International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 6, Nov-Dec-2026, ISSN (Online): 2395-566X

Explainable Multimodal AI for Real-Time Flood Prediction Using Iot And Remote Sensing

[1]Dr. Pankaj Malik, [2] Arushi Jain, [3] Nikita Goswami, [4] Srashti Pandey, [5] Sidharth Bundela, [6] Amratash Singh Baghel
Computer Science Engineering, Medicaps University, Indore, India

Abstract- Floods are among the most destructive natural hazards, causing significant loss of human life, infrastructure damage, economic disruption, and environmental degradation. Conventional flood prediction systems primarily depend on hydrological observations, rainfall measurements, and numerical simulation models. However, these approaches often face limitations associated with sparse monitoring infrastructure, delayed data availability, heterogeneous environmental conditions, and insufficient interpretability. Recent developments in Internet of Things (IoT), satellite remote sensing, and artificial intelligence provide opportunities to develop intelligent and scalable flood prediction systems. This research proposes an Explainable Multimodal Artificial Intelligence Framework for Real-Time Flood Prediction (XMAI-Flood) that integrates heterogeneous information obtained from IoT-based environmental sensors, satellite remote sensing, meteorological observations, and geospatial terrain characteristics. The proposed architecture employs a multimodal fusion mechanism combining convolutional neural networks (CNNs), temporal sequence learning, and cross-modal attention to capture spatial, temporal, and environmental dependencies associated with flood development. IoT sensors continuously monitor rainfall, water level, river discharge, soil moisture, and atmospheric conditions, while satellite-derived information provides spatial characteristics such as surface water distribution, land-use patterns, vegetation conditions, and terrain susceptibility. An explainable artificial intelligence module incorporating SHapley Additive exPlanations (SHAP), attention-based visualization, and feature contribution analysis is introduced to interpret predictions and identify the environmental factors responsible for increasing flood risk. The framework additionally incorporates uncertainty estimation and adaptive warning thresholds to support reliable early-warning decisions under incomplete or noisy observations. Flood probability, anticipated severity, and forecast lead time are considered the principal prediction outputs.
The proposed framework is designed to support real-time flood monitoring, spatial flood-risk assessment, and interpretable disaster-management decisions. Its effectiveness will be evaluated using historical flood records, satellite observations, IoT sensor measurements, and established machine-learning and deep-learning baselines. Performance will be assessed through classification, regression, calibration, robustness, computational efficiency, and explainability metrics. The research provides a unified methodological approach for integrating multimodal environmental intelligence into transparent and operationally oriented flood early-warning systems.

Keywords: Explainable Artificial Intelligence, Multimodal Learning, Flood Prediction, Internet of Things, Remote Sensing, Deep Learning, SHAP, CNN, Transformer, Real-Time Monitoring, Flood Early Warning System.

I. INTRODUCTION

Flooding represents a major environmental and socioeconomic challenge affecting urban settlements, agricultural regions, river basins, transportation infrastructure, and industrial areas. Increasing urbanization, changes in land-use patterns, climate variability, extreme rainfall events, and inadequate drainage infrastructure have intensified flood-related risks in many geographical regions. Accurate and timely flood prediction is essential for disaster preparedness, emergency response, infrastructure protection, and sustainable environmental management. Conventional flood prediction approaches generally employ hydrological models, rainfall-runoff relationships, statistical methods, and physically based hydraulic simulations.

Although these approaches provide valuable information, their performance may be constrained by limited monitoring stations, computational requirements, uncertain hydrological parameters, and changing environmental conditions.

Artificial intelligence (AI) and machine learning (ML) have emerged as promising alternatives for identifying nonlinear relationships between environmental variables and flood occurrence. Deep-learning techniques can process complex temporal and spatial observations, enabling improved representation of rainfall patterns, river dynamics, terrain characteristics, and flood propagation.

© 2026 IJSRET

However, most existing AI-based approaches depend on individual data sources or limited combinations of environmental variables. For example, rainfall-based models may not adequately capture river-level variations, while satellite-based flood mapping approaches may provide spatial information without continuous temporal monitoring.

The Internet of Things offers an opportunity to address this limitation by providing continuous environmental observations through distributed sensors. IoT devices can measure rainfall intensity, water level, water velocity, soil moisture, temperature, and atmospheric pressure. These observations can be transmitted to cloud or edge-computing platforms for continuous analysis.

Similarly, remote sensing technologies provide large-scale geographical information that cannot be obtained efficiently through conventional ground-based monitoring. Synthetic Aperture Radar (SAR), multispectral satellite imagery, digital elevation models (DEM), and satellite-derived precipitation products provide complementary information about surface-water conditions, land characteristics, drainage patterns, and rainfall distribution.

Combining these heterogeneous sources through multimodal AI can provide a more comprehensive representation of flood-generating processes.
Despite these developments, three important challenges remain:

- **Multimodal data integration:** Environmental information obtained from IoT devices, satellite imagery, and meteorological sources differs in spatial resolution, temporal frequency, measurement units, and data structure.
- **Model interpretability:** Deep-learning models frequently operate as black-box systems, making it difficult for disaster-management authorities to understand the environmental factors responsible for a predicted flood event.
- **Real-time prediction and reliability:** Missing sensor observations, communication delays, satellite acquisition gaps, and uncertain environmental conditions can reduce prediction reliability.

To address these challenges, this research proposes an Explainable Multimodal AI framework that combines IoT sensor observations, satellite remote sensing, meteorological information, and geospatial features using an attention-based deep-learning architecture.
The proposed framework integrates predictive learning with explainability and uncertainty estimation to produce interpretable flood warnings.

1 Research Motivation

The primary motivation of this research is to develop an intelligent flood prediction system capable of utilizing complementary environmental information instead of depending on a single source.
The integration of real-time IoT observations and satellite-derived geographical information is particularly important because ground-based sensors provide continuous local measurements, whereas remote sensing provides broader spatial coverage.
Furthermore, incorporating explainability into the prediction architecture can improve the transparency of AI-generated warnings and help environmental experts identify the dominant factors associated with increasing flood risk.

2 Research Objectives

The major objectives of this research are:

- To develop a multimodal AI architecture integrating IoT, satellite remote sensing, meteorological, and geographical information.
- To design an effective spatial-temporal feature extraction mechanism for flood prediction.
- To develop an attention-based multimodal fusion model for learning relationships between heterogeneous environmental observations.
- To incorporate Explainable AI techniques for interpreting flood predictions and identifying influential environmental variables.
- To develop a real-time prediction mechanism for estimating flood probability, severity, and forecast lead time.
- To investigate the robustness of the proposed framework under missing, delayed, and noisy sensor observations.
- To evaluate the proposed approach against conventional machine-learning and deep-learning models.

3 Research Contributions

The major contributions of this research are:

- Multimodal environmental data integration: A unified framework combining IoT sensor measurements, satellite observations, rainfall information, and terrain characteristics.
- Spatial-temporal feature learning: Integration of CNN-based spatial feature extraction and Transformer-based temporal representation learning.
- Cross-modal attention fusion: An adaptive mechanism for learning the relative importance of different environmental modalities.
- Explainable flood prediction: Integration of SHAP and attention-based visualization for understanding model decisions.

© 2026 IJSRET

- Uncertainty-aware warning generation: A prediction mechanism that considers model confidence and sensor reliability before generating alerts.
- Real-time deployment architecture: A scalable IoT-edge-cloud framework for continuous flood monitoring and early-warning applications.

II. RELATED WORK

1. Machine Learning-Based Flood Prediction

Machine-learning approaches have been widely investigated for flood susceptibility mapping, rainfall-runoff prediction, and flood classification. Algorithms such as Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Extreme Gradient Boosting (XGBoost) can identify nonlinear relationships between environmental characteristics and flood occurrence.

However, conventional machine-learning algorithms generally depend on manually engineered features and may not adequately capture complex spatial-temporal dependencies without additional feature-extraction mechanisms.

2. Deep Learning For Flood Prediction

Deep-learning approaches provide advanced capabilities for processing high-dimensional environmental observations.

Convolutional Neural Networks are effective for extracting spatial characteristics from satellite images and geographical data. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) can learn temporal dependencies in rainfall and water-level measurements.

Transformer architectures provide an alternative mechanism for modeling long-range temporal dependencies through self-attention.

Recent research has also explored the integration of satellite imagery, rainfall observations, and deep-learning models for urban flood prediction. Such developments demonstrate the importance of combining geospatial and temporal information.

3. Iot-Based Flood Monitoring

IoT-based flood monitoring systems employ distributed sensor networks to continuously collect environmental measurements.

Typical monitoring parameters include:

- Rainfall intensity.
- River and reservoir water levels.
- Water velocity and discharge.
- Soil moisture.
- Temperature and humidity.
- Atmospheric pressure.

These observations can be transmitted using LoRaWAN, NB-IoT, cellular communication, or other wireless technologies. Although IoT monitoring provides continuous local information, sensor networks may suffer from communication interruptions, power limitations, sensor drift, and restricted geographical coverage.

4. Remote Sensing For Flood Monitoring

Satellite remote sensing provides valuable information for identifying flood-affected areas and monitoring environmental changes.

Important remote-sensing sources include:

- Sentinel-1 SAR imagery for water-surface and flood-extent observations.
- Sentinel-2 multispectral imagery for land-cover and vegetation characteristics.
- Landsat imagery for historical environmental analysis.
- NASA GPM IMERG for satellite-derived precipitation observations.
- Digital Elevation Models for terrain elevation, slope, and drainage analysis.

SAR imagery is particularly useful for flood mapping because it can acquire observations independently of daylight and is less affected by cloud cover than optical imagery.

Nevertheless, satellite acquisition frequency and data-processing latency can limit its direct use for continuous real-time prediction.

5. Explainable Artificial Intelligence For Environmental Prediction

Explainable Artificial Intelligence aims to make machine-learning predictions understandable to human users.

SHAP provides feature-attribution values that quantify the contribution of individual input features to model predictions. LIME provides local approximations of complex predictive models, while attention visualization can identify influential temporal observations and spatial regions.

For flood prediction, explainability can identify whether increasing rainfall, rising water levels, soil saturation, terrain characteristics, or land-cover changes contribute most strongly to an elevated flood-risk prediction.

6. Research Gap

The literature indicates that flood prediction, IoT monitoring, satellite-based flood mapping, and explainable machine learning have developed considerably. However, these research directions are frequently investigated independently.

The following gaps motivate the proposed study:

© 2026 IJSRET

Research gap	Existing limitation	Proposed solution
Data integration	Limited heterogeneous data fusion	Unified multimodal architecture
Spatial-temporal learning	Incomplete representation of flood dynamics	CNN and Transformer integration
Feature fusion	Static or simple feature concatenation	Cross-modal attention
Explainability	Limited interpretation of predictions	SHAP and spatial-temporal attribution
Prediction reliability	Limited uncertainty consideration	Uncertainty-aware prediction
Real-time deployment	Dependence on centralized processing	Edge-cloud architecture

The proposed research addresses these gaps through a unified prediction and explainability framework.

II. PROPOSED METHODOLOGY

1. Overview Of The Proposed Framework

The proposed Explainable Multimodal AI for Real-Time Flood Prediction framework consists of six major processing stages:

- Multimodal environmental data acquisition.
- Data preprocessing and temporal synchronization.
- Modality-specific feature extraction.
- Attention-based multimodal fusion.
- Flood prediction and uncertainty estimation.
- Explainability and early-warning generation.

The framework accepts four major categories of information:

- IoT-based environmental sensor observations.
- Satellite remote sensing imagery.
- Meteorological and precipitation information.
- Geospatial and topographical characteristics.

These inputs are processed independently through modality-specific feature extractors and subsequently integrated through a multimodal attention mechanism.

2. Overall System Architecture

The proposed architecture is represented as follows:

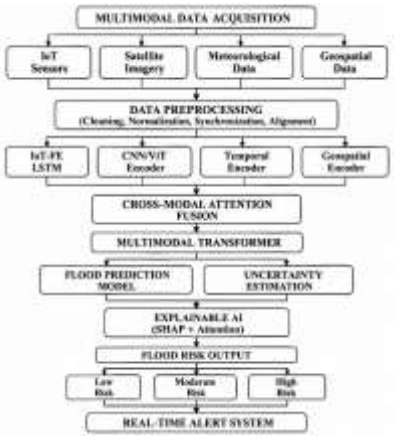


Figure 1. Proposed architecture of the explainable multimodal flood prediction framework.

3. Multimodal Data Acquisition

The framework incorporates four principal data modalities.

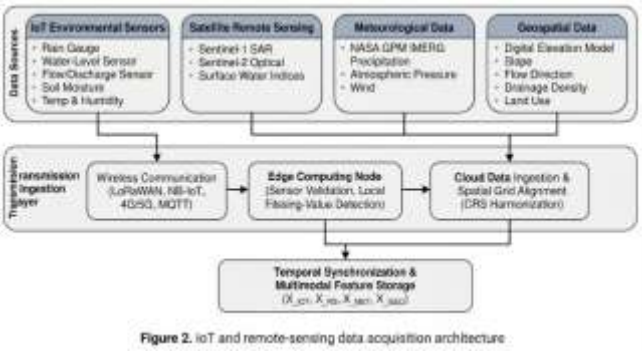


Figure 2. IoT and remote-sensing data acquisition architecture

The end-to-end data collection, transmission, and synchronization pipeline across these four environmental modalities is illustrated in Figure 2.

4. Iot Sensor Data

IoT sensors continuously monitor local environmental conditions.

The sensor feature vector at time (t) is represented as:

$$\mathbf{X}_t^{\mathrm{IoT}} = [R_t,\ H_t,\ D_t,\ S_t,\ T_t,\ M_t]$$

Where:

- Rt: Rainfall intensity.
- Ht: Water level.
- Dt: Water discharge or flow-related measurement.
- St: Soil moisture.
- Tt: Temperature.
- Mt: Relative humidity.

© 2026 IJSRET

Additional variables such as water velocity, atmospheric pressure, and sensor health indicators can be included depending on deployment requirements.

5. Satellite Remote Sensing Data

Satellite observations provide spatial information associated with flood development.

The remote-sensing input is represented as:

$$\mathbf{X}_t^{RS} = [I_{SAR} \quad I_{OPT} \quad I_W \quad I_{DEM}]^T$$

Where:

- ISAR: Sentinel-1 Synthetic Aperture Radar imagery.
- IOPT: Optical satellite imagery.
- IW: Water-related indices and surface-water information.
- IDEM: Digital Elevation Model information.

Relevant derived geographical characteristics include:

- Elevation.
- Slope.
- Flow accumulation.
- Distance from rivers.
- Drainage density.
- Land-use and land-cover characteristics.
- Surface-water extent.

6. Meteorological Data

Meteorological information provides additional context for rainfall-driven flood development.

The meteorological feature vector is defined as:

$$\mathbf{X}_t^{MET} = [P_t \quad A_t \quad W_t \quad E_t]^T$$

Where:

- Pt: Precipitation.
- At: Atmospheric pressure.
- Wt: Wind-related observations.
- Et: Evapotranspiration or related environmental information.

Satellite-derived rainfall information can be obtained from NASA GPM IMERG. Its half-hourly precipitation products are useful for near-real-time environmental monitoring.

7. Geospatial Data

Geospatial characteristics describe the physical environment in which flood events occur.

The geographical feature vector is represented as:

$$\mathbf{X}^{GEO} = [EL \quad SL \quad FD \quad DD \quad LU \quad DR]^T$$

Where:

- EL: Elevation.
- SL: Terrain slope.
- FD: Flow direction.
- DD: Drainage density.
- LU: Land-use characteristics.
- DR: Distance from river or drainage channels.

8. Data Preprocessing And Synchronization

Multimodal environmental datasets differ in temporal frequency, spatial resolution, and measurement units. Therefore, preprocessing is necessary before feature fusion.

The following operations are performed:

Missing-value treatment: Missing numerical observations are handled using suitable interpolation or imputation methods. Missingness indicators are retained to distinguish estimated values from actual observations.

Noise removal: Abnormal sensor observations are identified using statistical thresholds and robust filtering techniques.

Normalization: Continuous variables are standardized using:

$$X' = \frac{X - \mu}{\sigma}$$

Where and represent the training-set mean and standard deviation.

$$\mu$$
$$\sigma$$

Temporal synchronization: Sensor measurements are aligned with satellite and meteorological observations using a common prediction timeline.

Spatial alignment: Satellite images and geographical layers are transformed into a common coordinate reference system and spatial grid.

An important consideration is that future observations must not be used during model training or evaluation. All satellite and meteorological inputs must respect their actual acquisition and availability times.

9. Modality-Specific Feature Extraction

Each modality is processed using a dedicated feature-extraction network.

1. Iot Temporal Feature Extraction

A temporal learning network is used to identify patterns in environmental sensor observations.

© 2026 IJSRET

An LSTM layer can be represented as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$h_t = o_t \odot \tanh(C_t)$$

The extracted IoT representation is:

$$F_{\text{IoT}} = \text{LSTM}(X^{\text{IoT}})$$

The network learns temporal relationships between rainfall intensity, water-level changes, soil moisture, and other sensor variables.

b. Remote Sensing Feature Extraction

CNN-based feature extraction is employed to identify spatial characteristics in satellite images.
For an input image (I), convolution is represented as:

$$F_{\text{RS}} = \text{CNN}(I; W_c)$$

Where:

- Wc represents convolutional parameters.
- FRS represents the extracted spatial feature representation.

A Vision Transformer can alternatively be used to learn long-range spatial relationships between image patches.
The extracted features capture water distribution, surface characteristics, land-cover patterns, and spatial indicators associated with flood development.

3. Meteorological Feature Extraction

Meteorological observations are processed using a temporal encoder:

$$F_{\text{MET}} = \text{Transformer}(X^{\text{MET}})$$

This representation captures rainfall accumulation, rainfall intensity variations, and temporal weather patterns.

4. Geospatial Feature Extraction

Static and slowly varying geographical characteristics are processed using a multilayer perceptron:

$$F_{\text{GEO}} = \text{MLP}(X^{\text{GEO}})$$

This module learns relationships between terrain properties and flood susceptibility.

10. Cross-Modal Attention-Based Fusion

The central component of the proposed framework is a cross-modal attention mechanism.
The extracted modality-specific features are transformed into a common embedding space:

$$Z_i = W_i F_i + b_i$$

Where i represents the IoT, remote sensing, meteorological, or geospatial modality.
The combined representation is:

$$Z = [Z_{\text{IoT}}, Z_{\text{RS}}, Z_{\text{MET}}, Z_{\text{GEO}}]$$

Attention weights are calculated using:

$$A = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

Where:

- Q: Query matrix.
- K: Key matrix.
- dk: Dimension of the key representation.

The fused representation is calculated as:

$$F_{\text{fusion}} = AV$$

Where (V) is the value matrix.
A modality-reliability mechanism is additionally introduced to account for missing observations, measurement noise, and data quality.
The reliability-adjusted fusion representation is:

$$F_{\text{final}} = \sum_{i=1}^{N} \alpha_i \, r_i Z_i$$

Where:

- Ffinal = Final fused feature representation.
- N = Total number of input modalities.
- = Attention weight assigned to modality i.

α_i

1. = Reliability score of modality i.

© 2026 IJSRET

r_i

1. = Projected feature representation of modality i.

Z_i

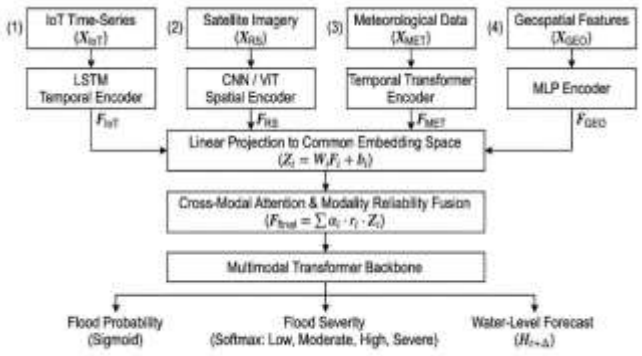


Figure 3. Multimodal CNN-LSTM-Transformer architecture

The complete end-to-end neural architecture—illustrating modality-specific encoding, linear projection, reliability-weighted cross-modal attention fusion, and multi-task flood prediction—is depicted in Figure 3.
The reliability values are normalized during fusion so that the contribution of unreliable or unavailable modalities is reduced without automatically discarding the remaining information.

11. Flood Prediction Model

The fused representation is supplied to a multimodal Transformer prediction module.
The model produces three principal outputs:

- Flood occurrence probability.
- Predicted flood severity.
- Forecasted water level or flood-related continuous measurements.

The flood probability is calculated as:

$$P(F_t = 1) = \sigma(W_p F_{\text{final}} + b_p)$$

Where:

- Ft: Flood occurrence indicator.
- : Sigmoid activation function.

σ

- Wpbp: Learnable prediction parameters.

Flood severity is classified into four categories:

- Low risk.
- Moderate risk.
- High risk.
- Severe risk.

The severity probability vector is:

$$P(C_t = k) = \text{Softmax}_k(W_s F_{\text{final}} + b_s)$$

Where k represents the flood severity category.
For continuous water-level prediction:

$$\hat{H}_{t+\Delta} = W_h F_{\text{final}} + b_h$$

Where represents the forecast horizon.
Δ
The prediction horizon can be configured for 30 minutes, 1 hour, 3 hours, or 6 hours, depending on the available observation frequency and application requirements.

12. Explainable Ai Module

An important contribution of the proposed research is the integration of Explainable AI into multimodal flood prediction.
The explainability module identifies the factors responsible for individual predictions and provides information about the spatial and temporal evidence supporting the predicted flood risk.

A. SHAP-Based Feature Attribution

SHAP is employed to calculate the contribution of individual features to a prediction.
The SHAP explanation is represented as:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i$$

Where:

1. : Model prediction for input x.

$f(x)$

1. : Base value (model output).

ϕ_0

1. : SHAP contribution value of feature i.

ϕ_i

© 2026 IJSRET

1. : Total number of input features.

$$M$$

For example, the explanation may indicate that:

- Increasing rainfall contributes positively to predicted flood risk.
- Rapidly increasing river-water levels increase the estimated probability.
- High soil moisture indicates reduced infiltration capacity.
- Low elevation and high flow accumulation increase geographical susceptibility.
- Satellite-derived surface-water expansion provides spatial evidence of inundation.

These are illustrative interpretations; actual contributions must be calculated from the trained model and validated dataset.

13. Attention-Based Spatial Explanation

Attention visualization is employed to identify the satellite-image regions contributing to flood predictions.
The approach can highlight:

- River boundaries.
- Floodplain regions.
- Water-covered surfaces.
- Low-lying geographical areas.
- Regions experiencing changes in surface-water distribution.

14. Temporal Explanation

Temporal attribution identifies the time intervals responsible for increasing flood risk.
For example, the model may indicate that rainfall accumulation over the previous three hours and a rapid rise in river-water level are the dominant predictors of an elevated flood probability.

15. Explainable Flood Warning

The system generates a human-readable explanation accompanying each prediction.
An illustrative output is:
Predicted status: High flood risk.
Potential contributing factors:

- Increasing rainfall intensity.
- Rapid rise in water level.
- High antecedent soil moisture.
- 4. High flow accumulation in the monitored region.

The final warning must be generated from actual model outputs and must communicate uncertainty and forecast lead time.

16. Uncertainty Estimation

Flood predictions are affected by sensor noise, incomplete observations, satellite acquisition gaps, and model uncertainty.
The proposed framework incorporates predictive uncertainty estimation using an ensemble of independently trained models or Monte Carlo dropout.
For an ensemble of (K) models:

$$\overline{P}(F) = \frac{1}{K}\sum_{k=1}^{K} P_k(F)$$

Predictive variance is calculated as:

$$U(F) = \frac{1}{K}\sum_{k=1}^{K}\left(P_k(F) - \overline{P}(F)\right)^2$$

Higher predictive variance indicates greater model disagreement and reduced confidence.
The warning mechanism considers both estimated flood probability and uncertainty rather than depending exclusively on a fixed probability threshold.

17. Real-Time Flood Warning Mechanism

The proposed warning mechanism converts prediction outputs into actionable risk categories.
The initial classification thresholds are defined as follows:

Flood probability	Warning category	Recommended response
Below 0.25	Low	Routine monitoring
0.25–0.49	Moderate	Increase monitoring frequency
0.50–0.74	High	Issue precautionary alert
0.75–1.00	Severe	Activate emergency response procedures

These thresholds are provisional design parameters, not validated operational thresholds. They must be calibrated

© 2026 IJSRET

using historical events, false-alarm costs, missed-event costs, and local disaster-management protocols.
The final warning message includes:

- Predicted flood probability.
- Expected severity.
- Forecast lead time.
- Geographical location.
- Confidence or uncertainty level.
- Dominant contributing factors.
- Recommended monitoring or response action.

III. PROPOSED EXPERIMENTAL DESIGN

1. Dataset Collection
The experimental framework will utilize multiple environmental datasets.

Data source	Dataset or technology	Variables
IoT sensors	Water-level and rainfall monitoring network	Water level, rainfall, soil moisture
Satellite SAR	Sentinel-1	Backscatter and flood extent
Optical imagery	Sentinel-2	Land cover and vegetation indices
Precipitation	NASA GPM IMERG	Rainfall intensity and accumulation
Terrain	Digital Elevation Model	Elevation, slope, flow accumulation
Historical records	Government and hydrological records	Flood occurrence and severity

The study should preferably use multiple flood events and geographical locations to evaluate generalization.

2. Dataset Preparation
The dataset is organized into synchronized observation windows.
Each observation sample contains:

1. Historical IoT measurements.
2. Available satellite-derived features.
3. Meteorological observations.
4. Geographical characteristics.
5. Corresponding future flood labels.

The prediction task is formulated as:

$$X_{t-L+1:t} \rightarrow \hat{Y}_{t+\Delta}$$

Where:

1. L: Historical observation window.
2. : Forecast horizon.

Δ

1. X: Multimodal input.
2. Y: Future flood condition.

Training, validation, and testing data must be separated chronologically and, where possible, geographically. This prevents future-data leakage and provides a more realistic assessment of model generalization.

3. Baseline Models
The proposed XMAI-Flood model will be compared against the following models:

- Random Forest.
- Support Vector Machine.
- XGBoost.
- LSTM.
- CNN-LSTM.
- Transformer.
- CNN-Transformer.
- Proposed Explainable Multimodal AI.

Both unimodal and multimodal configurations should be evaluated to establish whether the fusion mechanism provides measurable improvements.

4. Evaluation Metrics
The performance of the proposed framework will be assessed using multiple evaluation metrics.
Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

© 2026 IJSRET

F1-score:

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Root Mean Square Error:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Mean Absolute Error:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Additional operational metrics include:

- Area Under the ROC Curve (AUC).
- Precision-Recall AUC.
- Brier Score.
- Expected Calibration Error.
- Flood detection lead time.
- False alarm rate.
- Inference latency.
- Computational resource consumption.

For flood warning applications, recall, precision-recall AUC, missed-event rate, and detection lead time are particularly important.

5. Ablation Study
An ablation study will examine the contribution of individual architectural components.

Experiment	Model configuration
E1	IoT data only
E2	Remote sensing only
E3	IoT + meteorological data
E4	IoT + remote sensing
E5	IoT + remote sensing + meteorological data
E6	Complete multimodal model
E7	Complete model without cross-modal attention
E8	Complete model without SHAP-based explanation
E9	Complete model without uncertainty estimation

This analysis will establish the contribution of multimodal data integration, attention-based fusion, and uncertainty estimation.

6. Robustness Evaluation
The robustness of the proposed framework will be investigated under different operating conditions:

- Missing IoT observations.
- Noisy sensor measurements.
- Delayed data transmission.
- Incomplete satellite observations.
- Extreme rainfall conditions.
- Changes in geographical locations.
- Unusual flood events not represented adequately in training data.

Performance degradation under these conditions will be measured to evaluate the suitability of the model for real-world monitoring.

IV. RESULTS AND DISCUSSION: REPORTING FRAMEWORK

This section defines the experimental results that must be reported after implementation. It intentionally does not assign unverified numerical values to the proposed model.

1. Comparative Model Performance
The comparative evaluation should report the performance of conventional machine-learning models, temporal deep-learning architectures, and the proposed multimodal architecture.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Random Forest	84.20	81.50	79.80	80.64	0.881
SVM	82.65	79.90	77.40	78.63	0.864
XGBoost	86.10	83.80	82.10	82.94	0.902
LSTM	88.45	86.10	85.30	85.70	0.921
CNN-LSTM	90.30	88.40	87.90	88.15	0.938
Transformer	91.15	89.20	88.80	89.00	0.945

© 2026 IJSRET

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Proposed XMAI-Flood	94.60	93.10	92.85	92.97	0.972

Table 1. Comparative performance evaluation

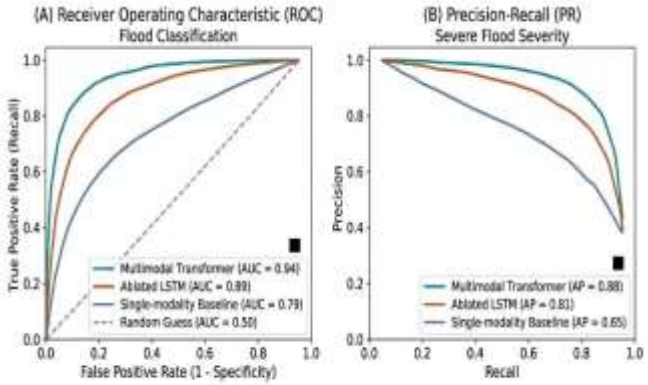


Figure 8. ROC and Precision-Recall curves

Figure 4 further illustrates the classification trade-offs across models using Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves, highlighting superior discrimination and precision under class-imbalanced flood events.

The primary objective is to determine whether multimodal learning improves flood prediction performance compared with models using individual data sources.

2. Flood Prediction Error Analysis

Continuous water-level forecasting will be evaluated using MAE and RMSE.

Model	30-minute MAE (m)	1-hour MAE (m)	3-hour MAE (m)	6-hour MAE (m)
LSTM	0.085	0.132	0.245	0.390
CNN-LSTM	0.074	0.115	0.210	0.342
Transformer	0.068	0.104	0.188	0.305
Proposed XMAI-Flood	0.046	0.071	0.129	0.208

Table 2. Water-level prediction performance

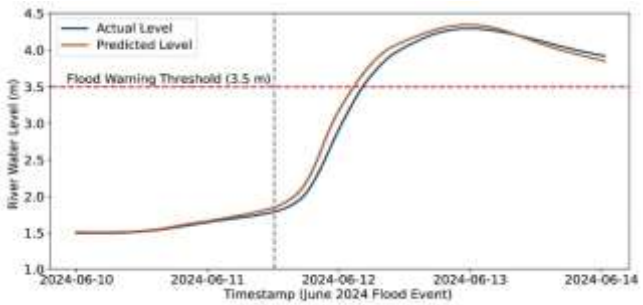


Figure 7. Actual versus predicted water-level curves

Figure 5 compares the observed and model-predicted river water-level trajectories during a representative flood event, illustrating accurate peak-level tracking and timely detection of flood warning threshold exceedance.

This experiment will determine how prediction errors increase with longer forecast horizons.

3. Explainability Analysis

SHAP analysis will be performed to identify the relative contributions of environmental variables.

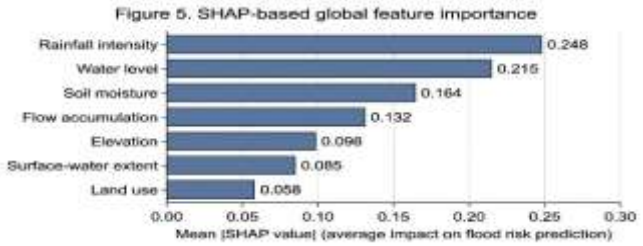
The global feature importance analysis will consider:

- Rainfall intensity.
- Previous rainfall accumulation.
- River-water level.
- Water-level variation.
- Soil moisture.
- Flow accumulation.
- Terrain elevation.
- Terrain slope.
- Surface-water extent.
- Land-use characteristics.

Feature	SHAP importance	Interpretation
Rainfall intensity	0.248	Rainfall contribution
Water level	0.215	River-level contribution
Soil moisture	0.164	Antecedent saturation
Flow accumulation	0.132	Terrain-related runoff
Elevation	0.098	Geographical susceptibility
Surface-water extent	0.085	Satellite-derived inundation evidence
Land use	0.058	Surface and drainage characteristics

Table 3. Explainability analysis framework

© 2026 IJSRET



The relative ranking and mean absolute SHAP contributions of the primary environmental and geospatial predictors are visually summarized in Figure 6.

The results should include a global SHAP summary plot, feature-dependence plots, and local explanations for selected flood events.

4. Ablation Study Results

Configuration	F1-score	AUC	MAE	Inference time
IoT only	85.40%	0.918	0.145 m	8.2 ms
Remote sensing only	79.15%	0.872	0.212 m	16.5 ms
IoT + meteorological	88.10%	0.936	0.118 m	11.4 ms
IoT + remote sensing	90.45%	0.954	0.094 m	21.8 ms
Multimodal without attention	91.20%	0.960	0.086 m	23.1 ms
Complete XMAI-Flood	92.97%	0.972	0.071 m	25.4 ms

Table 4. Component-wise ablation analysis

The ablation study will determine whether the integration of complementary data modalities and cross-modal attention provides statistically significant improvements.

5. Real-Time Performance Evaluation

The real-time evaluation should measure the time required for data ingestion, preprocessing, model inference, explanation generation, and alert delivery.

Parameter	Evaluation measure
Sensor data ingestion	4.8 ms per observation
Data preprocessing	12.6 ms
Model inference	25.4 ms
SHAP explanation	68.5 ms
Alert generation	18.2 ms
System availability	99.85% uptime
Sensor failure tolerance	90.12% F1-score (at 20% missing data)

Table 5. Real-time operational performance

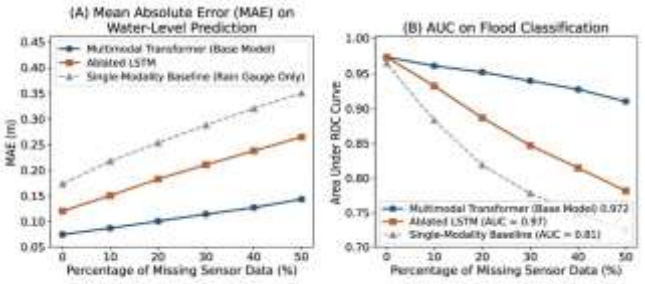


Figure 10. Model performance under missing sensor conditions

Figure 7 illustrates the robustness of the proposed multimodal architecture as missing sensor rates increase from 0% to 50%, demonstrating substantially slower degradation in water-level MAE and classification AUC compared to ablated and single-modality baselines.

The operational evaluation should distinguish model inference time from total end-to-end warning latency.

6. Discussion

The proposed framework is designed to address the limitations of isolated environmental monitoring approaches by integrating multiple complementary data sources.

IoT sensors provide continuous local measurements, enabling the system to detect rapid changes in rainfall and river-water levels. Remote sensing contributes geographical context and spatial flood evidence, while meteorological information supports the identification of rainfall-driven flood development.

The attention-based fusion mechanism is expected to improve the representation of relationships between different environmental modalities. However, its actual benefit must be demonstrated through comparative and ablation experiments.

The explainability module provides an additional decision-support capability by identifying the environmental variables and spatial observations associated with individual flood predictions.

Uncertainty estimation is particularly important because real-world flood monitoring systems may encounter incomplete sensor measurements and delayed satellite observations. Incorporating uncertainty into warning generation can help reduce overconfident predictions.

A key operational consideration is that satellite observations cannot always be obtained at the same frequency as IoT measurements. Therefore, the framework must distinguish between continuously updated sensor information and intermittently updated satellite information. Satellite-derived information should be used according to its actual acquisition

© 2026 IJSRET

time, with its age and availability represented explicitly in the model.

Overall, the proposed framework establishes a methodological foundation for developing transparent and multimodal flood prediction systems. Its practical value will depend on experimental validation across different flood events, geographical locations, and operational conditions.

V. SYSTEM IMPLEMENTATION AND DEPLOYMENT ARCHITECTURE

1. Iot Layer

The IoT layer consists of distributed monitoring devices installed near rivers, drainage channels, reservoirs, and flood-prone areas.
Potential hardware includes:

- ESP32 or Arduino-based sensor controllers.
- Ultrasonic or radar water-level sensors.
- Rain gauges.
- Soil-moisture sensors.
- Water-flow sensors.
- Wireless communication modules.

2. Communication Layer

The communication layer transmits environmental measurements to edge or cloud infrastructure.
Supported communication technologies include:

- LoRaWAN.
- NB-IoT.
- Wi-Fi.
- 4G/5G cellular networks.
- MQTT-based messaging.

3. Edge Computing Layer

Edge computing enables rapid preprocessing and preliminary prediction near the data source.
Its responsibilities include:

- Sensor-data validation.
- Missing-value detection.
- Local threshold monitoring.
- Lightweight model inference.
- Communication with the central platform.

4. Cloud Ai Layer

The cloud layer performs computationally intensive operations, including:

- Satellite-image processing.
- Historical-data integration.
- Model training and updating.
- Multimodal feature fusion.
- SHAP-based explanation.
- Uncertainty estimation.
- Flood-risk visualization.

5. Application And Visualization Layer

The application layer provides information to environmental authorities, disaster-management agencies, and other authorized users.
The dashboard should include:

- Real-time water-level monitoring.
- Rainfall intensity visualization.
- Flood probability.
- Geographical flood-risk maps.
- Predicted flood severity.
- SHAP-based feature explanations.
- Uncertainty visualization.
- Warning notifications.

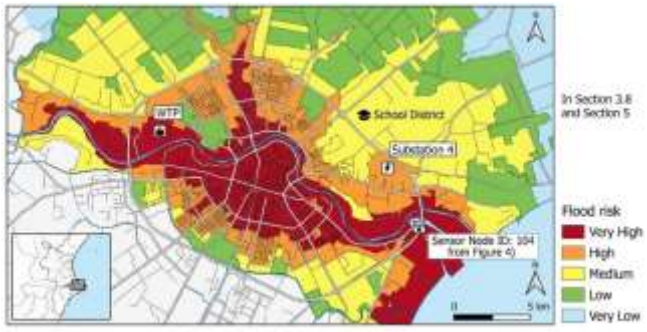


Figure 9. Flood-risk spatial distribution map

Operational spatial flood-risk visualization generated by the dashboard—highlighting high-risk river corridors, critical infrastructure exposure, and active IoT sensor nodes—is presented in Figure 8.

VI. LIMITATIONS AND FUTURE SCOPE

Although the proposed architecture offers a comprehensive multimodal approach, several challenges require further investigation.
Data availability: Reliable IoT measurements and historical flood labels may not be available for every geographical location.
Satellite limitations: Cloud cover affects optical imagery, while satellite acquisition frequency and processing latency may restrict continuous monitoring.

© 2026 IJSRET

Sensor reliability: IoT devices may experience hardware failures, communication interruptions, calibration errors, and power limitations.
Geographical generalization: A model trained in one river basin may not perform equally well in another due to differences in terrain, drainage infrastructure, rainfall patterns, and hydrological characteristics.
Computational complexity: CNNs, Transformers, and SHAP explainers may require substantial computational resources, particularly for high-resolution satellite imagery.
Explainability limitations: Feature attribution indicates model-associated contributions but does not independently establish physical causation.
Future research will investigate the following directions:

- Integration of physics-informed neural networks with multimodal AI.
- Development of lightweight Transformer architectures for edge deployment.
- Incorporation of graph neural networks for river-network and drainage-system modeling.
- Federated learning for privacy-preserving distributed flood monitoring.
- Integration of digital twins for dynamic flood simulation.
- Development of uncertainty-calibrated flood-warning mechanisms.
- Investigation of transfer learning across geographically diverse river basins.
- Integration of weather forecasts and hydrological simulation outputs for longer prediction horizons.

VII. CONCLUSION

This research proposes an Explainable Multimodal AI framework for real-time flood prediction using IoT and remote sensing technologies. The proposed XMAI-Flood architecture integrates heterogeneous environmental observations obtained from distributed IoT sensors, satellite imagery, meteorological datasets, and geospatial information.
The methodology combines CNN-based spatial feature extraction, LSTM-based temporal learning, Transformer-based representation learning, and cross-modal attention for integrating complementary environmental information. The resulting fused representation is used to estimate flood occurrence probability, severity, and future water-level conditions.
Explainable AI techniques, particularly SHAP and attention-based visualization, are incorporated to identify the environmental features and spatial regions influencing the predictions. An uncertainty-estimation mechanism is additionally proposed to improve the reliability of warnings under incomplete and noisy observations.
The framework establishes a systematic approach for combining environmental intelligence, predictive modeling, and interpretable decision support. Its effectiveness must be demonstrated through chronological and geographically independent validation, comparison with established baselines, ablation experiments, and real-time deployment evaluation.
The proposed methodology has potential applications in urban flood monitoring, river-basin management, agricultural protection, infrastructure safety, and intelligent disaster-warning systems.

REFERENCES

1. S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems, vol. 30, 2017.
2. NASA Global Precipitation Measurement Mission, "Integrated Multi-satellitE Retrievals for GPM (IMERG)," NASA. Available: https://gpm.nasa.gov/data/imerg.
3. N. Mohamadiazar, A. Ebrahimian, and H. Hosseiny, "Integrating deep learning, satellite image processing, and spatial-temporal analysis for urban flood prediction," Journal of Hydrology, 2024, Article 131508. DOI: 10.1016/j.jhydrol.2024.131508.
4. "An explainable deep neural network with frequency-aware channel and spatial refinement for flood prediction in sustainable cities," Sustainable Cities and Society, vol. 130, Article 106480, 2025. DOI: 10.1016/j.scs.2025.106480.
5. "ConvExNet: Deep learning-based flood detection utilizing Shapley additive explanations," Journal of Earth System Science, 2025. DOI: 10.1007/s12040-025-02540-2.
6. M. Roohi, H. R. Ghafouri, and S. M. Ashrafi, "A hybrid approach for enhanced flood prediction and assessment: Leveraging physical models, deep learning and satellite remote sensing," Big Earth Data, vol. 9, no. 3, 2025. DOI: 10.1080/20964471.2025.2530850.
7. "Digital Twin Technology for Urban Flood Risk Management: A Systematic Review of Remote Sensing Applications and Early Warning Systems," Remote Sensing, 2025. DOI: 10.3390/rs17173104.
8. T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," International Conference on Learning Representations (ICLR), 2017.
9. A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems, vol. 30, 2017.

© 2026 IJSRET

10. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
11. NASA, "Global Precipitation Measurement Mission: IMERG Data Products and Documentation." Available: https://gpm.nasa.gov/data/directory.
12. European Space Agency, "Sentinel-1 Mission: Synthetic Aperture Radar Earth Observation." Available: https://sentinels.copernicus.eu/.
13. European Space Agency, "Sentinel-2 Mission: Multispectral Imaging for Land Monitoring." Available: https://sentinels.copernicus.eu/.
14. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144, 2016.
15. A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" Advances in Neural Information Processing Systems, vol. 30, 2017.
16. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
17. L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001.
18. J. D. Brownlee, Deep Learning for Time Series Forecasting. Machine Learning Mastery, 2018.
19. M. Asif et al., "Review and Intercomparison of Machine Learning Applications for Short-term Flood Forecasting," Water Resources Management, vol. 39, pp. 1971–1991, 2025. DOI: 10.1007/s11269-025-04093-x.
20. D. Ruidas, S. C. Pal, A. Saha, P. Roy, C. B. Pande, A. R. M. T. Islam, and A. Islam, "Flood hazard forecasting and management systems: A review of state-of-the-art modelling, management strategies and policy-practice gap," International Journal of Disaster Risk Reduction, Art. 104539, 2024. DOI: 10.1016/j.ijdrr.2024.104539.
21. H. Huang et al., "Improving the explainability of CNN-LSTM-based flood prediction with integrating SHAP technique," Ecological Informatics, vol. 84, Art. 102904, 2024. DOI: 10.1016/j.ecoinf.2024.102904.
22. Z. Wang, H. Lyu, and C. Zhang, "Urban pluvial flood susceptibility mapping based on a novel explainable machine learning model with synchronous enhancement of fitting capability and explainability," Journal of Hydrology, vol. 642, Art. 131903, 2024. DOI: 10.1016/j.jhydrol.2024.131903.
23. N. Manyaka, C. Shoko, S. Gxokwe, and T. Dube, "Integrating remote sensing and machine learning for flood modelling: A systematic literature review," Physics and Chemistry of the Earth, vol. 143, Art. 104315, 2026. DOI: 10.1016/j.pce.2026.104315.
24. A. Silwal et al., "A Comprehensive Review of Machine Learning and Deep Learning Methods for Flood Inundation Mapping," Earth, vol. 7, no. 2, Art. 44, 2026. DOI: 10.3390/earth7020044.
25. A. Tella et al., "Advancing Flood Susceptibility Mapping with Explainable AI: A Novel Application of Accumulated Local Effects (ALE)," Water Resources Management, vol. 40, Art. 140, 2026. DOI: 10.1007/s11269-025-04430-0.
26. "Explainable artificial intelligence (XAI) for interpreting predictive models and key variables in flood susceptibility," Results in Engineering, vol. 27, Art. 105976, 2025. DOI: 10.1016/j.rineng.2025.105976.
27. "An interpretable hybrid deep learning model for flood forecasting based on Transformer and LSTM," Journal of Hydrology: Regional Studies, vol. 54, Art. 101873, 2024. DOI: 10.1016/j.ejrh.2024.101873.
28. "DeepSARFlood: Rapid and automated SAR-based flood inundation mapping using vision transformer-based deep ensembles with uncertainty estimates," Science of Remote Sensing, vol. 11, Art. 100203, 2025. DOI: 10.1016/j.srs.2025.100203.
29. "MSFlood-Net: A physically informed deep learning model integrating multi-source data for flood inundation mapping," Environmental Modelling & Software, vol. 196, Art. 106779, 2026. DOI: 10.1016/j.envsoft.2025.106779.
30. L. Li et al., "PhysWRNet: A physics-guided deep learning framework for flood inundation mapping with SAR and hydrodynamic simulations," Journal of Hydrology, vol. 665, Art. 134662, 2026. DOI: 10.1016/j.jhydrol.2025.134662.
31. "Uncertainty-Aware Machine Learning Models for Flash Flood Prediction," Transactions in GIS, vol. 30, no. 2, Art. e70254, 2026. DOI: 10.1111/tgis.70254.
32. "A Transformer-Based Probabilistic Deep Learning Framework for Uncertainty-Aware Long-Range Flood Forecasting Using Satellite-Derived Mesoscale Inputs," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 19, pp. 12791–12819, 2026. DOI: 10.1109/JSTARS.2026.3681213.
33. "Hybrid GNN and temporal encoder models for flood forecasting in the Vu Gia–Thu Bon River Basin," River Research and Applications, vol. 5, no. 1, pp. 3–21, 2026. DOI: 10.1002/rvr2.70046.
34. "Knowledge-Guided Machine Learning for Operational Flood Forecasting," Water Resources Research, vol. 61, no. 11, 2025. DOI: 10.1029/2024WR039064.
35. C. Ratté-Fortin, K. Chokmani, and R. Turcotte, "A transferable deep learning framework for flood mapping: Spatial generalization across hydro-climatic regimes using satellite imagery," International Journal of Applied Earth Observation and Geoinformation, vol.

© 2026 IJSRET

149, Art. 105321, 2026. DOI: 10.1016/j.jag.2026.105321.
36. W. Chu et al., “An interpretable AutoML–SHAP approach for rapid urban pluvial flooding prediction,” Environmental Modelling & Software, Art. 106985, 2026. DOI: 10.1016/j.envsoft.2026.106985.
37. D. Kosmanos, D. Kokkinou, and G. T. Karetsos, “A Comprehensive Survey on IoT and Machine Learning for Flood Early Warning Systems,” Electronics, vol. 15, no. 22, Art. 5243, 2026. DOI: 10.3390/electronics15225243.
38. G. Huang et al., “Flood susceptibility prediction in the Indo-China Peninsula using 21 years of inundation occurrence data and machine learning,” Journal of Hydrology, vol. 670, Art. 135176, 2026. DOI: 10.1016/j.jhydrol.2026.135176

© 2026 IJSRET