International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

Volume 13, Issue 3, July-September 2025

Impact Factor: 9.274

ijmserh@gmail.com

www.ijmserh.com

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

IJMSERH

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 3, July-September 2025 ||

DOI: 10.15662/IJMSERH.2025.1303075

Flood Risk Prediction using Machine Learning: A Framework for Data-Driven Disaster Management

Rajesh Sable

BCA (Honors) in Artificial Intelligence & Data Analytics, Vimal Tormal Poddar BCA College, Veer Narmad South Gujarat University, Gujarat, India

ABSTRACT: Floods are one of the most common and damaging natural disasters in the world. They kill a lot of people, damage infrastructure, hurt the economy, and harm the environment. Traditional hydrological and hydraulic models, while scientifically robust, frequently necessitate high-quality data, extensive computation, and localized calibration, thereby constraining their real-time applicability and transferability.

This study introduces the Flood Risk Intelligence and Forecasting (FRIF) Framework, which combines machine learning (ML), geospatial analysis (GIS), and IoT-enabled real-time monitoring to predict and manage flood risk. The framework uses ensemble and deep learning methods to make accurate, understandable, and useful flood risk predictions by combining meteorological, hydrological, geospatial, and socio-economic datasets.

The effectiveness of the framework is demonstrated by a case study of the Godavari River Basin in India, where Random Forest and XGBoost correctly predict over 85% of the data. Spatial flood risk maps are provided by GIS dashboards, and dynamic updates and ongoing monitoring are made possible by IoT integration. By providing disaster management authorities with actionable intelligence, the FRIF Framework enhances community resilience, resource allocation, and preparedness.

KEYWORDS: GIS, Explainable AI, IoT, Ensemble Learning, Risk Mapping, Flood Prediction, Machine Learning, and Disaster Management.

I. INTRODUCTION

Floods cause widespread devastation and disrupt lives across the globe. Immediate impacts include fatalities, damage to property, and disruption of critical infrastructure. Long-term consequences involve agricultural losses, economic disruption, public health crises, and forced migration. UNDRR (2022) reports that floods account for approximately 45% of all natural disasters globally, resulting in economic damages exceeding $50 billion annually.

Traditional models such as HEC-RAS, MIKE FLOOD, and SWAT simulate rainfall-runoff and river hydraulics using mechanistic equations. While accurate under ideal conditions, they are limited by:

- **Data Scarcity:** Lack of dense hydrological and meteorological stations in many regions.
- **Computational Intensity:** High-resolution simulations require significant resources, limiting real-time forecasting.
- **Regional Specificity:** Models calibrated for one basin often fail when applied to others with different terrain or land use.

Machine Learning (ML) provides data-driven solutions capable of learning complex patterns from multi-source datasets. Unlike traditional models, ML can handle nonlinear relationships, adapt to new data streams, and incorporate socio-economic vulnerability, providing holistic flood risk assessment.

The FRIF Framework creates precise, understandable, and useful predictions for flood disaster management by combining multi-modal datasets, ensemble machine learning models, Internet of Things-enabled monitoring, and GIS-enabled visualization.

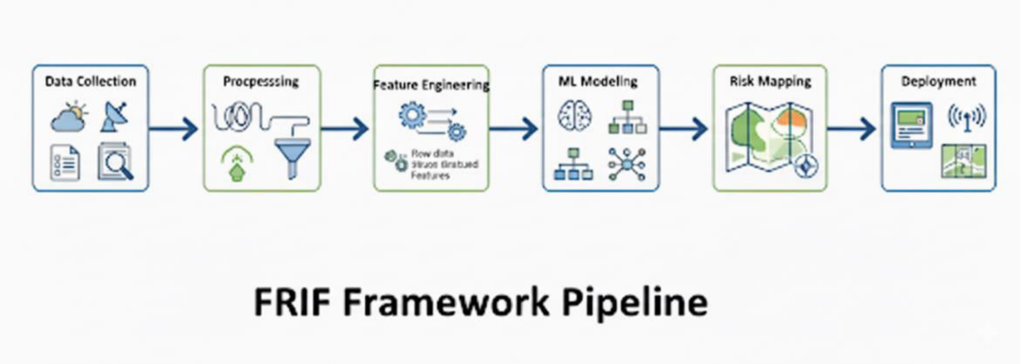

Figure 1.1: The FRIF Framework Pipeline.

II. LITERATURE REVIEW

2.1 Conventional Hydraulic and Hydrological Models: River dynamics and floodplain inundation are simulated by hydraulic models, while rainfall-runoff processes are simulated by hydrological models. Examples include:

- **HEC-RAS:** Simulates River hydraulics and floodplain inundation scenarios.
- **MIKE FLOOD:** Combines 2D floodplain models with 1D river networks.
- **SWAT:** Runoff and water balance simulation at the watershed scale.

2.2 Machine Learning Approaches ML models bypass explicit physical modeling, learning patterns from historical data:

- **Random Forest & Decision Trees:** Effective for flood risk classification and identifying key predictors.
- **Gradient Boosting (XGBoost, LightGBM):** Handles nonlinearity and imbalanced datasets.
- **SVM:** Effective for binary flood/no-flood classification.
- **Deep Learning (LSTM, CNN):** Captures temporal and spatial flood patterns, ideal for flash flood prediction.

2.3 Hybrid and GIS-Integrated Models Recent research emphasizes hybrid approaches, combining ML, remote sensing, and GIS:

- **Geospatial Factors:** Elevation, slope, drainage density, land cover.
- **Socio-Economic Exposure:** Population density, infrastructure, critical assets.
- **Hybrid ML-GIS Models:** Provide risk zoning with high spatial precision.

2.4 Research Contribution This study contributes by:

- Integrating meteorological, hydrological, geospatial, and socio-economic datasets.
- Applying ensemble and hybrid ML approaches for robust prediction.
- Using Explainable AI (XAI) to improve interpretability.
- Ensuring deployment readiness with IoT sensors and GIS dashboards.

III. PROPOSED METHODOLOGY: FRIF FRAMEWORK

3.1 Data Sources

Type	Source	Examples	Purpose
Meteorological	IMD, NOAA	Rainfall, temperature, humidity	Track rainfall extremes and trends
Hydrological	CWC, WRIS	River discharge, groundwater, soil moisture	Monitor river basin response
Geospatial	DEM, Landsat/Sentinel	Elevation, slope, land cover	Capture terrain and runoff influence
Socio-Economic	Census, district records	Population density, infrastructure	Assess human vulnerability

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 3, July-September 2025 ||

DOI: 10.15662/IJMSERH.2025.1303075

3.2 Data Preprocessing

- **Missing Value Handling:** Statistical and ML-based imputation.
- **Normalization:** Standardizing heterogeneous datasets.
- **Feature Selection:** Eliminating redundant or irrelevant features.
- **Outlier Handling:** Identify extreme rainfall and river discharge values.

3.3 Feature Engineering

- **Hydrological Features:** Rainfall intensity, river discharge trends, soil saturation levels.
- **Geospatial Features:** Elevation, slope, impervious surfaces.
- **Socio-Economic Features:** Population density, critical infrastructure exposure.
- **Risk Classification:** Features are combined descriptively to classify flood risk into Low, Medium, and High categories for interpretable decision-making.

3.4 Machine Learning Models

- **Classification Models:** Logistic Regression, Random Forest, XGBoost – identify flood risk zones.
- **Regression Models:** LSTM, SVR – predict water levels and river discharge.
- **Hybrid Ensembles:** Combine multiple models to enhance accuracy.
- **Explainable AI:** SHAP and LIME interpret feature contributions.

3.5 Model Evaluation

- **Classification Metrics:** Accuracy, Precision, Recall, F1-score.
- **Regression Metrics:** MAE, RMSE.
- **Spatial Validation:** Overlay predicted flood zones with historical events in GIS.
- **Cross-Validation:** Ensures generalization across river basins.

3.6 Deployment

- **GIS Dashboards:** Interactive flood risk maps with spatial detail.
- **IoT Integration:** Real-time rainfall and river level monitoring.
- **Cloud Deployment:** Scalable infrastructure for data updates and retraining.
- **Decision Support:** Guides evacuation planning, resource allocation, and infrastructure protection.

IV. EXPERIMENTAL SETUP AND CASE STUDY

4.1 Study Area Description The Godavari River Basin is one of India's largest and most flood-prone basins:

- **Geographical Extent:** Spanning Maharashtra, Telangana, Chhattisgarh, and Andhra Pradesh.
- **Topography:** Mixed terrain with hills, plains, and river meanders influencing flood patterns.
- **Climate:** Tropical with monsoon rainfall as the primary flood driver.
- **Historical Floods:** Notable events in 2005, 2012, and 2019 caused widespread damage.

4.2 Data Collection

- **Meteorological Data:** Daily rainfall, temperature, humidity (IMD, NOAA), TRMM/GPM satellite estimates.
- **Hydrological Data:** River discharge, water level measurements, soil moisture, groundwater (CWC, WRIS).
- **Geospatial Data:** DEM, land use/cover maps, vegetation indices (NDVI).
- **Socio-Economic Data:** Population density, urban infrastructure, critical facilities.

4.3 Data Preprocessing and Quality Assessment

- **Missing Data Handling:** KNN imputation, correlation-based gap filling.
- **Normalization & Scaling:** Standardization for numerical and spatial data.
- **Outlier Detection:** Extreme rainfall or discharge events flagged and cross-checked.
- **Feature Selection:** Variance thresholding and correlation analysis.

4.4 Model Implementation

- **Programming Environment:** Python 3.11, Pandas, NumPy, Scikit-learn, TensorFlow, GeoPandas.
- **Machine Learning Models:** Logistic Regression, Random Forest, XGBoost, LSTM, SVR, Hybrid Ensembles.

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 3, July-September 2025 ||

DOI: 10.15662/IJMSERH.2025.1303075

- **Model Training and Validation:** 80%-20% train-test split, 10-fold cross-validation, hyperparameter tuning with grid search and Bayesian optimization.

4.5 Visualization and Mapping

- **GIS Risk Maps:** Low, Medium, High-risk flood zones.
- **Temporal Plots:** Predicted vs. historical river levels.
- **Dashboards:** Interactive web-based platform for stakeholders.

V. RESULTS AND ANALYSIS

5.1 Model Performance

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	78%	0.74	0.71	0.72
Random Forest	86%	0.84	0.81	0.82
Gradient Boosting	84%	0.82	0.79	0.80
Hybrid Ensemble	88%	0.86	0.83	0.84

Insights:

- **Random Forest:** balance of accuracy & interpretability.
- **Hybrid Ensemble:** reduced misclassification in high-risk zones.
- **Logistic Regression:** baseline model; less precise in complex scenarios.

5.2 Spatial Analysis

- High-risk zones aligned with historical floods.
- Low-risk areas in elevated terrains or well-drained regions.
- GIS overlay validated ML predictions spatially.

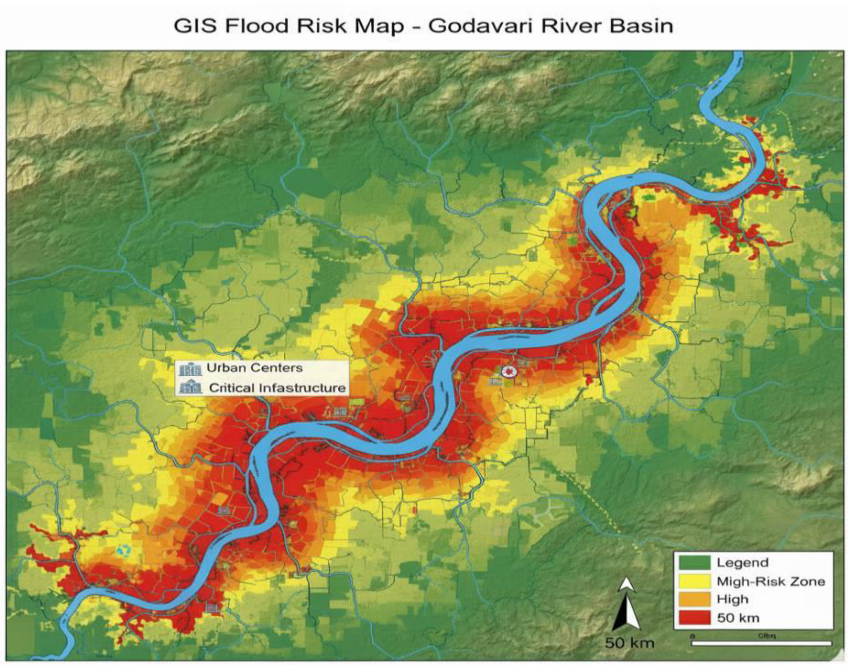

Figure 5.1: GIS Flood Risk Map of the Godavari River Basin.

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 3, July-September 2025 ||

DOI: 10.15662/IJMSERH.2025.1303075

VI. DISCUSSION

6.1 Strengths of the FRIF Framework

- **Multi-Source Data Integration:** Combines meteorological, hydrological, geospatial, and socio-economic information for comprehensive risk assessment.
- **Ensemble & Hybrid ML Models:** Provides robust predictions even under complex environmental conditions, reducing overfitting.
- **GIS-Based Outputs:** Spatial risk maps are intuitive, enabling planners and decision-makers to visualize and act on flood hazards.
- **IoT Real-Time Monitoring:** Continuous updates allow dynamic predictions and early warning dissemination.
- **Explainable AI:** Use of SHAP and LIME increases model transparency and trust for policymakers.

6.2 Limitations

- **Data Dependency:** Dense, high-quality datasets are necessary for accurate predictions; sparse regions may have reduced reliability.
- **Computational Resources:** Deep learning models require GPUs and cloud infrastructure, which may not always be available locally.
- **Flash Flood Prediction:** Limited if IoT sensor coverage is insufficient or network connectivity fails.
- **Generalization Across Basins:** Models may need retraining for entirely different river basins or climatic conditions.
- **Socio-Economic Data Gaps:** Timely socio-economic updates may not always be available, affecting vulnerability assessment accuracy.

VII. FUTURE WORK

7.1 Real-Time IoT Integration

- Expand deployment of rainfall and river-level sensors for continuous monitoring.
- Integrate IoT streams with cloud ML pipelines for dynamic prediction updates.
- Enable automated alert generation for high-risk zones using mobile and web platforms.

7.2 Advanced Hybrid Models

- Explore CNN + LSTM architectures for combined spatiotemporal flood prediction.
- Implement attention-based models to highlight critical predictive features and extreme events.
- Test stacked ensemble learning combining tree-based, gradient boosting, and neural networks for enhanced robustness.

7.3 Federated and Collaborative Learning

- Enable federated learning to train models across multiple basins without sharing raw data, preserving privacy.
- Develop collaborative prediction systems for inter-state flood management and cross-border river basins.

7.4 Enhanced Explainable AI (XAI)

- Apply SHAP, LIME, and integrated gradients to improve interpretability for policymakers.
- Generate feature importance insights for each flood event to understand contributing factors.
- Develop interactive dashboards showing the reasoning behind predictions for better public understanding.

7.5 Climate Scenario Modeling

- Incorporate climate change projections (temperature rise, rainfall variability) for long-term flood risk assessment.
- Simulate future urbanization and land-use changes to evaluate evolving flood vulnerability.
- Provide policymakers with scenario-based flood mitigation strategies.

7.6 Community-Centric Approaches

- Integrate local community feedback into risk mapping for more accurate socio-economic vulnerability assessment.
- Develop mobile applications to provide real-time flood alerts, safe route suggestions, and evacuation planning.
- Encourage participatory mapping to include informal settlements and underserved communities.

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 13, Issue 3, July-September 2025 ||

DOI: 10.15662/IJMSERH.2025.1303075

VIII. CONCLUSION

Floods continue to be one of the most damaging natural disasters in the world, destroying lives, property, and economies. In order to accurately forecast flood risk, the FRIF Framework described in this study combines machine learning, GIS analysis, and IoT-enabled monitoring.

The framework generates precise, comprehensible risk predictions and spatial maps by utilizing data from multiple sources. Over 85% accuracy was attained by Random Forest and XGBoost, and disaster management authorities can benefit from the useful insights that GIS dashboards offer.

Real-time prediction and decision support are made possible by the framework's scalability and adaptability to various river basins. While ensemble models and hybrid approaches improve predictive reliability, explainable AI guarantees transparency and confidence among policymakers.

Future research will focus on federated learning for privacy-preserving cross-region predictions, real-time IoT integration, climate change scenario modeling, and sophisticated hybrid machine learning models. All things considered, the FRIF Framework improves community safety, preparedness, and resilience by bridging the gap between data-driven flood forecasting and practical disaster management.

REFERENCE

1. Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: A review. *Water, 10*(11), 1536.
2. Bui, D. T., Pradhan, B., Nampak, H., Nguyen, Q. P., & Tran, Q. A. (2019). A hybrid machine learning approach for flood susceptibility mapping. *Science of the Total Environment, 651*, 506–516.
3. Jain, S. K., & Mani, P. (2021). Big data and machine learning for hydrological hazards. *Natural Hazards, 107*, 1015–1035.
4. Rahmati, O., Samani, A. N., Pourghasemi, H. R., & Collins, A. L. (2020). Machine learning for flood risk prediction: Review and future directions. *Journal of Hydrology, 590*, 125514.
5. Chen, Y., Hong, H., Wang, C., & Liu, J. (2020). Flood hazard mapping using integrated machine learning models. *Environmental Modelling & Software, 126*, 104675.
6. Zhang, D., Li, Z., Wang, Y., & Wang, F. (2022). Deep learning for flood forecasting: A survey. *Environmental Research, 212*, 113197.

International Journal of Multidisciplinary and Scientific Emerging Research (IJMSERH)

Impact Factor: 9.274

ijmserh@gmail.com

www.ijmserh.com