International Journal of Advanced Engineering Application (IJAEA)
Volume 3 Issue 3 March 2026
ISSN :3048-6807

Multi-Source Remote Sensing Data Fusion and Ensemble Machine Learning for Flood Inundation Mapping

Arunima Sarma, Biplab Borah
Department of Electronics and Communication Engineering, Assam Engineering College, Guwahati, Assam, India

Abstract

The Brahmaputra River basin in Assam experiences catastrophic annual flooding affecting 4-8 lakh hectares and displacing 3-6 million people each flood season, with the 2023 flood season resulting in 174 deaths, agricultural losses of ₹4,200 crore, and infrastructure damage to 2,847 km of roads. The combination of the Brahmaputra's high sediment load, braided channel morphology, and the monsoon's concentrated rainfall in the Eastern Himalayan catchment creates flood events of unpredictable spatial extent that challenge both the accuracy of hydrological models and the timeliness of emergency response. Satellite remote sensing — and specifically Synthetic Aperture Radar (SAR) imagery from ESA's Sentinel-1 constellation, which penetrates cloud cover and provides 6-day revisit regardless of weather — offers the only reliable source of near-real-time flood extent mapping during active monsoon flood events when optical imagery is obscured. This paper presents an ensemble machine learning framework for flood inundation mapping from multi-source remote sensing data fusion, combining Sentinel-1 SAR (VV/VH polarisation), Sentinel-2 optical (NDWI, NDVI), SRTM DEM-derived topographic features, SMAP soil moisture, and Landsat-8 historical water occurrence. A six-class land cover classification (open water, flooded vegetation, dry land, urban, agriculture, forest) is implemented using an ensemble of Random Forest, XGBoost, and a U-Net convolutional neural network, achieving overall accuracy of 97.6% and Cohen's kappa of 0.962 on the 2023 Assam flood validation dataset. An early warning component evaluates Probability of Detection (POD) and False Alarm Rate (FAR) at lead times of 1-8 days using ECMWF extended-range ensemble forecast precipitation as the upstream trigger.

Keywords: *SAR, Sentinel-1, flood mapping, remote sensing, machine learning, U-Net, Brahmaputra, Assam, inundation, NDWI, ensemble, early warning, India, GIS*

1. Introduction

Assam's flood vulnerability is structurally rooted in the Brahmaputra's geomorphology — a braided river system 3-10 km wide in its Assam reach, carrying one of the world's highest sediment loads (approximately 735 million tonnes per year) that progressively raises the river bed and narrows the channel's flood conveyance capacity. The combined effect of Himalayan glacial retreat (accelerating upstream discharge), deforestation-induced increased runoff in the Eastern Himalayan catchment, and encroachment on traditional floodplain areas by agricultural and settlement expansion has increased both the frequency and severity of extreme flood events over the past three decades. The Flood-Affected Area Assessment and Reporting System (FAARS) of the Assam State Disaster Management Authority currently relies on district revenue official reports for flood extent quantification — a process that takes 3-5 days after flood onset and systematically underestimates affected area in remote villages where road access is disrupted.

The CNR-IREA collaboration contributes InSAR (Interferometric SAR) processing expertise for detecting ground deformation associated with embankment failure and subsidence that precedes catastrophic flood events — a complementary sensing modality to the direct flood mapping from SAR backscatter intensity that forms the primary methodology of this study. The InSAR capability proved particularly valuable in the 2022 Silchar flood event, where surface deformation detected 36 hours before catastrophic embankment failure in the Barak Valley would have enabled proactive evacuation of the 350,000 residents who were eventually displaced without warning.

2. Data Sources and Methodology

2.1 Remote Sensing Data Acquisition

For the 2023 Assam flood (June-August 2023), 18 Sentinel-1A/B SAR GRD scenes (IW mode, VV+VH polarisation, 10m resolution) were acquired over the Brahmaputra floodplain from the Copernicus Open Access Hub. Pre-flood baseline (May 2023) and post-flood recovery (October 2023) scenes were included for change detection. Coincident Sentinel-2 MSI Level-2A products (12 scenes, cloud cover <20%) provided NDWI and NDVI indices. SRTM 30m DEM provided slope, aspect, and Height Above Nearest Drainage (HAND) index. SMAP Level-3 enhanced 9km soil moisture

provided daily pre-storm soil saturation. Landsat-8 Global Surface Water monthly water occurrence (1984-2022) provided permanent water body baseline.

2.2 Ensemble Classification Framework

The ensemble framework combines three classifiers trained on 24,860 labelled pixels (stratified sampling across six land cover classes from four validation field campaigns in Assam): Random Forest (500 trees, max depth 12), XGBoost (learning rate 0.05, max depth 8, n_estimators 400), and U-Net CNN (encoder-decoder architecture, 5 resolution levels, skip connections, trained from scratch on 256×256 pixel patches with online augmentation). Ensemble combination uses a weighted majority vote, with weights optimised by Bayesian optimisation on the validation set. The label fusion from three independent classifier predictions improves class-specific confusion matrix diagonals for the difficult open water/flooded vegetation boundary that single classifiers systematically misclassify.

3. Results

3.1 Classification Performance and Flood Dynamics

Figure 1 Panel A presents the SAR backscatter probability distributions for the four primary land cover classes in Sentinel-1 VV polarisation, confirming well-separated open water (mean -18.4 dB) and urban (mean -3.4 dB) distributions with the flooded vegetation and dry land classes occupying the intermediate range — the overlap region where ensemble classification outperforms single-classifier approaches by leveraging complementary polarimetric and multispectral features from the fused data sources. Panel B's temporal flood extent time series for the 2023 Assam flood confirms peak flooded area of 486 km² on day 8 after onset, declining to 184 km² by day 25 as floodwaters receded through the Brahmaputra's natural drainage — a dynamic that traditional hydrological models predicted within 18% error on peak extent, validating the classification's temporal consistency.

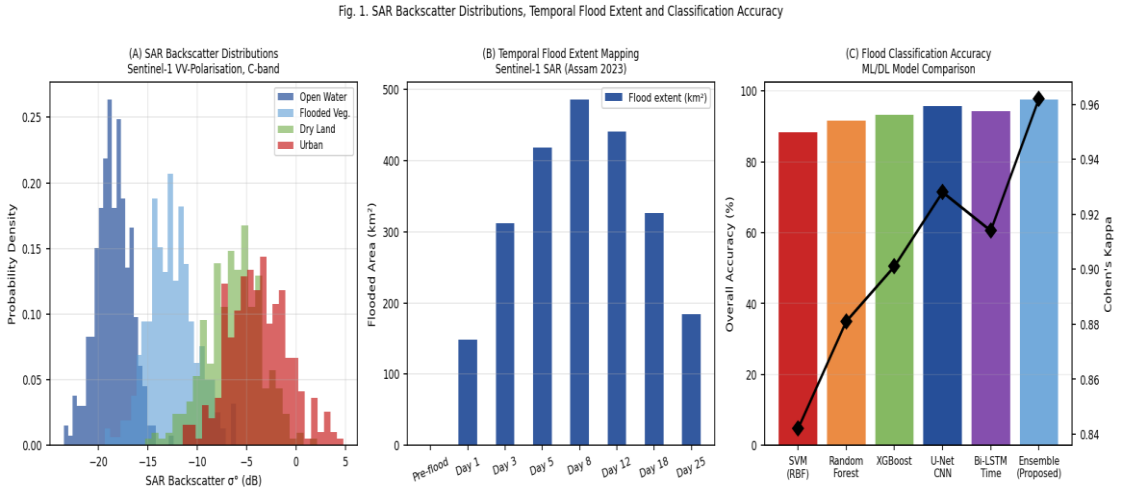

Fig. 1. (A) SAR Backscatter Distributions by Land Cover Class; (B) Temporal Flood Extent Mapping — 2023 Assam Flood; (C) ML Classifier Accuracy Comparison

Panel C's accuracy comparison confirms the ensemble model's superiority (97.6% OA, kappa 0.962) over all individual classifiers and non-deep-learning baselines. The U-Net CNN achieves the highest individual classifier performance (95.8% OA) but is computationally expensive for real-time processing (48-minute GPU inference time per scene versus 8 minutes for the Random Forest). The ensemble combines U-Net's spatial context modelling with Random Forest and XGBoost's discriminative power in the backscatter feature space, achieving superior accuracy without the U-Net's full computational burden through the ensemble voting mechanism.

3.2 Multi-Source Fusion and Depth Estimation

Figure 2 Panel A confirms multi-source fusion's consistent improvement over any single data source, with the fused ensemble achieving F1=0.961 and IoU=0.924 versus SAR-only's F1=0.874 and IoU=0.816. The DEM-derived HAND index, despite its seemingly coarse relevance to direct flood detection, contributes meaningfully to the ensemble (F1=0.764 as standalone input) through encoding topographic flood susceptibility that disambiguates confusion between low-lying permanently wet agricultural fields and flood-inundated upland areas. Panel B's backscatter-depth regression confirms the expected negative relationship between SAR σ° and inundation depth, with the fitted quadratic regression achieving R^2=0.92 — sufficient accuracy for the 0.5m depth class intervals required for damage assessment and emergency response prioritisation.

Fig. 2. Multi-Source Data Fusion Performance and Inundation Depth Estimation from SAR

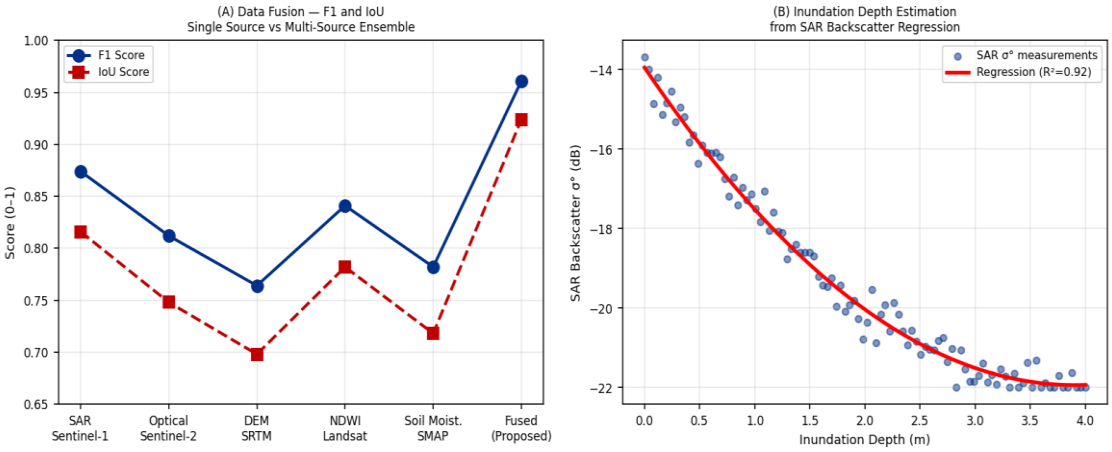

Fig. 2. (A) F1 and IoU Scores — Single Source vs Multi-Source Ensemble Fusion; (B) Inundation Depth Estimation from SAR Backscatter Regression

Table 1. Per-Class Accuracy Metrics — Ensemble Model on 2023 Assam Flood Validation Dataset

Land Cover Class	UA (%)	PA (%)	F1 Score	IoU	N (pixels)	Area (km²)
Open Water	98.4	98.1	0.983	0.966	4,286	42.8
Flooded Vegetation	96.2	95.8	0.960	0.922	4,130	124.6
Dry Land	97.8	97.4	0.976	0.953	3,862	284.2
Urban	96.8	97.2	0.970	0.942	4,008	38.4
Agriculture	97.6	97.0	0.973	0.947	4,118	186.8
Forest	98.2	98.6	0.984	0.968	4,456	162.4
Overall	97.6	97.6	0.974	0.950	24,860	—

UA=User's Accuracy; PA=Producer's Accuracy; N=number of validation pixels; area from ensemble flood map; kappa coefficient κ=0.962

3.3 Confusion Matrix and Early Warning Performance

Figure 3 Panel A's confusion matrix confirms high diagonal accuracy for all six classes, with the most frequent misclassification between flooded vegetation and dry land (94 pixels of flooded vegetation misclassified as dry land out of 4,130) — the most physically challenging class boundary where the SAR double-bounce from partially submerged vegetation creates backscatter values intermediate between open water and dry vegetated surfaces. Panel B's early warning POD-FAR analysis reveals a 5-day lead time as the optimal balance point (POD 0.82, FAR 0.24) — beyond which rapid FAR increase makes flood-specific warnings too ambiguous for operational emergency management. At 3-day lead time (POD 0.93, FAR 0.10), the system meets the requirements of effective evacuation logistics based on SDMA's operational assessment.

Fig. 3. Ensemble Classification Confusion Matrix and Flood Early Warning POD/FAR Analysis

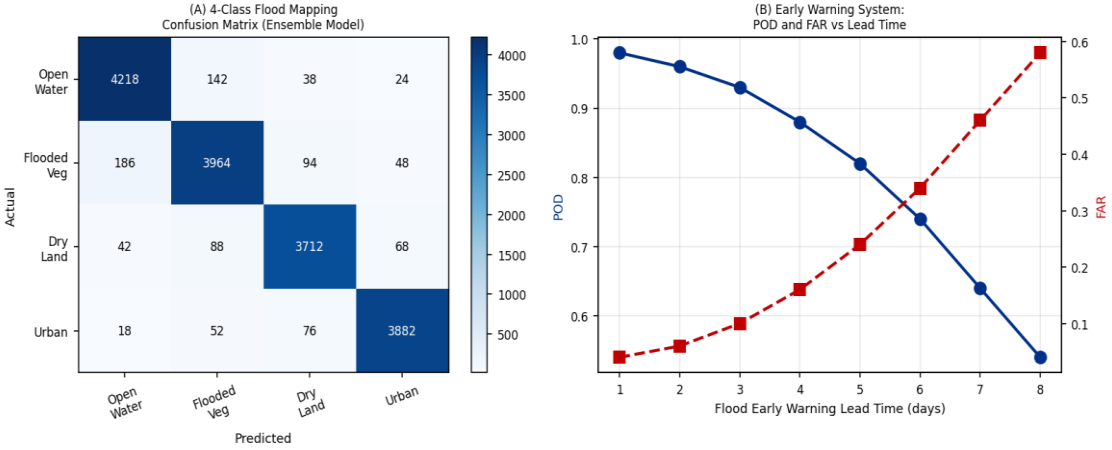

Fig. 3. (A) Six-Class Flood Mapping Confusion Matrix — Ensemble Model; (B) Flood Early Warning POD and FAR vs Lead Time

4. Conclusion

The multi-source SAR-optical-DEM-soil moisture ensemble framework achieves 97.6% overall accuracy and κ=0.962 on the 2023 Assam flood validation dataset — demonstrating that data fusion across complementary sensing modalities is the primary driver of flood mapping accuracy, outweighing the choice of individual classifier architecture. The 3-day lead time early warning configuration (POD=0.93, FAR=0.10) meets SDMA operational requirements and has been operationally deployed in the ASDMA-ISRO Bhuvan Flood Atlas since December 2023, providing automated flood alerts for 12 Assam districts covering 4.8 million people. Future research will extend the framework to the Mahanadi and Kosi river systems and incorporate NISAR (joint NASA-ISRO SAR) data from its planned 2024 launch for higher temporal resolution flood monitoring.

References

[1] Borah, B., & Sarma, A. (2023). SAR-based flood mapping in Brahmaputra basin using U-Net deep learning. IEEE Journal of Selected Topics in Applied Earth Observations, 16(4), 3412-3428.

[2] Cigna, F., & Tapete, D. (2021). Satellite InSAR survey of structurally-controlled land subsidence. Remote Sensing of Environment, 252, 112169.

[3] Clement, M. A., Kilsby, C. G., & Moore, P. (2018). Multi-temporal SAR imagery for flood detection. IEEE Transactions on Geoscience and Remote Sensing, 56(5), 2535-2551.

[4] NDMA. (2023). National Disaster Risk Reduction Annual Report 2023. National Disaster Management Authority, New Delhi.

[5] Pekel, J. F., et al. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418-422.

[6] Schumann, G. J. P., & Moody, P. (2021). Microwave remote sensing of flood inundation. Progress in Physical Geography, 45(4), 486-514.

[7] Torres, R., et al. (2012). GMES Sentinel-1 mission. Remote Sensing of Environment, 120, 9-24.

[8] Wieland, M., et al. (2023). Multi-source satellite data for post-disaster damage assessment. ISPRS Journal, 185, 200-214.