

RESEARCH

Open Access



# Development of an automated method for flood inundation monitoring, flood hazard, and soil erosion susceptibility assessment using machine learning and AHP–MCE techniques

A. Jaya Prakash<sup>1</sup>, Sazeda Begam<sup>2</sup>, Vít Vilímek<sup>3</sup>, Sujoy Mudi<sup>4</sup> and Pulakesh Das<sup>5\*</sup>

## Abstract

**Background** Operational large-scale flood monitoring using publicly available satellite data is possible with the advent of Sentinel-1 microwave data, which enables near-real-time (at 6-day intervals) flood mapping day and night, even in cloudy monsoon seasons. Automated flood inundation area identification in near-real-time involves advanced geospatial data processing platforms, such as Google Earth Engine and robust methodology (Otsu's algorithm).

**Objectives** The current study employs Sentinel-1 microwave data for flood extent mapping using machine learning (ML) algorithms in Assam State, India. We generated a flood hazard and soil erosion susceptibility map by combining multi-source data on weather conditions and soil and terrain characteristics. Random Forest (RF), Classification and Regression Tool (CART), and Support Vector Machine (SVM) ML algorithms were applied to generate the flood hazard map. Furthermore, we employed the multicriteria evaluation (MCE) analytical hierarchical process (AHP) for soil erosion susceptibility mapping.

**Summary** The highest prediction accuracy was observed for the RF model (overall accuracy [OA] > 82%), followed by the SVM (OA > 82%) and CART (OA > 81%). Over 26% of the study area indicated high flood hazard-prone areas, and approximately 60% showed high and severe potential for soil erosion due to flooding. The automated flood mapping platform is an essential resource for emergency responders and decision-makers, as it helps to guide relief activities by identifying suitable regions and appropriate logistic route planning and improving the accuracy and timeliness of emergency response efforts. Periodic flood inundation maps will help in long-term planning and policymaking, flood management, soil and biodiversity conservation, land degradation, planning sustainable agriculture interventions, crop insurance, and climate resilience studies.

**Keywords** Machine learning, Flood hazard, Multicriteria evaluation, Assam, Susceptibility

\*Correspondence:

Pulakesh Das

das.pulok2011@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Introduction

Abrupt changes in land use and land cover (LULC), anthropogenic activities, and alterations in climatic conditions are the primary drivers of hydrological extremes worldwide. Extreme precipitation, glacier lake outbursts, and dam failure often cause floods, which may increase with climate change (Acquaotta et al. 2019; Begam & Sen 2019; Harrison et al. 2018). Prolonged water inundation and flash floods damage natural resources and infrastructure, cause human life loss, reduce soil fertility, degrade agricultural and dependent socio-economy, ecosystems, biodiversity, habitat, etc. (Díaz et al. 2019; Weiskopf et al. 2020; Prakash et al. 2023). Anthropogenic landscape modifications via unsustainable land use practices, construction/developmental activities in the flood plains, alteration in river morphology, streams, floodplains, etc., determine the impact of floods on the course and consequences in different parts of a river basin (Dar et al. 2019). Several studies have assessed and predicted flood-related economic losses and damage to human well-being in India and elsewhere (Borah et al. 2023; Venkataramanan et al. 2019). Extreme flood events affect approximately 21 million people globally, which may increase to 54 million by 2030 due to climate change and socioeconomic growth (Luo et al. 2015). According to Patankar (2019), 278 flood events occurred between 1980 and 2017 in India, causing a loss of 58.7 billion USD and affecting 750 million people. Gangopadhyay et al. (2018) reported that flood cause around USD 7500 million in economic losses annually in India. Human life and economic losses due to catastrophic floods can be avoided using advanced early warning systems and modern techniques. The timely availability of flood inundation maps is essential for minimizing flood hazards, planning and providing emergency services, developing mitigation plans, and baseline data layers for policymakers. Satellite remote sensing effectively monitors surface water bodies and flood inundation by employing suitable spectral bands in the electromagnetic spectrum and providing a synoptic view in near real-time and past conditions (Das et al. 2021a). Simultaneously, geostatistical analysis employing the spatial layers of pre-flood LULC conditions, road networks, human settlements, topography, etc., enables flood hazard assessment and aids in relief, preparedness, and prevention efforts. Moreover, sophisticated analyses and hydrological models (e.g. 1D-2D MIKE FLOOD, HEC-RAS, and Global Flood Monitoring System [GFMS]) allow the simulation of the flood extent and quantify impact assessment (Kumar et al. 2020).

The accuracy and consistency of flood inundation mapping using remote sensing data relies on suitable sensors, operating wavelengths, and adopted approaches.

Although water identification and flood inundation assessment are easier with optical satellite data, there are inherent problems in obtaining cloud-free satellite images during the peak flood period, that is, the monsoon season. In contrast, microwave data have a relative advantage over optical data, as longer-wavelength microwave signals can penetrate clouds and enable surface feature mapping. Active microwave sensors capture backscatter signals from the terrain, which vary based on multiple terrain conditions and surface parameters, such as the roughness and dielectric constant. Waterbodies have a smooth surface (lower roughness) and a higher dielectric constant, which causes significantly lower backscatter from waterbodies than other land surface features and enables the identification of waterbodies in microwave data (Borah et al. 2018). The latest Sentinel-1 Synthetic Aperture Radar (SAR) data, freely available in the public domain, allows for systematic water inundation mapping at high spatial and temporal resolutions. The availability of pre-processed and rectified SAR data in platforms such as Google Earth Engine (GEE) enables bulk data processing for a large region in a shorter time (in minutes) without downloading the actual image tiles (Das et al. 2021b; Prakash et al. 2023). Several methods have been developed for water inundation area mapping using SAR data, such as thresholding, clustering, and deep learning, which discriminate water bodies from other land surface features (Mudi et al. 2022; Borah et al. 2018; Konapala et al. 2021; Pandey et al. 2022). Otsu's algorithm is one of the most robust methods for automated surface-water area mapping in the absence of reference observations. This approach identifies an intensity threshold value derived from the radiometric histogram, categorizing images into two classes: foreground and background (Otsu 1979). The threshold value was estimated by minimizing intraclass intensity or interclass variance. Several studies have employed this algorithm for surface water inundation area mapping and reported high accuracy (Mudi et al. 2022; Prakash et al. 2023).

Soil erosion is a complex process influenced by various geomorphological factors and has a detrimental effect on soil fertility and crop production. Many studies have shown that soil erosion in a watershed is primarily regulated by drainage characteristics such as stream, drainage density, flow accumulation, topography, and soil characteristics (Arabameri et al. 2020; Bhattacharya et al. 2020; Khatun et al. 2022). Additionally, intense agriculture, overgrazing, and LULC changes, including deforestation, urbanization, and loss of surface waterbodies (ponds and lakes), intensify surface runoff and lead to higher soil erosion (Das et al. 2018; Prashanth et al. 2023; Rather et al. 2017). Smolíková

et al. (2016) studied the debris flow in the Smědavská hora Mt, Czech Republic, and reported that antecedent precipitation index (API) and extreme precipitation regulate the nature of debris flow. The study of drainage characteristics has proven to be highly effective and has been widely employed in soil erosion studies. The morphometric features, such as stream orders, basin area, perimeter, and length of streams, are used to assess the impact of stream characteristics on land surface processes at a watershed or sub-watershed scale. Satellite data-derived digital elevation models (DEM) are widely used to characterize topographic complexity and derive drainage networks, watershed boundaries, and various morphometric indicators. Several studies have evaluated soil erosion susceptibility at the sub-watershed scale in order to develop sustainable watershed management plans (Bhattacharya et al. 2020; Mosavi et al. 2022). Precipitation, surface runoff, land surface features, soil type, and topographic variables mostly regulates the soil erosion and deposition or source-to-sink sediment transport in a watershed. In addition, flooding in river basins is one of the major drivers of soil erosion (Mishra et al. 2022; Bordoloi et al. 2020).

Previous studies have employed multicriteria analysis (MCA) for soil erosion susceptibility mapping, wherein they analyzed the LULC and morphometric indicators of the sub-watersheds for comparative assessment (Altaf et al. 2014; Rather et al. 2017). Bhattacharya et al. (2020) employed four multicriteria decision-making (MCDM) methods as *VlseKriterijumska optimizacija I Kompromisno Resenje* (VIKOR), technique for order preference by similarity to ideal solution (TOPSIS), simple additive weighting (SAW), compound factor (CF), and Soil and Water Assessment Tool (SWAT) model for soil erosion susceptibility analysis. Pradhan et al. (2020) assessed the soil erosion susceptibility in the Kosi River basin in the Indian state of Bihar using the Revised Universal Soil Loss Equation (RUSLE) model soil loss estimation followed by Analytical Hierarchical Process (AHP) multicriteria evaluation (MCE) for integration with other variables such as soil and morphometric indicators. In recent years, machine learning (ML) models have gained popularity in diverse data analyses, including natural hazard modeling (Chapi et al. 2017; Khosravi et al. 2019; Towfiqul Islam et al. 2021). ML algorithms, such as decision trees, random forests, and neural networks, can be trained to learn patterns and relations from historical data and make predictions based on these patterns. Mosavi et al. (2020) used multiple methods, including generalized linear model (GLM), flexible discriminate analyses (FDA), multivariate adaptive regression spline (MARS), random forest (RF), and their ensemble employing a set of drivers to assess

soil erosion susceptibility by analyzing the field data on flood events. Their study identified LULC, elevation, aspect, distance to the river, and soil depth as the most important variables for soil erosion.

In complex mountainous regions, such as the Himalayas, flood inundation area mapping and soil erosion susceptibility evaluation are critical because of their linkages to landslides and debris flow. This study was conducted in the Indian state of Assam, which experiences annual recurrent flooding. Previous studies aggregated multiple indicators related to soil, morphometric parameters, and LULC at the sub-watershed scale to assess soil erosion susceptibility. In contrast, this study attempted to analyze soil erosion susceptibility by adding a finer-resolution flood occurrence layer from 2017 to 2021. We employed the GEE platform for quick and near-real-time flood mapping and ML algorithms for flood hazard mapping. The AHP–MCE technique was applied to integrate multiple layers and estimate soil erosion susceptibility.

### Study area

The study area is located in the Indian state of Assam between 88.25° E and 96.00° E longitude and 24.50° N and 28.00° N latitude. Assam's major rivers include the Brahmaputra, Barak, Manas, and Subansiri rivers. The total area of Assam State is ~78,438 km<sup>2</sup>, of which 56,194 km<sup>2</sup> and 22,244 km<sup>2</sup> are occupied by the Brahmaputra and Barak River basins (Govt. for Assam and Water Resources). The Brahmaputra River is a snow- and rain-fed river that flows continuously throughout the year. The total population of Assam is more than 30 million, mostly living along the Brahmaputra River floodplain and primarily dependent on agriculture and allied sectors for sustenance. In addition, the study region is home to several national parks, such as Kaziranga National Park, a UNESCO World Heritage Site in the Eastern Himalayas, an ecologically important region. The Brahmaputra Basin experiences recurrent floods and is vulnerable to climate change due to its close association with the Himalayas, high population, and agricultural dependency (Sharma et al. 2018). Moreover, the higher rates of deforestation, shifting cultivation, and other LULC changes have significantly modified the hydrology of northeast India (Patidar et al. 2022). High river discharges cause significant changes in riverbank erosion (lateral erosion) and river channel morphology. The deposition of the sediment load of the Brahmaputra Delta is the largest in the world, covering an area of 1.76 million km<sup>2</sup>. Each year, flooding has caused the deposition of the sediment load by 1060 m into the Brahmaputra Delta and the Indian Ocean, signifying one of the largest fluvial sediment depositions in the

world (Milliman and Farnsworth 2011). According to Rashtriya Barh Ayog, the 10-year flood-prone area in Assam accounts for approximately 45.36% of the total geographic area of the state. The flood risk assessment conducted by Bhuyan et al. (2023) in the Nagaon district, located in the flood plain of the Brahmaputra River indicated above 90% of the population living in moderate to high-flood risk zones. Past studies reported changes in the socio-economy, reduction in annual income, and population migration in the Brahmaputra flood plain, including the indigenous community in Majuli Island, one of the world's largest inhabited river-made islands due to recurrent floods and erosion (Chaliha et al. 2012; Das 2016; Roy et al. 2020; Saikia 2022). The flood assessment conducted by Mudi et al. (2022) identified an average of 1600 km<sup>2</sup> of water inundation in cropland and 200 km<sup>2</sup> in settlement areas in the Assam state from 2018 to 2020. Whereas the flood impact assessment in three Indian states (Assam, Bihar, and West Bengal) and Bangladesh in the Ganga–Brahmaputra basin in 2020 reported above 23% of the total croplands and 5% of the total settlement areas were inundated.

## Materials and methods

### Satellite data and derived products and flood-causative factor variables

Sentinel-1 is a continuous all-weather, day-and-night, C-band imaging radar mission that operates at a centre frequency of 5.405 GHz with HH+HV, VV+VH, VV, and HH polarization. The interferometric wide-swath (250 km) mode provides a high spatial resolution of 10 m. At the equator, the two-satellite constellation provides a six-day precise repeat cycle, and the data are available in the public domain. Sentinel-1 SAR images (with VV and VH polarization bands) from July to September were accessed from 2016 to 2022. The selection of data from July to September in our study is primarily due to the convergence of factors related to high precipitation and flood susceptibility in India during this timeframe. This period corresponds to the monsoon season, marked by heavy and consistent rainfall, which significantly contributes to flood events by causing rapid runoff and increased river discharge. Historical flood records also indicate a heightened flood risk during these months.

Sentinel-2 is a constellation of two polar-operating satellite boarded sensors operating in the visual range of the electromagnetic spectrum. It provides images in 13 spectral bands: four at 10 m, six at 20 m, and three at 60 m spatial resolution. With two satellites, it provides images at five-day intervals at the equator. Complete cloud-free optical data are mostly unavailable in the monsoon season. One Sentinel-2 image with the least

cloud cover from July 2017 was used to validate the flood map generated using Sentinel-1 data.

The Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) is a 30+ year quasi-global rainfall dataset. CHIRPS incorporates 0.05° resolution satellite imagery with in situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring (Funk et al. 2015). CHIRPS uses several ground-based gauge networks, including the Global Summary of Day (GSOD), Global Historical Climate Network (GHEN), Southern African Science Service Center for Climate Change and Adaptive Land Management (SASSCAL), and the World Meteorological Organization's Global Telecommunication System (GTS) into CHIRP using a modified inverse distance weighting algorithm (Funk et al. 2015). CHIRPS is available on daily, pentad, and monthly scales (available at <ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/>). We used the latest monthly CHIRPS Version 2.0 dataset with a spatial resolution of 0.05°×0.05°.

The Shuttle Radar Topography Mission (SRTM) DEM was used to generate topographic layers and drainage characteristics. The Void Filled one arc-second available at 30 m resolution was accessed from the GEE platform. The National Bureau of Soil Survey and Land Use Planning (NBSS & LUP) gathered soil properties, such as depth, texture, and particle size. The ESRI 10 m global LULC map of 2020 created using Sentinel-2 imagery was used as an input LULC map. The data were generated using a deep learning model trained using over 5 billion hand-labeled Sentinel-2 pixels from over 20,000 sites spread across the globe's major biomes. The global tree canopy cover (TCC) percentage data with 30 m spatial resolution generated by Hansen et al. (2013) was used as input to indicate the percentage of tree cover above 5 m. The TCC% map for 2000 and tree canopy cover loss and gain from 2000 to 2019 were used to generate the TCC% map for 2019.

### Flood mapping methodology

The thresholding method is used for water inundation or flood mapping. Otsu's thresholding method, a nonparametric and unsupervised method that automatically detects the optimal threshold assuming a bimodal histogram of pixel values (Otsu 1979), was employed to determine the optimal threshold to separate water from non-water pixels image-by-image. This method relies on the principle that backscatter values are significantly different (lower) from those of non-water pixels (relatively higher). Otsu's method is primarily dependent on the histogram of the image pixels, which creates a binary image grouping of water

and non-water pixels. The pixels in a given image can be represented in L grey levels (1, 2, 3,...,L) written as Eq. (1)

$$N = n_1 + n_2 + n_3 + \dots + n_L \quad (1)$$

where 1, 2, 3,..., and L represent the pixels in a given image represented in L grey levels,  $n_i$  represents the number of pixels at level i, and N symbolises the total number of pixels (Tiwari et al. 2020). The normalised grey-level histogram and probability distribution are expressed by Eqs. (2), (3), (4). The optimum threshold (k) was obtained by restraining the weighted sum of the intraclass (k) variances of the foreground and background pixels. Thus, the criterion function  $\rho$  was introduced and defined in Eq. (5):

$$P_i = \frac{N_i}{N} \quad (2)$$

$P_i$  represents the probability that a pixel in the image belongs to a specific grey level "i."  $N_i$  represents the number of pixels in the image with the grey level "i," and N is the total number of pixels in the image

$$P_i \geq 0 \quad (3)$$

Equation (3) states that the probability of a pixel belonging to a particular grey level ( $P_i$ ) must be greater than or equal to zero, ensuring non-negative probabilities.

$$\sum_{i=1}^L P_i = 1 \quad (4)$$

Equation (4) states that the sum of all the probabilities ( $P_i$ ) for all possible grey levels "I" must equal 1, indicating that every pixel in the image falls into one of the grey levels.

$$\rho(k) = \frac{\sigma_B^2(k)}{\sigma_{BT}^2} \quad (5)$$

and  $\sigma_B^2$  and  $\sigma_{BT}^2$  are expressed as.

Equation (5), which introduces the variable "k," it represents the criterion function used to find the optimal threshold for separating water from non-water pixels. The objective of Otsu's thresholding method is to maximize this criterion function  $\rho(k)$ , which is the ratio of the interclass variance  $\sigma_B^2(k)$  and the intraclass variance  $\sigma_{BT}^2$ . The threshold "k" is selected to maximize the discriminability between water and non-water pixel values, automating the threshold determination process for water inundation mapping.

$$\sigma_B^2 = \omega_0 \omega_1 (\mu_1 - \mu_0)^2 \text{ and } \sigma_{BT}^2 = \sum_{i=1}^L (i - \mu_T)^2 n_i \quad (6)$$

Otsu's algorithm was implemented in GEE to automatise water inundation mapping using Sentinel-1 data. The permanent water bodies identified in the non-monsoon season were used as a reference to demarcate actual water inundation. The generated output was downloaded and pre-processed, wherein the scattered and isolated pixels were removed, retaining patches with four contiguous pixels as the minimum mapping unit (MMU).

The normalised difference water index (NDWI) was employed to map surface water bodies using Sentinel-2 optical data. The relative surface reflectance differences in the green and NIR bands were employed to highlight the surface water bodies.

$$NDWI = (\rho_{Green} - \rho_{NIR}) / (\rho_{Green} + \rho_{NIR}) \quad (7)$$

$\rho_{Green}$  and  $\rho_{NIR}$  are the reflectance in the green and NIR bands, respectively.

#### Data preparation for flood hazard and soil erosion susceptibility assessment

The selection of factors that influence or regulate flooding in a region is critical for flood hazard modeling. In the present investigation, six major parameters, including topography, drainage network, rainfall, soil, LULC, and TCC%, were considered to determine the characteristics of flood occurrence (Table 1). Topography, drainage patterns, and soil properties directly affect water flow and retention, while rainfall is the primary trigger for floods. LULC changes, such as urbanization, alter surface runoff patterns, and TCC, representing vegetation cover, influences water absorption. These factors are widely recognized in hydrology for their significance in flood risk assessment, ensuring a comprehensive analysis of both natural and anthropogenic elements influencing flood events. Vector soil data (with attributes such as depth, particle size, and texture) were converted into raster layers (Additional file 1: Fig. S1). DEM was used to create different topographic layers such as slope, aspect, terrain ruggedness index (TRI), and topographic wetness index (TWI) (Additional file 2: Fig. S2 and Additional file 3: Fig. S3). Moreover, using the hydrology tools in ArcGIS software, two drainage characteristic layers were created using DEM, such as the drainage network and flow accumulation (Additional file 4: Fig. S4). The Euclidean distance tool was employed to create layers on the distance between the drainage and drainage density (Additional file 5: Fig. S4). A fishnet with a grid size of 1 km was established for the flow accumulation, wherein the

**Table 1** List of data sources used in this study

Data	Source	Remarks
Flood maps	Sentinel-1 data Sentinel-2 data	Resolution: 100 m
Soil parameters (depth, particle size and texture)	NBSS & LUP	Vector data
Precipitation	CHIRPS	Resolution: 5 km
Elevation, slope, aspect, terrain ruggedness index (TRI), topographic wetness index (TWI), drainage network distance to drainage, and drainage density	Derived using SRTM-DEM	Resolution of SRTM-DEM data: 90 m
Flow accumulation		
Land use land cover (LULC)	ESRI	Resolution: 10 m
Tree canopy cover (TCC) percentage	Hansen et al. (2013)	Resolution: 30 m

Note: All data was resampled at 1 km spatial resolution for modeling. The vector data was converted into raster data with a pixel size of 1 km

mean flow accumulation was considered for each grid cell as the representative value (Additional file 5: Fig. S5). Furthermore, DEM data were used to demarcate the sub-basins, wherein the total precipitation from July to September was obtained for different years. The sub-basin precipitation layers were used to create the mean precipitation layers for 2016–2018 and 2019–2021 (Additional file 6: Fig. S6). The TCC% map of 2019 was constructed using the TCC% map of 2000, and the TCC loss and gain were recorded from 2000 to 2019. The TCC% loss was removed, and TCC% gains were added to the TCC% map for 2000 to generate the TCC% map for 2019 (Additional file 7: Fig. S7). All layers were resampled to a 1 km grid cell for further processing.

### Machine learning models

ML techniques are widely used for classification and regression analyses. We applied Random Forests (RF), Support Vector Machine (SVM), and Classification and Regression Trees (CART) in the current study.

Random Forest is an ensemble learning method proposed by Breiman (2001) that combines multiple decision trees to make predictions. In our study, we employed 500 trees in the RF model building. The input values for RF include a set of features related to topography, drainage network, rainfall, soil properties, land use/land cover, and Total Canopy Cover (TCC). Each tree in the forest was constructed using a random subset of the training data. During training, the model estimates the out-of-bag (OOB) error for each tree, allowing for the evaluation of model performance. RF prioritizes critical parameters, such as “ntree,” “mtry,” and variable importance, and provides an independent measurement of error prediction (Adam et al. 2014). “ntree” is the number of decision trees used in the model, whereas “mtry” controls number of input features available to consider at each node. The output values from RF are predictions of flood

susceptibility, with each tree contributing to the final prediction.

Support Vector Machine is a classifier that builds a hyperplane to separate data into multiple groups. We employed the radial basis function (RBF) kernel, known for its effectiveness in SVM applications (Cortes and Vapnik 1995). The input values for SVM consist of various features representing topographical, hydrological, and environmental attributes. SVM aims to maximize the margin between the hyperplane and the nearest data points of either class, known as support vectors. The output values of the model are class labels that indicate the predicted flood susceptibility of each pixel. The choice of kernel function and kernel parameter values, referred to as kernel configuration, significantly affects the SVM performance. In addition, the cost, regularization, and gamma parameters were used in model tuning.

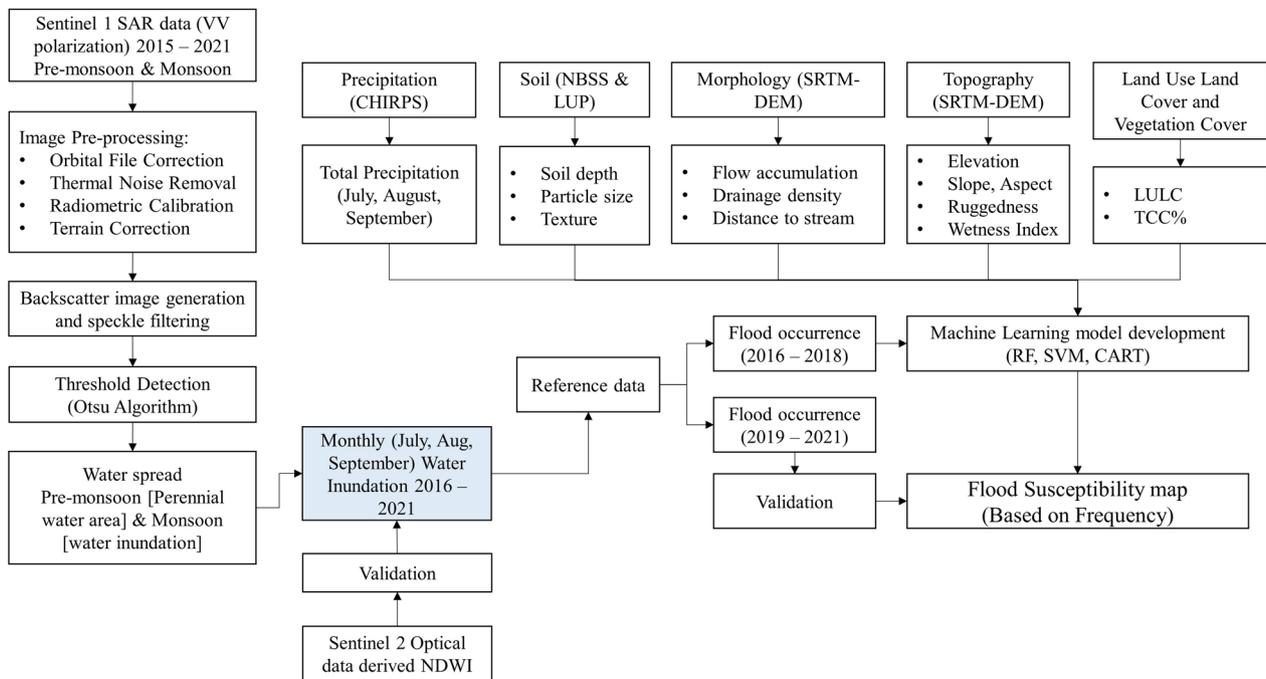
CART is a nonparametric method used for both classification and regression tasks. It partitions the input space into smaller, homogenous, and non-overlapping subregions by recursively splitting the data based on selected input features. In our study, the input values for CART include topographic attributes, drainage network characteristics, rainfall data, soil properties, land use/land cover information, and Total Canopy Cover (TCC). CART considers the interaction among important input factors and can capture nonlinear correlations by applying cascaded threshold values. Each tree in the CART model evaluates the influential input elements based on their contribution to the modeling process (Johnson et al. 2002). When dealing with categorical influencing factors, it is important to note that each category is treated as a separate input variable in tree-based models. The complexity parameter (cp) prevents excessive splitting of the decision tree, which is used to prune the CART model. A very low cp value leads to overfitting, while a large value leads to small tree.

**Flood hazard and soil erosion susceptibility mapping**

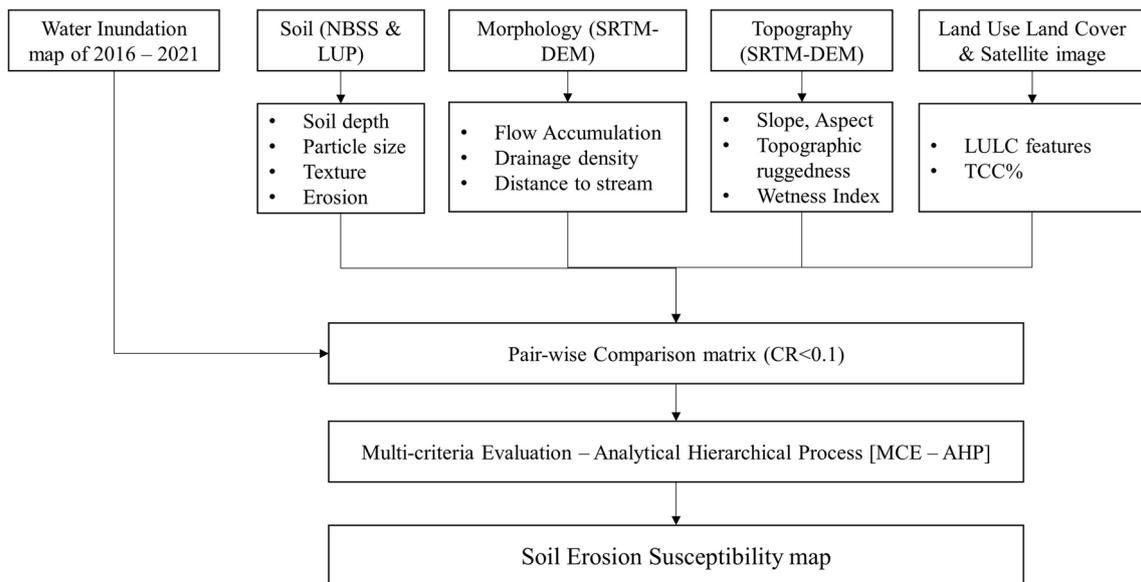
The overall methodology flowcharts for flood hazard and soil erosion susceptibility mapping are shown in Figs. 1 and 2, respectively.

The water-inundated areas mapped during 2016–2018 were used to develop ML models for flood hazard

analysis. Three machine learning algorithms, RF, SVM, and CART, were applied. The flood occurrence map for 2016–2018 was converted into point vector data, wherein 1000 flooded points and 1000 non-flooded points were randomly selected for model building. The approach considered the water inundated areas, wherein



**Fig. 1** Overall methodology flowchart for flood susceptibility mapping



**Fig. 2** Overall methodology flowchart for soil erosion susceptibility mapping

flooding frequency was not considered. The flood and non-flood occurrence data points were segregated into training and testing sets (70% and 30% of the total data points, respectively). The training data points were used to develop the models, and the testing data points were used to validate the modeling accuracy. A second round of model validation was executed. The three ML model predicted maps were validated with the flood occurrence data of 2019 to 2021, wherein 2000 random data points (1000 points for flooded and 1000 points for non-flooded regions) were collected from the flood occurrence map. The ML models were executed in R programming platform, wherein the ‘caret’ package was used for RF and SVM model building and ‘rpart’ for CART model building. tenfold cross-validation was applied. A maximum of 500 trees was employed in the RF modeling, wherein, the mtry value was iterated within a range of 0–20. The ‘radial’ kernal function was applied in the SVM model, wherein the gamma function was iterated for the highest prediction accuracy. Similarly, the cp parameter in the CART model was tuned to obtained the best prediction model.

MCE–AHP was applied to soil erosion susceptibility analysis. The input layers which caused soil erosion due to water inundation were integrated. The AHP method was used to assign weights to the factors evaluated in this study. The AHP method is a semi-quantitative MCDM technique in which pairwise assessment of diverse elements results in conclusions. The factors were scaled to a uniform range of 0–1, based on their potential for soil erosion. The data processing included the hierarchical ordering of the driving factors, assigning a score to each factor based on its relative relevance, creating a pairwise comparison matrix, computing the weight of each factor, and consistency checking. The factors were then

integrated by applying the derived weights to the factors using the following formula (Eq. 8):

$$SES = \sum_{i=1}^n F_i \times W_i \tag{8}$$

where F is a factor, W is the weight, and i is the factor number.

The soil erosion susceptibility map generated using MCE assigns a susceptibility score to different land areas based on various factors contributing to soil erosion. A pairwise comparison matrix was prepared, wherein the relative preferences were determined based on expert opinions and references to various past studies.

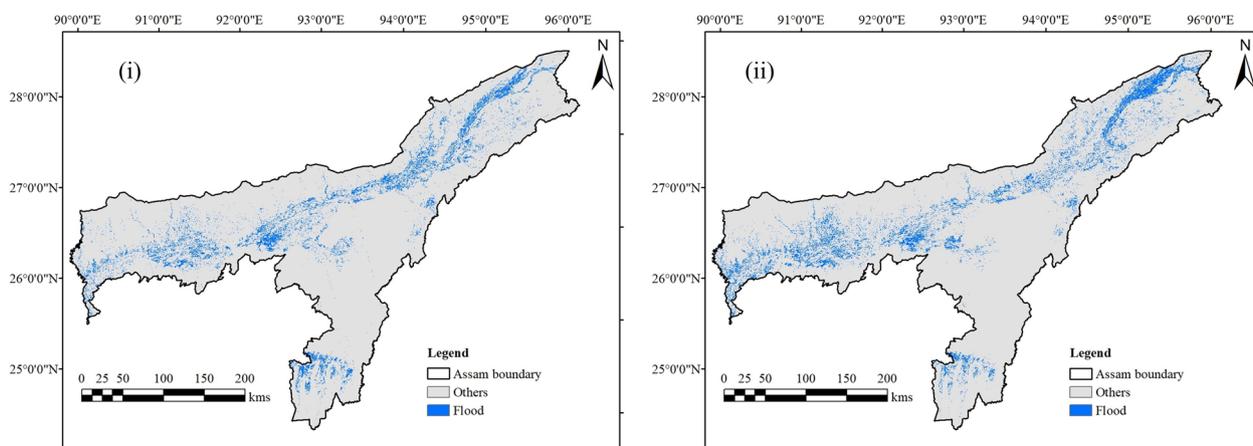
### Results

#### Flood inundation mapping

The water inundation areas were identified using SAR data from July to September 2016–2018 by applying Otsu’s method, as shown in Fig. 3. The total area that experienced water inundation is listed in Table 2. The maximum water inundation in Assam was recorded in 2020 (~3710 km<sup>2</sup>), followed by 2017 (~3035 km<sup>2</sup>).

**Table 2** Flood-affected areas in Assam in different years

Years	Flood inundated area (km <sup>2</sup> )
2016	2244.74
2017	3035.49
2018	1533.27
2019	2650.91
2020	3710.53
2021	1800.53



**Fig. 3** Flood occurrence maps (i) 2016–2018 and (ii) 2019–2021

Water inundation in 2016, 2019, and 2021 was identified as ~2244 km<sup>2</sup>, 2650 km<sup>2</sup>, and 1800 km<sup>2</sup>, respectively. The least water inundation was recorded in 2018, at ~1533 km<sup>2</sup>. The water inundation area identified by applying Otsu’s method to Sentinel-1 SAR data was compared with the Sentinel-2 optical data-derived normalized difference water index (NDWI) map for July 2017 (Additional file 8: Fig. S8). The comparison showed that Otsu’s method captured the water-inundated area well using Sentinel-1 SAR data. Moreover, Otsu’s method was applied to map the recent water inundation (third week of May 2022) and validate the performance of this method. Several ground data points collected from the flooded region in May 2022 were used to verify the identified water-inundated areas (Additional file 9: Fig. S9). Most of the flood-inundated areas were identified along the floodplains of the Brahmaputra River and are mainly croplands (Fig. 3 and Additional file 7: Fig. S7(i)). The flooded areas identified during 2016–2018 and 2019–2021 were overlaid on the LULC map to assess the flood impact (Table 3). The results showed that the majority of flood inundation affected croplands (>3000 km<sup>2</sup>), followed by bareland (including the sand deposit in the flood plain) (>1200 km<sup>2</sup>), scrub (>400 km<sup>2</sup>), and built-up areas (>130 km<sup>2</sup>). The percentage area showed that average 11.5% of the total cropland in Assam were flooded in 2016–2018, which increased to 13.26% in 2019–2021. Similarly, an average 1.31% of the total built-up area in

Assam were flooded in 2016–2018, which increased to 1.39% in 2019–2021. Most of the flooded vegetation class (vegetation cover in the floodplain) experienced water inundation (>23%). In comparison, forest cover (i.e. trees) was least impacted (<100 km<sup>2</sup>), as most of the total forest area in Assam is situated at a higher altitude. However, the bareland and scrubland along the river floodplain were flooded annually.

**Flood hazard mapping**

Three ML models, SVM, RF, and CART, were applied to generate the flood hazard map in Assam. In the RF model, the mtry value of 8 indicated the highest prediction accuracy. A gamma function value of 0.248 and cp value of 0.01 showed the highest prediction accuracy in the SVM and CART model, respectively. The model validation with testing data indicated the highest prediction accuracy for the RF model (overall accuracy [OA]: 82.91%; Kappa: 0.66) followed by SVM (OA: 82.23%; Kappa: 0.64) and CART (OA: 81.9%; Kappa: 0.64) (Table 4). Randomly collected training data points from the 2019 to 2021 flood occurrence maps were used to validate the performances of the three ML models. The assessment indicated the highest accuracy for CART (OA: 83.37%; Kappa: 0.67), followed by RF (OA: 82.36%; Kappa: 0.65) and SVM (OA: 81.15%; Kappa: 0.62). The variable importance of the RF model indicated the differential role of drivers in estimating flood hazards (Additional file 10: Fig. S10). Vegetation cover (TCC), LULC, and river network were identified as the three most important variables in flood hazard mapping. Comparatively, a moderate influence was observed from flow accumulation, elevation, precipitation, and distance from the river. In contrast, lower importance was observed for TRI, slope, particle size, TWI, soil depth, and soil texture (Additional file 10: Fig. S10).

The water inundation areas identified in 2019, 2020, and 2021, were 2651 km<sup>2</sup>, 3711 km<sup>2</sup>, and 1801 km<sup>2</sup>, respectively. The identified water-inundated areas during 2019–2021 were overlaid on the predicted flood hazard maps generated using RF, SVM, and CART (Figs. 4, 5, and 6). It can be observed that the flood risk areas

**Table 3** Flood-affected LULC areas in km<sup>2</sup> (in % of the class area)

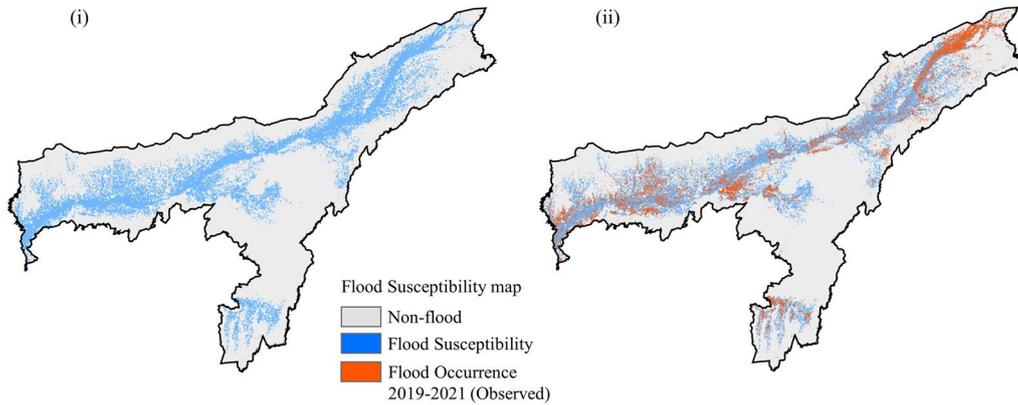
Class	Period	
	2016–2018	2019–2021
Bareland	1245.86 (34.49)	1673.63 (46.33)
Built-up	133.89 (1.31)	143.7 (1.39)
Cropland	3083.58 (11.53)	3545.37 (13.26)
Flooded Vegetation	53.76 (23.05)	74.02 (31.74)
Grass	70.36 (10.79)	77.96 (11.96)
Scrub	456.57 (15.08)	417.76 (13.8)
Trees	84.98 (0.28)	69.48 (0.23)

**Table 4** Modeling accuracy in predicting flood hazard mapping [OA: Overall Accuracy]

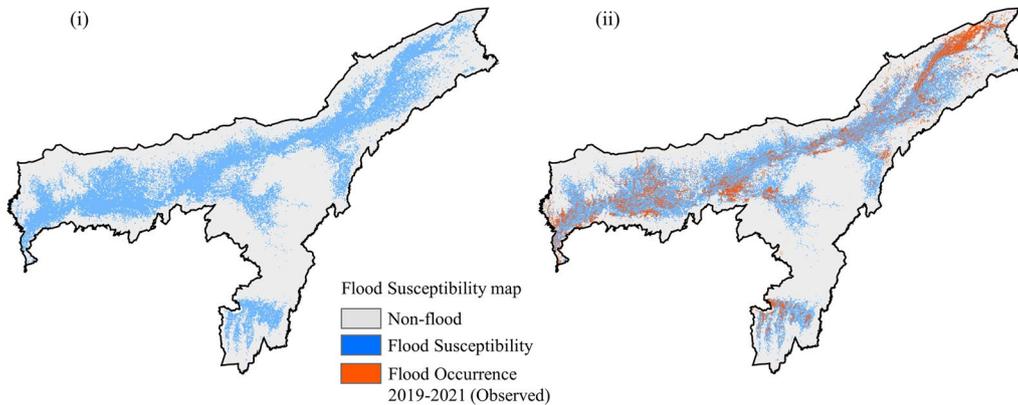
Training data (70% of flood occurrence points during 2016–18)	Testing data (30% of flood occurrence points during 2016–18)	Validation data (flood occurrence points during 2019–21)
Random forest (RF)	OA: 82.91% Kappa: 0.66	OA: 82.36% Kappa: 0.65
Classification and regression tool (CART)	OA: 81.9% Kappa: 0.64	OA: 83.37% Kappa: 0.67
Support vector machine (SVM)	OA: 82.23% Kappa: 0.64	OA: 81.15% Kappa: 0.62

identified by the three ML models are similar. It shows a nearly complete overlap between the predicted flood susceptible map and observed flood occurrence maps of 2019, 2020, and 2021. However, the flood susceptible area

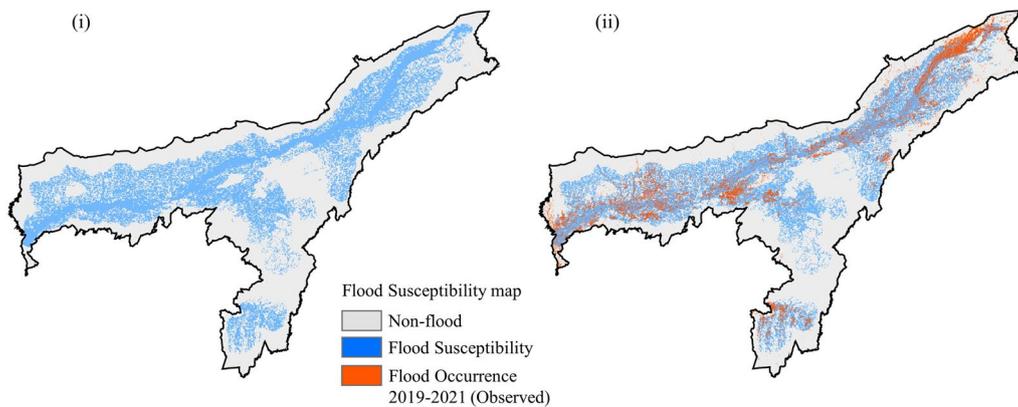
by the ML models overpredicted the actual flood occurrence area. The least flood susceptible area was predicted by the RF model (20,719 km<sup>2</sup>), which was comparatively higher in the SVM (27,510 km<sup>2</sup>) and significantly



**Fig. 4** (i) RF predicted flood susceptibility and (ii) Overlaid observed flood occurrence map of 2019–2021



**Fig. 5** (i) SVM predicted flood susceptibility and (ii) Overlaid observed flood occurrence map of 2019–2021



**Fig. 6** (i) CART predicted flood susceptibility and (ii) Overlaid observed flood occurrence map of 2019–2021

**Table 5** Flood hazard-prone areas predicted by the three ML models

ML models	Area (km <sup>2</sup> )	Area (in %)
RF	20,719	26.41
SVM	27,510	35.06
CART	50,192	63.97

**Table 6** Weight computed using the pair-wise comparison matrix to estimate the soil erosion susceptibility [Consistency Ratio (CR) ~0.06, i.e., <0.1]

Variable	Weight
Flood occurrence	0.20
TCC%	0.15
LULC	0.12
Distance to stream	0.1
Slope inclination	0.09
Flow accumulation	0.07
Soil texture	0.067
Soil particle size	0.066
Soil depth	0.065
River network density	0.032
Precipitation	0.015
TWI	0.013
TRI	0.012

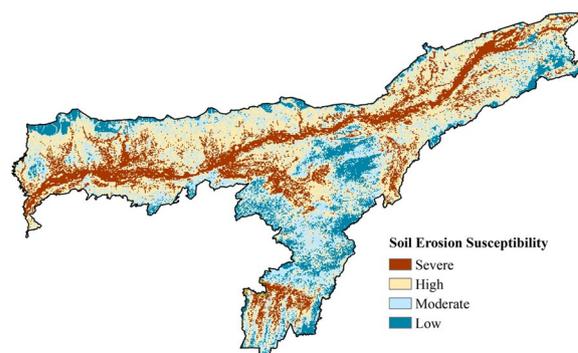
higher in the CART model (50,192 km<sup>2</sup>) (Table 5). The least flood-prone area identified by the RF model showed complete overlap with the predicted areas of the SVM and CART models. This study indicates that at least 26% of the area is flood-prone, mostly along the Brahmaputra River floodplain. The flood hazard map generated by the National Remote Sensing Centre (NRSC) is based on the flood occurrence identified in satellite data from 1998 to 2007 (<https://bhuvan-app1.nrsc.gov.in/thematic/thematic/index.php>), indicating a high resemblance with the current study.

**Soil erosion susceptibility mapping**

The relative weights of the determinant variables for the soil erosion susceptibility mapping are listed in Table 6. The weights were consistent with those of previous studies, with consistency ratios below 0.1. The soil erosion susceptibility map shows the relative susceptibility of different areas to soil erosion, with darker colors representing higher-risk areas (Fig. 7). The maximum importance was assigned to flood occurrence (0.20), TCC% (0.15), LULC (0.12), and Distance to Stream (0.1), as these factors contributed more to

**Table 7** Predicted areas under various soil erosion susceptibility classes

Class	Area in km <sup>2</sup> (in %)
Severe	17,638.07 (20.95)
High	34,646.18 (41.15)
Moderate	21,302.09 (25.31)
Low	10,598.99 (12.59)



**Fig. 7** Soil erosion susceptibility map generated using Multi-criteria Approach (MCE)

soil erosion. In contrast, variables such as Soil Depth (0.065), River Network Density (0.032), precipitation (0.015), TWI (0.013), and TRI (0.012) have a relatively lower contribution to soil erosion. Based on the resulting map, it is possible to identify areas at a higher risk of soil erosion and prioritize measures to prevent soil erosion in those areas. The maximum area (>41% of the total area) showed potential for high soil erosion, followed by moderate soil erosion in >25% of the total area (Table 7). Severe and low soil erosion were predicted in approximately 21% and 13% of the total area, respectively. The resultant map indicates a higher risk of soil erosion in croplands in the Brahmaputra River floodplain due to periodic flooding. In addition, the southern part of Assam (Cachar, Karimgunj, and Hailakandi districts) is vulnerable to severe soil erosion due to recurrent flooding. In comparison, lower soil erosion was predicted in forested regions with higher tree canopy density. The spatial distribution of soil erosion susceptibility maps can be valuable for land-use planning, soil erosion prevention and conservation, sustainable agricultural planning, land degradation studies, and other environmental management activities.

## Discussion

Satellite remote sensing, geospatial technologies, and the GEE platform were efficient and effective in identifying flood-inundated areas and risk assessment. This study used microwave and optical satellite images for flood mapping and validation at a high spatial resolution over the past seven years. The flood-inundated area identified in different years corroborates the areas reported in previous studies (Pandey et al. 2022; Mudi et al. 2022). The ML algorithms were competent in recognizing the factors influencing flooding in the study area. Overall, by comparing the results, it is evident that RF outperformed the other two models (CART and SVM). The superior performance of the RF model can be attributed to its ensemble of hundreds of decision trees. The flood hazard maps of the three ML models predicted similar patterns, with the lowest area predicted by RF and the highest by the CART model. The areas identified as susceptible to flood by the RF model may experience recurrent flooding during the monsoon season, which is pertinent for sustainable landscape management and flood mitigation planning. Most flood occurrences were observed in cropland floodplain areas of the Brahmaputra River, which have a low tree canopy density. The river network density was measured to show the impact of the Brahmaputra River on flood hazards in the region. Flow accumulation, elevation, precipitation, and river distance influence flood hazard mapping. In contrast, the influences of soil terrain ruggedness, slope, and soil characteristics were found to be less important. The flood hazard map of Assam developed by Gupta and Dixit (2022) indicated a similar pattern to that observed in the current study. They employed a pairwise comparison matrix for weight estimation of the MCA and assigned a higher weightage to precipitation, slope, elevation, and distance to the river. In comparison, the present study employed ML models that accurately identified the roles of various factors and their relative importance. Sachdeva and Kumar (2022) assessed flood hazards in the Hojai district, Assam, employing multiple ML models (including SVM and RF) and reported good accuracies. Their study indicated elevation, distance to the river, precipitation, and vegetation density as the major influencing factors. Maiti and Jana (2019) used ML models, including SVM, decision tree, and RF, for flood hazard mapping in the Mahanadi River Basin in India. They reported an overall accuracy of 81.3% for the RF model, similar to the accuracy reported in this study. Singha et al. (2022) also employed multiple ML models for flood hazard mapping in part of Assam state and reported a higher accuracy for the RF and gradient boosting model (GBM) than for other models. Their predicted flood hazard map using the RF model was

similar to the predicted map in this study. Moreover, the flood hazard map developed by the National Remote Sensing Centre (NRSC), Hyderabad, closely resembles that of the present study (Additional file 10: Fig. S10).

The accuracy of soil erosion susceptibility mapping is highly dependent on the selection of predictor variables, and identification of the most influential factors is essential for developing accurate models. Limited research has been conducted to assess soil erosion caused by flooding in Assam. This study estimated soil erosion susceptibility due to flooding, where a higher weighting was assigned to flood-inundated areas, followed by LULC, river network, topography, soil characteristics, and precipitation, as prescribed by previous studies in other similar landscapes (Sinshaw et al. 2021; Nekhay et al. 2009; Sajedi-Hosseini et al. 2018). The soil erosion susceptibility map indicates higher soil erosion along the floodplain of the Brahmaputra River and its tributaries owing to recurrent flooding. Mishra et al. (2022) recently published their findings on soil erosion and deposition at Majuli Island of the Brahmaputra River and reported that the sedimentary sequences developed by the Brahmaputra River are mud-dominated, whereas the Subansiri River has both mud and sand sequences. This is likely due to the differential source-to-sink transport of sediments, as the Brahmaputra River carries the sediment load for a longer distance than the Subansiri River. Their study also found that areas with sand-dominated facies were more vulnerable to erosion than mud-dominated facies. Bordoloi et al. (2020) studied the river bank erosion linking the recurrent flooding in the Subansiri River for the past three decades. Their study indicated significant riverbank erosion and a westward shift of the Subansiri River. Field visits revealed severe bank erosion in the Brahmaputra River and its tributaries in May 2022, which significantly altered the channel morphology by mobilizing channel sediments (Additional file 11: Fig. S11).

Understanding the relationship between predictor variables and soil erosion can help inform effective soil conservation and management strategies, which are crucial for sustainable land-use planning and agricultural practices. Studying predictor variables in soil erosion susceptibility mapping is a valuable tool for addressing soil degradation and promoting sustainable land management practices. Flood prevention, increasing the retention capacity of a catchment, and flood-related damage reductions are possible through sustainable land-use practices and management, improving forest cover, and suitably designing water harvesting (ponds, lakes, dams, etc.) and other artificial structures (culverts, bridges, etc.). The flood inundation mapping approach built into the GEE platform can be deployed for

near-real-time automated flood inundation mapping. Moreover, the adopted approach for flood hazard and soil erosion susceptibility assessment using novel ML techniques can be deployed for periodic assessment. Although the flood susceptibility map was validated with periodic water inundation data, the soil erosion potential couldn't be validated due to the lack of publicly available data for the Brahmaputra River basin. The adopted approach can be tested in other river basins, for which periodic soil erosion data is available in the public domain. Such validation will help in calibrating or tuning the assigned weights in soil erosion estimation.

The study outcome has diverse applications in mitigating flood damage and aid and rescue initiatives in this region. Moreover, the spatial layers developed in the current study are crucial inputs for long-term planning, conservation practices, agricultural and water resource development activities, and framing suitable policies. The study outcomes can also be deployed through mobile applications that can help inhabitants and managers during flooding. There is potential for improvements in flood mapping and hazard assessment using ML techniques to create more effective flood prevention and response measures. The study outcome can be integrated into wider-scale disaster management systems to improve community resilience to future flooding events.

## Conclusion

The publicly available Sentinel-1 SAR with frequent revisits and the GEE platform enabled the development of platforms for automated flood inundation area mapping in near-real-time (6-days intervals) using ML techniques. The maximum water inundation area was recorded in 2020 (~ 3710 km<sup>2</sup>). The developed approach can be operationalized for period monitoring and use in decision-making by overlapping with other important data layers such as transport (roads/rails), hospitals, and flood relief camp locations. The geostatistical analysis highlighted regions prone to soil erosion due to flooding. More than 26% of the area is vulnerable to flood hazards, and 41% shows the potential for high soil erosion. Using Artificial Intelligence (AI) technologies, this study predicts flood-prone areas in Assam to benefit resource managers and planners. The highest prediction accuracy was observed for the RF model (82.91%). The current study is also important for farmers, governments, and non-government entities related to flood prevention, infrastructure, water resource development, landslide studies, soil erosion and land degradation studies, climate resilience, sustainable and regenerative agricultural planning, crop and nutrition security, and crop and disaster insurance.

This study provides data suitable for assessing the impact of climate change on crop production, food and nutrition security, socioeconomic conditions, ecology, and the environment. This study contributes to advancing flood and soil erosion management practices and provides a valuable tool for decision-makers. Further studies are required to assess the deposition processes of eroded soil from higher altitudes transported by surface runoff and river flow.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40677-024-00275-8>.

**Additional file 1: Fig. S1.** (i) Soil depth, (ii) Particle size, and (iii) Soil Texture map.

**Additional file 2: Fig. S2.** Shuttle Radar Topographic Mission (SRTM) digital elevation model (DEM) data derived (i) elevation, (ii) Slope and (iii) Aspect map.

**Additional file 3: Fig. S3.** SRTM DEM data derived (i) Terrain Ruggedness Index (TRI) and (ii) Topographic Wetness Index (TWI) map.

**Additional file 4: Fig. S4.** SRTM DEM data derived (i) Drainage network, (ii) Drainage Density and (iii) Distance to Drainage.

**Additional file 5: Fig. S5.** Flow accumulation (at 1 km grid-scale) map.

**Additional file 6: Fig. S6.** Mean Annual Precipitation (July–September) (i) 2016–2018 and (ii) 2019–2021 (at sub-basin scale).

**Additional file 7: Fig. S7.** (i) Land use land cover (LULC) and (ii) Tree Canopy Cover (TCC) percentage map of 2020.

**Additional file 8: Fig. S8.** Validation of Sentinel-1 SAR data-derived water inundation map of 2017 using the Sentinel-2 optical data derived water index map.

**Additional file 9: Fig. S9.** Water Inundation map of May 2022 with a few field photographs (Photo Credit: Lakhyajit Baruah and Souvik Maity).

**Additional file 10: Fig. S10.** Variable importance plot derived using the RF algorithm. Flood hazard map of Assam (Source: <https://bhuvan-app1.nrsc.gov.in/thematic/thematic/index.php#>)

**Additional file 11: Fig. S11.** Field photos showing the soil river bank erosion in Brahmaputra River (Photo Credit: Suman Chetri).

## Acknowledgements

Not applicable.

## Author contributions

Conceptualisation: JP, SB, VV, SM, and PD; Data curation: JP and PD; Formal analysis: JP, VV, and PD; Funding acquisition: N/A; Investigation: JP, SB, VV, SM, and PD; Methodology: JP, SB, VV, SM, and PD; Project administration: JP and PD; Resources: JP and PD; Software: JP and PD; Supervision: VV and PD; Validation: JP, SB, VV, SM, and PD; Visualisation: JP and PD; Roles/Writing—original draft: JP and PD; Writing—review & editing: SB, VV and SM.

## Funding

Not applicable.

## Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

## Code availability

KN-ICGD-GEDI-0421.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Centre for Ocean, River, Atmosphere and Land Sciences, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India. <sup>2</sup>Department of Civil Engineering, University of Nottingham, Nottingham NG7 2RD, UK. <sup>3</sup>Charles University, Albertov 6, 128 00 Prague 2, Czech Republic. <sup>4</sup>BeZero Carbon Ltd, Discovery House, Banner St, London EC1Y8QE, UK. <sup>5</sup>School of Forest Resources, University of Maine, Orono 04469, USA.

Received: 19 June 2023 Accepted: 5 March 2024

Published online: 26 March 2024

## References

- Acquaotta F, Faccini F, Fratianni S, Paliaga G, Sacchini A, Vilímek V (2019) Increased flash flooding in Genoa Metropolitan area: A combination of climate changes and soil consumption? *Meteorol Atmos Phys* 131:1099–1110. <https://doi.org/10.1007/s00703-018-0623-4>
- Adam E, Mutanga O, Abdel-Rahman EM, Ismail R (2014) Estimating standing biomass in papyrus (*Cyperus papyrus* L.) swamp: exploratory of in situ hyperspectral indices and random forest regression. *Int J Remote Sens* 35:693–714. <https://doi.org/10.1080/01431161.2013.870676>
- Altaf S, Meraj G, Romshoo SA (2014) Morphometry and land cover based multicriteria analysis for assessing the soil erosion susceptibility of the western Himalayan watershed. *Environ Monit Assess* 186:8391–8412. <https://doi.org/10.1007/s10661-014-4012-2>
- Arabameri A, Tiefenbacher JP, Blaschke T, Pradhan B, Tien Bui D (2020) Morphometric analysis for soil erosion susceptibility mapping using novel GIS-based ensemble model. *Remote Sens* 12(5):874. <https://doi.org/10.3390/rs12050874>
- Begam S, Sen D (2019) Mapping of moraine dammed glacial lakes and assessment of their areal changes in the central and eastern Himalayas using satellite data. *J Mt Sci* 16(1):77–94. <https://doi.org/10.1007/s11629-018-5023-1>
- Bhattacharya RK, Chatterjee ND, Das K (2020) Sub-basin prioritisation for assessment of soil erosion susceptibility in Kangsabati, a plateau basin: a comparison between MCDM and SWAT models. *Sci Total Environ* 734:139474. <https://doi.org/10.1016/j.scitotenv.2020.139474>
- Borah SB, Sivasankar T, Ramya MNS, Raju PLN (2018) Flood inundation mapping and monitoring in Kaziranga National Park, Assam using Sentinel-1 SAR data. *Environ Monit Assess* 190:1–11. <https://doi.org/10.1007/s10661-018-6893-y>
- Borah G, Saikia N, Das S, Sharma S (2023) Wellbeing in the aftermath of floods: findings from a qualitative study in Bongaigaon District of Assam, India. *Wellbeing Space Soc* 4:100147. <https://doi.org/10.1016/j.wss.2023.100147>
- Bordoloi K, Nikam BR, Srivastav SK, Sahariah D (2020) Assessment of riverbank erosion and erosion probability using geospatial approach: a case study of the Subansiri River, Assam, India. *Appl Geomat* 12:265–280. <https://doi.org/10.1007/s12518-019-00296-1>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Chaliha S, Sengupta A, Sharma N, Ravindranath NH (2012) Climate variability and farmer's vulnerability in a flood-prone district of Assam. *Int J Clim Change Strateg Manag* 4(2):179–200
- Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ Model Softw* 95:229–245. <https://doi.org/10.1016/j.envsoft.2017.06.012>
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Dar RA, Mir SA, Romshoo SA (2019) Influence of geomorphic and anthropogenic activities on channel morphology of River Jhelum in Kashmir Valley, NW Himalayas. *Quat Int* 507:333–341. <https://doi.org/10.1016/j.quaint.2018.12.014>
- Das D (2016) Changing climate and its impacts on Assam, Northeast India. *Bandung* 2(1):1–13. <https://doi.org/10.1186/s40728-015-0028-4>
- Das P, Behera MD, Patidar N, Sahoo B, Tripathi P, Behera PR, Krishnamurthy YVN (2018) Impact of LULC change on the runoff, base flow and evapo-transpiration dynamics in eastern Indian river basins during 1985–2005 using variable infiltration capacity approach. *J Earth Syst Sci* 127:1–19. <https://doi.org/10.1007/s12040-018-0921-8>
- Das P, Mudi S, Behera MD, Barik SK, Mishra DR, Roy PS (2021a) Automated mapping for long-term analysis of shifting cultivation in Northeast India. *Remote Sens* 13(6):1066. <https://doi.org/10.3390/rs13061066>
- Das P, Pandey V, Dutta D (2021a) Land surface water resource monitoring and climate change. In: Mapping, monitoring, and modeling land and water resources. CRC Press, pp 311–326
- Díaz S, Settele J, Brondizio ES, Ngo HT, Guèze M, Agard J, Arneth A, Balvanera P, Brauman K, Butchart SH (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
- 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. *Sci Data* 2:150066. <https://doi.org/10.1038/sdata.2015.66>
- Gangopadhyay PK, Sharma BR, Pavelic P (2018) Co-solving groundwater depletion and seasonal flooding through an innovative managed aquifer recharge approach: converting pilot to a regional solution in the Ram Ganga Sub-basin. In: Clean and sustainable groundwater in India. Springer, pp 173–189
- Gupta L, Dixit J (2022) A GIS-based flood risk mapping of Assam, India, using the MCDA-AHP approach at the regional and administrative level. *Geocarto Int*. <https://doi.org/10.1080/10106049.2022.2060329>
- Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D, Stehman SV, Goetz SJ, Loveland TR (2013) High-resolution global maps of 21st-century forest cover change. *Science* 342:850–853. <https://doi.org/10.1126/science.1244693>
- Harrison S, Kargel JS, Huggel C, Reynolds J, Shugar DH, Betts RA, Vilímek V (2018) Climate change and the global pattern of moraine-dammed glacial lake outburst floods. *Cryosphere* 12(4):1195–1209. <https://doi.org/10.5194/tc-12-1195-2018>
- Johnson MA, Brown CH, Wells SJ (2002) Using classification and regression trees (CART) to support worker decision making. *Soc Work Res* 26(1):19–29
- Khatun M, Rahaman SM, Garai S, Das P, Tiwari S (2022) Assessing river bank erosion in the Ganges using remote sensing and GIS. In: Geospatial technology for environmental hazards: modeling and management in Asian countries, pp 499–512. [https://doi.org/10.1007/978-3-030-75197-5\\_22](https://doi.org/10.1007/978-3-030-75197-5_22)
- Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Dou J, Ly H-B, Gróf G, Ho HL, Hong H, Chapi K, Prakash I (2019) A comparative assessment of flood susceptibility modeling using multicriteria decision-making analysis and machine learning methods. *J Hydrol* 573:311–323. <https://doi.org/10.1016/j.jhydrol.2019.03.073>
- Konapala G, Kumar SV, Ahmad SK (2021) Exploring Sentinel-1 and Sentinel-2 diversity for flood inundation mapping using deep learning. *ISPRS J Photogramm Remote Sens* 180:163–173. <https://doi.org/10.1016/j.isprsjprs.2021.08.016>
- Kumar N, Kumar M, Sherring A, Suryavanshi S, Ahmad A, Lal D (2020) Applicability of HEC-RAS 2D and GFMS for flood extent mapping: a case study of Sangam area, Prayagraj, India. *Model Earth Syst Environ* 6(1):397–405. <https://doi.org/10.1007/s40808-019-00687-8>
- Luo T, Maddocks A, Iceland C, Ward P, Winsemius H (2015) World's 15 countries with the most people exposed to river floods. World Resources Institute. <https://www.wri.org/insights/worlds-15-countries-most-people-exposed-river-floods>
- Maiti S, Jana S (2019) Flood susceptibility mapping using machine learning techniques: a case study in the Mahanadi river basin, India. *J Environ Manag* 246:9–20
- Milliman JD, Farnsworth KL (2011) River discharge to the coastal ocean: a global synthesis. Cambridge University Press, Cambridge, UK, pp 1–383. <https://doi.org/10.1017/CBO9780511781247>

- Mishra S, Dutta D, Badekar A, Misra AK (2022) Sedimentation and erosion at Majuli Island, Assam, India: analysis of paleo-depositional environment. *Int J River Basin Manag*. <https://doi.org/10.1080/15715124.2022.2151606>
- Mosavi A, Golshan M, Janizadeh S, Choubin B, Melesse AM, Dineva AA (2020) Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins. *Geocarto Int*. <https://doi.org/10.1080/10106049.2020.1829101>
- Mosavi A, Golshan M, Janizadeh S, Choubin B, Melesse AM, Dineva AA (2022) Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins. *Geocarto Int* 37:2541–2560. <https://doi.org/10.1080/10106049.2020.1829101>
- Mudi S, Prakash J, Das P (2022) Flood hazard mapping in Assam using Sentinel-1 SAR data. In: *Geospatial technology for environmental hazards*. Springer, pp 459–473. [https://doi.org/10.1007/978-3-030-75197-5\\_20](https://doi.org/10.1007/978-3-030-75197-5_20)
- Nekhay O, Arriaza M, Boerboom L (2009) Evaluation of soil erosion risk using analytical network process and GIS: a case study from Spanish mountain olive plantations. *J Environ Manag* 90(10):3091–3104. <https://doi.org/10.1016/j.jenvman.2009.04.022>
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9:62–66
- Pandey AC, Kaushik K, Parida BR (2022) Google Earth Engine for large-scale flood mapping using SAR data and impact assessment on agriculture and population of Ganga-Brahmaputra basin. *Sustainability* 14(7):4210. <https://doi.org/10.3390/su14074210>
- Patankar A (2019) Impacts of natural disasters on households and small businesses in India. *Asian Development Bank Economics Working Paper Series*. <https://doi.org/10.2139/ssrn.3590902>
- Patidar N, Das P, Tripathi P, Behera MD (2022) Covariation between land use change and hydrological balance in river basin scale. In: *Geospatial technologies for land and water resources management*, pp 279–294
- Pradhan RK, Srivastava PK, Maurya S, Kumar Singh S, Patel DP (2020) Integrated framework for soil and water conservation in Kosi River Basin. *Geocarto Int* 35:391–410. <https://doi.org/10.1080/10106049.2018.1520921>
- Prakash AJ, Kumar S, Behera MD, Das P, Kumar A, Srivastava PK (2023) Impact of extreme weather events on cropland inundation over Indian subcontinent. *Environ Monit Assess* 195(1):50. <https://doi.org/10.1007/s10661-022-10553-3>
- Prashanth M, Kumar A, Dhar S, Verma O, Rai SK, Kouser B (2023) Land use/land cover change and its implication on soil erosion in an ecologically sensitive Himachal Himalayan watershed, Northern India. *Front Glob Change* 6:1124677. <https://doi.org/10.3389/fgc.2023.1124677>
- Rather MA, Satish Kumar J, Farooq M, Rashid H (2017) Assessing the influence of watershed characteristics on soil erosion susceptibility of Jhelum basin in Kashmir Himalayas. *Arab J Geosci* 10(3):59. <https://doi.org/10.1007/s12517-017-2847-x>
- Roy N, Pandey BW, Rani U (2020) Protecting the vanishing geo-cultural heritage of India: case study of Majuli Island in Assam. *Int J Geoh Heritage Parks* 8(1):18–30. <https://doi.org/10.1016/j.ijgeop.2020.03.003>
- Sachdeva S, Kumar B (2022) Flood susceptibility mapping using extremely randomised trees for Assam 2020 floods. *Eco Inform* 67:101498. <https://doi.org/10.1016/j.ecoinf.2021.101498>
- Saikia L (2022) Flood hazard of the Brahmaputra River in Assam: current mitigation approaches, challenges and sustainable solution options. In: *Riverine systems: understanding the hydrological, hydrosocial and hydro-heritage dynamics*. Springer, Cham, pp 99–111
- Sajedi-Hosseini F, Choubin B, Solaimani K, Cerdà A, Kaviani A (2018) Spatial prediction of soil erosion susceptibility using a fuzzy analytical network process: application of the fuzzy decision making trial and evaluation laboratory approach. *Land Degrad Dev* 29(9):3092–3103. <https://doi.org/10.1002/ldr.3058>
- Sharma SV, Roy PS, Chakravarthi V, Srinivasa Rao G (2018) Flood risk assessment using multicriteria analysis: a case study from Kopili River Basin, Assam, India. *Geomat Nat Haz Risk* 9:79–93. <https://doi.org/10.1080/19475705.2017.1408705>
- Singha C, Swain KC, Meliho M, Abdo HG, Almohamad H, Al-Mutiry M (2022) Spatial analysis of flood hazard zoning map using novel hybrid machine learning technique in Assam, India. *Remote Sens* 14(24):6229. <https://doi.org/10.3390/rs14246229>
- Sinshaw BG, Belete AM, Mekonen BM, Wubetu TG, Anley TL, Alamneh WD, Birhanu MM (2021) Watershed-based soil erosion and sediment yield modeling in the Rib watershed of the Upper Blue Nile Basin, Ethiopia. *Energy Nexus* 3:100023. <https://doi.org/10.1016/j.nexus.2021.100023>
- Smolíková J, Blahut J, Vilímek V (2016) Analysis of rainfall preceding debris flows on the Smědávská hora Mt., Jizerské hory Mts, Czech Republic. *Landslides* 13:683–696. <https://doi.org/10.1007/s10346-015-0601-6>
- Towfiqul Islam ARM, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, Kuriqi A, Linh NTT (2021) Flood susceptibility modeling using advanced ensemble machine learning models. *Geosci Front* 12:101075. <https://doi.org/10.1016/j.gsf.2020.09.006>
- Venkataramanan V, Packman AI, Peters DR, Lopez D, McCuskey DJ, McDonald RI, Miller WM, Young SL (2019) A systematic review of the human health and social well-being outcomes of green infrastructure for stormwater and flood management. *J Environ Manage* 246:868–880. <https://doi.org/10.1016/j.jenvman.2019.05.028>
- Weiskopf SR, Rubenstein MA, Crozier LG, Gaichas S, Griffis R, Halofsky JE, Hyde KJW, Morelli TL, Morissette JT, Muñoz RC, Pershing AJ, Peterson DL, Poudel R, Staudinger MD, Sutton-Grier AE, Thompson L, Vose J, Weltzin JF, Whyte KP (2020) Climate change effects on biodiversity, ecosystems, ecosystem services, and natural resource management in the United States. *Sci Total Environ* 733:137782. <https://doi.org/10.1016/j.scitotenv.2020.137782>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.