# FLOOD DAMAGE PREDICTION USING PRE & POST-EVENT SATELLITE DATA AND MACHINE LEARNING IN COASTAL BANGLADESH

**Rahatul Hasan *[1], Maksudur Rahman [2], Nabiha Tahsin[3], Seam Sikder[4], Asir Foysal[5], Showmitra Kumar Sarkar[6]**

[1] *Rahatul Hasan, Khulna University of Engineering & Technology, Bangladesh, e-mail: hasan2117048@stud.kuet.ac.bd **
[2] *Maksudur Rahman, Khulna University of Engineering & Technology, Bangladesh, e-mail mrfhost@gmail.com*
[3] *Nabiha Tahsin, Khulna University of Engineering & Technology, Bangladesh, e-mail: tahsin2117043@stud.kuet.ac.bd*
[4] *Seam Sikder, Khulna University of Engineering & Technology, Bangladesh, e-mail: sikder2117040@stud.kuet.ac.bd*
[5] *Asir Foysal, Khulna University of Engineering & Technology, Bangladesh, e-mail: foysal2117036@stud.kuet.ac.bd*
[6] *Showmitra Kumar Sarkar, Khulna University of Engineering & Technology, Bangladesh, e-mail: showmitrasarkar@urp.kuet.ac.bd*

***Corresponding Author**

## ABSTRACT

Bangladesh is a country that is periodically affected by flooding caused by tropical cyclones (TC), which is devastating to the livelihood of people, destroys infrastructure, and destroys agricultural production. Our experiment in this paper combines flood prediction and damage evaluation by applying a variety of remote-sensing data resources with machine learning algorithms. Post and pre-event satellite data were used to analyze five major TC events, which include Bulbul (2019), Amphan (2020), Yaas (2021), Sitrang (2022), and Remal (2024) events satellite data. Sentinel-1 SAR (Synthetic Aperture Radar) data, including Vertical-Vertical polarization (VV) before, VV after, and VV difference, were combined with Sentinel-2 Multispectral Instrument(MSI) data, which includes Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI), along with rainfall, elevation, slope, land use-land cover (LULC), and proximity to coast and rivers, to generate flood-related predictors for analysis. Flood labels were created using a threshold-based change detection approach on VV difference values. Random Forest (RF) and Light Gradient Boosting Machine (LightGBM) models were trained on a balanced dataset of 250,000 samples and achieved high predictive accuracy (≈0.83). Feature importance analysis identified VV_before, VV_after, and proximity to water bodies as the most influential predictors. The final flood probability maps were overlaid with LULC data to assess land use specific flood damage. Results indicate that fishponds (357.5 km²) and croplands (244.9 km²) were the most affected, followed by built-up areas (83.6 km²). As the framework demonstrates, remote sensing combined with AI can provide quite decent forecasts of floods and locate the areas of damage, and it is a huge success in disaster management and long-term planning of the coast in Bangladesh.

***Keywords:*** *Flood prediction, Machine learning, Sentinel imagery, Damage estimation, Coastal Bangladesh.*

## 1. INTRODUCTION

Bangladesh is among the most vulnerable locations globally in terms of floods due to the low-lying land, rainy, and deltaic landform at the point where the Ganges, Brahmaputra, and Meghna rivers meet (Shampa et al., 2025). Compound flooding due to storm surges, tidal backwater, and excessive monsoon rains is very prone at the coastal locations, particularly at Khulna, Satkhira, and Bagerhat (Uddin et al., 2019; Rahman et al., 2019). Huge storms such as TC Sidr (2007), Aila (2009), Amphan (2020), and Remal (2024) demonstrate that life on the coast is very dangerous and results in people, farms being destroyed, and displaced (Subhani & Ahmad, 2019). Rising sea level and land sinking have aggravated these extreme weather bursts, and this has highlighted the dire need to establish smart flood monitoring and predictive data systems that will reduce the harm (Abdo et al., 2025).

The traditional flood evaluation methods that were typically utilized in Bangladesh, including field surveys and hydrological models, were approaching the limits due to data gaps, time lags, and high prices (Choudhury & Yabar, 2025). Certain mapping techniques, such as threshold leveling and change detection were tried and compared in a recent study (Lang et al., 2024). Other threshold-setting tricks such as active contour (Soudagar et al., 2025), Kittler-Illicworth and Otsus algorithms (Lou et al., 2024; Zhang et al., 2025)have been employed by people in earlier mapping applications. Although they are working, they have not been effective in automation and scaling of real time mapping. An alternative approach could be to connect machine learning to geospatial applications like Google Earth Engine (Rahman et al., 2025).

As recent studies established, a combination of remote sensing and geospatial analysis actually assists us to comprehend TC risk in such areas as coastal Bangladesh. It divides exposure, sensitivity, and adaptive capacity into parts through Principal Component Analysis (PCA) and it is extremely beneficial to disaster preparation and resilience development (Sarkar et al., 2024). The Earth Observation (EO) and Machine Learning (ML) have provided technological advances that allow us today to create data-driven tools to enhance flood detection, mapping, and impact prediction (Rifath et al., 2024; Toma et al., 2024). Synthetic Aperture Radar (SAR) is particularly useful in flood monitoring since it can view during clouds and darkness providing us with full weather coverage (Shampa et al., 2025). We obtain significantly superior classification and size estimation of floods in dynamic coastal settings when we combine SAR with such ML algorithms as RF, Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost) (Abdo et al., 2025). By introducing GIS to that equation, it becomes possible to overlay several different sources of data to create maps of flood-susceptibility in detail: digital elevation models, rainfall, land use, vegetation indices, and proximity to rivers (Kader et al., 2024). Such combined models do not only increase the accuracy of space but also simplify the interpretation of flood drivers using explainable ML models such as SHapley Additive exPlanations (SHAP) analysis (Abdo et al., 2025). As it has already been demonstrated by the previous studies, a combination of Sentinel-1 SAR and Sentinel-2 optical data with the use of ML can significantly enhance the resolution of flood mapping of both TC and monsoonal events ( Rahman et al., 2019; Shampa et al., 2025). Despite all these advances, the reality in coastal Bangladesh is that land-use alteration, changes in rainfall and tidal surges react in intricate manners regarding real-time monitoring and the evaluation of flood effects(Hoque et al., 2022). A combination of SAR, GIS, and ML is the most efficient way to optimize flood mapping and risk management due to three main reasons (Sampurno et al., 2024). First, SAR has the capability of taking high-resolution images at any weather with a short revisit period to provide us with timely data when floods occur (Giustarini et al., 2016). Second, SAR data combined with GIS spatial analysis are more precise in the observation and prediction of future floods, which could minimize their possible human and economic losses (Singha et al., 2024). Third, ML also enhances this ability further since it searches through massive amounts of SAR and GIS data and identifies patterns and dependencies that are invisible to more traditional methods. The result of this data-driven method is an improved flood risk management and reduction of its effects, as decision makers will have a reliable tool to protect the flood prone regions. This study combines environmental, topographic, and socioeconomic factors with multi-temporal Sentinel-1 and Sentinel-2 satellite datasets to create a comprehensive flood damage prediction framework. LightGBM and RF models are used to estimate the amount of damage and forecast the occurrence of floods in various land-use categories, including

vegetation, aquaculture, cropland, and built-up areas. This study on the coastal area of Bangladesh, in which the five TC floods of 2019-2024 have identified the trends of the damage and flood patterns. The aim of this strategy is to make flood risk modeling more accurate and aid in managing the disaster in coastal regions based on data.

In general, the method is supposed to provide us with fast and large-scale maps of floods based on open Earth -observation data through ensemble ML, however, there are several caveats. First, the flood labels are based on Sentinel-1 VV change, Change Detection and Thresholding (CDAT) with a world background and an insignificant visual verification. That gives a lot of uncertainty, as we do not have much information on-ground water level, particularly on a green and rough surface. Second, the predictor layers were all interpolated to a 10m grid by resampling coarser products such as Shuttle Radar Topography Mission (SRTM) and Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) that can introduce a scale error into the models. Third, we only repeated a single LULC epoch at all events hence we are likely to under-capturing rapid land-cover transitions as well as seasonal crop cycles. Fourth, the point predictions were smoothed with inverse distance weighting and that has the propensity to smooth out the extreme values locally and can spew out artefacts where we barely have samples. Fifth, the set ±10 day event window may fail to capture brief floods or water, which persists afterwards. Finally, the models were solely trained on five coastal TC, hence we cannot guarantee that they will be useful on river or pluvial floods and in other locations unless we retrain and undertake a select transfer validation.

## 2. METHODOLOGY

### 2.1 Study Area

The study was conducted in the southwestern coastal region of Bangladesh, encompassing the districts of Khulna, Bagerhat, and Satkhira, which are among the most disaster-prone areas of the country due to their geographical proximity to the Bay of Bengal.

This low-lying deltaic region of Bangladesh is a hot spot for TC, tidal flooding and intrusion of salty water in its water bodies all of which put the livelihoods and infrastructure of local people are at huge risk. The Area of Interest (AOI) is bounded between 22°30′N to 23°00′N latitude and 89°00′E to 90°00′E longitude, covering an approximate area of 6103.46 sq.km. The topography is generally flat and not far above mean sea level with elevations ranging from just 2 to 5 metres above mean sea level, which makes it highly susceptible to storm surges and the long-term rise in sea level (Ashrafuzzaman et al., 2022).

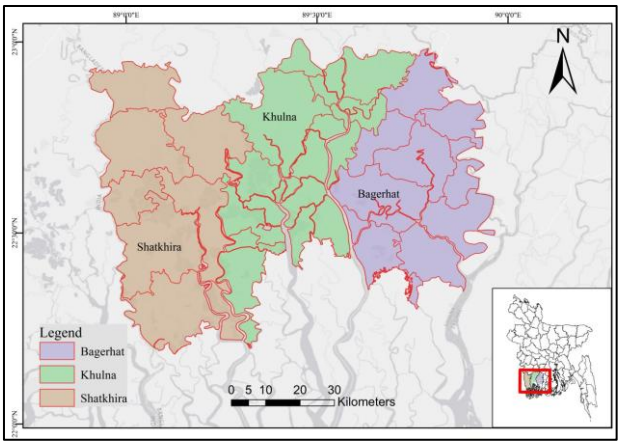



Figure 1: Study area map

The population of Khulna, Bagerhat, and Satkhira districts is 2,663,209, 1,779,804, and 2,332,344, respectively, highlighting the high population density and human vulnerability in this coastal zone(BBS,

2022). We created LULC info for the study area using Sentinel-2A imagery that is classified using a RF machine learning algorithm. The study area map (Figure 1) represents the spatial extent and administrative boundaries of the selected districts, showing major river networks and the location of the study region in Bangladesh.

### 2.2 Event Selection and Data Collection

This study focuses on five TC induced flood events that severely affected the southwestern coastal zone of Bangladesh including Bulbul (2019), Amphan (2020), Yaas (2021), Sitrang (2022), and Remal (2024). These cases provide a good mixture of surge-rain interplay and behavior of the land surface in reaction to these factors, which is critical to the development of robust models for flood detection and associated damage. The TC events were selected based on their documented large-scale impacts. Bulbul displaced over two million people (Aljazeera, 2019). Amphan affected more than ten million people and breached embankments (Brac, 2020). Yaas caused widespread wind, storm surge and rainfall impacts (ReliefWeb, 2021). Sitrang impacted approximately 2.3 million people, including 1.1 million children (Unicef, 2022). Remal affected around 4.6 million people across 19 districts and caused severe flooding of agricultural and aquaculture areas (Acaps, 2024). These cases provide us the baseline data required to train and test our flood occurrence and land use based damage model along Khulna, Bagerhat and Satkhira belt. For each TC, fixed pre and post analysis windows to ensure consistency in the data analyses among the sensors:

(1) Bulbul (Pre: 31 Oct–09 Nov 2019; Post: 10–20 Nov 2019),
(2) Amphan (Pre: 10–19 May 2020; Post: 20–30 May 2020),
(3) Yaas (Pre: 16–25 May 2021; Post: 26 May–05 Jun 2021),
(4) Sitrang (Pre: 14–23 Oct 2022; Post: 24 Oct–03 Nov 2022) and
(5) Remal (Pre: 16–25 May 2024; Post: 26 May–05 Jun 2024).

Inundation cues were derived from Sentinel-1 SAR (VV) by computing the backscatter difference $\Delta VV = VV_{post} - VV_{pre}$; sustained VV decreases over low-relief terrain were treated as flood indicators following established practice (Twele et al., 2016). Sentinel-2A MSI provided NDVI and NDWI within the same windows to capture vegetation stress and surface-water expansion, consistent with applications in Bangladesh (Singha et al., 2020). CHIRPS daily precipitation was aggregated to 3-day and 10-day totals aligned to event windows (Funk et al., 2015). Topography was represented by SRTM 30 m Digital Elevation Model (DEM) and a derived slope layer to encode flow paths and ponding potential (Farr et al., 2007). Exposure context was provided by a study-specific LULC map generated from Sentinel-2A using a RF classifier in Google Earth Engine (GEE) with 2,355 training samples across waterbody, vegetation, built-up, cropland, and fishpond (mid-interval: 22 Dec 2022–03 Mar 2023), leveraging RF's robustness for remote sensing (Belgiu & Drăgu, 2016). The Euclidean distance was taken from the rivers and the coastline and used as a proximity predictor. All rasters were resampled to 10m, reprojected to WGS84/UTM Zone45N our study default and corrected for co-registration of pixels to perfectly line up over each other for subsequent training of our models, flood mapping and damage assessment.

### 2.3 Data Preprocessing and Feature Generation

The preprocessing and feature-generating workflow has combined GEE, ArcGIS Pro, and Python in preparing the satellite derived prediction to be used in predicting flood in coastal Bangladesh. Sentinel-2 Level-2A surface reflectance imagery was used for atmospheric correction and cloud masking applications using the Scene Classification Layer (SCL) in order to construct median composite images for the pre-flood and post-flood periods. This step was necessary in order to remove cloud and atmospheric noise in the optical data (Twele et al., 2016). To detect the flood-affected region, Sentinel-1 SAR VV polarization data in Interferometric Wide (IW) mode were used.
(Table 1) summarizes the input features that were considered in this study, including their data sources, spatial resolution and relevance to the assessment of flood impact. The selected features collectively represent vegetation condition, surface water dynamics, flood-related backscatter changes, rainfall intensity, terrain characteristics, land use patterns, and proximity to major hydrological and coastal

features factors that are widely recognized as key determinants of flood damage in coastal environments.

Table 1: Input features, sources and relevance

| Feature Name | Source | Resolution | Description | Relevance |
|---|---|---|---|---|
| NDVI | Sentinel-2 | 10m | Vegetation index | It indicates vegetation health. |
| NDWI | Sentinel-2 | 10m | Water index | Reflects water distribution. |
| VV_pre, VV_post | Sentinel-1 | 10m | Pre/post-event backscatter | Measures flood impact. |
| Rain_3/10day | CHIRPS | 5km | Cumulative rainfall data | Important for hydrometeorology. |
| Elevation/Slope | SRTM | 30m | Elevation and terrain data | Identifies flood-prone areas. |
| LULC | Sentinel-2 | 10m | Land use/land cover classification | For land-use specific damage. |
| Dist_to_river, Dist_to_Coast | Derived | 10m | Distance to rivers and coast | Crucial for flood vulnerability. |

A focal median filter with a 30 m radius was applied to Sentinel-1 SAR data to reduce speckle noise and improve surface homogeneity, thereby enhancing flood detection accuracy (Chini et al., 2019). All data sets were re-projected to WGS 84 / UTM Zone 45N and all data sets were resampled to a common 10 m resolution using Snap Raster tool in ArcGIS Pro. NDVI, NDWI, and VV backscatter differences were derived using Raster Calculator, while Extract by Mask and Reclassify tools were used for AOI clipping and LULC and flood threshold coding. All the features were stacked into multi-band raster for each event including spectral indices, SAR metrics, rainfall, topography, and proximity variables. Random sampling generated 50,000 points per event, and feature values were extracted and exported as CSV files. Data from five flood events 250,000 samples (5*50,000) were combined and cleaned using Python (Pandas, NumPy) to create a working and balanced dataset for machine learning modeling. This workflow ensured that all predictors would be consistent and balanced spatially between events to make the data as much predictive as possible.

### 2.4 Label Preparation

Flood masks were created by using the CDAT algorithm (Long et al., 2014), which was applied on Sentinel-1 VH backscatter data. The difference of post-flood and pre-flood backscatter ($D = VV_{\text{post}} - VV_{\text{pre}}$) was calculated and statistical threshold was further applied to distinguish flooded (1) pixels from unflooded (0) pixels. Pixels were considered to be flooded when the difference value (D) between the two radar images was less than -3 dB indicating that the radar backscatter intensity was affected dramatically. The thresholding formula used was:

$$PF < (\mu[D] - f_c \times \sigma[D] \quad (1)$$

Where $PF$ represents the pixel value, $\mu[D]$ and $\sigma[D]$ are the mean and standard deviation of the difference image, and $f_c$ is an empirical constant (1.5–2.0). This method of separating flooded areas is effective because the high reduction in radar reflectance separates the flooded areas from the unaffected areas. The resulting binary flood mask was checked against Sentinel-2 NDWI difference maps to make spatial validation (Long et al., 2014).

### 2.5 Sampling and Dataset Formation

After flood label generation, the extracted dataset exhibited serious class imbalance where the number of non-flooded samples was much more than flooded ones. To prevent a biased model and so enhance the classification, we applied a stratified resampling strategy. The data set was balanced by randomly

picking the same number of flooded (label = 1) and non-flooded (label = 0) samples, in order to be exactly 50:50, respectively. This balancing step ensured that the model does not show preference towards the majority class and ensured that the model is predictive-i.e. balance the training and validation. Then, the machine learning modeling was done in Python taking the resultant balanced dataset (flood and non-flood pixels) only as the input.

### 2.6 Machine Learning Modeling

The balanced multi-event rainfall dataset was used as the training dataset and two ensemble classifiers, RF and LightGBM, were applied to predict the classes of pixels as flooded (1) and non-flooded (0). RF builds many decorrelated decision trees on bootstrap resamples and aggregates their votes, which reduces variance and overfitting while handling non-linear interactions typical of EO predictors (Belgiu & Drăgu, 2016). LightGBM is a gradient-boosting framework that adds trees sequentially, fitting each to the residuals of a differentiable loss (binary log-loss), and uses histogram binning plus leaf-wise growth for speed and accuracy on large geospatial tables (Ke et al., 2017). We performed an 80/20 stratified split (by class and event) for training/testing and ran 5-fold stratified cross-validation on the training set. Hyperparameters were tuned to maximize Area Under the Receiver Operating Characteristic Curve (ROC-AUC) using randomized/Bayesian search and a short grid around the best region: for RF n_estimators (≈200–1000), max_depth, min_samples_split/leaf, max_features (sqrt/log2); for LightGBM learning rate (0.01–0.1), trees (≈300–2000) with early stopping, num_leaves, max_depth, feature/bagging fractions, and L1/L2 regularization. Class weights were kept balanced; we verified robustness with class_weight='balanced' (RF) or scale_pos_weight (LightGBM) when slight drift appeared.

Performance was computed on the held-out test set and summarized (mean±SD) across Cross-Validation (CV) folds. With TP, FP, TN and FN from the confusion matrix:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \tag{5}$$

ROC curves plot True Positive Rate (TPR) vs. False Positive Rate (FPR) as the threshold varies. AUC is the integral of ROC and reflects ranking ability. We report confusion matrices, Precision, Recall, F1-score, Accuracy and ROC-AUC for both RF and LightGBM on the test set. Feature importance was extracted as Gini importance RF and gain LightGBM to interpret drivers; SAR VV_after and VV_before consistently ranked highest, followed by distance-to-coast/river and NDWI/NDVI. Feature importance was extracted as Gini importance RF and gain LightGBM; VV_after and VV_before consistently dominated, followed by proximity (distance-to-coast/river) and spectral indices (NDWI/NDVI). The best model (by AUC/F1) produced pixel-level flood probabilities, which we thresholded to 0/1 for the binary flood map used in damage overlays and optionally reclassified (Low/Moderate/High) for planning maps; a light 3×3 majority filter removed isolated artefacts prior to LULC intersection.

### 2.7 Flood Prediction Mapping and Damage Assessment

The predicted flood extent was generated by running the trained ML models on the prepared dataset to classify flooded and non-flooded zones. The model outputs were aggregated based on their geographic coordinates (X, Y) and imported into ArcGIS Pro for spatial processing. Using the Inverse Distance Weighted (IDW) interpolation method, a continuous flood extent surface was created to represent flood distribution across the study area. This interpolated flood surface was then overlaid with the LUCL

containing cropland, fishpond, built-up, vegetation and waterbody. Through spatial intersection and zonal analysis, the total flooded area was quantified for each land-use class, enabling the assessment of flood impact across different land-use types.

## 3. RESULTS

### 3.1 NDVI and NDWI Variations Before and After Floods

Before and after flood events, the NDVI and NDWI maps provide us with a good indication about the health of the vegetation and about distribution of the water. In the pre-flood NDVI (Figure 2) all the values are typically centered around 0 to 1, which shows a moderate to dense vegetation cover. After the flood, NDVI values decreased significantly and especially for the low-lying and flood-prone areas,

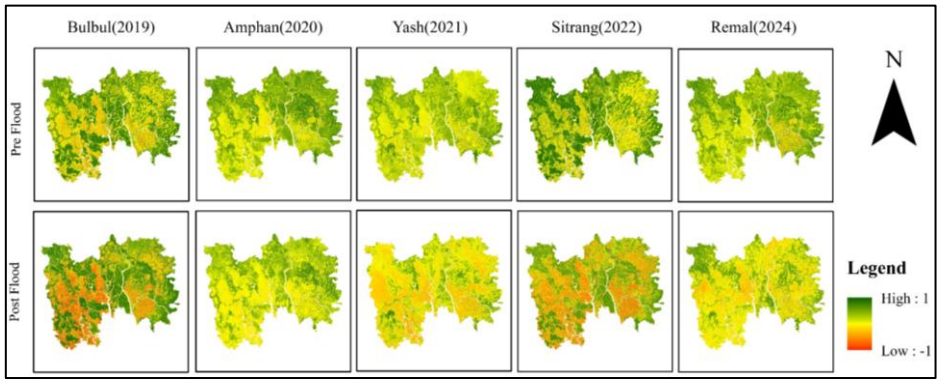



Figure 2: NDVI (Pre vs. Post) by event.

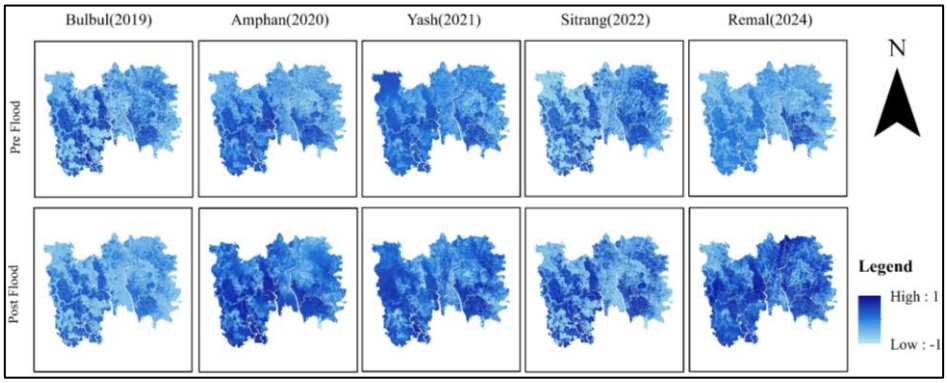
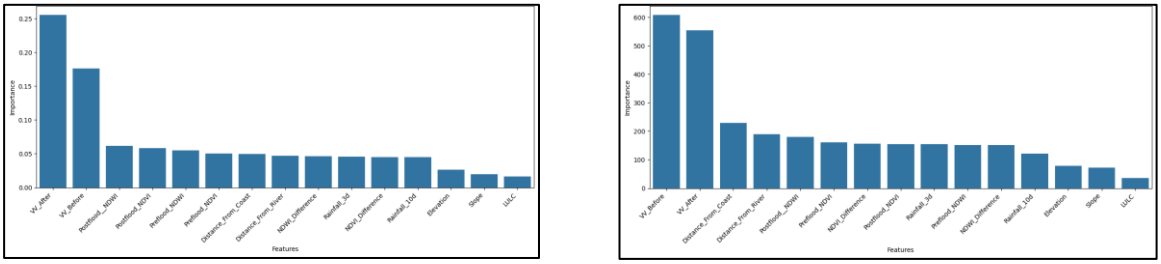



Figure 3: NDWI (Pre vs. Post) by event.

indicating vegetation stress or loss due to prolonged flooding. Similarly, the NDWI (Figure 3) maps illustrate the distribution of water, with the values after the floods depicting that there is strong flooding all over the impacted areas.

### 3.2 Key Predictors in Flood Mapping with SAR and Geospatial Data

Our feature-importance analysis (Figures 4(a) & 4(b)) indicates that VV_after and VV_before from Sentinel-1 are the strongest variables for flood mapping that highlight the importance of changes in the radar backscatter before and after flooding. In addition, other variables of interest were NDWI and NDVI capturing water presence and vegetation response, distance from the coast and river that estimates geographic flood susceptibility. LightGBM represents values that indicate proximity and models backscatter values; while RF model slightly emphasize spectral indices (NDVI, NDWI) more. Finally, radar-based features are the dominants in terms of predictive performance and validates that they are very much important to capture the presence and outlines of the flood-affected areas in coastal Bangladesh.

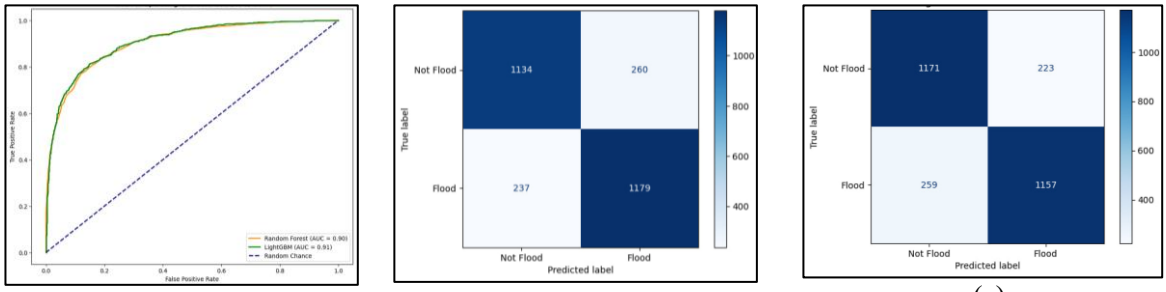


(a) (b)

Figure 4: Feature importance (a) RF and (b) LightGBM.

### 3.3 Comparison of Model Performance for Flood Detection

Figure 5(a) reveals that both the models: RF (AUC = 0.90) and LightGBM (AUC = 0.91) are highly discriminative, although LightGBM has a small advantage. These two curves are also far above the 1:1 line and this proves performance more than chance.




(a) (b) (c)

Figure 5: (a) ROC curves; Confusion Matrices (b) RF (c) LightGBM

Table 2: Model performance

| Metric | RF | LightGBM |
|---|---|---|
| Accuracy | 0.8231 | 0.8285 |
| Precision | 0.8193 | 0.8384 |
| Recall | 0.8326 | 0.8171 |
| F1-Score | 0.8259 | 0.8276 |
| ROC AUC | 0.9049 | 0.9093 |

According to the confusion matrices (Figure 5(b) & 5(c)), the two classifiers can correctly differentiate between flood and non-flood pixels. In the case of RF, TP = 1179, TN = 1134, FP = 260, FN = 237; in the case of LightGBM, TP = 1157, TN = 1171, FP = 223, FN = 259. Therefore, there are fewer false positives (223 vs. 260) with LightGBM, i.e. it is more likely to not mislabel non-flooded areas, whereas RF has fewer false negatives (237 vs. 259), i.e. it is slightly more likely to get flooded pixels. Briefly, both the models offer a strong predictive power: LightGBM is better at specificity (less FPR/ more accuracy) and RF at sensitivity (more recall). Which is the best model will thus be based on the application (preventing false alarms vs. overlooking floods). (Table 2) compares the performance of the RF and LightGBM models for flood prediction. Both models performed well, with accuracy of 0.8231 (RF) and 0.8285 (LightGBM). LightGBM achieved slightly higher precision (0.8384) and F1-score (0.8276), while RF showed a marginally better recall (0.8326), indicating both models are reliable, with LightGBM performing marginally better overall.

### 3.4 Model Outcome

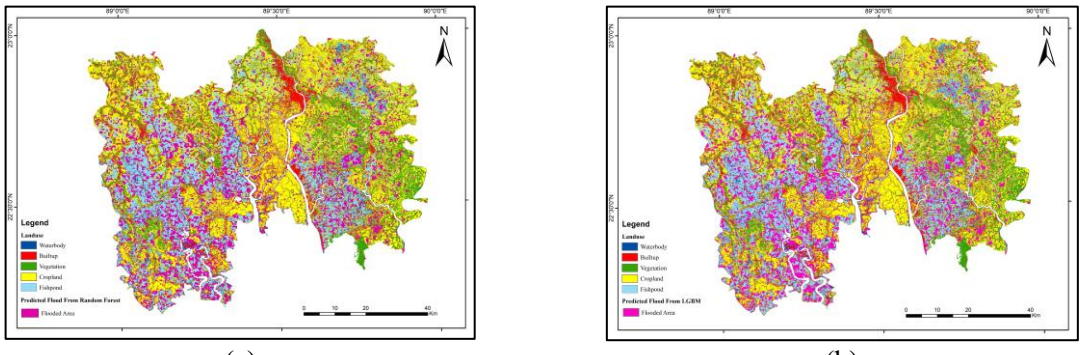
RF (Figure 6(a)) mapped 14.081% of the AOI as flooded, yielding a broad, continuous footprint along tidal rivers and coastal polders with additional speckled patches in low-lying cropland/aquaculture mosaics consistent with RF's higher recall and a tendency to include fringe/transitional pixels so fewer flooded areas are missed. LightGBM (Figure 6(b)) classified 14.2818% of the AOI as flooded (≈ +0.20 percentage points vs. RF) and produced a more compact, coherent map that concentrates on high-confidence flood corridors along major channels and the coast, with fewer isolated inland pixels reflecting higher precision and reduced false positives while still delineating slightly larger, continuous inundation zones. Overall, both models agree on the dominant river–coast pathways and estimate very similar total flooded area (~14.1–14.3%); RF is preferable for sensitive screening (minimizing misses), whereas LightGBM is preferable for conservative mapping (minimizing false alarms) with cleaner, more contiguous patches.


(a) (b)

Figure 6: (a) Predicted flood zones by RF, (b) Predicted flood zones by LightGBM

### 3.5 Flood Damage Mapping Across Land Use Types in Coastal Areas

The LULC classification identified five main categories: cropland (41.06%), fishpond (24.18%), built up (13.34%), vegetation (12.78%) and water bodies (8.64%). It indicates a heavy impact on agriculture and aquaculture, which are especially sensitive to the damage caused by floods and to saltwater intrusion. The land-use flood damage estimation (Figure 7) reflects the way the flood impacts various sections of the coastlines. Figure 7(a) is the RF overlay and (Figure 7(b)) is the LightGBM overlay. As LightGBM proved to be more accurate and more precise, we are using it as our final flood-extent model, and the rest of the area calculations. The LightGBM map and land-use layer intersect informs us (Table 3) that fishponds and croplands are the most vulnerable, as approximately 357.50km$^2$ and 244.92km$^2$ are flooded, respectively- showing the extent to which aquaculture and farming are exposed to these low-lying, polder terrain around tidal channels and distributaries. Constructed areas also become severely damaged (~83.64 -km$^2$ flooded), which may cause a serious inconvenience to housing, services, and road systems. Water bodies exhibit approximately 85.73km$^2$ of flooded land the water dispersing in the events and the least affected area belongs to vegetation (approximately 48.08km$^2$) which is understandable due to the elevation and drainage that is more effective in most of the vegetated or forested areas.




(a) (b)

Figure7: Land-use overlay with predicted flood (a) RF (b) LightGBM

Summing up, the findings indicate that the losses of flooding are concentrated on the productive land-use (particularly aquaculture and agriculture), and underscores the significance of enhancing embankment, controlling polder drainage, and land use planning to reduce recurrent losses in the area, at least in coastal Bangladesh.

Table 3: Land-use wise flood-affected area (km²)

| Land Use | Area (Sq.km) | Flooded (Sq.km) | Non-Flooded Area (Sq.km) |
|---|---|---|---|
| Waterbody | 475.6528 | 85.73344448 | 389.9193369 |
| Builtup | 768.5379 | 83.6410972 | 684.8968324 |
| Vegetation | 775.2673 | 48.08124755 | 727.1860561 |
| Cropland | 2557.794 | 244.9247784 | 2312.869664 |
| Fishpond | 1481.996 | 357.5006458 | 1124.495374 |

## 4. CONCLUSION

This paper demonstrates that multi source remote sensing images and machine learning algorithms RF and LightGBM can be successfully used in predicting floods and damage evaluation in the coastal areas of Bangladesh. Using Sentinel-1, SAR and Sentinel-2 data, with rainfall, topography, and land-use variables, the framework developed was found to have a high classification accuracy (greater than 82 per cent) and is very reliable in model classification that is, distinguishing between flooded and non-flooded regions. The analysis disclosed that the most significant predictors were radar backscatter difference (VVbefore and VVafter), which supported the significance of SAR data to detect a flood. Spatial superimposition of the forecasted flood area and land-use type depicted that fishponds and croplands were the most affected, and this indicates how the aquaculture and agricultural industries were vulnerable in the low-lying areas. This combined strategy offers a scalable and data-driven framework of operational flood monitoring and land-use specific estimation of damage, making disaster preparedness and sustainable land use in the coastal areas of Bangladesh a reality. The future research might involve the incorporation of real-time satellite information and hydrodynamic models in order to improve the early warning systems and responsibilities.

## DECLARATION OF USE OF AI

The authors declare that artificial intelligence (AI) tools were used only for language polishing, grammar correction, and improving the clarity of the manuscript. AI tools were not used for data analysis, model development, result generation, interpretation of results, or drawing scientific conclusions. All technical content, methodology, analysis, and interpretations were conceived, executed, and verified by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the work.

## REFERENCES

Abdo, H. G., Richi, S. M., Bindajam, A. A., Zerouali, B., Katipoğlu, O. M., Prasad, P., Alharbi, M. M., Ramadan, M. S., Mallick, J., & Ghribi, M. (2025). Machine learning-driven flood susceptibility mapping: comparing model performances and feature influences in a coastal watershed of the Eastern mediterranean. *Journal of Coastal Conservation*, *29*. https://doi.org/10.1007/s11852-025-01138-6

acaps. (2024). *Bangladesh impact of Tropical Cyclone Remal*. https://www.Acaps.Org/En/Countries/Archives/Detail/Bangladesh-Impact-of-Tropical-Cyclone-Remal.

Aljazeera. (2019). *Cyclone kills at least 14 in India and Bangladesh*. https://www.Aljazeera.Com/News/2019/11/10/Deadly-Cyclone-Bulbul-Weakens-over-Bangladesh?

Ashrafuzzaman, M., Santos, F. D., Dias, J. M., & Cerdà, A. (2022). Dynamics and Causes of Sea Level Rise in the Coastal Region of Southwest Bangladesh at Global, Regional, and Local Levels. *Journal of Marine Science and Engineering*, *10*(6), 779. https://doi.org/10.3390/JMSE10060779/S1

BBS. (2022). *Population and Housing Census 2022*. https://bbs.portal.gov.bd/sites/default/files/files/bbs.portal.gov.bd/page/b343a8b4_956b_45ca_872f_4cf9b2f1a6e0/2024-01-31-15-51-b53c55dd692233ae401ba013060b9cbb.pdf

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 114, pp. 24–31). Elsevier B.V. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Brac. (2020). *Cyclone Amphan Situation Report*. Https://Response.Brac.Net/Wp-Content/Uploads/2020/08/4_Situation-Update_Cyclone-Amphan_23-May-2020.Pdf.

Chini, M., Pelich, R., Pulvirenti, L., Pierdicca, N., Hostache, R., & Matgen, P. (2019). Sentinel-1 InSAR coherence to detect floodwater in urban areas: Houston and hurricane harvey as a test case. *Remote Sensing*, *11*. https://doi.org/10.3390/rs11020107

Choudhury, K. N., & Yabar, H. (2025). Flood Hazard Assessment and Monitoring in Bangladesh: An Integrated Approach for Disaster Risk Mitigation. *Earth (Switzerland)*, *6*. https://doi.org/10.3390/earth6030090

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. E. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, *45*. https://doi.org/10.1029/2005RG000183

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific Data*, *2*. https://doi.org/10.1038/sdata.2015.66

Giustarini, L., Hostache, R., Kavetski, D., Chini, M., Corato, G., Schlaffer, S., & Matgen, P. (2016). Probabilistic Flood Mapping Using Synthetic Aperture Radar Data. *IEEE Transactions on Geoscience and Remote Sensing*, *54*, 6958–6969. https://doi.org/10.1109/TGRS.2016.2592951

Hoque, M. Z., Ahmed, M., Islam, I., Cui, S., Xu, L., Prodhan, F. A., Ahmed, S., Rahman, M. A., & Hasan, J. (2022). Monitoring Changes in Land Use Land Cover and Ecosystem Service Values of Dynamic Saltwater and Freshwater Systems in Coastal Bangladesh by Geospatial Techniques. *Water (Switzerland)*, *14*. https://doi.org/10.3390/w14152293

Kader, Z., Islam, M. R., Aziz, M. T., Hossain, M. M., Islam, M. R., Miah, M., & Jaafar, W. Z. W. (2024). GIS and AHP-based flood susceptibility mapping: a case study of Bangladesh. *Sustainable Water Resources Management*, *10*. https://doi.org/10.1007/s40899-024-01150-y

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

Lang, F., Zhu, Y., Zhao, J., Hu, X., Shi, H., Zheng, N., & Zha, J. (2024). Flood Mapping of Synthetic Aperture Radar (SAR) Imagery Based on Semi-Automatic Thresholding and Change Detection. *Remote Sensing*, *16*. https://doi.org/10.3390/rs16152763

Long, S., Fatoyinbo, T. E., & Policelli, F. (2014). Flood extent mapping for Namibia using change detection and thresholding with SAR. *Environmental Research Letters*, *9*. https://doi.org/10.1088/1748-9326/9/3/035002

Lou, L., Chen, C., Li, M., & Liu, K. (2024). Comparative Analysis of Water Body Extraction Accuracy Based on Thresholding Method. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-4–2024*, 331–336. https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-331-2024

Rahman, M. M., Kamruzzaman, M., Deb, L., & Islam, H. M. T. (2025). Flood mapping, damage assessment, and susceptibility zonation in northeastern Bangladesh in 2022 using geospatial datasets. *Progress in Disaster Science*, *25*. https://doi.org/10.1016/j.pdisas.2024.100402

Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T. (2019). Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. *Earth Systems and Environment*, *3*, 585–601. https://doi.org/10.1007/s41748-019-00123-y

reliefweb. (2021). *Bangladesh: Cyclone YAAS - Final Report (n° MDRBD027) - Bangladesh*. https://Reliefweb.Int/Report/Bangladesh/Bangladesh-Cyclone-Yaas-Final-Report-n-Mdrbd027.

Rifath, A. R., Muktadir, M. G., Hasan, M., & Islam, M. A. (2024). Flash flood prediction modeling in the hilly regions of Southeastern Bangladesh: A machine learning attempt on present and future climate scenarios. *Environmental Challenges*, *17*. https://doi.org/10.1016/j.envc.2024.101029

Sampurno, J., Putra, M. G. E., Faryuni, I. D., & Adriat, R. (2024). Flood impact assessment in remote areas using machine learning, SAR, and GIS: a case study of Ngabang District, Indonesia. *Journal of Hydroinformatics*, *26*, 2928–2938. https://doi.org/10.2166/hydro.2024.324

Sarkar, S. K., Rudra, R. R., & Santo, M. M. H. (2024). Cyclone vulnerability assessment in the coastal districts of Bangladesh. *Heliyon*, *10*(1). https://doi.org/10.1016/j.heliyon.2023.e23555

Shampa, Nasir, N. N., Winey, M. M., Dey, S., Zahid, S. M. T., Tasnim, Z., Islam, A. K. M. S., Hussain, M. A., Hossain, M. P., & Muktadir, H. M. (2025). Integration of Remote Sensing and Machine Learning Approaches for Operational Flood Monitoring Along the Coastlines of Bangladesh Under Extreme Weather Events. *Water (Switzerland)*, *17*. https://doi.org/10.3390/w17152189

Singha, C., Rana, V. K., Pham, Q. B., Nguyen, D. C., & Łupikasza, E. (2024). Integrating machine learning and geospatial data analysis for comprehensive flood hazard assessment. *Environmental Science and Pollution Research*, *31*, 48497–48522. https://doi.org/10.1007/s11356-024-34286-7

Soudagar, R., Chowdhury, A., & Bhardwaj, A. (2025). Enhanced large-scale flood mapping using data-efficient unsupervised framework based on morphological active contour model and single synthetic aperture radar image. *Journal of Environmental Management*, *380*. https://doi.org/10.1016/j.jenvman.2025.124836

Subhani, R., & Ahmad, M. M. (2019). Socio-economic impacts of cyclone aila on migrant and non-migrant households in the southwestern coastal areas of Bangladesh. *Geosciences (Switzerland)*, *9*. https://doi.org/10.3390/geosciences9110482

Toma, A., Șandric, I., & Mihai, B. A. (2024). Flooded area detection and mapping from Sentinel-1 imagery. Complementary approaches and comparative performance evaluation. *European Journal of Remote Sensing*, *57*. https://doi.org/10.1080/22797254.2024.2414004

Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing*, *37*, 2990–3004. https://doi.org/10.1080/01431161.2016.1192304

Uddin, K., Matin, M. A., & Meyer, F. J. (2019). Operational flood mapping using multi-temporal Sentinel-1 SAR images: A case study from Bangladesh. *Remote Sensing*, *11*. https://doi.org/10.3390/rs11131581

Unicef. (2022). *UNICEF BANGLADESH Cyclone SITRANG*. https://www.Unicef.Org/Media/129631/File/Bangladesh-Humanitarian-SitRep-%28Cyclone-SITRANG%29-25-October-2022.Pdf.

Zhang, Y., Liu, X., Sun, W., You, T., & Qi, X. (2025). Multi-Threshold Remote Sensing Image Segmentation Based on Improved Black-Winged Kite Algorithm. *Biomimetics*, *10*. https://doi.org/10.3390/biomimetics10050331