Journal of Geoscience and Environment Protection, 2026, 14(11), 43-68
https://www.scirp.org/journal/gep
ISSN Online: 2327-4344
ISSN Print: 2327-4336

Multi-Temporal Sentinel-1 Synthetic Aperture Radar and Machine Learning for Flood Susceptibility Assessment of Urban Roads in Nairobi City County, Kenya

Patrick Kathiari Gitile, Josphat K. Z. Mwatelah

Department of Geomatic Engineering and Geospatial Information Science, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya
Email: pakevi07@gmail.com

How to cite this paper: Gitile, P. K., & Mwatelah, J. K. Z. (2026). Multi-Temporal Sentinel-1 Synthetic Aperture Radar and Machine Learning for Flood Susceptibility Assessment of Urban Roads in Nairobi City County, Kenya. *Journal of Geoscience and Environment Protection, 14,* 43-68.
https://doi.org/10.4236/gep.2026.1411003

Received: September 30, 2026
Accepted: November 7, 2026
Published: November 10, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

Urban flooding increasingly disrupts transport systems in rapidly urbanizing cities, yet road-oriented flood-susceptibility assessments that integrate recurrent flood observations with geographically separated model evaluation remain limited in Nairobi City County, Kenya. This study developed a road-oriented flood-susceptibility framework integrating multi-temporal Sentinel-1 Synthetic Aperture Radar (SAR), Geographic Information Systems (GIS), and supervised machine learning. Fourteen selected flood events from 2017 to 2023 were mapped using pre-event and during-event VV backscatter changes, event-specific statistical thresholds, and permanent-water masking. The resulting event maps were combined to generate a flood-frequency surface ranging from 0 to 14 events, which was converted into a binary modelling response. A road-corridor sampling framework generated 4,372,735 observations, of which 3,076,758 from 10 constituencies were used for training, and 1,295,977 from seven geographically separated constituencies were reserved for testing. Sixteen predictors representing terrain, hydrology, rainfall, land cover, urban development, road characteristics, and background Sentinel-1 radar behaviour were evaluated using Logistic Regression, Random Forest, Gradient Boosting, HistGradientBoosting, and XGBoost. XGBoost achieved a test accuracy of 0.7505, F1-score of 0.7269, balanced accuracy of 0.7493, and ROC-AUC of 0.8257. SHAP analysis identified SAR Mean, SAR Standard Deviation, digital elevation, building density, and rainfall among the most influential predictors. The resulting model classified 7235.952 km of road network into five relative susceptibility classes. Temporal assessment using separately mapped 2024 flood observations identified 944.609 km of flooded road length, of which 423.584 km (44.84%) occurred in

the High and Very High classes and 634.780 km (67.20%) occurred in the Moderate-to-Very-High classes. These findings demonstrate the potential of integrating recurrent SAR observations and machine learning for screening urban roads for further flood investigation. The resulting susceptibility map represents relative spatial differences within the assessed road network rather than calibrated flood probabilities or predictions of flood depth, velocity, or duration.

Keywords

Urban Flooding, Flood Susceptibility, Sentinel-1 SAR, Machine Learning, Road Network, Nairobi City County

1. Introduction

Urban flooding is a major infrastructure and mobility challenge because intense or prolonged rainfall can exceed drainage capacity, generate surface runoff and inundate transport corridors. The consequences extend beyond the flooded locations themselves by disrupting mobility, public transport, emergency response, economic activity and access to essential services. Urban flood occurrence is spatially heterogeneous because rainfall interacts with impervious surfaces, drainage systems, terrain and land-use patterns (Bloch et al., 2012; Fletcher et al., 2015; IPCC, 2022).

Road networks are particularly important targets for flood-susceptibility assessment because they form interconnected infrastructure systems whose disruption can propagate through urban mobility networks. Identifying road sections with recurrent flood susceptibility can support infrastructure agencies in screening locations for drainage assessment, maintenance and resilience planning. Global assessments have demonstrated substantial exposure of transport infrastructure to natural hazards, including flooding (Koks et al., 2019). In Nairobi City County, recurrent urban flooding occurs within a rapidly developing environment where rainfall, terrain, drainage, impervious surfaces and urban form interact (Oluchiri, 2025; Ongaga et al., 2024).

Satellite remote sensing provides an opportunity to characterize flood occurrence repeatedly over large areas. Sentinel-1 Synthetic Aperture Radar (SAR) is particularly useful because C-band observations can be acquired during cloudy conditions and independently of daylight. Multi-temporal SAR permits flood events to be mapped from changes in backscatter and provides a basis for characterizing the normal temporal behaviour of radar backscatter before an event (Torres et al., 2012; Twele et al., 2016; Schlaffer et al., 2015). Sentinel-1 multi-temporal observations can also provide statistical descriptors of background radar behaviour. Temporal mean backscatter provides a measure of the typical radar response of a surface over a reference period, whereas temporal standard deviation describes the variability of that response through time. Such statistics are useful in change-detection frameworks because flood-related observations can be evaluated against

the expected temporal behaviour of the underlying surface rather than against a single reference acquisition (Schlaffer et al., 2015; Tupas et al., 2023). In urban environments, where radar backscatter is influenced by surface roughness, vegetation, buildings and double-bounce interactions, temporal statistics can therefore provide additional information about the radar characteristics of locations that may not be represented adequately by conventional terrain and hydrological predictors (Twele et al., 2016; Oakes et al., 2023). In this study, SAR Mean and SAR Standard Deviation were therefore incorporated as background radar descriptors, calculated from a temporally separate non-event Sentinel-1 reference dataset rather than from the observations used to construct the flood-response variable.

Flood-susceptibility modelling can relate observed flood occurrence to terrain, hydrological, climatic, land-cover, urban and infrastructure characteristics. Machine-learning approaches are capable of representing nonlinear relationships and interactions among predictors and have been widely applied in flood susceptibility and flood prediction studies (Mosavi et al., 2018; Tehrany et al., 2014, 2019).

Despite these developments, much of the existing flood-susceptibility literature focuses on general flood-prone areas rather than road-specific susceptibility, while relatively fewer studies integrate recurrent SAR-derived flood observations with road-oriented sampling and geographically separated machine-learning evaluation. In addition, SAR variables are commonly used either to detect individual flood events or as environmental predictors, but the distinction between event observations used to construct the response and temporally separate observations used to characterize background radar behaviour is not always explicit. This study addresses these methodological gaps by developing a road-oriented modelling framework in which recurrent Sentinel-1 observations are used to construct the flood-response variable, while temporally separate Sentinel-1 observations are used to derive background SAR Mean and SAR Standard Deviation predictors.

However, environmental observations often exhibit spatial and temporal dependence. Randomly partitioning closely spaced observations can therefore place highly similar observations in both training and testing datasets and produce optimistic estimates of generalization (Roberts et al., 2017; Meyer et al., 2019).

This study develops a road-oriented flood-susceptibility framework for Nairobi City County using a multi-temporal Sentinel-1 SAR flood inventory, GIS-derived conditioning factors, geographically separated training and testing data, comparative evaluation of five supervised machine-learning algorithms, SHAP-based model interpretation, and independent temporal validation using 2024 flood observations. The study also evaluates the spatial correspondence of the resulting susceptibility classes with documented historical flood-prone locations and government-documented flood hotspots. The resulting product represents relative susceptibility within the assessed road network and is intended as a screening layer rather than a hydraulic prediction of flood depth, velocity or duration.

2. Study Area and Data

The study was conducted within Nairobi City County, Kenya, focusing specifically on the urban road network. The analytical units were road-corridor sampling observations associated with 50 m road sections. The modelling period covered selected flood events between 2017 and 2023, including events from both the March-May long-rain season and the October-December short-rain season. Events were retained when suitable Sentinel-1 pre-flood and during-flood observations were available and provided sufficient temporal coverage to support event-specific change detection. They should therefore be interpreted as a selected empirical event record rather than a complete inventory of all floods during 2017-2023. The predictor variables, source datasets, native resolutions, and operational derivation settings used in the analysis are summarized in Table 1.

Table 1. Data sources.

Predictor Category	Variables	Primary Source Product	Native Resolution	Derivation Window/Operational Setting
Terrain	DEM, Slope, TPI	JAXA ALOS PALSAR	12.5 m	Filled DEM; Slope in degrees; TPI using a 3 × 3 neighborhood window.
Hydrology	Flow Accumulation, TWI, HAND, Distance to River	Derived from DEM/Vector River Layer	12.5 m	HAND derived following Nobre et al. (2011); Flow Accumulation transformed via ln({FlowAccum} + 1).
Climate	Cumulative seasonal rainfall	CHIRPS (Funk et al., 2015)	Approximately 5 km	Derived by summing the seven MAM seasonal rainfall rasters and seven OND seasonal rainfall rasters for 2017-2023 in ArcGIS Pro. Values were extracted at road-corridor sampling points.
Urban	LULC, NDBI, Building Density, Road Density	ESA WorldCover/ Sentinel-2/Vector Maps	10 m/Vector	NDBI following Zha et al. (2003); Urban densities derived via a 500 m circular search neighborhood.
Road Corridor	Legal Road Width, Buffer Distance	KRB Roads Register/Vector Network	Vector	Width represents gazetted corridor boundaries; Buffer represents perpendicular node offsets.
Baseline Radar	SAR Mean, SAR Standard Deviation	Copernicus Sentinel-1 GRD VV Stack	10 m	Calculated over dry reference subsets (exclusively February and September 2017-2023).

Sixteen predictor variables were used: DEM, Flow Accumulation, Slope, Topographic Position Index (TPI), Topographic Wetness Index (TWI), Height above Nearest Drainage (HAND), Distance to River, Rainfall, Land Use/Land Cover (LULC), Normalized Difference Built-up Index (NDBI), Building Density, Road Density, Legal Road Width, Buffer Distance, SAR Mean, and SAR Standard Deviation. Raster-derived predictors were extracted at sampling points from their source grids; no bilinear interpolation was applied. The terrain and hydrological layers were derived from the JAXA ALOS PALSAR DEM (reported native resolution 12.5 m). CHIRPS rainfall has a native grid of approximately 5 km (Funk et al., 2015). ESA WorldCover provides LULC at 10 m resolution (Zanaga et al.,

2021). Building Density and Road Density were calculated using a 500 m-radius circular neighbourhood. LULC classes were represented as Vegetation, Grassland/Sparse Vegetation, Built-up/Impervious, Bare/Sparse Soil, and Water/Wetlands. The Sentinel-1 GRD VV layers have a nominal 10 m pixel spacing; this does not mean that individual urban roads narrower than a pixel are independently resolved.

SAR Mean and SAR Standard Deviation describe background radar behaviour and were calculated from a separate non-event Sentinel-1 VV reference stack acquired during February and September in 2017-2023. These months were reserved for the background radar reference dataset and were not used as the selected flood-event seasons in the MAM (March-May) and OND (October-December) inventory. Sentinel-1 GRD scenes were processed with orbit correction, thermal-noise removal, radiometric calibration, speckle filtering and terrain correction, then aligned to a common 10 m output grid. A refined Lee filter with a 7 × 7 window and the ALOS PALSAR DEM for terrain correction were used in the reported processing workflow. Pixel-wise temporal mean and standard deviation were calculated from the reference stack to produce one SAR Mean and one SAR Standard Deviation raster.

The overall study area, analytical framework, and sequence of the flood-susceptibility assessment are summarized in **Figure 1**.

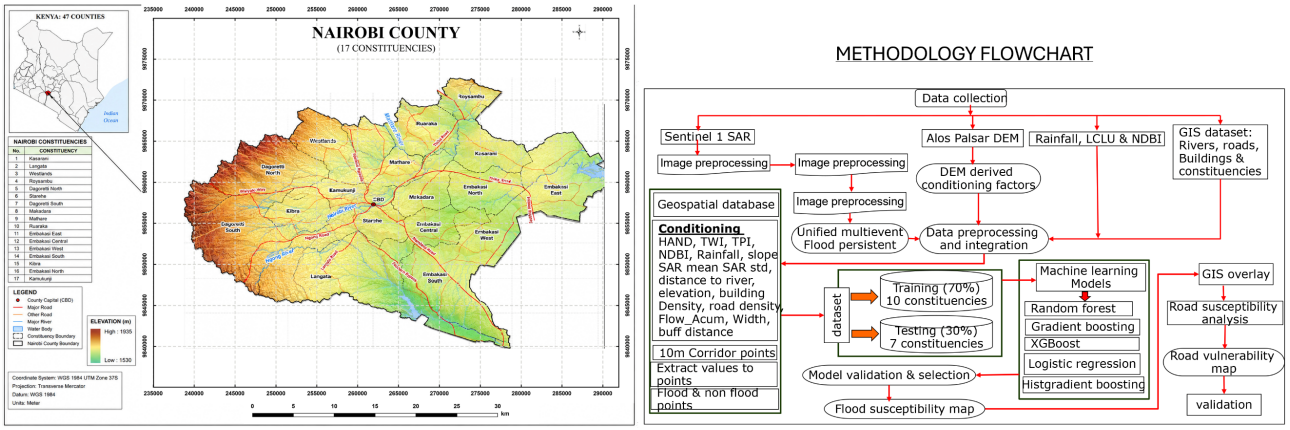


Figure 1. Study area and analytical framework for flood-susceptibility assessment of urban roads in Nairobi City County, Kenya.

3. Methods

3.1. Multi-Temporal Flood Inventory Development

For each of the 14 selected flood events, Sentinel-1 VV observations acquired before and during the flood period were processed using orbit correction, thermal-noise removal, radiometric calibration, speckle filtering and terrain correction. Nairobi City County was covered by multiple Sentinel-1 scenes, as a single scene did not necessarily provide complete spatial coverage of the study area. Therefore, Sentinel-1 scenes acquired on the same date were grouped and mosaicked to produce spatially continuous coverage of the county. Before mosaicking, the scenes

were checked for their acquisition dates, orbit information, polarization, and spatial coverage. Compatible scenes were combined to generate county-wide pre-event and during-event VV backscatter mosaics.

The pre-event and during-event mosaics were then used to calculate the change in VV backscatter in decibels (dB). For each candidate flood event, the backscatter difference was calculated by subtracting the pre-event VV backscatter from the during-event VV backscatter. An event-specific statistical threshold was subsequently applied to identify potential flood-affected areas, after which permanent water bodies were masked. The resulting event maps were combined to derive a flood-frequency surface representing the number of mapped flood events at each pixel.

Spatial coverage and compatibility between the paired mosaics were assessed to minimize gaps and ensure that backscatter comparisons were based on suitable observations. The final flood-frequency surface was used to derive the binary flood response for subsequent road-oriented flood susceptibility modelling.

The pre-flood/reference image used for each event was distinct from the non-event reference-image stack used to derive the SAR Mean and SAR Standard Deviation predictors.

Backscatter values were expressed in decibels (dB). For event *j*, the pixel-wise backscatter change was calculated as the during-flood observation minus the corresponding pre-flood/reference observation:

$$\mathrm{LogDiff}_{i,j} = \sigma^0_{i,j,\mathrm{during}} - \sigma^0_{i,j,\mathrm{pre}} \tag{1}$$

where LogDiff is the event-specific backscatter change; σ^0 (during) is the calibrated VV backscatter during the event; and σ^0 (pre) is the corresponding pre-flood/reference backscatter.

For each event, the distribution of LogDiff was summarized, and a threshold of 1.5 standard deviations below the mean was adopted as an empirical change-detection criterion to identify unusually negative backscatter changes associated with potential inundation. This 1.5-standard-deviation criterion was an operational rule applied consistently across the selected events; it is not a calibrated inundation probability threshold. Consequently, the resulting flood-frequency and binary response variables should be interpreted as recurrence-based empirical labels rather than direct measurements of flood depth, duration or hydraulic severity.

$$T_j = \mu\left(\mathrm{LogDiff}_j\right) - 1.5\sigma\left(\mathrm{LogDiff}_j\right) \tag{2}$$

where T_j is the threshold for event *j*, μ (LogDiff$_j$) is the mean event-specific backscatter change, and σ (LogDiff$_j$) is its standard deviation. Pixels with LogDiff values equal to or below T_j were classified as potentially flooded, whereas pixels above the threshold were classified as non-flooded. The event-specific LogDiff statistics, standard deviations, threshold multipliers, and resulting operational thresholds are presented in Table 2.

Table 2. Event-specific Sentinel-1 LogDiff statistics and processing limits.

S/N	Flood Event/Season Label	Minimum (dB)	Maximum (dB)	Statistical Mean (μ)	Standard Deviation (σ)	Multiplier (k)	Operational Threshold T = μ − 1.5σ
1	Long Rains 2017	−30.346	27.072	0.142	2.214	1.5	−3.1789
2	Long Rains 2018	−19.352	25.511	0.939	1.916	1.5	−1.9353
3	Long Rains 2019	−25.018	36.885	2.685	2.953	1.5	−1.7448
4	Long Rains 2020	−23.374	19.989	0.258	1.616	1.5	−2.1657
5	Long Rains 2021	−23.578	26.508	0.891	1.720	1.5	−1.6881
6	Long Rains 2022	−20.905	21.801	−0.308	1.605	1.5	−2.7153
7	Long Rains 2023	−27.307	25.346	2.601	2.840	1.5	−1.6584
8	Short Rains 2017	−21.902	20.953	2.062	2.347	1.5	−1.4590
9	Short Rains 2018	−26.070	30.459	−0.468	1.981	1.5	−3.4401
10	Short Rains 2019	−19.370	23.201	1.050	1.945	1.5	−1.8668
11	Short Rains 2020	−20.116	22.993	0.111	1.503	1.5	−2.1436
12	Short Rains 2021	−18.669	21.520	−0.784	1.646	1.5	−3.2525
13	Short Rains 2022	−18.143	21.674	−0.386	1.290	1.5	−2.3206
14	Short Rains 2023	−22.931	20.697	0.932	1.703	1.5	−1.6218

The same rule was applied to all 14 selected events. Permanent-water areas were then masked to reduce confusion between recurrent flood inundation and permanent surface water. The use of statistical backscatter-change thresholds is consistent with established Sentinel-1 flood change-detection approaches (Twele et al., 2016; Li et al., 2018; Tupas et al., 2023).

The 14 binary event maps were summed up to obtain a persistent flood-frequency surface. For each pixel *i*, flood frequency was calculated as:

$$F_i = \sum_{j=1}^{14} I_{i,j} \tag{3}$$

where $I_{i,j}$ = 1 if pixel *i* was classified as potentially flooded during event *j* and 0 otherwise.

The resulting frequency surface ranged from 0 to 14. Pixels affected during 0 - 2 selected events were assigned to the non-flood class (0), whereas pixels affected during 3 - 14 events were assigned to the flood class (1). The three-event cut-off was used as an operational recurrence definition for supervised modelling rather than as a calibrated probability threshold. No claim is made that the cut-off represents a universal definition of flood susceptibility. The resulting persistent flood-frequency surface for the 14 selected Sentinel-1 flood events is shown in Figure 2.

3.2. Predictor-Variable Preparation and Road-Corridor Sampling

3.2.1. Rainfall Predictor Preparation

Seasonal rainfall data for 2017-2023 were used to develop the rainfall predictor.

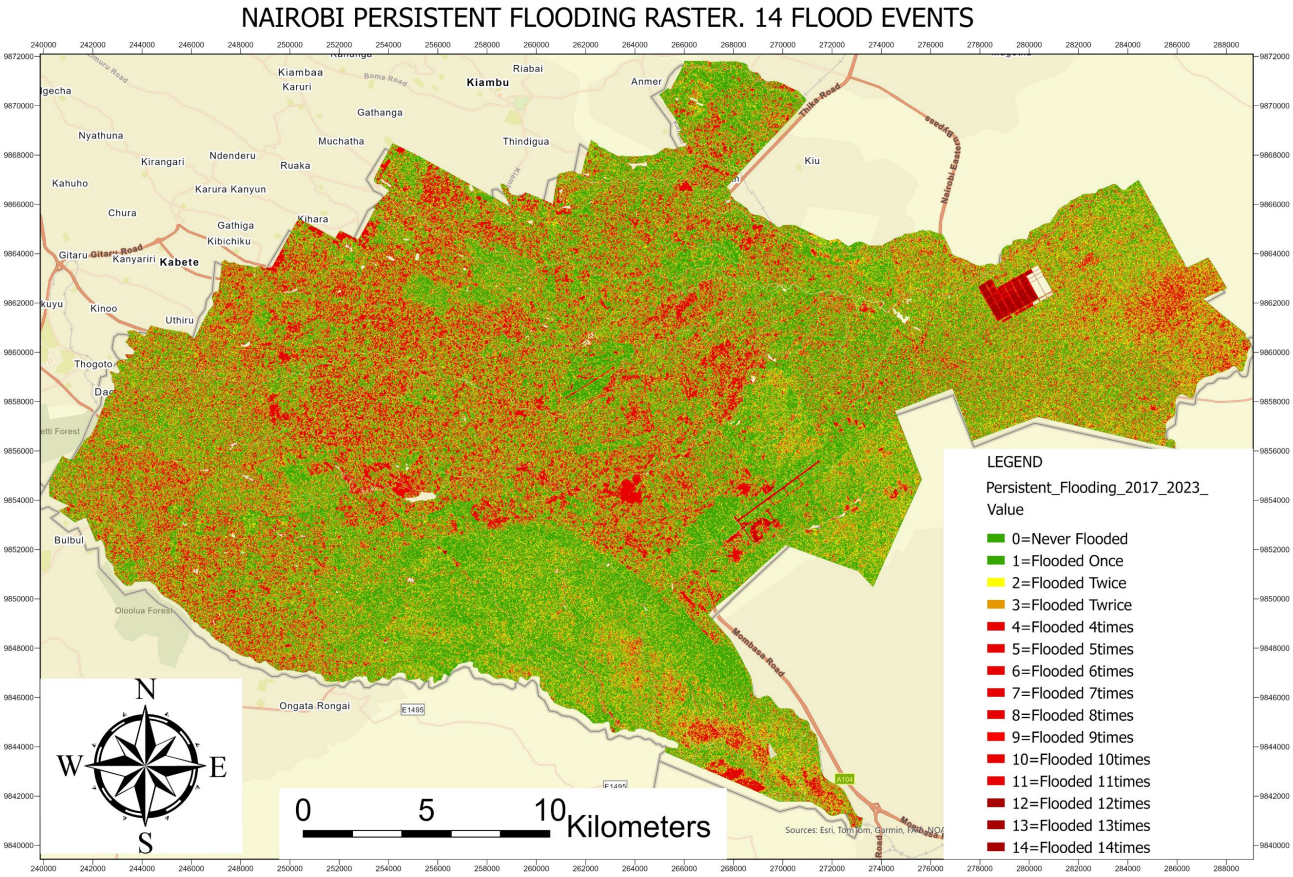


Figure 2. Persistent flood-frequency surface derived from the 14 selected Sentinel-1 SAR flood events, 2017-2023. The colour gradient represents increasing flood recurrence frequency.

The dataset comprised 14 seasonal rainfall rasters: seven representing the March-May long-rain seasons and seven representing the October-December short-rain seasons. In ArcGIS Pro, the seven long-rainfall rasters and seven short-rainfall rasters were combined by summing their rainfall totals to produce a cumulative seasonal rainfall surface for the study period. Rainfall values were subsequently extracted from the resulting raster at the road-corridor sampling points and incorporated into the predictor dataset for supervised machine-learning model development. The resulting predictor represents cumulative seasonal rainfall across the 14 seasons and is distinct from the three-day and seven-day antecedent rainfall totals used in the Sentinel-1 flood-event assessment and verification workflow.

3.2.2. Road-Corridor Sampling Procedure

Road-corridor sampling points were generated along road centerlines at 10 m intervals. At each centerline node, a perpendicular transect was created across the associated road-corridor extent, and points were placed along the transect at 10 m spacing. The transect extent was defined using the available Legal Road Width attribute, which represents the mapped/gazetted road-reserve width where that attribute was available; it should not be interpreted as the paved carriageway width. Buffer Distance records the perpendicular offset of a sampled point from the road

centerline and was retained as a road-corridor position variable. The point-generation process produced 4,372,735 observations.

Raster values were extracted at each point using the source-cell value at that location, without raster interpolation. Therefore, different points can share the same predictor values when they fall within the same raster cell; this is expected and does not create additional independent sensor information. Point coordinates and vector-derived road-corridor attributes were retained. The resulting observations are spatially clustered and should not be interpreted as statistically independent samples.

To control for spatial dependence and autocorrelation, the point features were partitioned by administrative constituency borders before model execution. The 4,372,735 total observations were allocated into geographically distinct training and testing zone blocks:

- Training Constituency Block (N = 3,076,758 points): Kasarani, Langata, Roysambu, Dagoretti South, Embakasi East, Embakasi Central, Makadara, Kamukunji, Ruaraka, and Mathare.
- Geographically Held-Out Testing Block (N = 1,295,977 points): Westlands, Starehe, Dagoretti North, Embakasi South, Kibra, Embakasi North, and Embakasi West.

The five machine-learning models were trained at the point level. After prediction, point-level class-1 scores were assigned to the associated 50 m road sections. For each section, the arithmetic mean of the available point scores within its geometry was calculated, and the resulting continuous score was grouped into five relative susceptibility classes using Jenks Natural Breaks.

3.3. Label Generation Parameter Sensitivity Analysis

To evaluate the stability and structural robustness of our heuristic labeling framework, a systematic sensitivity analysis was conducted by varying two core data-engineering parameters around their operational baselines: the event-specific SAR change-detection multiplier (k) used to isolate flooded pixels, and the multi-event recurrence cut-off value used to define the binary target variable.

The baseline processing pipeline utilized an event threshold of $\mu - 1.5\sigma$ combined with a recurrence cut-off of ≥3 events to categorize a sampling observation as flooded. We systematically evaluated alternative threshold settings using a grid configuration of SAR threshold multipliers (k = 1.0, 1.5, and 2.0) and recurrence cut-offs of ≥2, ≥3, and ≥4 events. For each unique parameter scenario, we tracked the resulting response-class prevalence within the training database alongside the geographically held-out predictive performance (ROC-AUC) of the XGBoost model. The sensitivity analysis was used to assess the stability of the response labels and to determine whether reasonable changes in the event-detection threshold and recurrence definition materially altered model discrimination.

The sensitivity scenarios indicate that the positive response-class prevalence and held-out ROC-AUC varied with the event threshold and recurrence definition. Across the seven scenarios reported in **Table 3**, ROC-AUC ranged from 0.8113

Table 3. Label generation parameter sensitivity analysis matrix.

Scenario ID	SAR Threshold Multiplier (k)	Recurrence Cut-Off Event Count	Response Class Prevalence (Flooded %)	Held-Out Model Performance (ROC-AUC)	Spatial Footprint Characterization
Relaxed 1	1.0	≥2	54.82%	0.8142	Captures minor ephemeral ponding; prone to false positives from surface moisture.
Relaxed 2	1.0	≥3	49.15%	0.8190	Intermediate threshold; captures high-frequency dampness variations.
Moderate 1	1.5	≥2	48.33%	0.8215	Lowered recurrence threshold; includes singular extreme inundations.
Baseline	1.5	≥3	43.49%	0.8257	Optimal trade-off; isolates structurally recurrent road-flooding zones.
Moderate 2	1.5	≥4	37.12%	0.8231	Strict recurrence; restricts target class to chronic low-lying channels.
Strict 1	2.0	≥3	35.04%	0.8184	High radar-change constraint; filters out shallow overbank overflows.
Strict 2	2.0	≥4	29.65%	0.8113	Highly restricted target profile; risks omitting critical pluvial flash zones.

to 0.8257. The baseline (k = 1.5; recurrence ≥3 events) was retained as the operational specification for the main analysis. These results show sensitivity to label-generation choices and do not establish that the baseline is universally optimal; the selection should be understood as a study-specific modelling decision.

3.4. Machine-Learning Modelling and Evaluation

Five supervised classification algorithms were evaluated: Logistic Regression, Random Forest, Gradient Boosting, HistGradientBoosting and XGBoost. Random Forest is based on ensemble decision trees (Breiman, 2001), Gradient Boosting follows the additive boosting framework described by Friedman (2001), and XGBoost implements a scalable gradient-boosted tree framework (Chen & Guestrin, 2016).

Model performance was assessed using accuracy, precision, recall (sensitivity), specificity, F1-score, balanced accuracy and ROC-AUC to characterize overall classification and class-specific discrimination (Sokolova & Lapalme, 2009; Fawcett, 2006; Brodersen et al., 2010).

Continuous predictors were standardized using parameters estimated from the training data. Flow Accumulation was transformed as $\ln(x + 1)$ to reduce right-skewness. LULC was encoded using one-hot encoding. Hyperparameter selection was conducted using grid search with five-fold cross-validation within the training data, and a probability threshold of 0.50 was used for binary predictions. The preprocessing pipeline and selected settings were then held fixed during evaluation on the geographically held-out constituencies.

The model settings are summarized in Table 4. Preprocessing parameters were

Table 4. Reproducible machine learning implementation configurations.

Parameter Domain	XGBoost Settings	Random Forest Settings	HistGradientBoosting Settings	Gradient Boosting Settings	Logistic Regression Settings
Core Base Learners	500 gradient-boosted decision trees	500 parallel bootstrap decision trees	250 successive histogram binned trees	300 sequentially boosted decision trees	Single parametric log-odds model function
Tree Depth Control	max_depth = 6	max_depth = 15	max_leaf_nodes = 31	max_depth = 4	N/A (Linear parameter optimization)
Learning Step Rate	learning_rate = 0.05	N/A	learning_rate = 0.05	learning_rate = 0.05	N/A (No learning step optimization)
Regularization Flags	reg_alpha = 1.0, reg_lambda = 5.0	min_samples_leaf = 4	l2_regularization = 1.5	min_samples_split = 10	L2 Ridge penalty regularization (C = 1.0)
Feature Sub-Sampling	colsample_bytree = 0.8	max_features = "sqrt"	N/A	subsample = 0.8	N/A (All variables evaluated globally)
Data Binning Settings	N/A	N/A	max_bins = 255	N/A	N/A
Optimization Solver	N/A	N/A	N/A	N/A	Lbfgs iterative quasi-Newton solver
Classification Rule	Fixed at 0.50 probability score	Fixed at 0.50 vote tally share	Fixed at 0.50 probability score	Fixed at 0.50 probability score	Fixed at 0.50 probability score

fitted on the training partition only and then applied unchanged to the geographically held-out test partition. The classification threshold was fixed at 0.50 for all models.

The geographically held-out test dataset was used to compare the predefined metrics and select the model used for final susceptibility mapping. Therefore, it was not an entirely untouched dataset for post-selection inference. XGBoost was selected because it achieved the highest observed ROC-AUC (0.8257), accuracy (0.7505) and F1-score (0.7269). Its reported performance should be interpreted within this model-comparison framework rather than as an independently confirmed optimum.

For the final XGBoost model, the reported settings were a learning rate of 0.05, maximum tree depth of 6, subsample ratio of 0.8, and 500 estimators. The model used regularization parameters reg_alpha = 1.0 and reg_lambda = 5.0, with colsample_bytree = 0.8. These settings are reported here consistently with the final model description and should correspond to the configuration used to generate the results.

3.5. Susceptibility Mapping and Model Interpretation

The continuous XGBoost output was treated as a modelled class-1 susceptibility score within the assessed road network. Jenks Natural Breaks was then used to classify the continuous output into five relative classes: Very Low, Low, Moderate, High and Very High. These classes represent relative spatial differences within the

study area and are not calibrated probabilities of flooding, nor do they represent flood depth, velocity, duration or hydraulic connectivity (Jenks & Caspall, 1971). SHAP was used to quantify the contribution of individual predictors to the XGBoost predictions (Lundberg & Lee, 2017).

3.6. Independent and External Validation

Temporal hold-out assessment used mapped 2024 flooded-road observations, which were temporally separated from the 2017-2023 modelling inventory. The 2024 flood extent was represented as a binary raster at 10 m output resolution in WGS 1984 UTM Zone 37S. Flood pixels were converted to polygons without simplification and intersected with the 50 m road-segmentation layer.

The 2024 flood footprint was derived from separate Sentinel-1 imagery using the same log-difference change-detection workflow as the 2017-2023 inventory. Thus, the validation is temporally independent but not fully independent in measurement method. Shared SAR detection limitations—including layover, radar shadow, double-bounce scattering, surface roughness and vegetation effects—may affect both the modelling labels and the 2024 validation layer. The results are therefore interpreted as temporal-spatial consistency evidence, not as validation against independent hydraulic measurements or field-mapped flood truth.

To provide external verification, two point-based spatial datasets were held out from model development:

Historical Flood-Prone Locations (N = 23): Compiled from historical multi-agency response archives and disaster records.

Official Government Hotspots (N = 37): Obtained directly from the Ministry of Interior and National Administration (Nairobi Rivers Regeneration Programme) city diagnostics mapped on March 15, 2026.

These coordinates were overlaid directly onto our final classified road susceptibility maps using spatial point-in-polygon tools to calculate the exact capture rates within each susceptibility zone.

4. Results

4.1. Flood Inventory and Modelling Dataset

The multi-temporal workflow produced a 14-event flood inventory covering selected Sentinel-1-observed flood events between 2017 and 2023. Combining the event maps produced a persistent flood-frequency surface ranging from 0 to 14, which was subsequently converted to the binary response used for machine learning. The road-oriented modelling dataset contained 4,372,735 observations: 3,076,758 for training and 1,295,977 for geographically separated testing.

The flooded and non-flooded class proportions were broadly comparable between training and testing. The training dataset contained 43.49% flooded and 56.51% non-flooded observations, while the testing dataset contained 45.08% flooded and 54.92% non-flooded observations (**Table 5**).

Table 5. Composition of the geographically separated modelling dataset.

Dataset	Constituencies	Observations	Flooded Observations	Non-Flooded Observations	Flood (%)	Non-Flood (%)	Share of Dataset (%)
Training	10	3,076,758	1,338,184	1,738,574	43.49	56.51	70.36
Testing	7	1,295,977	584,285	711,692	45.08	54.92	29.64
Total	17	4,372,735	1,922,469	2,450,266	43.97	56.03	100.00

4.2. Predictor Diagnostics

Exploratory analysis showed substantial variation among the environmental and urban predictors. Flow Accumulation displayed the strongest right-skewness, while Distance to River, HAND, Building Density and Road Density also contained long upper tails. Extreme observations were retained because they represented plausible environmental or urban conditions rather than confirmed data-entry errors. The cumulative seasonal rainfall predictor represented rainfall totals accumulated across the 14 MAM and OND seasons from 2017 to 2023. Its distribution and relationship with the other conditioning factors were examined alongside the terrain, hydrological, urban, road and SAR-derived predictors. Rainfall's contribution to the XGBoost predictions was subsequently examined using SHAP analysis.

Pearson correlation analysis showed the strongest linear relationships among several elevation-derived variables, particularly DEM, HAND, Slope and TPI. Building Density and Road Density were positively associated, reflecting their shared representation of urban development. SAR Mean and SAR Standard Deviation showed weaker linear relationships with the conventional terrain and urban predictors, indicating that the temporal radar descriptors contributed information not fully represented by those variables.

Variance inflation factor analysis indicated that DEM and Rainfall had the highest VIF values, 7.130 and 7.069, respectively, whereas all remaining predictors had VIF values below 2.01. Both variables were retained because they represent distinct dimensions of the flood-susceptibility problem and because the tree-based models can accommodate nonlinear relationships and interactions. VIF was, therefore, treated as a diagnostic of predictor structure rather than as an automatic variable-elimination criterion.

The five supervised classification models were evaluated using the geographically separated testing dataset (Total testing N = 1,295,977 observations; consisting of exactly 584,285 true flooded records and 711,692 true non-flooded records). Table 6 provides the mathematically verified metric distributions derived from the raw confusion matrices.

XGBoost achieved the highest observed values for the principal performance measures, with a true accuracy of 0.7505, balanced accuracy of 0.7493, verified F1-score of 0.7269, and ROC-AUC of 0.8257. The receiver operating characteristic curves for the five evaluated models are shown in Figure 3, illustrating their comparative discrimination performance on the geographically held-out testing dataset.

Table 6. Comparative performance of the five supervised machine-learning algorithms on the geographically separated testing dataset.

Supervised Learning Algorithm	Test Accuracy	Balanced Accuracy	Precision	Recall (Sensitivity)	Specificity	F1-Score	ROC-AUC
XGBoost	0.7505	0.7493	0.7175	0.7366	0.7619	0.7269	0.8257
Gradient Boosting	0.7494	0.7480	0.7163	0.7353	0.7608	0.7257	0.8246
HistGradientBoosting	0.7476	0.7468	0.7119	0.7395	0.7542	0.7255	0.8204
Random Forest	0.7454	0.7438	0.7138	0.7268	0.7607	0.7202	0.8159
Logistic Regression	0.7435	0.7433	0.7049	0.7415	0.7451	0.7229	0.8223

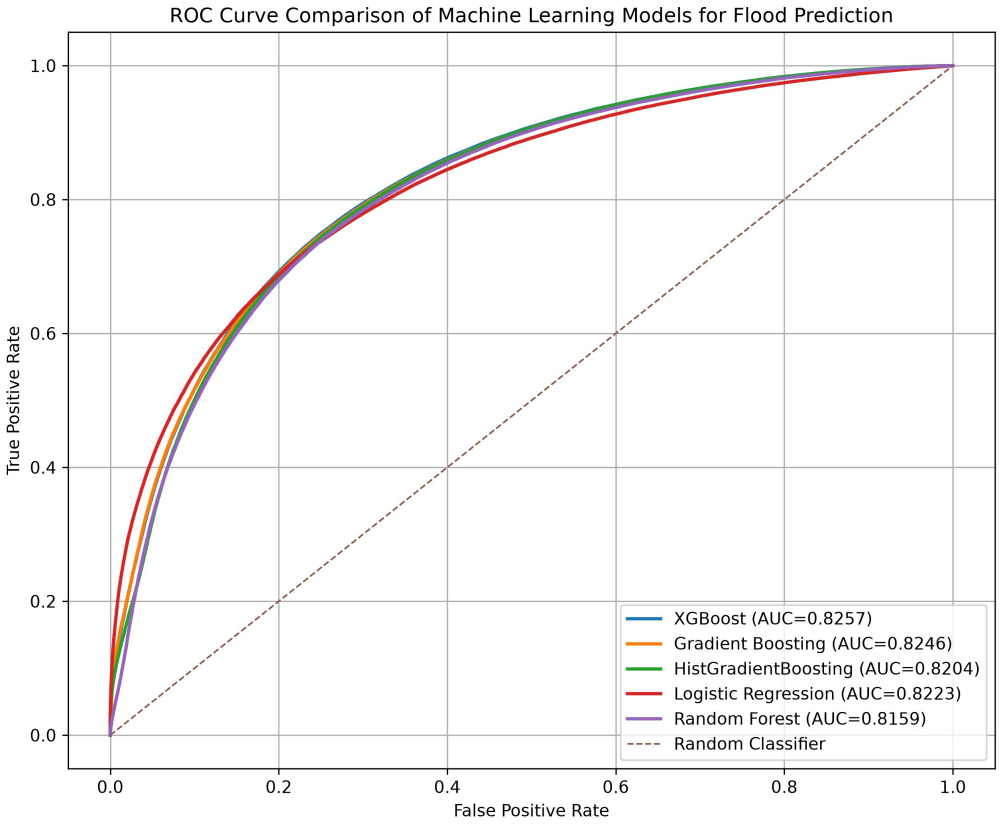


Figure 3. Receiver operating characteristic curves for the five evaluated machine-learning algorithms on the geographically separated testing dataset.

Gradient Boosting produced closely comparable performance, with accuracy of 0.7494 and ROC-AUC of 0.8246. Pairwise McNemar tests indicated differences in paired categorical predictions among the ten model pairs after Bonferroni adjustment (adjusted $p < 0.001$). These tests do not establish that differences in ROC-AUC or other continuous performance metrics are practically large. Given the narrow metric range and spatial clustering of observations, model differences should be interpreted cautiously (McNemar, 1947).

Pairwise McNemar tests identified statistically detectable differences in

paired categorical predictions among the ten model pairs after Bonferroni adjustment (adjusted $p < 0.001$). These tests assess paired categorical prediction differences and do not directly establish statistically significant differences between continuous performance metrics such as ROC-AUC or F1-score. Given the large testing sample and the spatially structured nature of the observations, the results were interpreted together with the magnitude of the observed metric differences rather than as evidence of a large practical separation among models (McNemar, 1947).

4.3. Feature Importance and Model Interpretation

Global SHAP analysis identified SAR Mean, SAR Standard Deviation, DEM, Building Density and Rainfall among the five most influential predictors in the selected XGBoost model. The rainfall predictor represents cumulative seasonal rainfall across 2017-2023 rather than rainfall immediately preceding individual flood events. Its SHAP contribution therefore describes how the model used the spatial variation in cumulative rainfall when producing road-flood susceptibility scores. SHAP importance measures model-based predictive contribution and does not, by itself, establish a causal relationship between rainfall and road flooding. The relative contribution of the predictors to the selected XGBoost model is illustrated by the global SHAP feature-importance ranking in **Figure 4**.

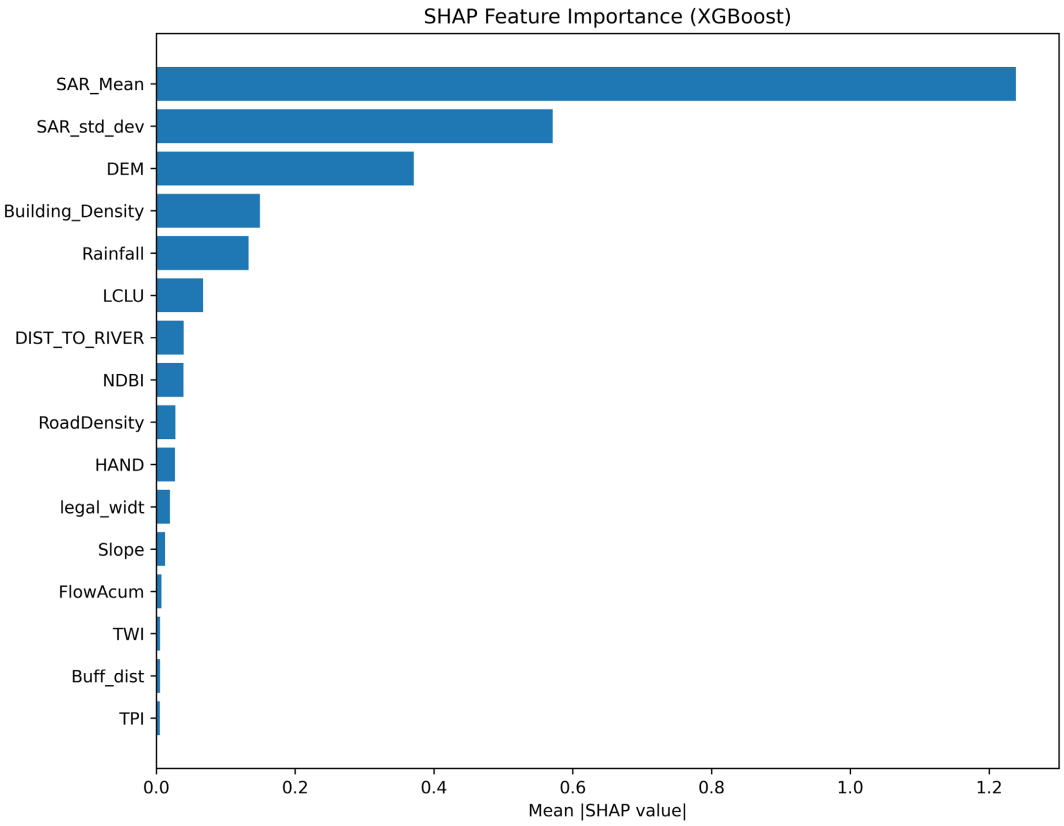


Figure 4. Global SHAP feature-importance ranking for the selected XGBoost model. Predictors are ordered according to their mean absolute SHAP contribution.

SAR Mean and SAR Standard Deviation describe background radar behaviour rather than event-specific flood change. They were calculated from the separate February and September Sentinel-1 VV reference observations collected during 2017-2023. The observations were combined into a temporal stack to produce one SAR Mean raster and one SAR Standard Deviation raster for the predictor dataset. Such temporal statistics can help characterize expected radar behaviour and variability under reference conditions (Schlaffer et al., 2015; Tupas et al., 2023; Oakes et al., 2023).

The SAR predictors and the flood-response variable were derived from different Sentinel-1 acquisition sets, which reduces direct reuse of the event observations in predictor construction. They are nevertheless physically related because both describe the radar response of the same landscape. The SAR predictors should therefore be interpreted as background radar descriptors rather than as independent observations of flooding. In dense urban environments, SAR backscatter can also be affected by buildings, vegetation, layover, radar shadow, surface roughness and double-bounce scattering (Twele et al., 2016; Schlaffer et al., 2015). SHAP values should consequently be interpreted as model-based predictive contributions rather than causal effects.

4.4. Road Susceptibility Mapping

Continuous XGBoost susceptibility scores were converted into five relative susceptibility classes using Jenks Natural Breaks. The classes were designated Very Low, Low, Moderate, High and Very High. The classification represents relative spatial differences in the modelled output within Nairobi City County and should not be interpreted as absolute probabilities of flooding. Because Jenks Natural Breaks is data-dependent, the class boundaries are specific to the modelled Nairobi road network and should not be assumed to represent transferable probability thresholds for other study areas. The spatial distribution of the five relative flood-susceptibility classes across the assessed urban road network is presented in **Figure 5**.

The final susceptibility assessment covered 7,235.952 km of road network (**Table 7**). The High and Very High classes comprised 2,556.707 km, corresponding to 35.33% of the assessed road length. Higher susceptibility was spatially concentrated in selected portions of the road network rather than being uniformly distributed across the county. The susceptibility score ranges and their corresponding operational meanings for road screening and follow-up investigation are summarized in **Table 8**.

The five susceptibility classes represent relative categories derived from the modelled susceptibility scores using the Jenks natural breaks classification method. The operational definitions indicate relative priorities for screening and further investigation; they do not represent direct measurements of flood depth, flow velocity, drainage capacity or hydraulic connectivity. Site-specific inspections and additional hydraulic or drainage assessments are required before selecting or implementing interventions.

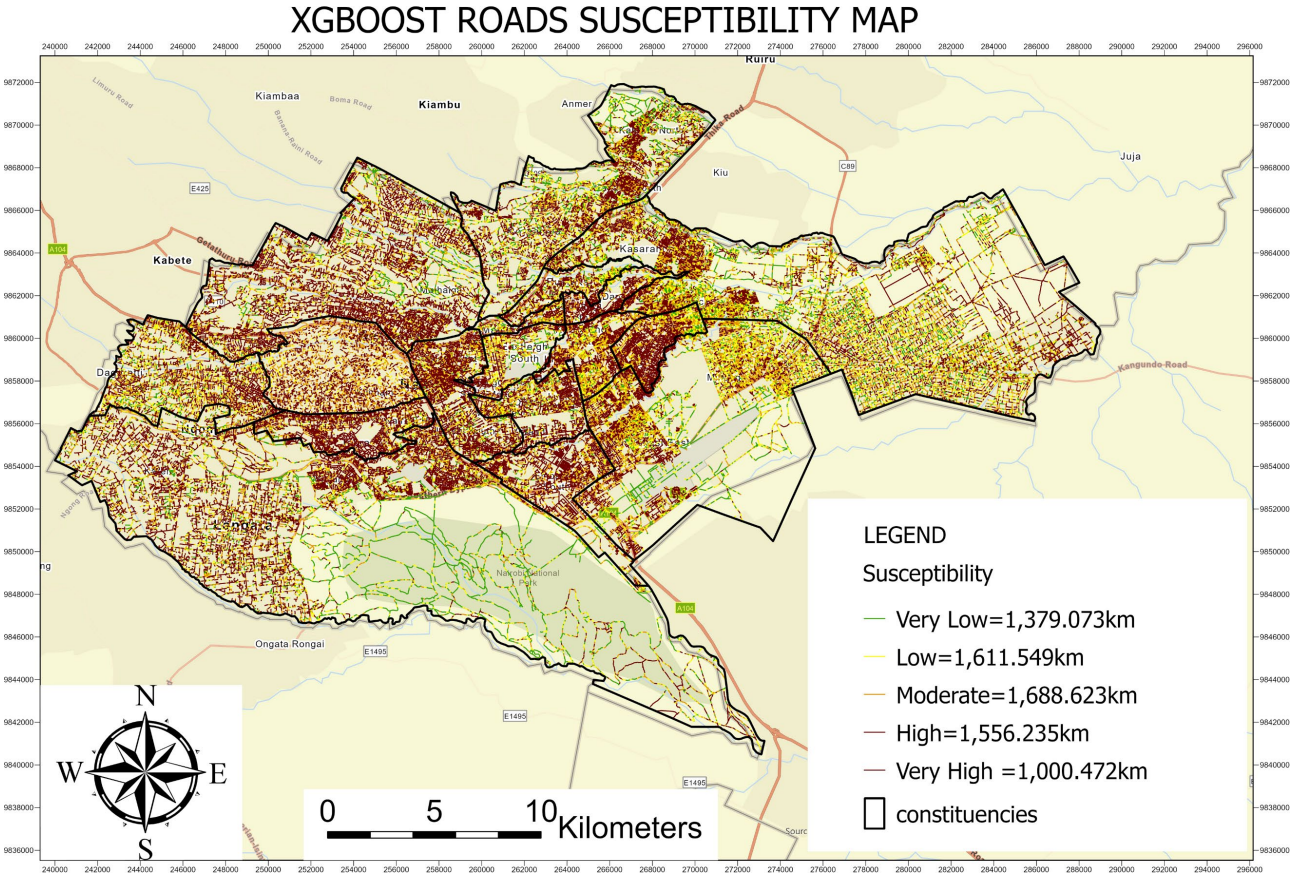


Figure 5. XGBoost-based relative flood-susceptibility classification of the urban road network in Nairobi City County. Colours represent the five ordered susceptibility classes according to the map legend.

Table 7. Road-network distribution by susceptibility class.

Susceptibility Class	Road Length (km)	Share of Network (%)	Road Segments
Very Low	1379.073	19.059	34,750
Low	1611.549	22.271	41,064
Moderate	1688.623	23.337	43,627
High	1556.235	21.507	40,520
Very High	1000.472	13.826	27,896
Total	7235.952	100.000	187,857

Table 8. Operational mapping meanings of the five susceptibility classes.

Susceptibility Class	Susceptibility Score Range	Revised Operational Definition
Very Low	0.006 - 0.094	Road sections with the lowest relative modelled flood susceptibility scores. These sections represent a lower priority for initial flood-related screening compared with sections in higher susceptibility classes.
Low	0.095 - 0.192	Road sections with relatively low modelled flood susceptibility scores. Routine monitoring may be appropriate, while recognising that localised flooding may still occur.

Continued

Moderate	0.193 - 0.586	Road sections with intermediate modelled flood susceptibility scores. These sections may warrant further screening to identify local drainage constraints and other site-specific factors that could contribute to flooding.
High	0.587 - 0.843	Road sections with relatively high modelled flood susceptibility scores. These sections are a higher priority for follow-up assessment, including inspection of drainage conditions, local topography and potential flood pathways.
Very High	0.844 - 0.999	Road sections with the highest relative modelled flood susceptibility scores. These sections are the highest priority for detailed site investigation and verification of local flood and drainage conditions.

4.5. Independent 2024 Validation

Independent temporal validation used the 2024 flooded-road observations. A total of 944.609 km of the assessed road network intersected the mapped 2024 flood extent, equivalent to 13.054% of the 7235.952 km network. The distribution of the observed 2024 flooded-road length across the five susceptibility classes is summarized in Table 9.

Table 9. Susceptibility and 2024 validation matrix.

Class	Road Segments per Class	Total Road (km) per Class	% of Network per Class	2024 Flooded Road (km)	Flood Exposure %	Share of 2024 Flooded Road (%)
Very Low	34,750	1379.073	19.06	132.731	9.625	14.051
Low	41,064	1611.549	22.27	177.098	10.989	18.748
Moderate	43,627	1688.623	23.34	211.196	12.507	22.358
High	40,520	1556.235	21.51	230.547	14.814	24.407
Very High	27,896	1000.472	13.83	193.037	19.295	20.436
Total	187,857	7235.952	100	944.609	13.054	100.000

The proportion of road length intersecting the 2024 flooded extent increased from 9.625% in the Very Low class to 19.295% in the Very High class. The High and Very High classes contained 423.584 km of independently observed flooded-road length, representing 44.84% of the total 2024 flooded-road length. When Moderate was included, the Moderate-to-Very-High classes contained 634.780 km, equivalent to 67.20% of the observed flooded-road length. The spatial relationship between the independently observed 2024 flooded-road extent and the modelled susceptibility classes is illustrated in Figure 6.

Flood exposure (%) is the flooded-road length within a susceptibility class divided by the total road length of that class, whereas the share of flooded road (%) is the flooded-road length within that class divided by the total flooded-road length across all classes.

The observed gradient indicates increasing flood exposure across the ordered susceptibility classes, but it also shows that substantial flooded-road length occurred

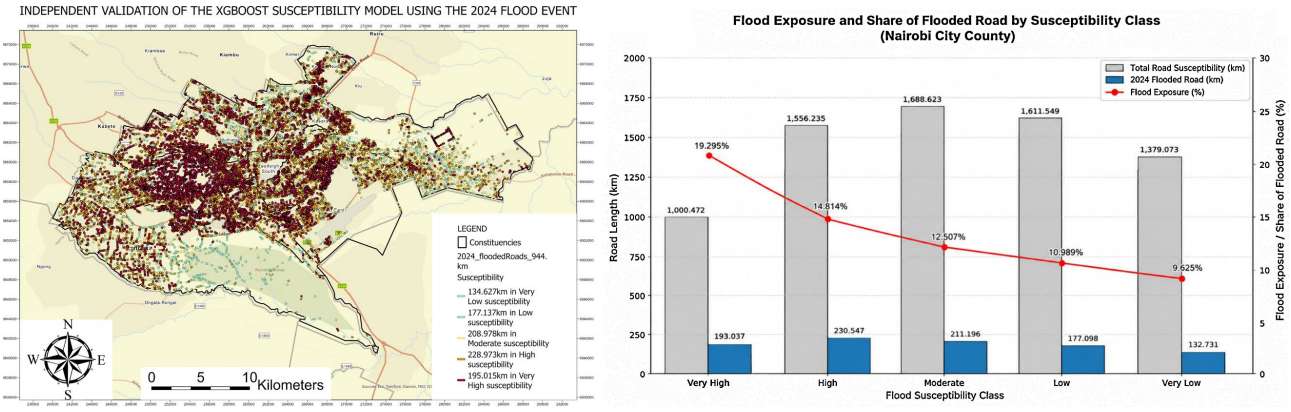


Figure 6. Independent temporal validation of the susceptibility classification using 2024 flood observations. Independently observed flooded-road extent is overlaid on the five modelled susceptibility classes.

in the lower classes. The susceptibility map should therefore be interpreted as a relative screening product rather than as a deterministic prediction of flooding during an individual rainfall event.

4.6. Statistical and External Validation

The independent 2024 validation indicated an association between susceptibility class and observed flooded-road exposure. The Kruskal-Wallis test (Kruskal & Wallis, 1952) indicated differences among the five classes ($H = 2667.854$, $p < 0.001$), while the chi-square test indicated an association between susceptibility class and flood occurrence ($\chi^2(4) = 2225.537$, $p < 0.001$; Cramer's V = 0.109). Spearman's rank correlation was positive and statistically significant ($\rho = 0.117$, $p < 0.001$), indicating a positive monotonic relationship. However, the small magnitude of ρ and Cramer's V indicates that the strength of association was modest despite statistical significance. The ordered-trend permutation test produced an empirical p-value of 0.00020 from 5,000 permutations, providing additional evidence of an ordered increase in observed flood exposure across the susceptibility classes. Because the underlying road observations are spatially clustered, these p-values should be interpreted as supporting evidence of association rather than as estimates based on fully independent observations. The statistical test results, corresponding p-values, and estimated effect sizes are summarized in Table 10.

The statistical results therefore support the presence of an ordered association,

Table 10. Statistical assessment of the relationship between susceptibility class and exposure.

Test Procedure	Observed Metric Statistic	Calculated p-Value	Inferred Association Effect Size
Kruskal-Wallis	H = 2667.854	<0.001	Statistically Significant Difference across Groups
Spearman Rank Correlation	$\rho = 0.117$	<0.001	Weak Positive Monotonic Trend
Chi-Square Independence	$\chi^2(4) = 2225.537$	<0.001	Modest Categorical Link (Cramer's V = 0.109)
Ordered Permutation (5000 runs)	T = 11,855,921,271.500	0.00020	Significant Positive Monotonic Ordered Sequence

but the effect sizes indicate that susceptibility class explains only a modest component of the observed variation in 2024 road-flood exposure.

4.7. Historical Flood-Prone Locations and Government-Documented Flood Hotspots

The historical validation dataset comprised 23 documented flood-prone locations. Four locations (17.39%) occurred within the Very High class, eight (34.78%) within the High class and nine (39.13%) within the Moderate class. Two locations (8.70%) occurred within the Low or Very Low classes. Consequently, 21 of the 23 documented historical flood-prone locations (91.30%) occurred within the Moderate-to-Very-High susceptibility classes.

The government-documented hotspot dataset contained 37 locations. The Ministry of Interior and National Administration/Nairobi Rivers Regeneration Programme mapping exercise is reported as dated 15 March 2026. Thirty-four locations (91.89%) occurred within the Moderate-to-Very-High susceptibility classes, while 19 (51.35%) occurred within the High and Very High classes. A consolidated summary of the temporal and external spatial validation results is provided in Table 11.

Table 11. Summary of independent temporal and external spatial validation results.

Validation Dataset	Observed Agreement	Observed Correspondence
2024 flood observations	423.584 km in High + Very High	44.84% of flooded road length
Historical flood-prone locations	21 of 23 in Classes 3 - 5	91.30%
Government hotspots (2026)	34 of 37 in Classes 3 - 5	91.89%

The spatial correspondence of documented historical flood-prone locations and government-identified flood hotspots with the modelled susceptibility classes is illustrated in Figure 7.

5. Discussion

The findings demonstrate the utility of repeated Sentinel-1 SAR observations for characterizing recurrent flood occurrence along urban roads. Rather than relying on a single event, the study integrated 14 selected events between 2017 and 2023 into a persistent flood-frequency representation. This captures recurrence within the available observation record and provides a more stable modelling response than a single-event inventory. However, the selected events represent only those for which suitable satellite observations were available and should not be interpreted as a complete record of all floods during the study period.

SAR Mean and SAR Standard Deviation were among the most influential predictors in the selected XGBoost model, together with DEM, Building Density and Rainfall. Their contribution is consistent with the use of multi-temporal Sentinel-1

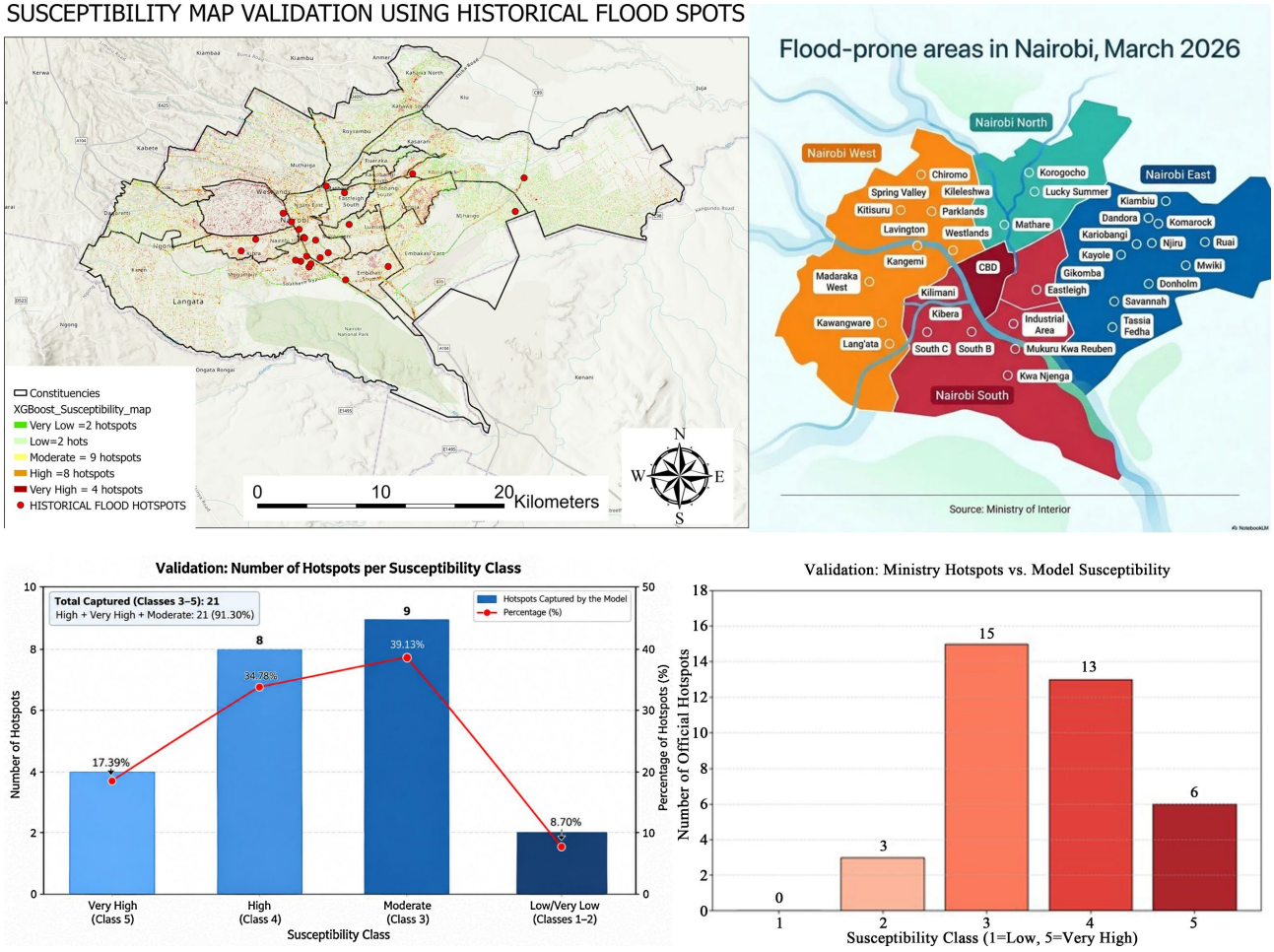


Figure 7. External spatial consistency assessment using documented historical flood-prone locations and government-identified flood hotspots. The locations are overlaid on the modelled susceptibility classes to examine their spatial correspondence with the susceptibility assessment.

statistics to describe background radar response and temporal variability.

The two SAR predictors were calculated from a separate non-event reference stack comprising February and September observations from 2017-2023. This reference dataset was not used to construct the event-level flood inventory, which was derived independently from paired pre-flood and during-flood observations. The separation reduces direct information overlap between predictor construction and response generation, while acknowledging that both datasets are physically based on Sentinel-1 backscatter. The predictors therefore provide background radar descriptors, not independent measurements of flooding. Previous SAR flood-mapping studies have similarly used temporal backscatter statistics to characterize expected surface behaviour and detect deviations associated with inundation (Schlaffer et al., 2015; Tupas et al., 2023; Oakes et al., 2023).

The five evaluated algorithms produced broadly comparable performance on the geographically separated testing dataset. XGBoost achieved the highest observed values for the principal reported metrics, but its ROC-AUC advantage over Gradient Boosting was only 0.0011. The results therefore indicate a narrow per-

formance range rather than a large separation among algorithms. The geographic partition provides a more demanding assessment of geographic generalization than a conventional random split because entire constituencies were withheld from model development. Nevertheless, geographic separation does not eliminate spatial dependence; neighbouring environmental and hydrological conditions can remain correlated.

Independent 2024 validation showed increasing observed flood exposure across the ordered susceptibility classes, from 9.625% in Very Low to 19.295% in Very High. The Moderate-to-Very-High classes contained 67.20% of the observed flooded-road length. The inferential analyses supported an ordered association, but the Spearman correlation ($\rho = 0.117$) and Cramer's V (0.109) indicate modest effect sizes. The substantial flooded-road length observed in lower classes further demonstrates that susceptibility classes should not be interpreted deterministically. Historical and government-documented locations provided additional spatial consistency evidence, with 91.30% and 91.89%, respectively, occurring within the Moderate-to-Very-High classes. These results support use of the map as a screening layer for infrastructure assessment rather than as a substitute for hydraulic modelling or site-specific engineering investigation.

The susceptibility map provides a spatial screening tool for identifying road sections that warrant further investigation. Road sections in the High and Very High classes can be examined in greater detail for drainage capacity, culvert performance, road crossfall, inlet condition, pavement condition and upstream runoff pathways. Integration with road-asset inventories, drainage-network datasets, maintenance records and historical flood reports can strengthen risk-informed road management.

Limitations

The assessment represents relative flood susceptibility and does not simulate flood depth, velocity, duration or hydraulic connectivity.

The flood inventory is based on 14 selected Sentinel-1 events between 2017 and 2023 and therefore does not represent all flood-generating conditions.

The 1.5-standard-deviation event threshold and the three-event recurrence cutoff are empirical operational choices; their outputs should not be interpreted as calibrated probabilities or universal flood definitions.

SAR flood detection in dense urban environments can be affected by buildings, vegetation, radar shadow, layover, surface roughness and double-bounce scattering.

Detailed drainage capacity, blockage and maintenance conditions were not comprehensively represented.

The SAR Mean and SAR Standard Deviation summarize a multi-year non-event reference record and describe background radar behaviour rather than year-specific hydrological conditions.

The road-oriented sampling design produces many spatially clustered observations, so conventional inferential p-values may be sensitive to spatial dependence.

Predictor layers may represent different spatial and temporal reference conditions, which can introduce additional mismatch between historical flood labels and predictor values.

Model selection was based on comparative performance within the available geographically separated hold-out framework; therefore, the reported ranking should not be interpreted as evidence of universal algorithmic superiority.

6. Conclusion

This study developed a road-oriented flood-susceptibility framework for Nairobi City County by integrating multi-temporal Sentinel-1 SAR observations, GIS-derived conditioning factors and supervised machine learning. Fourteen selected flood events from 2017 to 2023 were used to construct a recurrent flood inventory, which was combined with 16 terrain, hydrological, climatic, urban, road and SAR-derived predictors. Five machine-learning algorithms were evaluated using geographically separated training and testing datasets. The findings indicate that Sentinel-1 background radar descriptors, particularly SAR Mean and SAR Standard Deviation, provided useful predictive information when combined with terrain, hydrological, climatic, urban and road characteristics.

XGBoost achieved an accuracy of 0.7505, F1-score of 0.7269, balanced accuracy of 0.7493 and ROC-AUC of 0.8257 on the geographically separated test dataset. The differences among the algorithms were relatively small. SHAP analysis identified SAR Mean, SAR Standard Deviation, DEM, Building Density and Rainfall among the most influential predictors.

The contribution of SAR Mean and SAR Standard Deviation indicates that background radar characteristics provided useful predictive information when combined with the other conditioning factors. These predictors describe model-relevant variation in Sentinel-1 backscatter and should not be interpreted as causal evidence of flood occurrence.

The resulting susceptibility assessment covered 7,235.952 km of road network, with 35.33% classified as High or Very High susceptibility. Independent 2024 flooded-road observations showed an increase in observed flood exposure from 9.625% in Very Low to 19.295% in Very High. The Moderate-to-Very-High classes contained 67.20% of the independently observed flooded-road length. Historical and government-documented hotspot datasets provided additional spatial consistency checks, with 91.30% and 91.89%, respectively, occurring within the Moderate-to-Very-High classes.

Overall, the results demonstrate the potential of integrating recurrent SAR observations, GIS conditioning factors and geographically separated machine-learning evaluation for road-oriented flood-susceptibility assessment. The resulting map should be interpreted as a relative screening product for prioritising further investigation, drainage assessment and resilience planning rather than as a replacement for detailed hydraulic modelling or site-specific engineering assessment.

7. Recommendations

For operational use, the susceptibility map should be combined with field inspection, drainage condition, asset criticality, traffic importance, observed event-specific rainfall and historical flood reports.

Future studies should incorporate more detailed drainage and hydraulic variables, higher-resolution terrain and rainfall information, additional observed flood events, and spatially structured validation or block-based resampling to better account for dependence among nearby road observations.

Updating the Sentinel-1 reference archive as new observations become available would also allow the SAR background statistics to be recalculated and the susceptibility model to be periodically reassessed.

Acknowledgements

The authors thank the Department of Geomatic Engineering and Geospatial Information Science, Jomo Kenyatta University of Agriculture and Technology, for the academic environment and support provided for this research.

Author Contributions

Patrick Kathiari Gitile prepared the original draft and contributed to data processing, analysis and visualization. Josphat K. Z. Mwatelah contributed to supervision and critical review of the manuscript. Both authors read and approved the submitted version.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Bloch, R., Jha, A. K., & Lamond, J. (2012). *Cities and Flooding: A Guide to Integrated Urban Flood Risk Management for the 21st Century*. World Bank.

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The Balanced Accuracy and Its Posterior Distribution. In *2010 20th International Conference on Pattern Recognition* (pp. 3121-3124). IEEE. https://doi.org/10.1109/icpr.2010.764

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters, 27,* 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Fletcher, T. D., Shuster, W., Hunt, W. F., Ashley, R., Butler, D., Arthur, S. et al. (2015). SUDS, LID, BMPs, WSUD and More—The Evolution and Application of Terminology Surrounding Urban Drainage. *Urban Water Journal, 12,* 525-542. https://doi.org/10.1080/1573062x.2014.916314

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics, 29,* 1189-1232. https://doi.org/10.1214/aos/1013203451

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S. et al. (2015). The Cli-

mate Hazards Infrared Precipitation with Stations—A New Environmental Record for Monitoring Extremes. *Scientific Data, 2,* Article 150066. https://doi.org/10.1038/sdata.2015.66

IPCC (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability*. Cambridge University Press.

Jenks, G. F., & Caspall, F. C. (1971). Error on Choroplethic Maps: Definition, Measurement, Reduction. *Annals of the Association of American Geographers, 61,* 217-244. https://doi.org/10.1111/j.1467-8306.1971.tb00779.x

Koks, E. E., Rozenberg, J., Zorn, C., Tariverdi, M., Vousdoukas, M., Fraser, S. A. et al. (2019). A Global Multi-Hazard Risk Analysis of Road and Railway Infrastructure Assets. *Nature Communications, 10,* Article 2677. https://doi.org/10.1038/s41467-019-10442-3

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association, 47,* 583-621. https://doi.org/10.1080/01621459.1952.10483441

Li, Y., Martinis, S., Plank, S., & Ludwig, R. (2018). An Automatic Change Detection Approach for Rapid Flood Mapping in Sentinel-1 SAR Data. *International Journal of Applied Earth Observation and Geoinformation, 73,* 123-135. https://doi.org/10.1016/j.jag.2018.05.023

Lundberg, S. M., & Lee, S.-I. (2017). A Unified approach to Interpreting Model Predictions. In U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, & R. Fergus (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4765-4774). Curran Associates, Inc.

McNemar, Q. (1947). Note on the Sampling Error of the Difference between Correlated Proportions or Percentages. *Psychometrika, 12,* 153-157. https://doi.org/10.1007/bf02295996

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of Spatial Predictor Variable Selection in Machine Learning Applications—Moving from Data Reproduction to Spatial Prediction. *Ecological Modelling, 411,* Article 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water, 10,* Article 1536. https://doi.org/10.3390/w10111536

Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A. et al. (2011). Height above the Nearest Drainage—A Hydrologically Relevant New Terrain Model. *Journal of Hydrology, 404,* 13-29. https://doi.org/10.1016/j.jhydrol.2011.03.051

Oakes, G., Hardy, A., & Bunting, P. (2023). RadWet: An Improved and Transferable Mapping of Open Water and Inundated Vegetation Using Sentinel-1. *Remote Sensing, 15,* Article 1705. https://doi.org/10.3390/rs15061705

Oluchiri, S. O. (2025). Urban Flooding in the Cities of Kisumu, Mombasa, and Nairobi, Kenya: Causes, Vulnerability Factors, and Management. *African Journal of Empirical Research, 6,* 342-351. https://doi.org/10.51867/ajernet.6.1.29

Ongaga, C. O., Makokha, M., Obiero, K., Kipkemoi, I., & Diang'a, J. (2024). Urbanization and Hydrological Dynamics: A 22-Year Assessment of Impervious Surface Changes and Runoff in an Urban Watershed. *Frontiers in Water, 6,* Article ID: 1455763. https://doi.org/10.3389/frwa.2024.1455763

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G. et al. (2017). Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography, 40,* 913-929. https://doi.org/10.1111/ecog.02881

Schlaffer, S., Matgen, P., Hollaus, M., & Wagner, W. (2015). Flood Detection from Multi-Temporal SAR Data Using Harmonic Analysis and Change Detection. *International Journal of Applied Earth Observation and Geoinformation, 38,* 15-24. https://doi.org/10.1016/j.jag.2014.12.001

Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing & Management, 45,* 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

Tehrany, M. S., Jones, S., & Shabani, F. (2019). Identifying the Essential Flood Conditioning Factors for Flood Prone Area Mapping Using Machine Learning Techniques. *CATENA, 175,* 174-192. https://doi.org/10.1016/j.catena.2018.12.011

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood Susceptibility Mapping Using a Novel Ensemble Weights-of-Evidence and Support Vector Machine Models in Gis. *Journal of Hydrology, 512,* 332-343. https://doi.org/10.1016/j.jhydrol.2014.03.008

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E. et al. (2012). GMES Sentinel-1 Mission. *Remote Sensing of Environment, 120,* 9-24. https://doi.org/10.1016/j.rse.2011.05.028

Tupas, M. E., Roth, F., Bauer-Marschallinger, B., & Wagner, W. (2023). An Intercomparison of Sentinel-1 Based Change Detection Algorithms for Flood Mapping. *Remote Sensing, 15,* Article 1200. https://doi.org/10.3390/rs15051200

Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-Based Flood Mapping: A Fully Automated Processing Chain. *International Journal of Remote Sensing, 37,* 2990-3004. https://doi.org/10.1080/01431161.2016.1192304

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brock-mann, C., Quast, R. et al. (2021). *ESA WorldCover 10 m 2020 v100.* Zenodo.

Zha, Y., Gao, J., & Ni, S. (2003). Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery. *International Journal of Remote Sensing, 24,* 583-594. https://doi.org/10.1080/01431160304987

Breiman, L. (2001). Random Forests. *Machine Learning, 45,* 5-32. https://doi.org/10.1023/a:1010933404324