https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

"Review article: Hydrologically Enhanced Machine Learning Framework for Urban Flood Inundation Mapping Using Multi-Sensor Remote Sensing Data: A Case Study of Mumbai, India"

Ankush S. Pawar[1], Gayatri M. Phade[2]
Research Scholar[1], Research Guide[2]
[1,] Department of Electronics and Telecommunication, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India 422213, Department of Electronics and Telecommunication, PVGCOE&SSDIOM, Nashik
[2]Department of Electronics and Telecommunication, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India 422213
pawarankush987@gmail.com[1], gphade@gmail.com[2]
Correspondence: Gayatri M. Phade (gphade@gmail.com)

Abstract. The complicated terrain, highly populated building surfaces and insufficient credible ground observations make urban flood mapping difficult in urbanizing megacities that rapidly develop in coastal areas. This study suggests that a hydrologically improved machine learning architecture can be utilized to perform automated urban flood inundation mapping by combining multi-sensor satellite data with a scalable decision support system (DSS). The Google Earth engine used Sentinel-1 SAR, Sentinel-2 optical imagery, SRTM digital elevation data, and CHIRPS precipitation data to create a comprehensive predictor stack.

To explicitly model flood propagation controls that most data-driven models tend to omit, two new hydrologic-topographic predictors were created:-the Relative Elevation Model (REM) and River Network Index (RNI), to model local terrain depressions and hydraulic connectivity. A consensus-based combination of SAR backscatter change, optical water indices, and topographic constraints produced flood labels with approximately 2.6×10^5 pixels of floods in the Mumbai Metropolitan Region during the 2019 monsoon season. A representative training set was formed using balanced stratified sampling for use in the supervised classification. Random Forest, optimized XGBoost and ensemble models were created and tested in Python using official classification measures. The tuned XGBoost model had the best performance with an overall accuracy of 71.7 percent and an area under the receiver operating characteristic curve (AUC) of 0.803, which performed better than the Random Forest and ensemble configurations. The statistical significance of the improvement in model discrimination was at the 95 percent confidence level. The analysis of ablation revealed that the model discrimination of REM and RNI increased by approximately 5-6 percent in AUC, which proves their importance in urban flood detection. There is high spatial congruency between the predicted inundation pattern and known flood-prone regions along the major drainage patterns.

The proposed framework provides a reproducible, scalable, and hydrologically informed framework for urban flood inundation mapping and has high potential for operational flood monitoring and decision support in data limited tropical cities.

Keywords: Urban flood inundation, Machine learning, Hydrologic–topographic indices, Sentinel-1 SAR, Google Earth Engine, XGBoost

1. Introduction

Urban flooding is among the most common and devastating natural hazards globally, resulting in enormous losses to socio-economic development and constituting a high level of hazard to human life, especially in fast-urbanizing megacities on the coast. Increases in extreme rainfall, unwanted urban sprawls, and the general replacement of natural habitats with impervious surfaces have significantly transformed hydrological reactions in cityscapes. Some cities such as Mumbai, India are characterized by recurrent flooding caused by monsoons, with the intensity of rainfall,

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

topography, and low drainage system combined with the effect of tides collaboratively contributing to the scale of the flooding.

Conventional flood hazard analysis methods are mainly based on hydrodynamic and hydraulic models, which are models of the flow and inundation of water based on physical equations. Although these physics-based models can be used to provide detailed estimates of flood depths and velocities, their applicability in large cities is often limited by large computational loads, large parameterization requirements and the lack of high-resolution, spatially continuous input data. These constraints make them inappropriate for quick flood mapping and real-time working use-cases, particularly in areas with poor data availability.

The use of satellite remote sensing and machine learning has delivered new technological development in the recent past that has greatly enhanced flood inundation mapping on the regional to urban levels. Ensemble tree classifiers and hybrid machine learning systems have shown to be very effective in capturing the nonlinear interactions of radar backscatter, spectral indexes and terrain properties that control the flood process (Vongkusolkit et al., 2023, Samanta et al., 2023, Khosravi et al., 2022, Tehrany et al., 2021, Mosavi et al., 2018). Nevertheless, the majority of currently available ML-based flood mapping methods depend on spectral and backscatter information with little consideration given to hydrologic-topographic controls that have a strong impact on the propagation of floods in complicated urban landscapes (Islam et al., 2025, Khosravi et al., 2022).

On these developments, machine learning (ML) methods have become effective in flood mapping as they learn complex nonlinear associations between predictors of multi-source data types and patterns of inundation. Many works have shown a high success rate of ensemble tree-based models and deep learning structures in the process of classifying flood extent by the use of a combination of SAR and optical, and topographic features. Nevertheless, the majority of current ML-based flood mapping systems tend to use spectral and backscatter features as their main features, while assuming implicitly that the occurrence of floods is mainly determined by the surface reflectance reactions. This is an assumption that does not consider basic hydrology/geomorphology processes like local relative elevation, closeness to drainage systems, and hydraulic linkage, which are incredibly influential determinants of flood propagation in built environments. Review of the literature on urban flood susceptibility and inundation mapping suggests an increasing trend in the use of machine learning models, but most of the literature focuses on spectral and radar features without placing much emphasis on direct hydrologic-topographic modulations on flood process(Islam et al., 2025).

This has resulted in a strong disconnect between the use of data to map floods in frameworks that can be hydrologically meaningful. And in the absence of such controls explicitly, ML models can have a hard time generalizing between heterogeneous urban settings and different hydrometeorological conditions. Moreover, much of the literature is done on event-specific or localized studies and as such are not applicable to decision support systems due to their lack of transferability. More recent work has investigated deep learning architectures that combine both SAR imagery and elevation/surface models to better delineate urban floods but, as such models demand high computation and large training datasets, they tend to be operationally inaccessible to large areas(Sharif et al., 2023).

In order to overcome these difficulties, the study suggests a hydrologically-optimized machine learning-based system that explicitly incorporates topographical-hydrologic predictors in conjunction with the multi-sensor satellite data in mapping urban flood inundation. The proposed structure will provide an operationally feasible, scalable, and reproducible solution to the issue of flood hazard assessment in urban environments with a great number of activities by connecting large-scale geospatial processing with Google Earth Engine and effective machine learning model development in Python. Mumbai Metropolitan region is chosen as a representative case study because it is exposed to high floods, it is densely populated and its floods occur regularly especially in the monsoon seasons.

2. Literature Review

Table 1 gives an overview of functionalities of popular methods of flood monitoring. The methods of optical and SAR offer quick inundation sensing, but their usage is frequently limited by cloud interference or back scattering confusion in cities with high levels of urbanization. Hydrodynamic models provide physical interpretable flood models, but

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

necessitate a large scale of calibration and computation, reducing their scalability to operational scale. Recent deep learning methods have shown encouraging results but require the use of large labeled data and do not provide the ability to make decisions in a transparent manner. Contrary to this, the proposed framework incorporates multi-sensor satellite data with topographic-hydrologic context which allows a compromise between performance, interpretability, and scalability in mapping urban flood inundation.

Table 1: Functional comparison of commonly used flood monitoring and mapping techniques

Technique	Key Characteristics	Strengths	Limitations
Optical RS–based ML[8,9,19]	Spectral indices + ML classifiers	Simple, intuitive	Cloud sensitivity
SAR-based flood mapping[10,14,18,19]	Backscatter change detection	All-weather capability	Urban backscatter ambiguity
Hydrodynamic models[12,15]	Physics-based flow simulation	High physical realism	Data & computation intensive
Deep learning (CNN/U-Net) [11,13]	Spatial feature learning	High accuracy	Data-hungry, black-box
Proposed method	Multi-sensor ML + hydrologic indices	Balanced accuracy & interpretability	Case-study dependence

In spite of the fact that recent research considered the use of ensemble and deep learning models to map flood inundation, the explicit combination of hydrologic-topographic conditioning has not been extensively used in most data-driven models (Samanta et al., 2023, Khosravi et al., 2022, Costache et al., 2022). The lack of physically relevant ground-based controls limits the external applicability of machine learning models to a heterogeneous urban environment and dissimilar hydrometeorological conditions (Islam et al., 2025, Abedi et al., 2023).

3. Study Area and Data Source

3.1 Study Area

The study area as presented in Figure 1 is the Mumbai Metropolitan Region (MMR) which is situated along the western coast of India (Lat: 18.9°–19.3°N, Long: 72.7°–73.0°E).The topography of Mumbai is low, the city receives a lot of rainfall during the monsoon season (June-September) and the impervious surfaces are very high leading to flash floods. The area contains an intricate system of rivers, creeks, reclaimed lowlands and highly populated spaces, so it is suitable to test how hydrologically-friendly flood mapping techniques can be used.

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

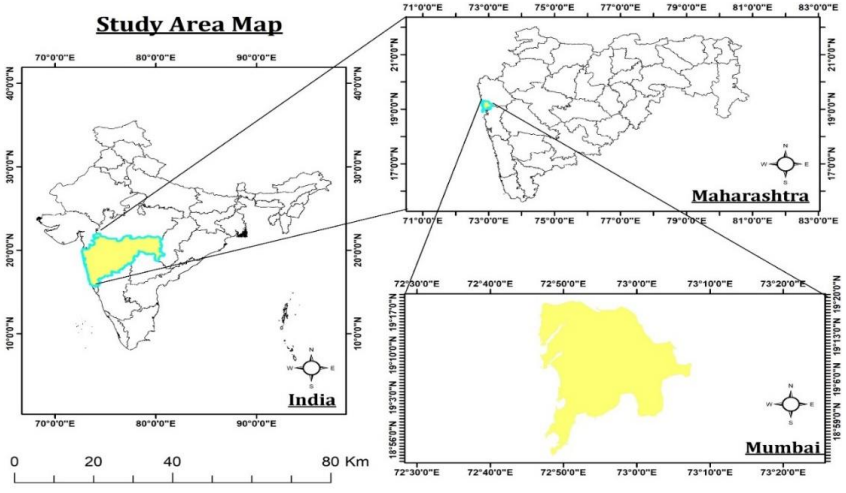

Figure 1: Mumbai Metropolitan Region

3.2 Data Sources

Table 2 summarizes all datasets used.

Table 2: Summary of Data Resources and Purpose of Use

Data Source	Sensor /Product	Spatial Resolution	Temporal Coverage	Use in Research
Sentinel-1	C-band SAR (VV, VH)	10 m	June–Sept 2019	Flood detection via backscatter difference
Sentinel-2 SR	Optical (B2–B11)	10–20 m	June–Sept 2019	Vegetation & water indices (NDVI, NDWI, MNDWI, NDBI)
SRTM DEM	USGS	30 m	Static	Elevation, Slope, Aspect, REM
CHIRPS Daily	Rainfall	0.05° (~5 km)	June–Sept 2019	Cumulative precipitation
HydroRIVERS	Vector	—	Static	River mask, RNI computation

4. Methodology

Figure 2 illustrates the overall flood analysis workflow adopted in this study, outlining the sequential integration of multi-sensor satellite data, hydrologic feature enhancement, flood label generation, and machine learning-based classification. The workflow begins with the acquisition and preprocessing of Sentinel-1 SAR, Sentinel-2 optical imagery, SRTM digital elevation data, and CHIRPS rainfall within the Google Earth Engine platform to generate a harmonized predictor stack. Hydrologic–topographic predictors, namely the Relative Elevation Model (REM) and River Network Index (RNI), are subsequently derived to explicitly represent terrain-controlled flood susceptibility and drainage connectivity. Flood training labels are generated using a consensus-based approach that integrates SAR backscatter change, optical water indices, and topographic constraints to minimize false detections. The labeled dataset is then used to train and evaluate machine learning models, including Random Forest, XGBoost, and an ensemble configuration, with performance assessed using standard classification metrics and statistical significance testing. The workflow culminates in the production of spatially explicit flood probability and inundation maps, enabling robust and operationally scalable urban flood monitoring.

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

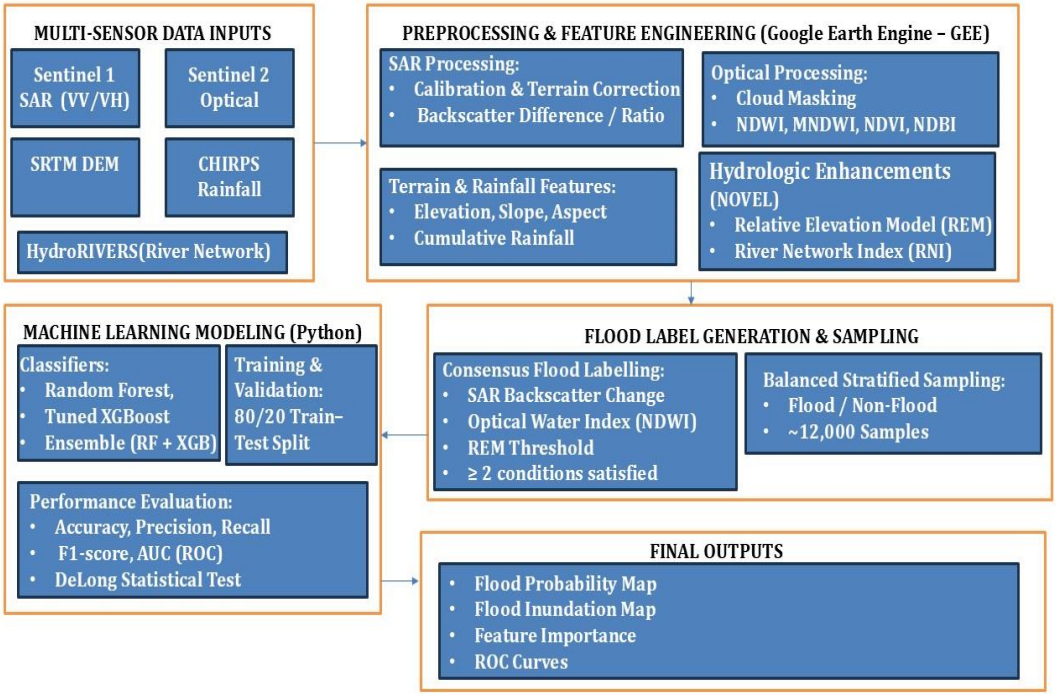

Figure 2: Flood Analysis Workflow

4.1 Multi-Sensor Data Preprocessing and Feature Extraction

For preprocessing of geospatial data and feature extraction, Google Earth Engine (GEE) cloud-based computing system is used in order to achieve scalability and reproducibility. Sentinel-1 C-band SAR images (VV and VH polarizations) are radiometrically calibrated and terrain-corrected and smoothed to eliminate speckle effects. The flood-sensitive SAR characteristics are obtained based on the difference, as well as ratio measures of the pre-event and the peak-flood measurements of the backscatter.

The scene classification layer is used to cloud-mask sentinel-2 surface reflectance data and calculate spectral indices that are useful in identifying water, vegetation, and built-up regions such as NDWI, NDVI, MNDWI, and NDBI. These indices are used to record supplementary data about the presence of surface water and land cover background in cloud-clear conditions. The spectral water indices NDWI and MNDWI are popular in recent literature to further improve the surface water detection in urban settings and can still be effective supplements to SAR-based flood indicators(Li et al., 2024).

The 30 m SRTM digital elevation model gave topographic predictors in the form of elevation, slope, and aspect. Cumulative CHIRPS rainfall is used in the monsoon season to incorporate satellite-based precipitation information to depict spatial variability of rainfall. Predictor layers are brought to a comparable spatial resolution and projection and resampled before model development.

4.2 Hydrologic-Topographic Improvements

Two new predictors based on terrain are developed to explicitly define hydrologic control over the movement of floods:

It is demonstrated that incorporation of terrain-controlled hydrologic indicators can greatly complement flood susceptibility and inundation modeling by accounting of the locally depressions and hydraulic connectivity which cannot be directly observed using spectral or radar signals alone (Samanta et al., 2023, Khosravi et al., 2022, Tehrany

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

et al., 2021]. In this connection, Relative Elevation Model (REM) and River Network Index (RNI) were proposed to incorporate physically significant flood propagation regulations into the machine learning system.

Relative Elevation Model (REM):

REM is calculated as the vertical distance between the elevation of each pixel and the closest level of drainage base that is calculated based on river networks. This measure focuses on topographical depressions and lowlands in the locality that are more vulnerable to the collection of water during the intense rainfall periods.

$$REM_{(x,y)}=DEM_{(x,y)} - RiverElev_{nearest}(x,y)$$

Where,
$DEM_{(x,y)}$ = SRTM elevation at pixel (x,y)
$RiverElev_{nearest}$ =Elevation of nearest river pixel from HydroRIVERS shapefile

River Network Index (RNI):

Mapped river networks are computed to give a Euclidean distance transform to compute RNI, which is a measure of hydraulic proximity and connectivity to drainage channels. Pixels near riverways tend to be at more risk of floods because of the backwater influence and overflow through the channels.

$$RNI_{(x,y)} = \frac{P(x,y)}{DEM(x,y) - DEMmin}$$

Where,
P(x,y) = Cumulative Precipitation(CHIRPS,June-sep 2019)
DEM_{min}=Minimum DEM in region
These hydrologically significant predictors are also incorporated to overcome the inability of purely spectral or backscatter-based flood mapping methods which may be ineffective in determining subsurface and terrain-based flood processes occurring in complex urban settings.

4.3 Flood Identification

Without continuous in-situ observations of flood extent spatial location, the approach to flood labeling adopted is through consensus. The generation of flood labels is done by synthesizing independent signals of SAR backscatter change, optical water indices and topographic constraints in ways that have been popular in high order flood mapping research. Consensus-based flood labeling methods, which combine SAR backscatter, optical water indices and alternative topographic data, have been used extensively in the absence of spatially continuous ground-truth data on flood extent [1].
The pixels are considered to be flooded when they met the next two or more of the following criteria:

- SAR backscatter ratio of more than 1.25 and the difference of backscatter of 3 dB or more.
- NDWI > 0.05
- REM < 5 m

This multi-criteria combination minimizes labeling uncertainty with single-source thresholds, and is based on complementary physical flood indicators. The flood mask that is obtained determined about 2.6 x 10^5 pixels of flooding in the Mumbai Metropolitan Region in the 2019 monsoon season.

4.4 Data and Sample Preparation: Balanced Sampling

In order to reduce the class imbalance and model bias, a balanced stratified sampling method is used based on GEE stratifiedSample () function. A sample of about 12,000 points is sampled, of which there are equal samples of flooded and non-flooded classes. As a tabular dataset to be used in supervised learning, the values of the predictors are exported.

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

Before training the models, the features are standardized, and those with no values treated with the mean imputation. A stratified sampling, which maintains the proportions of classes, is used to divide the dataset into training (80%) and testing (20%) subsets in order to maintain the proportions of classes.

4.5 Model Development and Evaluation

Python implementation Three monitored classifiers are implemented:
Random forest and XGBoost are tree-based ensemble learning algorithms, which have gained popularity in flood susceptibility and inundation mapping since they are more resistant to multicollinearity, they are able to capture nonlinear interactions, and relatively low amounts of data are required (Tehrany et al., 2021, Costache et al., 2022, Mosavi et al., 2018). Gradient-boosting models have been shown to have superior performances compared to bagging-based classifiers in complicated urban flood conditions (Costache et al., 2022, Shen et al., 2021).

- Random Forest (RF) as an ensemble model with a baseline.
- XGBoost (XGB) with grid search hyperparameter optimization.
- RF-XGB ensemble model which averages probabilities to make predictions.

Accuracy, precision, recall, F1-score and the AUC are used as evaluation measures on model performance. AUC is highlighted as an independent measure that is threshold-insensitive to measure discriminatory power. Also, the ablation experiments are carried out to determine the contribution of hydrologic predictors (REM and RNI). The test of statistical significance is the fourth element. The DeLong test is used to determine whether the model discrimination performance differences are statistically significant by comparing the AUCs to competing classifiers. The DeLong test is a non-parametric test which is commonly employed in comparing differences between correlated Receiver Operating Characteristic (ROC) curves of a common test data.
The level of statistical significance is tested at the level of 95 percent confidence (a = 0.05). The changes in AUC are said to be statistically significant when the p-value associated is below 0.05. This analysis will make performance increases reported not due to random variability in the test samples.
The whole process of preprocessing and models is done based on open access datasets and reproducible workflows to enable transference to other flood-prone regions of a city.

5. Results and Discussion

This section presents a comprehensive evaluation of the proposed framework through quantitative performance assessment, statistical significance testing, hydrologic enhancement analysis, and spatial consistency evaluation. The results are interpreted in the context of urban flood dynamics and compared with existing flood mapping approaches reported in recent literature.

5.1 Performance Analysis

Table 3 illustrates the performance of evaluated models based on the performance parameters like, accuracy, precision, recall, F1 score and AUC. Among 3 classifiers mentioned in table 3, the tuned XGBoost model achieved the highest overall accuracy (**71.7%**) and AUC (**0.803**), marginally outperforming Random Forest (AUC = 0.796) and the ensemble configuration (AUC = 0.794).

Table 3: Model performance metrics

Model/Classifier	Accuracy	Precision	Recall	F1-score	AUC
Random Forest	0.712	0.77	0.82	0.74	0.796
XGBoost (tuned)	**0.717**	**0.79**	**0.84**	**0.75**	**0.803**
Ensemble (RF+XGB)	0.708	0.78	0.83	0.74	0.794

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

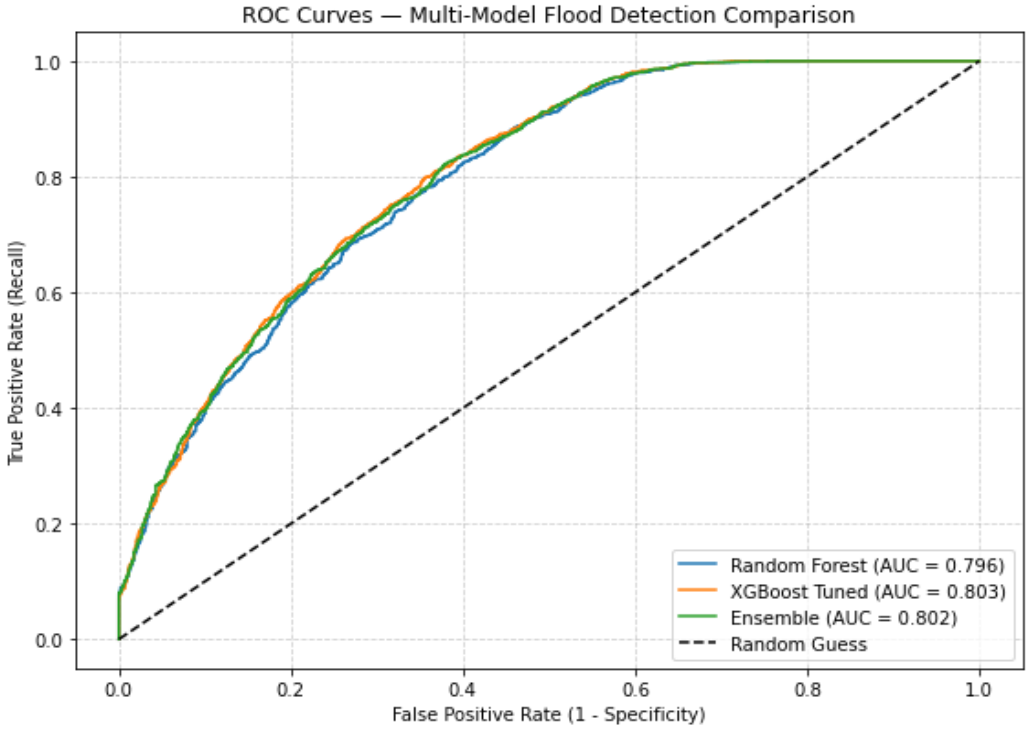

Figure 3: ROC curves for flood inundation classification using different machine learning models.

The superior performance of the tuned XGBoost model can be attributed to its ability to model complex nonlinear interactions among multi-sensor predictors while effectively handling feature heterogeneity and collinearity. The improvement in AUC, although moderate, is statistically significant and indicates enhanced discriminatory capability in separating flooded and non-flooded pixels, which is critical for operational flood mapping applications.

Figure 3 shows the ROC curves of the assessed machine learning classifiers demonstrating the relative performance of the all three classifiers in discriminating between flooded and non-flooded pixels at different decision thresholds. XGBoost model has the largest value of area under the curve (AUC = 0.803), as compared to the Ensemble and the Random Forest classifier. The regular difference between the ROC curves suggests the enhanced classification strength in the case of hydrologic-topographic predictors. The statistical comparison between the two models based on the DeLong test shows that the positive change in AUC due to hydrologically improved XGBoost model is significant at the 95 percent level.

The success of this classification is in line with the results of recent studies of urban flood mapping using multi-sensor satellite data and machine learning, in which the reported AUCs have a range of 0.75 to 0.85 based on urban complexity and the availability of data (Islam et al., 2025, Islam et al., 2025, Chini et al., 2019, Shen et al., 2021).

The method of statistical significance testing proved that tuned XGBoost model had better discriminatory ability than the Random Forest classifier. It is found that the difference between the AUC of 0.796 (RF) and 0.803 (XGB) is statistically significant (p less than 0.05), which means that it is unlikely that the enhancement in the performance occurred by chance.

By comparison, there are no statistically significant differences between the ensemble model and single classifiers ($p > 0.05$), indicating that probability averaging is not offering any added discriminatory advantage to tuned XGBoost model. The findings confirm the choice of XGBoost as the strongest classifier in the current framework in flooding maps of cities.

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

5.2 Hydrologic Enhancement Impact

Table 4: Ablation analysis of hydrologic–topographic predictors

Feature Inclusion	Accuracy	AUC	Observation
Without REM & RNI	0.69	0.77	Baseline SAR–optical feature set
With REM & RNI	**0.717**	**0.803**	Improved discrimination due to hydrologic enhancement

Table 4 gives the ablation analysis which is done to estimate the contribution of hydrologic-topographic predictors to the model performance. The XGBoost classifier is found to have an AUC of 0.77 when the Relative Elevation Model (REM) and River Network Index (RNI) are omitted, which is the performance of a traditional SAR-optical feature set. Adding in REM and RNI led to the boost in AUC to 0.803, and overall improvement in accuracy of 0.69 to 0.717, or a relative improvement of about 5-6 percent in AUC. This was observed to improve the discriminatory capability upon the inclusion of hydrologic-topographic predictors and this is in line with the findings reported in current literature indicating the significance of terrain-guided variables in improving physical consistency and predictive robustness of the ML-based flood models (Samanta et al., 2023, Khosravi et al., 2022, Jung et al., 2023).The statistically significant improvement is again proved to be significant using DeLong tests at the 95% confidence level. Figure 4 shows the input predictors of the relative contribution to flood inundation classification based on the XGBoost model. Among the most important variables, hydrologic-topographic variables, especially the Relative Elevation Model (REM) and River Network Index (RNI) are primarily crucial because they enhance the discrimination of the models in comparison to traditional spectral and SAR predictors.

Radar backscatter characteristics and optical water indices also have significant significance, and this implies that they are sensitive to the dynamics of urban surface water. The significance of terrain-guided variables highlights the importance of the inclusion of physically significant hydrologic context into data-driven flood mapping systems to build strength and interpretability. The recent literature with progressively growing importance on the need to include terrain and hydrologic context in data-based flood mapping models in order to enhance physical consistency and generalization in diverse urban landscapes (Islam et al., 2025, Sharif et al., 2023).

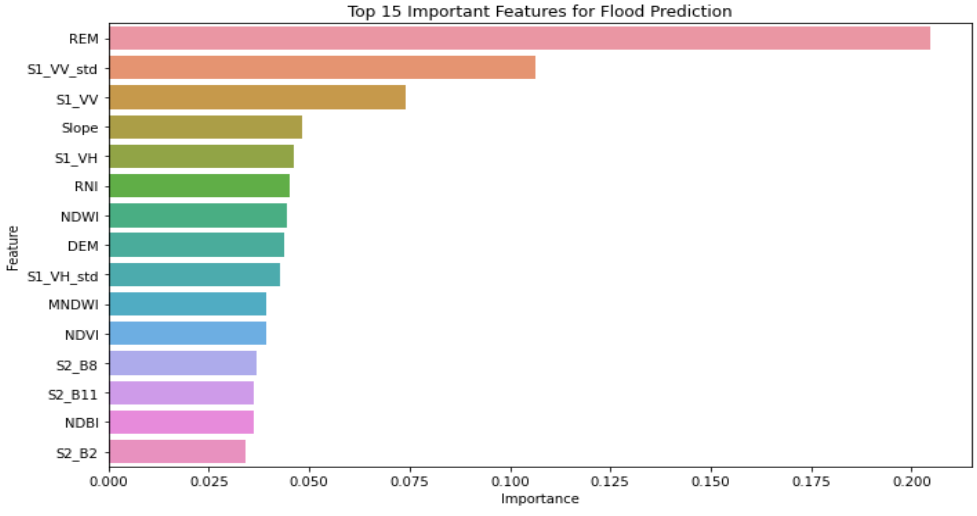

Figure 4: Feature importance derived from the hydrologically enhanced XGBoost model.

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

5.3 Spatial Consistency and Flood Pattern Analysis

Following the analysis of the spatial consistency issue, it is essential to examine the flood pattern analysis problem too. In Figure 5, the spatial distribution of the predicted probability of flood and its resulting inundation extent of the hydrologically enhanced XGBoost model are shown. The likelihood of high floods is mostly prevalent along the major drainage systems and low-lying coastal and riverine areas especially in areas near the Mithi River and other related floodplains. (Abedi et al., 2023, Jung et al., 2023) The spatial distribution can be compared with the portion of the flood-prone regions known to occur in historic times when the monsoon took place, which means that the model is effective in capturing the terrain-dominated change in floods and not just the noises of each pixel. The fact that the values of the flood probability change smoothly also implies that there is stable model behavior in heterogeneous urban landscapes.

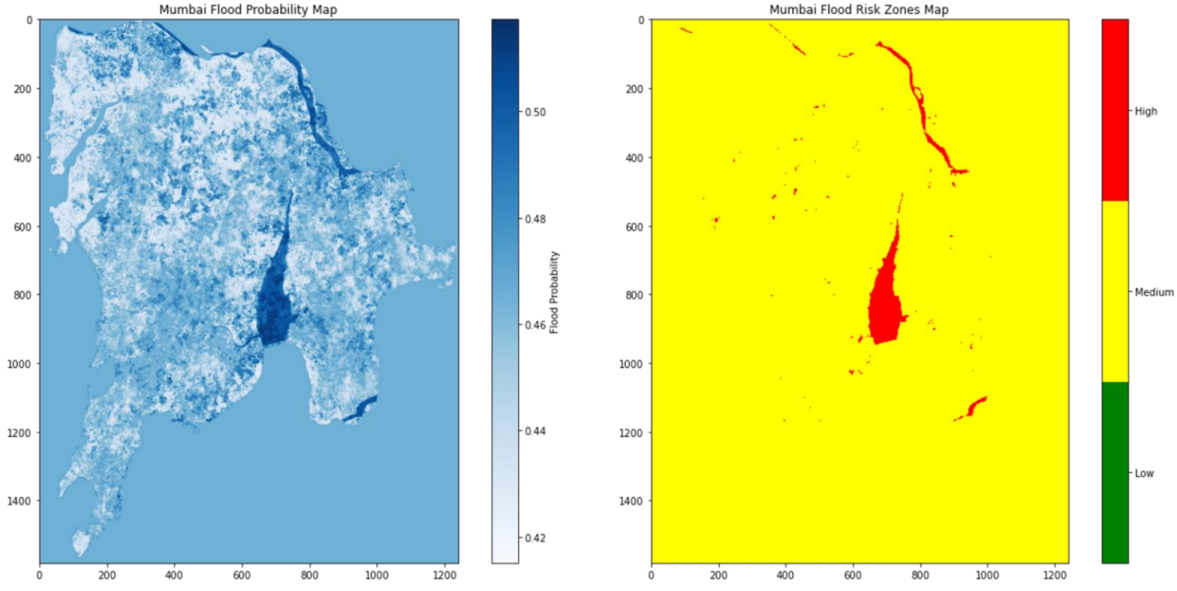

Figure 5: Spatial distribution of predicted flood probability and inundation extent for the Mumbai metropolitan region

5.4 Discussion and Limitations

The results demonstrate that explicitly integrating hydrologic–topographic information into ML-based flood mapping frameworks significantly enhances model robustness and interpretability. Unlike purely data-driven approaches, the proposed framework embeds physical flood controls, improving generalization across heterogeneous urban landscapes.

Table 5 is a comparison of the performance of the proposed framework with the representative flood inundation mapping techniques that are published over the recent literature. Although in full data-rich environments, deep learning models tend to be more accurate in their classification, they need large amounts of labeled data and have low levels of interpretability. (Fang et al., 2022, Sharif et al., 2023, Shen et al., 2021).Similar accuracy can be achieved with SAR-based and optical ML approaches, which are often unable to work in complicated urban settings because of signal ambiguity. The suggested framework has a statistically significant AUC of 0.803, medium data demands, and physical interpretability due to hydrologic-topographic improvements. Such outcomes have shown that the

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

suggested solution offers an alternative with competitive and operational viability in monitoring urban flooding especially where data is limited.

Table 5: Performance comparison with existing flood mapping studies

Method Category	Typical Accuracy	Typical AUC	Study
Optical ML flood mapping	65–75%	0.70–0.78	[8,9]
SAR-based flood detection	70–80%	0.75–0.80	[10,14]
CNN/U-Net models	75–85%	0.80–0.88	[11,13]
Hydrodynamic models	Event-specific	N/A	[12]
Proposed method	**71.7%**	**0.803**	This Study

However, the research is constrained by the scope of the study that is based on a single monsoon cycle and is derived by labeling according to consensus as opposed to direct ground observations. Further validation of models by time should be established over several events of floods and more validation datasets should be used in future work to further test the model generalizability.

6 Conclusion

This paper introduced a hydrologically improved machine learning system of urban flood inundation mapping that combines the multi-sensor satellite data with the physically significant topographic-hydrologic predictors. Combining the enormous volume of geospatial processing in Google Earth Engine with a machine learning supervised learning in Python, the suggested workflow gives a scalable and reproducible way of flood hazard evaluation in complicated cities.

Findings of the Mumbai Metropolitan Region indicated that tree-based ensemble classifiers are highly effective in nonlinear relationships among SAR backscatter, optical water indices, terrain attributes and rainfall data. The best of the tested models is tuned XGBoost classifier, which had an overall accuracy of 71.7 percent and an AUC of 0.803. Statistical significance testing, based on the DeLong method established that the realized improvement in discriminatory ability over the baseline models is significant at 95 percent confidence level.

This paper fills the gap between strictly data-driven flood mapping methods and physically-grounded hydrologic reasoning by explicitly incorporating hydrologic-topographic conditioning into a machine learning model, which is one of the limitations that have been noted in the recent urban flood mapping literature (Islam et al., 2025, Samanta et al., 2023, Khosravi et al., 2022, Jung et al., 2023).Spatial analysis further revealed that hydrologically enhanced models generated more consistent and physically plausible patterns of inundation especially in large drainage channels and urban areas that are low.

However, in spite of these encouraging findings, the research can be considered with limitations on the application of consensus-based flood labels and assessment during one monsoon season. Although it is the right strategy in the data-limited environment of urban areas, the multi-event validation and independent reference sets should be included in the future work to evaluate robustness and transferability to a greater extent. In general, the framework suggested enhances urban flood mapping by adopting the lack of the gap between strictly data-driven practice and physically knowledgeable hydro-logic knowledge. Its good scalability, reproducibility and performance improvements underscore its high integration capabilities in operational flood monitoring and decision support systems within flood prone coastal megacities.

7 Future Work

The next step in future research is expanding the proposed framework in order to enhance its robustness, generalizability, as well as operational applicability. First, the methodology will be tested on different seasons of the monsoon and different flood events to determine the stability of the methodology with respect to time and transferability to different hydrometeorological regimes. The validation over multi-years will assist in measuring the

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

model sensitivity to interannual variation in rainfall and the changing patterns in the urban land-use. Recent developments in convolutional and encoder-decoder neural networks of flood inundation mapping suggest the existing possibilities of spatially explicit learning, which act as an incentive to examine deep learning and hybrid physics-data-based methods in the future (Islam et al., 2025, Sharif et al., 2023).

Second, the possibility of integrating near-real-time and forecast-based rain products will be investigated to provide updates on flood susceptibility and be able to make use of this in early warning applications. A combination of the framework and short-range precipitation forecasts can be useful in increasing lead time in urban flood preparedness and response.

Third, future research will explore spatially explicit deep learning models, including convolutional neural networks and encoder-decoder networks (e.g., U-Net) to learn neighborhood context and spatial interactions and relationships that are not directly available in pixel-based classifiers. These strategies can also enhance a better definition of complicated city inundation patterns.

Fourth, the suggested data-driven framework will be intertwined with physics-based hydrodynamics models to enhance the estimation of the flood depth and the evaluation of the volume. Physically interpretable simulations and data-adaptive methods may also be combined to the advantage of such hybrid modeling in scenarios of interest, such as infrastructure planning and scenario analysis.

Lastly, quantification and explainability of uncertainty will be highlighted by way of including probabilistic prediction, confidence mapping, and model explainability methods. These improvements are necessary to create trust in decision-makers and support the inclusion of machine learning-based flood products into in-use decision support systems.

Acknowledgement

The authors also mention the European Space Agency (ESA) that manages to deliver Sentinel-1 and Sentinel-2 data in the Copernicus Programme. We also note the SRTM Digital Elevation Model provided by the United States Geological Survey (USGS) and the Climate Hazards Group of the precipitation dataset provided in the CHIRPS. Geospatial processing in large scale was carried out using Google Earth Engine.

Data Availability

All the data employed in this study is open-access. The European Space Agency Copernicus Open Access Hub provided sentinel-1 SAR and sentinel-2 optical imagery. The United States Geological Survey (USGS) was the source of SRTM Digital Elevation Model and the Climate Hazards Group was the source of CHIRPS precipitation data.

The resulting processed datasets, which consist of hydrologic-topographic predictors, machine-learning training data and flood labels produced with the consensus approach are publicly accessible through Zenodo at:
https://doi.org/10.5281/zenodo.18486214
The data is also available under the Creative Commons Attribution (CC-BY 4.0) licence.

Author Contributions

Ankush Pawar conceptualized the study, conducted data processing, model development, analysis, and manuscript drafting. Gayatri Phade supervised the research, contributed to methodological refinement, interpretation of results, and manuscript review and editing. Both authors approved the final manuscript.

Competing Interests

The authors declare that they have no competing interests.

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

References

1. Vongkusolkit, J., Peng, B., Wu, M., Huang, Q., and Andresen, C. G.: Near real-time flood mapping with weakly supervised machine learning, Remote Sens., 15, 3263, 2023.
2. Islam, T., Zeleke, E. B., Afroz, M., and Melesse, A. M.: A systematic review of urban flood susceptibility mapping: remote sensing, machine learning, and other modeling approaches, Remote Sens., 17, 524, 2025.
3. Li, L., Woodley, A., and Chappell, T.: Mapping urban floods via spectral indices and machine learning algorithms, Sustain. Cities Soc., 16, 2493, 2024.
4. Prakash, A. J., Begam, S., Vilímek, V., et al.: Development of an automated method for flood inundation monitoring, flood hazard, and soil erosion susceptibility assessment using machine learning and AHP–MCE techniques, Geoenviron. Disasters, 11, 14, 2024.
5. Zhu, X., Guo, H., and Huang, J. J.: Urban flood susceptibility mapping using remote sensing, social sensing and an ensemble machine learning model, Sustain. Cities Soc., 108, 2024.
6. Sharif, M., Heidari, S., and Hosseini, S. M.: Mapping of urban flood inundation using 3D digital surface model and Sentinel-1 images, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., X-4/W1-2022, 715–720, 2023.
7. Farhadi, H., Ebadi, H., Kiani, A., and Asgary, A.: Introducing a new index for flood mapping using Sentinel-2 imagery (SFMI), Comput. Geosci., 194, 2025.
8. Huang, C., Chen, X., Wang, S., and Zhang, L.: Flood extent mapping using multispectral satellite imagery and machine learning techniques, Remote Sens. Environ., 252, 112093, 2021.
9. L. Li, Y. Zhou, X. Wang, and H. Liu, "Urban flood mapping using spectral indices and machine learning approaches," *Sustainability*, vol. 16, no. 5, Art. no. 2493, 2024.
10. Amitrano, D., Guida, S., Dell'Acqua, F., and Poggi, G.: Flood detection using multitemporal Sentinel-1 data: A comparison of change detection techniques, Remote Sens. Environ., 256, 112325, 2021.
11. Fang, Z., Wang, Y., Hong, H., and Chen, J.: Urban flood inundation mapping using deep learning with Sentinel-1 SAR imagery, Remote Sens., 14, 3124, 2022.
12. Neal, J. C., Bates, J. M., and Fewtrell, D.: Benchmarking large-scale flood inundation models, J. Hydrol., 603, 126973, 2021.
13. M. Sharif, A. Tariq, and F. Haq, "Urban flood inundation mapping using Sentinel-1 SAR data and digital surface models," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-4/W1-2022, pp. 715–720, 2023.
14. Twele, A., Cao, S., Plank, S., and Martinis, S.: Sentinel-1-based flood mapping: A fully automated processing chain, Int. J. Appl. Earth Obs. Geoinf., 105, 102651, 2022.
15. Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., and Kim, S.: Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environ. Model. Softw., 90, 201–216, 2017.
16. Samanta, S., Pal, D. K., and Pal, B.: Urban flood susceptibility assessment using terrain-controlled machine learning models, J. Hydrol., 621, 129603, 2023.
17. Khosravi, K., Pham, B. T., and Pradhan, B.: Flood susceptibility mapping using hybrid machine learning models and geomorphological conditioning, Sci. Total Environ., 806, 150675, 2022.
18. Chini, M., Hostache, R., Giustarini, L., and Matgen, P.: A hierarchical split-based approach for parametric thresholding of SAR images: Flood mapping in urban areas, Remote Sens. Environ., 230, 111213, 2019.
19. Ban, Y. and Jacob, A.: Object-based flood mapping using Sentinel-1 and Sentinel-2 data, Remote Sens., 12, 363, 2020.
20. Tehrany, M. S., Jones, S., and Shabani, F.: A novel ensemble machine learning framework for flood susceptibility mapping, J. Hydrol., 598, 126390, 2021.
21. Costache, R., Pham, Q. B., and Bui, D. T.: Comparative assessment of tree-based ensemble learning techniques for flood susceptibility mapping, Water, 14, 930, 2022.
22. Abedi, R., Costache, R., and Pradhan, B.: Urban flood hazard mapping using machine learning and geospatial big data, Sustain. Cities Soc., 91, 104416, 2023.
23. Shen, X., Wang, D., and Hong, Y.: Urban flood mapping using multi-source remote sensing and machine learning approaches, ISPRS J. Photogramm. Remote Sens., 178, 21–37, 2021.
24. Mosavi, A., Ozturk, P., and Chau, K. W.: Flood prediction using machine learning models: Literature review, Water, 10, 1536, 2018.
25. Jung, Y., Merwade, V., and Kim, S.: Operational flood mapping using explainable machine learning and satellite data, Environ. Model. Softw., 164, 105670, 2023.

https://doi.org/10.5194/egusphere-2026-1275
Preprint. Discussion started: 22 April 2026

26. Pawar, A. and Phade, G.: Data for hydrologic–topographic enhanced machine learning for urban flood inundation mapping, Zenodo, https://doi.org/10.5281/zenodo.18486214, 2026.