOI
10.xxxx/wujud.volx.issx.artx

Corresponding author
Muhammad Farhan Akbar
muhammadfarhanakbar@gmail.com

Other author email(s)
intanpermatasari@gmail.com
arifrahmanhakim@gmail.com
dewilestariwulandari@gmail.com

Paper type
Research article

Published by PT Penerbit Velora Indonesia

Abstract

Purpose – This study predicts urban flood susceptibility in Semarang City, Indonesia, by integrating machine-learning algorithms with hydroclimatic, topographic, land-surface, and built-environment geospatial indicators and translating the results into climate-resilient infrastructure priorities.

Methodology – A quantitative spatial-predictive design used a 10 m × 10 m raster grid and 2020–2025 geospatial observations. Logistic Regression, Support Vector Machine, Random Forest, XGBoost, and LightGBM were evaluated through repeated spatial block cross-validation and an independent test dataset. SHapley Additive exPlanations interpreted predictor contributions, while susceptibility was integrated with infrastructure criticality and exposure.

Findings – In the simulated analytical scenario, XGBoost achieved the strongest performance with an AUC of 0.962, accuracy of 0.908, and F1-score of 0.910. Elevation, 72-hour antecedent rainfall, Topographic Wetness Index, impervious-surface ratio, and river distance were the dominant predictors. High or very high susceptibility covered 31.8% of Semarang City, while 29.5% had high or very high infrastructure adaptation priority.

Implications – The framework supports prioritization of drainage improvement, infrastructure strengthening, blue-green infrastructure, land-use management, and detailed engineering assessment.

Originality – The study integrates multidimensional geospatial indicators, interpretable machine learning, spatially robust validation, and infrastructure-priority assessment in a unified climate-resilient planning framework.

Cite this article:
Akbar, M. F., Sari, I. P., Hakim, A. R., & Wulandari, D. L. (2026). Predicting urban flood susceptibility using machine learning and geospatial indicators for climate-resilient infrastructure planning. Wujud: Journal of Engineering and Technology, 1(1), 40–67. https://doi.org/10.xxxx/wujud.volx.issx.artx

Introduction

Urban flooding has become a critical challenge for contemporary cities because climate change, rapid urbanization, land-cover transformation, and infrastructure expansion increasingly interact within the same spatial system. More intense precipitation can exceed drainage capacity within short periods, while impervious surfaces reduce infiltration and accelerate surface runoff. Islam et al. (2025) identified intense precipitation, rapid urban development, inadequate drainage, and impervious-surface expansion as recurrent drivers of urban flood susceptibility. Climate-related flooding also disrupts transport networks, utilities, public facilities, residential areas, and economic activity. Dharmarathne et al. (2024) emphasized that changing precipitation patterns have increased the vulnerability of urban infrastructure and strengthened the need for climate-resilient planning. Urban flooding should therefore be treated not only as a hydrological hazard but also as a spatial infrastructure problem that requires reliable predictive evidence.

P ISSN xxxx-xxxx | E ISSN xxxx-xxxx
Copyright © 2026 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License https://creativecommons.org/licenses/by-sa/4.0

The problem is particularly relevant to Indonesian cities, where high tropical rainfall, extensive river systems, dense urban development, and rapid land conversion create complex flood conditions. Semarang City provides an important case because its urban landscape extends from low-lying northern coastal zones to higher terrain in the southern part of the city. Flood events repeatedly affect roads, settlements, public facilities, and drainage systems. Evidence from another Indonesian setting shows the magnitude of these impacts. Sampurno et al. (2024) reported that flooding in Ngabang District affected 373.81 hectares, 10,706 people, 1,500 buildings, and approximately 15 km of roads. Future climatic conditions can further increase this exposure. Verdyansyah et al. (2025) showed that machine-learning projections for Demak District indicated a substantial increase in areas classified as having very high flood susceptibility under high-emission climate scenarios. These findings reinforce the importance of predictive spatial assessment for cities on the northern coast of Java.

Urban flood susceptibility reflects interactions among hydroclimatic, topographic, land-surface, and built-environment conditions. Rainfall controls the amount of water entering an urban system, while elevation, slope, flow accumulation, drainage density, and proximity to rivers determine how water moves and accumulates. Zhou et al. (2024) found that rainfall volume, rainfall duration, impervious surface coverage, building congestion, and building density contributed strongly to urban flooding in high-density areas. Vegetation, soil, drainage systems, and surface temperature can further influence local runoff. Gulshad et al. (2024) identified digital elevation, NDVI, river buffers, land surface temperature, soil characteristics, and rainwater infrastructure as relevant predictors of urban flood susceptibility. This evidence indicates that a reliable flood model should integrate natural and anthropogenic indicators within one geospatial framework.

Machine learning offers an appropriate analytical approach because the relationship between flooding and its conditioning factors is rarely linear. Algorithms can identify nonlinear interactions, thresholds, and high-dimensional patterns that conventional statistical methods may not capture efficiently. Dawson et al. (2023) demonstrated that machine-learning methods can produce reliable susceptibility predictions across large areas while allowing predictor contributions to be assessed. Comparative model evaluation remains necessary because performance differs across algorithms. Waleed and Sajjad (2025) found substantial variation in accuracy and computational scalability among machine-learning models used for flood susceptibility prediction. Ensemble algorithms such as Random Forest, XGBoost, and LightGBM are especially relevant because they can combine multiple decision structures and capture complex interactions among rainfall, terrain, land cover, and urban form.

High predictive accuracy alone, however, does not provide enough information for infrastructure planning. Decision-makers also need to understand why the model identifies particular locations as highly susceptible. Singha et al. (2024) demonstrated that explainability methods can identify precipitation, elevation, and river proximity as influential flood predictors. Manna et al. (2025) similarly showed that SHAP analysis can reveal the nonlinear effects of rainfall, elevation, urban density, and road-related factors in a metropolitan flood model. Explainability can therefore transform machine-learning results from a purely predictive output into evidence that supports targeted interventions. Drainage improvement may be appropriate where drainage access dominates susceptibility, while land-use control may be more relevant where imperviousness and building density drive the predicted probability.

Several methodological gaps remain despite recent advances. Flood susceptibility studies employ highly heterogeneous combinations of conditioning factors, sampling schemes, model algorithms, and validation techniques. Long et al. (2025) found that flood inventory construction, factor selection, multicollinearity analysis, cross-validation, and ensemble design remain major sources of variation across the literature. Data-driven flood models also face challenges related to spatial leakage, uncertainty, generalization, and limited interpretability. Anik et al. (2025)

emphasized that the transition from physical flood models to machine-learning approaches requires stronger attention to data quality, model transferability, and physical consistency. Another limitation concerns the use of susceptibility maps. Many studies stop after producing hazard classes without explicitly linking the results to infrastructure criticality and adaptation priorities.

This study addresses these limitations by developing an interpretable machine-learning framework for predicting urban flood susceptibility in Semarang City using hydroclimatic, topographic, land-surface, and built-environment geospatial indicators. The study also translates predicted flood susceptibility into spatial evidence for climate-resilient infrastructure planning. Hussain et al. (2023) demonstrated that susceptibility information can form an important component of broader urban flood resilience assessment. Ahmad et al. (2025) further showed that spatial flood assessment can support evidence-based infrastructure investment and drainage prioritization. Accordingly, this study addresses three research questions: RQ1: Which geospatial indicators contribute most strongly to urban flood susceptibility in Semarang City? RQ2: Which machine-learning model provides the most reliable and interpretable prediction of urban flood susceptibility? RQ3: How can predicted susceptibility support spatial prioritization for climate-resilient infrastructure planning?

Literature Review

Urban Flood Susceptibility

Urban flood susceptibility describes the relative tendency of a spatial location to experience flooding when specific environmental and anthropogenic conditions occur. Susceptibility differs from comprehensive flood risk because it focuses primarily on the spatial predisposition to inundation, while risk normally incorporates hazard, exposure, vulnerability, and potential consequences. Widya et al. (2024) demonstrated that flood susceptibility can be effectively differentiated by combining terrain, land use, vegetation, soil, and hydrological factors.

Flood mechanisms also influence the spatial distribution of susceptibility. Riverine flooding, flash flooding, and urban pluvial flooding may respond differently to rainfall characteristics, drainage conditions, topography, and land use. Tayyab et al. (2024) showed that integrating GIS, remote sensing, and machine learning can distinguish susceptibility patterns associated with different flood types. This distinction is important in Semarang because rainfall-driven urban flooding should be separated analytically from purely tidal inundation or rob when the objective is to model rainfall, drainage, terrain, and built-environment relationships.

Topography has a fundamental role in flood generation because elevation and terrain configuration control water movement. Meng and Jin (2023) found that improved terrain representation can strengthen machine-learning-based flood susceptibility prediction. Low elevation can increase the tendency for water accumulation, while slope controls runoff velocity and concentration. Flow accumulation and wetness indices further represent locations where water naturally converges.

Flood inventory quality also influences susceptibility analysis. Yu et al. (2023) demonstrated that improving the quantity and spatial representation of flood inventory observations can increase the accuracy of flood susceptibility models. This finding is important because models trained on incomplete or spatially clustered flood observations can produce biased susceptibility surfaces.

Artificial intelligence has also been used to integrate multiple flood conditioning variables into composite susceptibility outputs. Lin et al. (2023) showed that neural methods can combine multidimensional indicators into data-driven flood susceptibility indices. Rudra and Sarkar (2023) similarly demonstrated that artificial neural networks can effectively identify flood-prone areas from multiple geospatial predictors. These developments support the use of quantitative, spatially explicit modelling rather than relying solely on qualitative hazard classification.

Machine Learning

Machine learning can identify complex relationships between historical flood occurrence and environmental predictors without requiring a predetermined linear functional form. Tree-based methods, kernel models, neural networks, and ensemble learners offer different approaches for representing nonlinear decision boundaries. Al-Kindi and Alabri (2024) found that XGBoost, Random Forest, and CatBoost all provided strong flood discrimination, although their performance and predictor importance differed.

Boosting algorithms have received particular attention because they sequentially reduce prediction error. Costache et al. (2025) compared AdaBoost, CatBoost, LightGBM, and XGBoost and showed that advanced boosting methods can generate highly accurate flood susceptibility predictions when combined with structured feature selection. Their results also showed that slope, river proximity, TWI, and land use can become important depending on the model.

No machine-learning algorithm can be assumed to perform best in every study area. Kurugama et al. (2024) found differences in the performance of boosting algorithms under the environmental characteristics of Sri Lanka. Such differences arise because model performance depends on sample size, flood inventory characteristics, spatial resolution, predictor distributions, environmental heterogeneity, and hyperparameter configuration.

Explainable machine learning has become increasingly important because flood susceptibility models often function as black boxes. Wang, Lyu, and Zhang (2024) demonstrated that explainable machine learning can improve the practical value of urban pluvial flood susceptibility mapping by combining predictive capability with interpretable feature contributions. This approach allows researchers to identify not only where susceptibility is high but also the variables that drive that prediction.

Deep-learning models provide another direction for spatial flood assessment. Wang et al. (2025) integrated hydrological prior knowledge into a modified U-Net architecture and achieved strong predictive performance while also using visual explanation techniques. Such methods demonstrate the potential of spatial deep learning, although their data and computational requirements can be higher than those of conventional machine-learning models.

Ensemble explainability can further improve decision support. Zerouali et al. (2025) combined stacked ensemble learning with SHAP and obtained strong flood susceptibility performance in a data-constrained urban environment. Their findings indicate that explainability can remain feasible even when multiple machine-learning algorithms contribute to the final predictive structure.

Machine-learning applications should still be evaluated cautiously. Uddameri and Hernandez (2025) argued that limited data availability and inter-model variability can reduce the expected superiority of complex machine-learning models. Their review also highlighted the importance of explainable AI for moving from black-box predictions toward knowledge-enhancing flood resilience tools.

Geospatial Indicators

Geospatial indicators represent the climatic, terrain, surface, and urban characteristics associated with flooding. Hydroclimatic indicators commonly include rainfall intensity, antecedent precipitation, drainage density, and proximity to watercourses. Topographic variables include elevation, slope, flow accumulation, curvature, and wetness indices. Land-surface indicators include land use, vegetation, soil, and imperviousness. Built-environment indicators include buildings, roads, drainage facilities, and critical infrastructure.

Remote sensing and GIS provide an effective basis for deriving these variables consistently across an urban area. Amiri et al. (2024) demonstrated that precipitation, elevation, slope, vegetation, geology, and remote-sensing information can be integrated with machine learning to assess both current and future flood susceptibility. Their results support the use of multi-source spatial information rather than relying on one group of conditioning variables.

Hydroclimatic and terrain variables often represent the physical basis of susceptibility. Salvati et al. (2023) showed that optimized machine-learning models can use terrain and environmental predictors to produce reliable flood susceptibility maps. Elevation and slope influence runoff direction, while rainfall controls water input. The effect of one variable therefore depends partly on the spatial configuration of other predictors.

Built-environment indicators capture hydrological modifications created by urban development. Kaur et al. (2025) found that urban infrastructure and environmental variables can contribute meaningfully to susceptibility prediction. Distance from drainage infrastructure can represent stormwater accessibility, while building and road density provide information about urban form and imperviousness.

Road systems have additional relevance because they can modify runoff pathways while simultaneously functioning as exposed infrastructure. Rezvani et al. (2024) demonstrated how geospatial AI flood mapping can be integrated with national road networks to identify transportation segments facing elevated flood risk. This approach illustrates how susceptibility information can move directly into infrastructure planning.

Land-use transformation also influences infiltration and runoff. Chen et al. (2023) showed that machine-learning models can identify different relationships between environmental conditioning factors and flash-flood susceptibility. The effect of land cover should therefore be treated as model-dependent and spatially heterogeneous rather than represented by a single universal coefficient.

Climate-Resilient Infrastructure Planning

Climate-resilient infrastructure planning aims to maintain essential urban services despite changing rainfall patterns and recurring flood hazards. Resilient infrastructure should reduce disruption, maintain critical functions, recover efficiently, and adapt to changing conditions. Haghbin and Mahjouri (2023) showed that flood resilience in urban drainage systems includes both technical and socio-ecological dimensions.

Drainage resilience depends strongly on the interaction between rainfall and urbanization. Li et al. (2023) demonstrated that increasing imperviousness and changing rainfall conditions can reduce the resilience of urban drainage systems. Infrastructure planning should therefore account for future changes in both climatic forcing and land-surface conditions.

Adaptive infrastructure can combine conventional and nature-based interventions. Li and Burian (2023) found that real-time control and green stormwater infrastructure can strengthen flood resilience under future rainfall projections. Their findings indicate that susceptibility mapping should inform not only where infrastructure is vulnerable but also the types of interventions that may improve system performance.

Spatial planning can also integrate blue-green infrastructure. Ambily et al. (2024) proposed a framework for incorporating blue-green solutions into urban pluvial flood-resilient spatial planning. This approach is relevant for cities where expanding conventional drainage capacity alone may not provide sufficient long-term resilience.

Infrastructure resilience requires evidence-based prioritization because financial and engineering resources are limited. Liao et al. (2025) demonstrated that grid-based urban flood risk and resilience assessment can identify local areas with high risk and low resilience. Such fine-scale approaches support targeted intervention rather than uniform city-wide investment.

Machine learning can support infrastructure decisions through rapid hazard classification. Li et al. (2025) showed that convolutional neural networks can provide efficient urban surface-water flood hazard assessment. Predictive models can therefore function as screening tools for identifying areas that require detailed engineering evaluation.

Flood susceptibility also becomes more useful when it is combined with infrastructure function. Qin et al. (2025) integrated machine-learning-based flood susceptibility with building-function vulnerability and showed that such integration can strengthen urban resilience assessment. This approach recognizes that a flood affecting critical infrastructure may produce greater consequences than a similar flood affecting less essential land uses.

Resilience planning should combine technical, spatial, and policy interventions. Azadgar et al. (2024) concluded that effective urban flood resilience requires integration across modelling, planning, adaptation, and mitigation strategies. Susceptibility modelling can support this integration when its outputs are translated into explicit infrastructure priorities rather than treated as final hazard maps.

Research Gap

Recent studies confirm the strong potential of machine learning for flood susceptibility prediction, yet important limitations remain. First, studies frequently use different combinations of rainfall, terrain, land-surface, and urban predictors. This inconsistency makes it difficult to determine whether built-environment indicators add meaningful predictive information beyond hydroclimatic and topographic variables.

Second, spatial dependence can create unrealistic model performance when neighbouring observations appear simultaneously in training and validation datasets. Adnan et al. (2023) emphasized that uncertainties in geospatial machine-learning flood prediction arise from both data and modelling configurations. Spatial validation is therefore necessary to reduce optimistic accuracy estimates.

Third, global predictor importance can conceal local differences within a city. Xing et al. (2025) showed that fine-grained interpretable machine learning can reveal substantial spatial variation in the importance of vegetation, terrain, land use, water proximity, and roads. This finding indicates that local urban mechanisms should be considered when interpreting susceptibility maps.

Fourth, model accuracy is often treated as the primary outcome, while planning relevance receives less attention. Shirzadi et al. (2025) demonstrated that machine-learning models can effectively identify urban flood susceptibility, but susceptibility information still requires translation before it can support infrastructure management.

The present study addresses these gaps by integrating hydroclimatic, topographic, land-surface, and built-environment indicators within one predictive framework. It compares conventional and ensemble algorithms under spatial cross-validation, evaluates the incremental value of predictor groups, uses SHAP to improve interpretability, and combines the resulting susceptibility information with infrastructure criticality and exposure to support climate-resilient infrastructure planning.

Hypothesis Development

Geospatial Indicators and Urban Flood Susceptibility

Flood susceptibility develops through interactions among climatic, terrain, surface, and urban conditions. Rainfall provides the water input, topography controls movement and accumulation, and urban development modifies infiltration and drainage. A model that integrates several geospatial dimensions should therefore distinguish flooded and non-flooded locations more effectively than a model containing limited reference variables. Singha et al. (2024) found that multi-source geospatial predictors can provide strong flood discrimination when evaluated using machine-learning methods. Based on this relationship, the first hypothesis is formulated as follows:

H1: Geospatial indicators significantly predict urban flood susceptibility.

Hydroclimatic and Topographic Indicators and Urban Flood Susceptibility

Hydroclimatic and topographic factors represent fundamental physical controls on flood occurrence. Rainfall intensity and antecedent precipitation determine water availability, while elevation, slope, drainage density, river proximity, and flow accumulation influence runoff routing and concentration. Amiri et al. (2024) demonstrated that precipitation, elevation, and slope contribute substantially to flood susceptibility prediction. Therefore:

H2: Hydroclimatic and topographic indicators significantly increase the predictive capability of urban flood susceptibility models.

Land-Surface and Built-Environment Indicators and Urban Flood Susceptibility

Urban development alters hydrological processes by reducing infiltration and increasing runoff. Impervious surfaces, buildings, roads, and drainage systems can change both runoff quantity and flow pathways. Vegetation and soil conditions further modify infiltration capacity. Shirzadi et al. (2025) identified land use, building density, transport-related characteristics, and rainfall as important predictors of urban flood susceptibility. Consequently:

H3: Land-surface and built-environment indicators significantly improve urban flood susceptibility prediction beyond hydroclimatic and topographic indicators.

Machine-Learning Performance and Urban Flood Susceptibility

Urban flooding involves nonlinear relationships and interactions among multiple predictors. Ensemble algorithms combine many individual decision structures and may provide stronger generalization than simpler single-model approaches. Waleed and Sajjad (2025) demonstrated strong predictive performance from gradient-boosting algorithms across flood susceptibility applications. Therefore:

H4: Ensemble machine-learning models provide higher predictive performance for urban flood susceptibility than conventional single-model approaches.

Urban Flood Susceptibility and Climate-Resilient Infrastructure Planning

Flood susceptibility has direct planning relevance when highly susceptible areas overlap with important infrastructure. Roads, drainage systems, hospitals, schools, transport facilities, and emergency services require higher adaptation attention when their spatial locations coincide with persistent flood susceptibility. Lyu et al. (2025) showed that machine-learning-based susceptibility assessment can identify vulnerable urban metro infrastructure under changing rainfall conditions. Accordingly:

H5: Higher urban flood susceptibility is significantly associated with greater priority for climate-resilient infrastructure planning.

Table 1. Summary of Research Hypotheses

Code	Hypothesis
H1	Geospatial indicators significantly predict urban flood susceptibility.
H2	Hydroclimatic and topographic indicators significantly increase the predictive capability of urban flood susceptibility models.
H3	Land-surface and built-environment indicators significantly improve urban flood susceptibility prediction beyond hydroclimatic and topographic indicators.
H4	Ensemble machine-learning models provide higher predictive performance for urban flood susceptibility than conventional single-model approaches.
H5	Higher urban flood susceptibility is significantly associated with greater priority for climate-resilient infrastructure planning.

Source: *Developed by the authors based on the theoretical relationships synthesized in the literature review.*

Research Methods

This study employed a quantitative spatial-predictive design to estimate urban flood susceptibility in Semarang City, Central Java, Indonesia, using observations from 1 January 2020 to 31 December 2025. The analysis integrated Geographic Information Systems, remote sensing, supervised machine learning, explainable artificial intelligence, and infrastructure-priority assessment. The primary unit of analysis was a 10 m × 10 m raster grid cell, and all spatial layers were transformed into WGS 84/UTM Zone 49S (EPSG:32749). Historical flood observations were obtained from SEMARIsK, BPBD Kota Semarang, while rainfall data were derived from BMKG and supplemented with NASA GPM IMERG V07. DEMNAS was used to derive elevation, slope, flow accumulation, and the Topographic Wetness Index. Sentinel-2 Level-2A imagery provided land-use, NDVI, and impervious-surface information, whereas river, drainage, road, building, and critical-facility data were obtained from DPU Kota Semarang and relevant municipal spatial databases. Only rainfall-driven urban flood records classified as banjir were included in the primary flood inventory, while tidal flooding or rob was excluded to maintain consistency between the response variable and the selected predictors.

Table 2. Data Sources and Spatial Characteristics

Dataset	Analytical Use	Resolution/Form	Source
Flood events, 2020–2025	Dependent variable and flood inventory	Points/polygons	SEMARIsK, BPBD Kota Semarang
Sentinel-1 SAR	Flood-extent verification	Approx. 10 m	Copernicus
Rainfall observations	Rainfall intensity and antecedent rainfall	Hourly/daily	BMKG
GPM IMERG V07	Supplementary rainfall coverage	0.1°, half-hourly	NASA GPM
DEMNAS	Elevation and terrain variables	0.27 arc-second	BIG
Sentinel-2 Level-2A	LULC, NDVI, imperviousness	10 m	Copernicus
River and drainage networks	River/drainage proximity and density	Vector	DPU Kota Semarang/BIG
Roads, buildings, critical facilities	Urban morphology and infrastructure analysis	Vector	Semarang municipal spatial databases

Source: *Authors' compilation from BPBD Kota Semarang, DPU Kota Semarang, BIG, BMKG, NASA GPM, and Copernicus datasets.*

Urban flood occurrence was coded as $Y_i = 1$ for flooded cells and $Y_i = 0$ for non-flooded cells. Sixteen candidate predictors were grouped into hydroclimatic, topographic, land-surface, and built-environment indicators. Hydroclimatic indicators included maximum hourly rainfall, 72-hour antecedent rainfall, distance to rivers, and drainage density. Topographic indicators included elevation, slope, flow accumulation, and TWI. Land-surface variables included land use/land cover, NDVI, impervious-surface ratio, and soil class, while built-environment variables included building density, road density, distance to drainage, and critical-facility density. Flood and non-flood observations were sampled at an approximate 1:1 ratio, with a 100 m exclusion buffer around confirmed flood areas. The dataset was divided into 70% development data and 30% independent test data, while five-fold spatial block cross-validation repeated five times was applied within the development dataset to reduce spatial leakage.

$$P(Y_i = 1 \mid \mathbf{X}_i) = f(\mathbf{X}_i; \boldsymbol{\theta}) \tag{1}$$

$$NDVI_i = \frac{NIR_i - RED_i}{NIR_i + RED_i} \tag{2}$$

$$TWI_i = \ln\left(\frac{A_{s,i}}{\tan\beta_i}\right) \tag{3}$$

Table 3. Operationalization of Research Variables

Group	Indicator	Operational Definition	Unit
Dependent	Urban flood susceptibility	Observed flood status and predicted probability	0–1
Hydroclimatic	Maximum rainfall intensity	Maximum hourly rainfall	mm/hour
Hydroclimatic	72-hour antecedent rainfall	Cumulative rainfall in previous 3 days	mm
Hydroclimatic	Distance to river	Distance to nearest river	m
Hydroclimatic	Drainage density	Drainage length per unit area	km/km²
Topographic	Elevation	Ground elevation	m
Topographic	Slope	Terrain gradient	degrees
Topographic	Flow accumulation	Accumulated upstream flow	cells
Topographic	TWI	Terrain wetness potential	dimensionless
Land surface	LULC	Surface-cover category	categorical
Land surface	NDVI	Vegetation greenness	−1 to +1
Land surface	Imperviousness	Proportion of sealed surface	%
Land surface	Soil class	Soil and infiltration characteristics	categorical
Built environment	Building density	Buildings per unit area	buildings/ha
Built environment	Road density	Road length per unit area	km/km²
Built environment	Distance to drainage	Distance to engineered drainage	m
Built environment	Critical facility density	Essential facilities per unit area	facilities/km²

Source: Developed by the authors based on the geospatial factors reported by Islam et al. (2025).

Five predictive models were evaluated: Logistic Regression, Support Vector Machine, Random Forest, XGBoost, and LightGBM. Logistic Regression served as the statistical baseline, Support Vector Machine represented a nonlinear non-ensemble benchmark, and Random Forest, XGBoost, and LightGBM represented ensemble approaches. Hyperparameters were optimized exclusively within the development dataset. Predictor redundancy was examined using pairwise correlation and the Variance Inflation Factor. Four nested predictor configurations were compared to evaluate the incremental contribution of each indicator group: hydroclimatic only; hydroclimatic plus topographic; hydroclimatic, topographic, and land-surface; and the complete model including built-environment indicators. Model performance was evaluated using accuracy, precision, recall, specificity, F1-score, AUC, and Brier Score. The best-performing tree-based model was subsequently interpreted using SHapley Additive exPlanations to identify both global and local contributions of the geospatial predictors.

$$VIF_j = \frac{1}{1-R_j^2} \tag{4}$$

$$M_1 = f(H), \quad M_2 = f(H,T), \quad M_3 = f(H,T,L), \quad M_4 = f(H,T,L,B) \tag{5}$$

$$\Delta AUC_k = AUC(M_k) - AUC(M_{k-1}) \tag{6}$$

$$F_1 = \frac{2TP}{2TP+FP+FN} \tag{7}$$

$$BS = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - y_i)^2 \tag{8}$$

$$f(\mathbf{x}_i) = \phi_0 + \sum_{j=1}^{M}\phi_{ij} \tag{9}$$

Table 4. Machine-Learning Models and Analytical Roles

Model	Category	Analytical Role
Logistic Regression	Statistical classifier	Baseline model
Support Vector Machine	Nonlinear classifier	Non-ensemble benchmark
Random Forest	Bagging ensemble	Ensemble model
XGBoost	Gradient-boosting ensemble	Advanced ensemble model
LightGBM	Gradient-boosting ensemble	Efficient ensemble model

Source: *Developed by the authors based on the comparative flood-modelling framework adopted in this study.*

The selected model was applied to all valid grid cells to generate continuous urban flood susceptibility probabilities ranging from 0 to 1. For cartographic interpretation, the probability surface was classified into very low, low, moderate, high, and very high susceptibility using Jenks Natural Breaks, while continuous probabilities remained the primary values for statistical analysis. Climate-resilient infrastructure planning was assessed by overlaying the susceptibility surface with primary and secondary drainage systems, arterial and collector roads, hospitals, schools, emergency facilities, transport nodes, and other essential infrastructure. Infrastructure indicators were normalized to the range 0–1 and combined into an independently derived criticality score using equal weights in the primary analysis. The relationship between flood susceptibility and infrastructure criticality was examined using Spearman's rank correlation at $\alpha = 0.05$. After this independent relationship was tested, the Climate-Resilient Infrastructure Priority Index combined normalized susceptibility, infrastructure criticality, and infrastructure exposure. All spatial and statistical procedures were conducted using QGIS 3.x and Python with scikit-learn, XGBoost, LightGBM, GeoPandas, Rasterio, and SHAP, with a fixed random seed of 42 to support reproducibility.

$$Z_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \tag{10}$$

$$C_i = \sum_{j=1}^{m} w_j Z_{ij}, \qquad \sum_{j=1}^{m} w_j = 1 \tag{11}$$

$$\rho_s = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(Q_i - \bar{Q})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(Q_i - \bar{Q})^2}} \tag{12}$$

$$CRIPI_i = 0.40S_i + 0.35C_i + 0.25E_i \tag{13}$$

Results and Discussion

Note: The numerical results in this section are a coherent simulated analytical scenario developed to demonstrate the proposed modelling framework. They must be replaced or verified with actual outputs before empirical submission.

Descriptive Characteristics of Flood and Geospatial Indicators

The final analytical dataset consisted of 12,000 spatial observations, comprising 6,000 flooded cells and 6,000 non-flooded cells after duplicate removal, spatial verification, and application of the 100 m exclusion buffer. A total of 8,400 observations (70%) were allocated to model development, while 3,600 observations (30%) were retained as an independent test dataset. Flooded observations were predominantly concentrated in the northern and northeastern parts of Semarang City, particularly in low-lying and densely developed areas associated with Genuk, Semarang Utara, Gayamsari, Pedurungan, and Tugu. In contrast, non-flooded observations were more frequently distributed across the elevated southern areas, where slopes were steeper, vegetation cover was greater, and impervious development was relatively lower.

Table 5. Descriptive Statistics of Continuous Geospatial Indicators

Indicator	Flooded Cells Mean ± SD	Non-Flooded Cells Mean ± SD	Minimum	Maximum
Maximum rainfall intensity (mm/hour)	41.82 ± 12.47	31.36 ± 10.18	8.40	87.60
72-hour antecedent rainfall (mm)	186.54 ± 54.21	142.73 ± 47.66	41.20	328.40
Distance to river (m)	287.36 ± 214.51	612.48 ± 386.27	4.20	2,184.70
Drainage density (km/km²)	3.84 ± 1.23	2.71 ± 1.08	0.42	7.91
Elevation (m)	8.94 ± 9.71	47.63 ± 51.38	0.80	348.60
Slope (degrees)	2.83 ± 2.46	9.62 ± 8.21	0.04	39.74
Flow accumulation (cells)	1,284.73 ± 836.42	412.56 ± 365.81	1	5,864
TWI	9.24 ± 1.73	6.78 ± 1.61	2.36	14.81
NDVI	0.21 ± 0.13	0.38 ± 0.17	−0.12	0.79

Indicator	Flooded Cells Mean ± SD	Non-Flooded Cells Mean ± SD	Minimum	Maximum
Impervious surface (%)	72.64 ± 17.83	48.31 ± 21.72	3.40	98.70
Building density (buildings/ha)	31.87 ± 11.42	19.63 ± 10.36	0.80	58.40
Road density (km/km²)	12.36 ± 4.81	8.72 ± 4.19	0.94	24.76
Distance to drainage (m)	143.82 ± 106.73	287.46 ± 194.38	2.70	1,246.30

Source: *Authors' simulated analysis based on the proposed integration of SEMARIsK BPBD Kota Semarang, BMKG/NASA GPM, DEMNAS-BIG, Sentinel-2, and DPU Kota Semarang datasets.*

Flooded locations were characterized by substantially lower elevation and slope but higher TWI and flow accumulation. These differences indicate that terrain configuration plays an important role in controlling the concentration and persistence of surface water. Flooded cells also exhibited considerably higher impervious-surface ratios and building densities, indicating that urban development intensified the hydrological effects of rainfall. Zhou et al. (2024) similarly demonstrated that impervious-surface expansion and dense building configurations can increase urban flood vulnerability.

Vegetation showed the opposite pattern, with lower mean NDVI values in flooded areas. Reduced vegetation can decrease interception and infiltration while increasing rapid surface runoff. Gulshad et al. (2024) also identified vegetation condition, digital elevation, soil characteristics, and urban drainage as important determinants of flood susceptibility. Collectively, these descriptive patterns indicate that urban flooding in Semarang cannot be explained by rainfall alone but emerges from the interaction of hydroclimatic, topographic, land-surface, and built-environment conditions.

Machine-Learning Model Performance

All five predictive models were able to discriminate between flooded and non-flooded observations, although substantial differences were observed in their predictive performance. XGBoost achieved the strongest independent-test performance, with an accuracy of 0.908, recall of 0.919, F1-score of 0.910, and AUC of 0.962. LightGBM ranked second with an AUC of 0.956, followed by Random Forest with 0.947. Support Vector Machine and Logistic Regression achieved lower AUC values of 0.902 and 0.873, respectively.

Table 6. Independent-Test Performance of Machine-Learning Models

Model	Accuracy	Precision	Recall	Specificity	F1-Score	AUC	Brier Score
Logistic Regression	0.812	0.804	0.826	0.798	0.815	0.873	0.142
Support Vector Machine	0.842	0.835	0.855	0.829	0.845	0.902	0.121
Random Forest	0.891	0.883	0.904	0.878	0.893	0.947	0.089
XGBoost	0.908	0.901	0.919	0.897	0.910	0.962	0.076
LightGBM	0.901	0.895	0.911	0.891	0.903	0.956	0.081

Source: *Authors' simulated machine-learning analysis using the proposed 70:30 development-test partition and repeated spatial block cross-validation.*

XGBoost demonstrated the strongest combination of discrimination and calibration, as indicated by both the highest AUC and the lowest Brier Score. The consistently stronger results of Random Forest, XGBoost, and LightGBM indicate that ensemble algorithms were better able to represent nonlinear relationships among environmental and urban predictors. Waleed and Sajjad (2025) similarly reported strong performance from XGBoost and LightGBM in comparative flood-susceptibility modelling.

The performance difference also indicates that increasing algorithmic complexity does not automatically guarantee better results. Model superiority should instead be established through independent validation. In this study, XGBoost maintained strong performance under spatial cross-validation and on the independent test dataset. The results therefore support H4, which proposed that ensemble machine-learning models provide higher predictive performance than conventional single-model approaches.

Contribution of Geospatial Indicator Groups

The nested predictor analysis showed a progressive increase in predictive performance as additional geospatial indicator groups were incorporated. The hydroclimatic-only model achieved an AUC of 0.844. Adding topographic indicators increased the AUC to 0.907, representing an improvement of 0.063. Incorporating land-surface indicators further increased the AUC to 0.936, while the complete model containing built-environment indicators achieved the strongest performance, with an AUC of 0.962.

Table 7. Incremental Performance of Geospatial Predictor Groups

Model	Predictor Groups	AUC	F1-Score	Brier Score	ΔAUC
M1	Hydroclimatic	0.844	0.786	0.159	–
M2	Hydroclimatic + Topographic	0.907	0.846	0.124	+0.063
M3	Hydroclimatic + Topographic + Land Surface	0.936	0.878	0.101	+0.029
M4	Hydroclimatic + Topographic + Land Surface + Built Environment	0.962	0.910	0.076	+0.026

Source: *Authors' simulated nested predictor-group analysis using the XGBoost modelling framework.*

The improvement between M1 and M2 demonstrates that rainfall and hydrological characteristics become considerably more informative when terrain characteristics are included. Elevation, slope, flow accumulation, and TWI describe the physical pathways through which rainfall is transformed into localized inundation. Amiri et al. (2024) similarly found that precipitation, elevation, and slope contributed substantially to flood susceptibility. The AUC improvement of 0.063, with an estimated 95% confidence interval of 0.052–0.074, supports H2.

The introduction of land-surface and built-environment variables increased AUC from 0.907 to 0.962. The total improvement of 0.055 indicates that vegetation, imperviousness, buildings, roads, and engineered drainage provide predictive information beyond rainfall and topography. This result supports H3 and demonstrates the importance of explicitly representing urbanization in urban flood models.

The complete geospatial model also substantially outperformed the label-permutation benchmark, which achieved an AUC close to random classification at 0.501. The substantial difference between the complete model and the null benchmark supports H1, confirming that the integrated set of geospatial indicators contains meaningful information for predicting urban flood susceptibility.

Predictor-Group Ablation Analysis

Ablation analysis confirmed the contribution of each predictor group by measuring model performance after removing one group at a time. Removing hydroclimatic indicators produced the largest decline, reducing AUC from 0.962 to 0.918. Removal of topographic indicators reduced AUC to 0.927, while exclusion of land-surface and built-environment predictors reduced AUC to 0.944 and 0.949, respectively.

Table 8. Predictor-Group Ablation Analysis

Predictor Group Removed	AUC After Removal	Change from Full Model	Interpretation
Hydroclimatic	0.918	−0.044	Very strong contribution
Topographic	0.927	−0.035	Strong contribution
Land surface	0.944	−0.018	Moderate additional contribution
Built environment	0.949	−0.013	Meaningful additional contribution

Predictor Group Removed	AUC After Removal	Change from Full Model	Interpretation
None: Full model	0.962	–	Best-performing configuration

Source: *Authors' simulated ablation analysis of the final XGBoost model.*

Hydroclimatic variables represented the largest individual predictor-group contribution, which is reasonable because rainfall provides the primary water input responsible for urban flooding. Nevertheless, removing topographic, land-surface, or built-environment indicators also reduced predictive performance. The results therefore demonstrate that no single indicator group adequately explains urban flood susceptibility. Instead, susceptibility reflects interactions between climatic forcing, terrain configuration, land-surface characteristics, and urban development.

Explainability and Dominant Flood-Susceptibility Factors

SHAP analysis identified elevation as the most influential individual predictor, followed by 72-hour antecedent rainfall, TWI, impervious-surface ratio, and distance to river. Building density, distance to drainage, and NDVI also made meaningful contributions to model predictions.

Table 9. Global SHAP Importance of the Final XGBoost Model

Rank	Predictor	Mean Absolute SHAP Value	Main Relationship with Flood Susceptibility
1	Elevation	0.182	Lower elevation increased susceptibility
2	72-hour antecedent rainfall	0.158	Higher rainfall increased susceptibility
3	TWI	0.131	Higher wetness increased susceptibility
4	Impervious surface ratio	0.114	Greater imperviousness increased susceptibility
5	Distance to river	0.096	Shorter distance increased susceptibility
6	Building density	0.083	Greater density increased susceptibility
7	Distance to drainage	0.067	Greater distance generally increased susceptibility
8	NDVI	0.054	Greater vegetation generally reduced susceptibility

Source: *Authors' simulated SHAP analysis of the final XGBoost model.*

Elevation emerged as the dominant predictor because the northern part of Semarang contains extensive low-lying areas where gravitational drainage is relatively limited and surface-water accumulation is more likely. Singha et al. (2024) similarly identified elevation as one of the most important factors in machine-learning-based flood assessment.

The importance of 72-hour antecedent rainfall indicates that flood susceptibility depends not only on rainfall intensity during an individual event but also on preceding hydrometeorological conditions. Accumulated rainfall can reduce available storage and increase drainage-system loading before the peak rainfall event occurs. TWI further captured areas where terrain characteristics favour water accumulation.

Impervious-surface ratio and building density were among the strongest anthropogenic predictors. Their importance confirms that urban development modifies the relationship between rainfall and surface runoff. The result also explains why the inclusion of built-environment indicators improved the predictive performance of the complete model.

Spatial Distribution of Urban Flood Susceptibility

The final XGBoost model classified 18.4% of Semarang City as very low susceptibility, 24.7% as low, 25.1% as moderate, 19.3% as high, and 12.5% as very high. Combined high and very high susceptibility covered approximately 118.84 km², corresponding to 31.8% of the total study area.

Table 10. Spatial Distribution of Urban Flood Susceptibility in Semarang City

Susceptibility Class	Area (km^2)	Percentage
Very Low	68.76	18.4
Low	92.30	24.7
Moderate	93.80	25.1
High	72.12	19.3
Very High	46.72	12.5
Total	373.70	100.0

Source: *Authors' simulated spatial classification of XGBoost flood probabilities using Jenks Natural Breaks.*

High and very high susceptibility zones were concentrated primarily across Genuk, Semarang Utara, Gayamsari, Pedurungan, and parts of Tugu. In contrast, lower susceptibility occurred more frequently in elevated southern areas. This spatial differentiation reflects the strong influence of elevation, terrain wetness, rainfall accumulation, impervious surfaces, and urban density.

Independent spatial validation showed that 88.6% of observed flooded test cells were located within high or very high susceptibility classes. This correspondence indicates that the final map captured the general spatial pattern of observed flooding rather than merely generating strong statistical performance. Xing et al. (2025) similarly demonstrated that fine-scale machine-learning analysis can identify pronounced spatial heterogeneity in urban flood susceptibility.

Infrastructure Exposure to Flood Susceptibility

Overlay analysis indicated that several critical infrastructure systems intersected high and very high flood-susceptibility zones. Approximately 113.4 km of major roads and 101.2 km of primary and secondary drainage channels were located within these two classes. The same zones also contained 35 health facilities, 89 schools, 12 emergency facilities, and 10 major transport nodes.

Table 11. Infrastructure Exposure Across Urban Flood Susceptibility Classes

Infrastructure Type	Very Low	Low	Moderate	High	Very High
Major roads (km)	58.4	72.7	81.3	69.8	43.6
Drainage network (km)	45.2	63.1	74.8	61.7	39.5
Health facilities (n)	9	13	18	21	14
Schools (n)	24	31	44	53	36
Emergency facilities (n)	3	4	5	7	5
Major transport nodes (n)	2	3	4	6	4

Source: *Authors' simulated infrastructure overlay analysis based on the proposed DPU Kota Semarang and municipal critical-facility datasets.*

The exposure of major roads is particularly important because flooding on strategic road segments can create indirect disruption by reducing accessibility to workplaces, hospitals, schools, and emergency services. Rezvani et al. (2024) showed that integrating geospatial flood assessment with road-network information can reveal infrastructure vulnerabilities that are not visible from flood mapping alone.

Critical-facility exposure also demonstrates why susceptibility should not be interpreted only according to inundated area. A relatively small highly susceptible area may become a major resilience priority when it contains infrastructure essential to public health, emergency response, or urban mobility.

Climate-Resilient Infrastructure Priority

Spearman's rank analysis demonstrated a strong positive association between flood susceptibility and independently calculated infrastructure adaptation need, with $\rho_s = 0.68$ and $p < 0.001$. The result indicates that areas with greater predicted flood susceptibility tended to coincide with areas containing more critical or exposed infrastructure. The statistically significant positive relationship supports H5.

After the independent relationship was evaluated, flood susceptibility, infrastructure criticality, and infrastructure exposure were integrated into the Climate-Resilient Infrastructure Priority Index. The resulting CRIPI classified 18.5% of Semarang City as high priority and 11.0% as very high priority. Together, these classes covered approximately 110.24 km², representing 29.5% of the study area.

Table 12. Climate-Resilient Infrastructure Priority Index in Semarang City

CRIPI Class	Area (km²)	Percentage	Planning Interpretation
Very Low	75.11	20.1	Routine infrastructure management
Low	87.45	23.4	Periodic flood monitoring
Moderate	100.90	27.0	Preventive adaptation recommended
High	69.13	18.5	Priority infrastructure strengthening
Very High	41.11	11.0	Immediate detailed engineering and resilience assessment
Total	373.70	100.0	–

Source: *Authors' simulated CRIPI analysis based on integrated flood susceptibility, infrastructure criticality, and infrastructure exposure.*

The priority classification should be interpreted as a screening mechanism rather than an automatic prescription for engineering intervention. Areas where high susceptibility is primarily associated with impervious surfaces and limited vegetation may be suitable for infiltration enhancement, detention systems, permeable surfaces, and blue-green infrastructure. Ambily et al. (2024) emphasized that blue-green infrastructure can provide an effective component of urban pluvial flood-resilience planning.

In locations where susceptibility is more strongly associated with drainage limitations, interventions may instead require rehabilitation of drainage channels, capacity improvement, pumping-system assessment, detention storage, or more intensive maintenance. Li and Burian (2023) showed that adaptive drainage management combined with green stormwater infrastructure can improve resilience under future rainfall conditions. The CRIPI therefore provides a spatial basis for determining where detailed engineering, financial, and social assessments should be prioritized.

Hypothesis Testing Results

The integrated predictive, explainability, and spatial analyses supported all five hypotheses in the simulated analytical scenario.

Table 13. Summary of Hypothesis Testing Results

Hypothesis	Main Result	Decision
H1	Full model AUC = 0.962 versus permutation benchmark AUC = 0.501; $p < 0.001$	Supported
H2	ΔAUC = 0.063; 95% CI = 0.052–0.074	Supported
H3	Combined ΔAUC = 0.055; 95% CI = 0.043–0.067	Supported
H4	XGBoost AUC = 0.962 versus SVM AUC = 0.902; ΔAUC = 0.060; 95% CI = 0.047–0.073	Supported
H5	$\rho_s = 0.68$; $p < 0.001$	Supported

Source: *Authors' simulated statistical, machine-learning, SHAP, and spatial analyses based on the proposed research framework.*

Overall, the findings indicate that urban flood susceptibility in Semarang is best explained through the integration of natural and anthropogenic processes. Hydroclimatic variables represented the primary forcing mechanisms, while terrain characteristics determined where runoff accumulated. Land-surface and built-environment variables provided additional information by capturing the hydrological consequences of urbanization. The strong performance of XGBoost further demonstrates the usefulness of nonlinear ensemble modelling for integrating these multidimensional relationships.

The principal contribution of the study extends beyond predictive accuracy. By combining machine-learning-based susceptibility assessment with explainability, infrastructure criticality, and infrastructure exposure, the analysis provides a direct link between flood prediction and climate-resilient infrastructure planning. The framework identifies not only where urban flooding is more likely to occur but also where infrastructure adaptation should receive greater priority, thereby transforming flood susceptibility modelling into a more decision-oriented tool for urban resilience.

Conclusion

Within the simulated analytical scenario, urban flood susceptibility in Semarang City was shaped by the combined influence of hydroclimatic, topographic, land-surface, and built-environment conditions. Among the evaluated algorithms, XGBoost produced the strongest predictive performance, with an AUC of 0.962, outperforming Logistic Regression, Support Vector Machine, Random Forest, and LightGBM. Elevation, 72-hour antecedent rainfall, Topographic Wetness Index, impervious-surface ratio, and distance to rivers emerged as the most influential predictors. The progressive improvement in model performance after incorporating topographic, land-surface, and built-environment indicators confirms that urban flooding cannot be explained by rainfall alone. Approximately 31.8% of the study area was classified as having high or very high flood susceptibility, with the most vulnerable areas concentrated predominantly across the low-lying and densely developed northern and northeastern parts of Semarang.

The integration of flood susceptibility with infrastructure criticality and exposure further demonstrates the practical value of machine learning for climate-resilient infrastructure planning. The Climate-Resilient Infrastructure Priority Index identified approximately 29.5% of Semarang City as requiring high or very high adaptation priority, particularly where flood-prone areas intersect with major roads, drainage networks, health facilities, schools, emergency services, and transportation nodes. These findings indicate that interpretable machine learning can function not only as a flood-prediction tool but also as a spatial decision-support framework for prioritizing drainage improvement, infrastructure strengthening, blue-green infrastructure, land-use management, and detailed engineering assessment. Future applications should replace the simulated values with verified empirical outputs, incorporate updated flood inventories and higher-resolution infrastructure data, and integrate future climate scenarios to strengthen the long-term reliability of urban resilience planning.

Acknowledgements

The authors gratefully acknowledge the institutions and agencies whose geospatial, meteorological, disaster, and infrastructure data support the proposed study framework, particularly BPBD Kota Semarang, BMKG, Badan Informasi Geospasial, Dinas Pekerjaan Umum Kota Semarang, NASA GPM, and the Copernicus Earth Observation Programme. The authors also acknowledge the availability of open geospatial and remote-sensing resources that facilitate the integration of flood observations, terrain characteristics, land-surface indicators, and infrastructure information. The contributions of all parties who support data preparation, spatial analysis, and methodological development are sincerely appreciated.

Author Contributions

Conceptualization: First Author.

Data curation: First Author.

Formal analysis: First Author.

Investigation: First Author and Second Author.

Methodology: First Author.

Project administration: Second Author.
Supervision: Second Author.
Validation: Second Author.
Visualization: First Author.
Writing - original draft: First Author.
Writing - review and editing: First Author and Second Author.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted without external financial support, and the authors retained full independence in the design, analysis, interpretation, and preparation of the manuscript.

Conflict of Interest

The authors declare that they have no financial, professional, personal, or other conflicts of interest that could have influenced the conduct, interpretation, or reporting of this research.

Statement on the Use of Artificial Intelligence

The authors used ChatGPT (OpenAI) and Claude AI (Anthropic) as AI-assisted tools to support language refinement, manuscript organization, improvement of academic clarity, and assistance in structuring selected methodological and explanatory sections of the manuscript. The AI tools were not used as authors and did not independently determine the research findings or conclusions. All AI-assisted output was reviewed, edited, and verified by the authors, including the accuracy of scientific statements, methodological descriptions, citations, references, and final interpretation. The authors take full responsibility for the originality, accuracy, research integrity, confidentiality, ethical compliance, and final content of the manuscript.

References

Adnan, M. S. G., Siam, Z. S., Kabir, I., Kabir, Z., Ahmed, M. R., Hassan, Q. K., Rahman, R. M., & Dewan, A. (2023). A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction. Journal of Environmental Management, 326, 116813. https://doi.org/10.1016/j.jenvman.2022.116813

Ahmad, S., Peng, X., Ashraf, A., Yin, D., Chen, Z., Ahmed, R., Israr, M., & Jia, H. (2025). Building resilient urban drainage systems by integrated flood risk index for evidence-based planning. Journal of Environmental Management, 374, 124130. https://doi.org/10.1016/j.jenvman.2025.124130

Al-Kindi, K. M., & Alabri, Z. (2024). Investigating the role of the key conditioning factors in flood susceptibility mapping through machine learning approaches. Earth Systems and Environment, 8, 63–81. https://doi.org/10.1007/s41748-023-00369-7

Ambily, P., Chithra, N. R., & Mohammed Firoz, C. (2024). A framework for urban pluvial flood resilient spatial planning through blue-green infrastructure. International Journal of Disaster Risk Reduction, 103, 104342. https://doi.org/10.1016/j.ijdrr.2024.104342

Amiri, A., Soltani, K., Ebtehaj, I., & Bonakdari, H. (2024). A novel machine learning tool for current and future flood susceptibility mapping by integrating remote sensing and geographic information systems. Journal of Hydrology, 632, 130936. https://doi.org/10.1016/j.jhydrol.2024.130936

Anik, M. S. B. M., An, C., & Li, S. S. (2025). Evolution from the physical process-based approaches to machine learning approaches to predicting urban floods: A literature review. Environmental Systems Research, 14, 15. https://doi.org/10.1186/s40068-025-00409-3

Azadgar, A., Nyka, L., & Salata, S. (2024). Advancing urban flood resilience: A systematic review of urban flood risk mitigation model, research trends, and future directions. Land, 13(12), 2138. https://doi.org/10.3390/land13122138

Chen, Y., Zhang, X., Yang, K., Zeng, S., & Hong, A. (2023). Modeling rules of regional flash flood susceptibility prediction using different machine learning models. Frontiers in Earth Science, 11, 1117004. https://doi.org/10.3389/feart.2023.1117004

Costache, R., Bărbulescu, A., Fanti, R., & Segoni, S. (2025). Flood susceptibility assessment using multi-tier feature selection and ensemble boosting machine learning models. Water, 17(14), 2041. https://doi.org/10.3390/w17142041

Dawson, G., Butt, J., Jones, A., & Fraccaro, P. (2023). Flood susceptibility mapping at the country scale using machine learning approaches. Applied AI Letters, 4(4), e88. https://doi.org/10.1002/ail2.88

Dharmarathne, G., Waduge, A. O., Bogahawaththa, M., Rathnayake, U., & Meddage, D. (2024). Adapting cities to the surge: A comprehensive review of climate-induced urban flooding. Results in Engineering, 22, 102123. https://doi.org/10.1016/j.rineng.2024.102123

Gulshad, K., Yaseen, A., & Szydłowski, M. (2024). From data to decision: Interpretable machine learning for predicting flood susceptibility in Gdańsk, Poland. Remote Sensing, 16(20), 3902. https://doi.org/10.3390/rs16203902

Haghbin, S., & Mahjouri, N. (2023). Quantifying and improving flood resilience of urban drainage systems based on socio-ecological criteria. Journal of Environmental Management, 339, 117799. https://doi.org/10.1016/j.jenvman.2023.117799

Hussain, M., Tayyab, M., Ullah, K., Ullah, S., Rahman, Z. U., Zhang, J., & Al-Shaibah, B. (2023). Development of a new integrated flood resilience model using machine learning with GIS-based multi-criteria decision analysis. Urban Climate, 50, 101589. https://doi.org/10.1016/j.uclim.2023.101589

Islam, T., Zeleke, E. B., Afroz, M., & Melesse, A. M. (2025). A systematic review of urban flood susceptibility mapping: Remote sensing, machine learning, and other modeling approaches. Remote Sensing, 17(3), 524. https://doi.org/10.3390/rs17030524

Kaur, B., Binns, A., McBean, E., Sandink, D., Castro, K., & Gharabaghi, B. (2025). Urban flood susceptibility mapping for Toronto, Canada, using supervised regression and machine learning models. Journal of Flood Risk Management, 18(2), e70051. https://doi.org/10.1111/jfr3.70051

Kurugama, K. M., Kazama, S., Hiraga, Y., & Samarasuriya, C. (2024). A comparative spatial analysis of flood susceptibility mapping using boosting machine learning algorithms in Rathnapura, Sri Lanka. Journal of Flood Risk Management, 17(2), e12980. https://doi.org/10.1111/jfr3.12980

Li, J., & Burian, S. J. (2023). Evaluating real-time control of stormwater drainage network and green stormwater infrastructure for enhancing flooding resilience under future rainfall projections. Resources, Conservation and Recycling, 198, 107123. https://doi.org/10.1016/j.resconrec.2023.107123

Li, J., Strong, C., Wang, J., & Burian, S. (2023). An event-based resilience index to assess the impacts of land imperviousness and climate changes on flooding risks in urban drainage systems. Water, 15(14), 2663. https://doi.org/10.3390/w15142663

Li, Z., Liu, H., Fu, Z., & Fu, G. (2025). Assessment of rainfall-driven urban surface water flood hazards using convolutional neural networks. Journal of Flood Risk Management, 18(3), e70102. https://doi.org/10.1111/jfr3.70102

Liao, Z., He, X., Tian, W., Zhang, Z., Wang, H., & Xie, W. (2025). An integrated assessment of urban flooding risk and resilience based on spatial grids. Urban Water Journal, 22(2), 174–185. https://doi.org/10.1080/1573062X.2024.2443895

Lin, J. L., Sreng, C. S., Oare, E., & Batarseh, F. A. (2023). NeuralFlood: An AI-driven flood susceptibility index. Frontiers in Water, 5, 1291305. https://doi.org/10.3389/frwa.2023.1291305

Long, G., Tantanee, S., Nusit, K., & Sooraksa, P. (2025). A comprehensive review of flood-prone area zonation using ensemble and hybrid machine learning models with a framework proposal for modelling. Geocarto International, 40(1), 2492372. https://doi.org/10.1080/10106049.2025.2492372

Lyu, H.-M., Yin, Z., Shen, S.-L., Chen, X.-S., & Su, D. (2025). A novel framework incorporating machine learning into GIS for flood susceptibility prediction of urban metro systems. Science China Technological Sciences, 68(6), 1620701. https://doi.org/10.1007/s11431-024-2925-8

Manna, H., Das, M., Pramanik, M., Sarkar, S., Mahato, S., Talukdar, S., Alkhuraiji, W. S., & Zhran, M. (2025). Ensemble intelligence for urban resilience: Flood susceptibility modeling in Mumbai using advanced machine learning. Geomatics, Natural Hazards and Risk, 16(1), 2588718. https://doi.org/10.1080/19475705.2025.2588718

Meng, C., & Jin, H. (2023). A comparison of machine learning models for predicting flood susceptibility based on the enhanced NHAND method. Sustainability, 15(20), 14928. https://doi.org/10.3390/su152014928

Qin, X., Wang, S., Meng, M., Long, H., Zhang, H., & Shi, H. (2025). Enhancing urban resilience through machine learning-supported flood risk assessment: Integrating flood susceptibility with building function vulnerability. npj Urban Sustainability, 5, 19. https://doi.org/10.1038/s42949-025-00208-w

Rezvani, S. M. H. S., Silva, M. J. F., & Almeida, N. (2024). Mapping geospatial AI flood risk in national road networks. ISPRS International Journal of Geo-Information, 13(9), 323. https://doi.org/10.3390/ijgi13090323

Rudra, R. R., & Sarkar, S. K. (2023). Artificial neural network for flood susceptibility mapping in Bangladesh. Heliyon, 9(6), e16459. https://doi.org/10.1016/j.heliyon.2023.e16459

Salvati, A., Nia, A. M., Salajegheh, A., Ghaderi, K., Asl, D. T., Al-Ansari, N., Solaimani, F., & Clague, J. J. (2023). Flood susceptibility mapping using support vector regression and hyper-parameter optimization. Journal of Flood Risk Management, 16(4), e12920. https://doi.org/10.1111/jfr3.12920

Sampurno, J., Putra, M. G. E., Faryuni, I. D., & Adriat, R. (2024). Flood impact assessment in remote areas using machine learning, SAR, and GIS: A case study of Ngabang District, Indonesia. Journal of Hydroinformatics, 26(11), 2928–2938. https://doi.org/10.2166/hydro.2024.324

Shirzadi, A., Salvati, A., Tahan, M. H., Shahabi, H., Nodoushan, E. J., Ramezani, M., Hashim, M., & Clague, J. J. (2025). Urban flood susceptibility mapping using deep and machine learning algorithms as a management tool: A case study of Sanandaj City, Iran. Ecological Indicators, 178, 113886. https://doi.org/10.1016/j.ecolind.2025.113886

Singha, C., Rana, V. K., Abu El-Magd, S. A., Nguyen, D. C., & Łupikasza, E. (2024). Integrating machine learning and geospatial data analysis for comprehensive flood hazard assessment. Environmental Science and Pollution Research, 31, 48497–48522. https://doi.org/10.1007/s11356-024-34286-7

Tayyab, M., Hussain, M., Zhang, J., Ullah, S., Tong, Z., Rahman, Z. U., Al-Aizari, A. R., & Al-Shaibah, B. (2024). Leveraging GIS-based AHP, remote sensing, and machine learning for susceptibility assessment of different flood types in Peshawar, Pakistan. Journal of Environmental Management, 371, 123094. https://doi.org/10.1016/j.jenvman.2024.123094

Uddameri, V., & Hernandez, E. A. (2025). Machine learning for flood resiliency: Current status and unexplored directions. Environments, 12(8), 259. https://doi.org/10.3390/environments12080259

Verdyansyah, A., Chang, Y.-L., Wang, F.-C., Tsai, F., & Lin, T.-H. (2025). Integrating CMIP6 and remote sensing datasets for current and future flood susceptibility projections using machine learning under climate change scenarios in Demak District for future sustainable planning. Sustainability, 17(18), 8188. https://doi.org/10.3390/su17188188

Waleed, M., & Sajjad, M. (2025). Advancing flood susceptibility prediction: A comparative assessment and scalability analysis of machine learning algorithms via artificial intelligence in high-risk regions of Pakistan. Journal of Flood Risk Management, 18(1), e13047. https://doi.org/10.1111/jfr3.13047

Wang, Z., Lyu, H., & Zhang, C. (2024). Urban pluvial flood susceptibility mapping based on a novel explainable machine learning model with synchronous enhancement of fitting capability and explainability. Journal of Hydrology, 642, 131903. https://doi.org/10.1016/j.jhydrol.2024.131903

Wang, J., Sanderson, J. E., Saify Iqbal, S. M., & Woo, W. L. (2025). Accelerated and interpretable flood susceptibility mapping through explainable deep learning with hydrological prior knowledge. Remote Sensing, 17(9), 1540. https://doi.org/10.3390/rs17091540

Widya, L. K., Rezaie, F., Lee, W., Lee, C.-W., Nurwatik, N., & Lee, S. (2024). Flood susceptibility mapping of Cheongju, South Korea based on the integration of environmental factors using various machine learning approaches. Journal of Environmental Management, 364, 121291. https://doi.org/10.1016/j.jenvman.2024.121291

Xing, Z., Lyu, G., Yao, Y., Liu, Z., & Zhang, X. (2025). Fine-grained analysis and mapping of urban flood susceptibility with interpretable machine learning: A case study of Hefei, China. Journal of Hydrology: Regional Studies, 60, 102501. https://doi.org/10.1016/j.ejrh.2025.102501

Zerouali, B., Almaliki, A. H., & Santos, C. A. G. (2025). Flood susceptibility mapping in arid urban areas using SHAP-enhanced stacked ensemble learning: A case study of Jeddah. Journal of Environmental Management, 393, 127128. https://doi.org/10.1016/j.jenvman.2025.127128

Zhou, S., Zhang, D., Wang, M., Liu, Z., Gan, W., Zhao, Z., Xue, S., Müller, B., & Ni, X. (2024). Risk-driven composition decoupling analysis for urban flooding prediction in high-density urban areas using Bayesian-optimized LightGBM. Journal of Cleaner Production, 457, 142286. https://doi.org/10.1016/j.jclepro.2024.142286