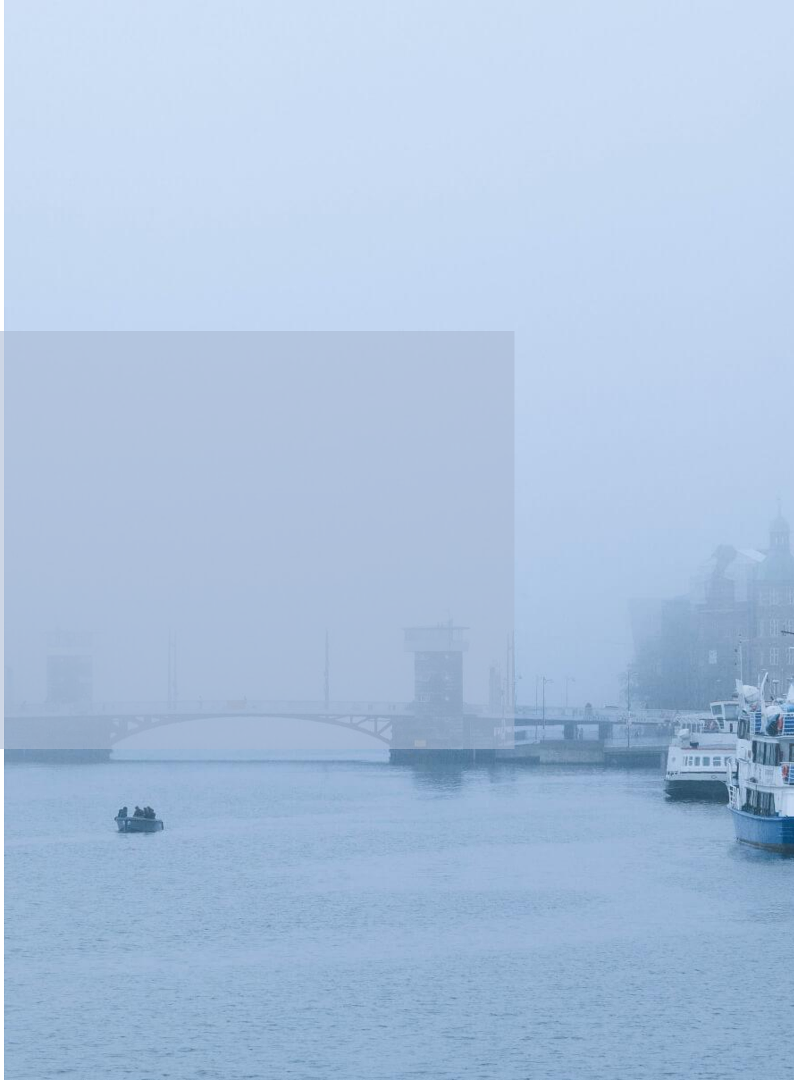
FLOOD SUSCEPTIBILITY MAPPING

TEAM MEMBERS

SAI VENKATA SOHITH GUTTA (IMT2022042)

MARAMREDDY SIDDHARTH REDDY (IMT2022031)

SHREYAS S (IMT2022078)



Background

Natural Hazards Research 3 (2023) 420–436

Contents lists available at ScienceDirect

Natural Hazards Research

journal homepage: www.keaipublishing.com/en/journals/natural-hazards-research

GIS-based machine learning algorithm for flood susceptibility analysis in the Pagla river basin, Eastern India

Nur Islam Saikh, Prolay Mondal *

Department of Geography, Raiganj University, Raiganj, West Bengal, 733134, India

ARTICLE INFO

Keywords:
Flood susceptibility
Pagla river
Machine learning algorithm
GIS
Receiver Operating Characteristic

ABSTRACT

The unique characteristics of drainage conditions in the Pagla river basin cause flooding and harm the socio-economic environment. The main purpose of this study is to investigate the comparative utility of six machine learning algorithms to improve flood susceptibility and ensemble techniques' capability to elucidate the underlying patterns of floods and make a more accurate prediction of flood susceptibilities in the Pagla river basin. In the present scenario, the frequency of flood conditions in this study area becomes high with heavy and sudden rainfall, so it is essential to study flood mitigation and measure. At first, a spatial flood database was built with 200 flood locations and sixteen flood influencing factors, and its process with the help of the Geographic Information System (GIS) environment and build up different models applying the machine learning techniques. It has found different flood susceptibility zone using machine learning-based Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), Reduced Error Pruning Tree (REPTree), Logistic Regression (LR), and Bagging helping GIS environment and the model validation using the Receiver Operating Characteristic (ROC) Curve. Afterward, ensemble all the models to gate comparative accuracy of the flood zone. The calculated areas are under the very high flood susceptibility zone 8.69%, 14.92%, 14.17%, 12.98%, 14.65%, 13.24% and 13.41% for ANN, SVM, RF, REPTree, LR and Bagging, respectively. Finally, ROC curve, the Standard Error (SE), and the Confidence Interval (CI) at 95 per cent were used to assess and compare the performance of the models. The obtained results indicate that all models are highly accepted Area Under Curve (AUC) of ROC between 0.889 (LR) to 0.926 (Ensemble). From the estimation of the accuracy of the applied methods using ROC, it is found that the Ensemble model has the higher capability compared to the other applied models in projecting flood susceptibility in the study area. It has the highest area under the ROC curve the AUC values are 0.918 and 0.926, the SE (0.023, 034), and the narrowest CI (95 per cent) (0.873–0.962, 0.859–0.993) whereas highest area under Bagging (the ROC) curve (AUC) value (0.914, 0.919), for both the training and validation datasets. After ensembling, the result shows that the result is a highly flood susceptible area located at the lower part of the study area. In this area, the very high flood susceptibility zone values lie between 4.46 and 6.00 in the ensemble result. The areas comprise the low height and belong to Murarai I, Murarai II, Suti I and Suti II C.D. block of West Bengal. The current study will help the policymakers and the researcher determine the flood conditioning problems for prospects.

sustainability

MDPI

Article

Application of GIS and Machine Learning to Predict Flood Areas in Nigeria

Eseosa Halima Ighile *, Hiroaki Shirakawa and Hiroki Tanikawa

School of Environmental Studies, Nagoya University, Nagoya 464-8601, Japan; sirakawa@urban.env.nagoya-u.ac.jp (H.S.); tanikawa@nagoya-u.jp (H.T.)
* Correspondence: ighileeseosa1@gmail.com

check for updates

Citation: Ighile, E.H.; Shirakawa, H.; Tanikawa, H. Application of GIS and Machine Learning to Predict Flood Areas in Nigeria. *Sustainability* **2022**, *14*, 5039. https://doi.org/10.3390/

Abstract: Floods are one of the most devastating forces in nature. Several approaches for identifying flood-prone locations have been developed to reduce the overall harmful impacts on humans and the environment. However, due to the increased frequency of flooding and related disasters, coupled with the continuous changes in natural and social-economic conditions, it has become vital to predict areas with the highest probability of flooding to ensure effective measures to mitigate impending disasters. This study predicted the flood susceptible areas in Nigeria based on historical flood records from 1985~2020 and various conditioning factors. To evaluate the link between flood incidence and the fifteen (15) explanatory variables, which include climatic, topographic, land use and proximity information, the artificial neural network (ANN) and logistic regression (LR) models were trained and tested to develop a flood susceptibility map. The receiver operating characteristic curve (ROC) and area under the curve (AUC) were used to evaluate both model accuracies. The results show that both techniques can model and predict flood-prone areas. However, the ANN model produced a higher performance and prediction rate than the LR model, 76.4% and 62.5%, respectively. In addition, both models highlighted that those areas with the highest susceptibility to flood are the low-lying regions in the southern extremities and around water areas. From the study, we can establish that machine learning techniques can effectively map and predict flood-prone areas and serve as a tool for developing flood mitigation policies and plans.

Keywords: machine learning; artificial neural networks; logistic regression; flood prediction; nigeria

This study utilizes insights from two research papers published in different journals. One paper analyzes flood susceptibility for the Pagla River Basin in West Bengal, while the other assesses flood risk across Nigeria. Both studies were valuable for generating new features from existing data and enhancing the use of machine learning models for flood risk analysis

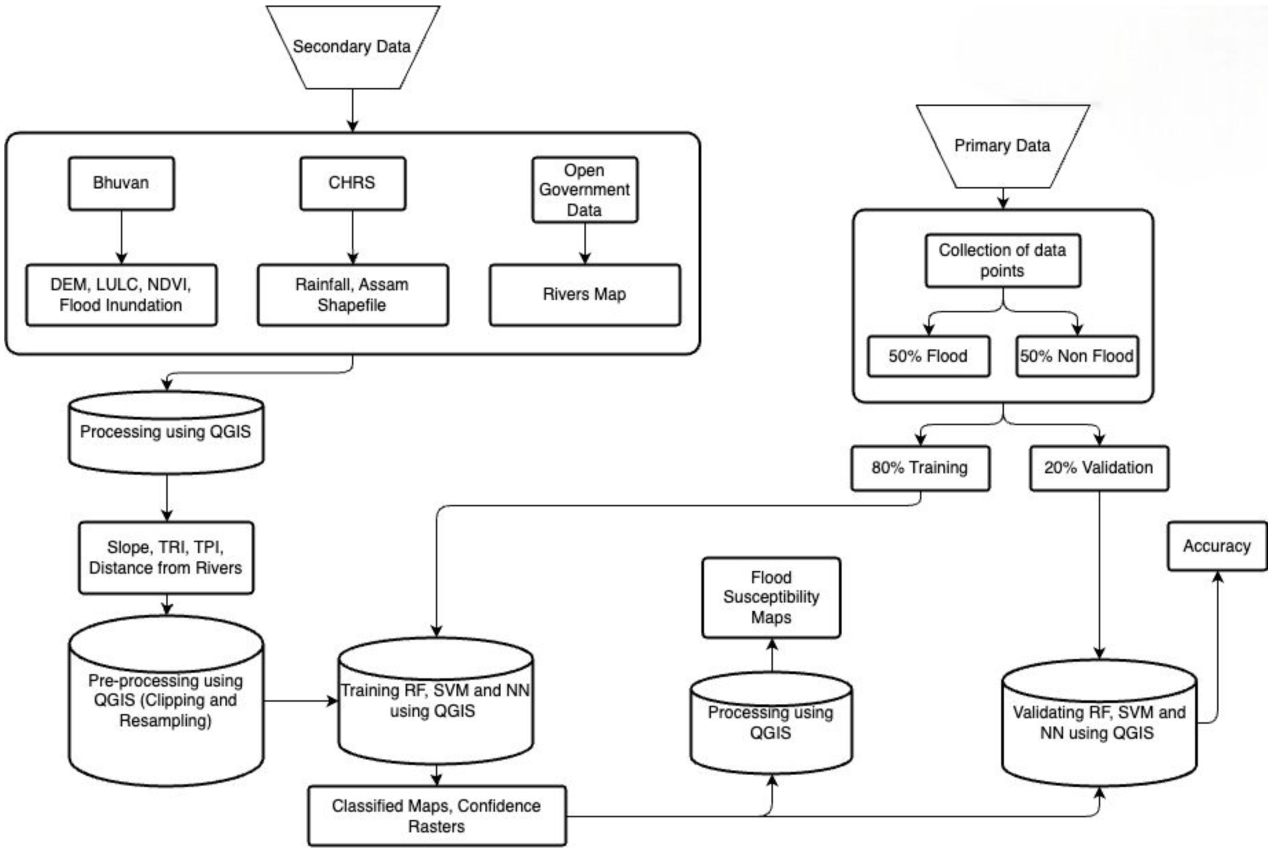
Problem Definition:

Assam, situated in northeastern India, faces significant flood risks due to the Brahmaputra River's annual monsoon-induced flooding. Historical events, such as the devastating floods of 1954, 1962, and more recent floods in 2004, 2020 and 2024, have caused severe damage, impacting millions and disrupting livelihoods.

Objective:

The objective of this study is to optimize the flood susceptibility model by maximizing predictive accuracy and minimizing error in identifying flood-prone areas. This helps ensure reliable risk assessment and enhances flood management strategies for the state of Assam.

Workflow


Secondary Data
Bhuvan
CHRS
Open Government Data
DEM, LULC, NDVI, Flood Inundation
Rainfall, Assam Shapefile
Rivers Map
Processing using QGIS
Slope, TRI, TPI, Distance from Rivers
Pre-processing using QGIS (Clipping and Resampling)
Training RF, SVM and NN using QGIS
Classified Maps, Confidence Rasters
Primary Data
Collection of data points
50% Flood
50% Non Flood
80% Training
20% Validation
Flood Susceptibility Maps
Processing using QGIS
Validating RF, SVM and NN using QGIS
Accuracy

Digital Elevation Model (DEM)

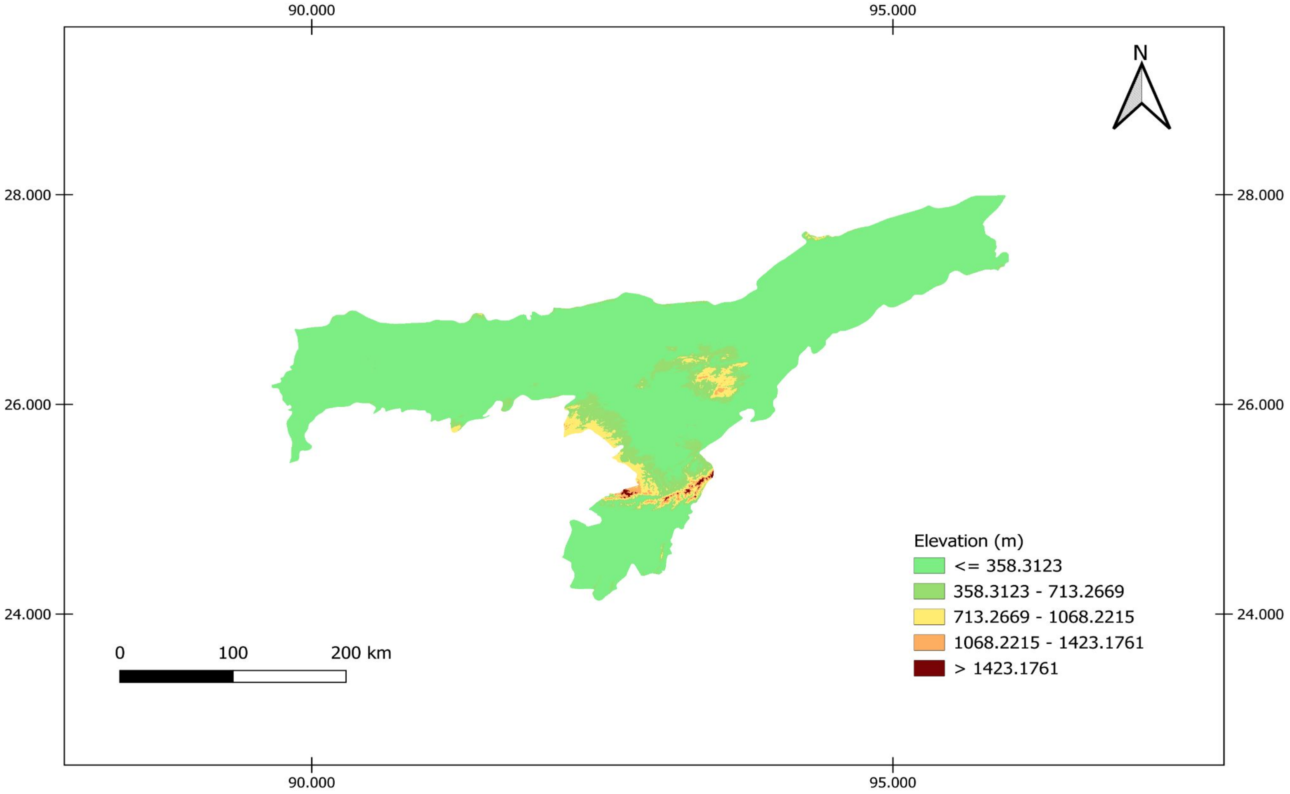
DEM is a digital representation of the Earth's surface topography, typically represented as a grid or raster where each cell has an elevation value.

Importance of DEM in Flood Analysis:

High elevation: Increases surface runoff, reducing flood risk.

Flat areas: More prone to flooding due to high water discharge.

Digital Elevation Model (DEM)



Slope

- Slope represents the rate of elevation change between neighboring cells in a DEM.
- It is expressed in degrees and is calculated from DEM

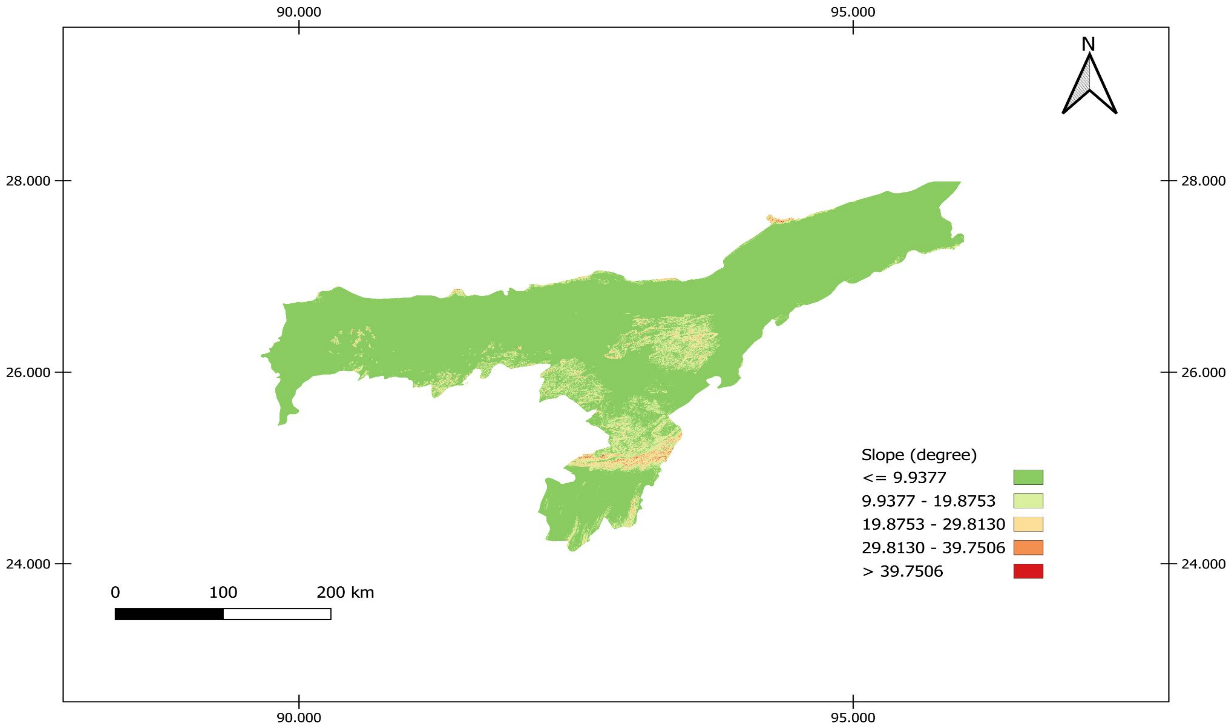
Importance of Slope in Flood Analysis:

Steep slopes: Facilitate rapid water runoff, which minimizes water accumulation and significantly reduces the risk of flooding.

Gentle slopes or Flat areas: Slows down water movement, increasing the likelihood of water pooling and making these areas more prone to flooding.

.

Slope


90.000
95.000
N
28.000
26.000
24.000
Slope (degree)
<= 9.9377
9.9377 - 19.8753
19.8753 - 29.8130
29.8130 - 39.7506
> 39.7506
0
100
200 km
90.000
95.000

Topographic Roughness Index (TRI)

- The Topographic Roughness Index (TRI) is a measure of the variability in elevation within a given area.

$$TRI = \sqrt{\sum_{i=1}^{8}(x_i - E)^2}$$

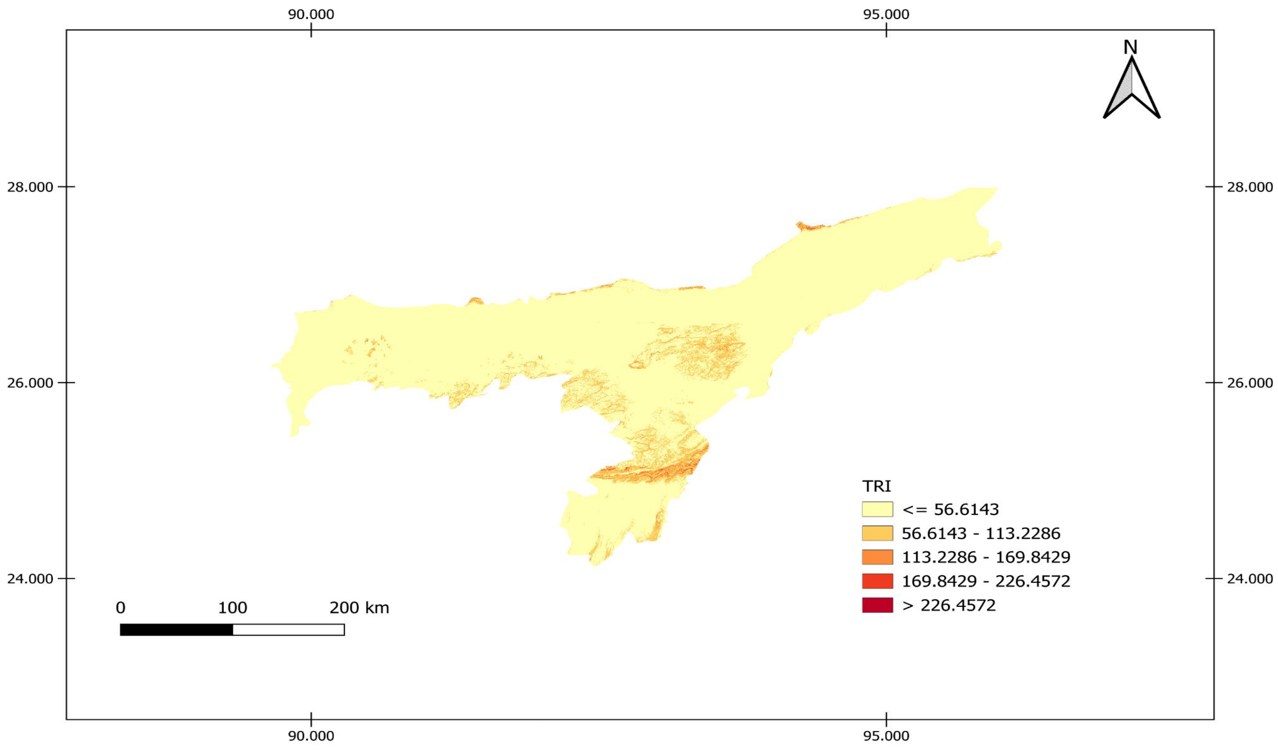
where x_i refers to each of the eight neighbors of the center cell E.

Importance of TRI in Flood Analysis:

High TRI Values: Indicate more rugged and uneven terrain, which can result in faster water drainage and reduced flood risk.

Low TRI Values: Suggest flatter areas with less variation in elevation, which can lead to slower water movement and an increased risk of water pooling and flooding.

Topographic Roughness Index (TRI)



Topographic Position Index (TPI)

- The Topographic Position Index (TPI) measures the relative position of a cell within its landscape, indicating whether a location is on a peak, ridge, valley, or flat area.

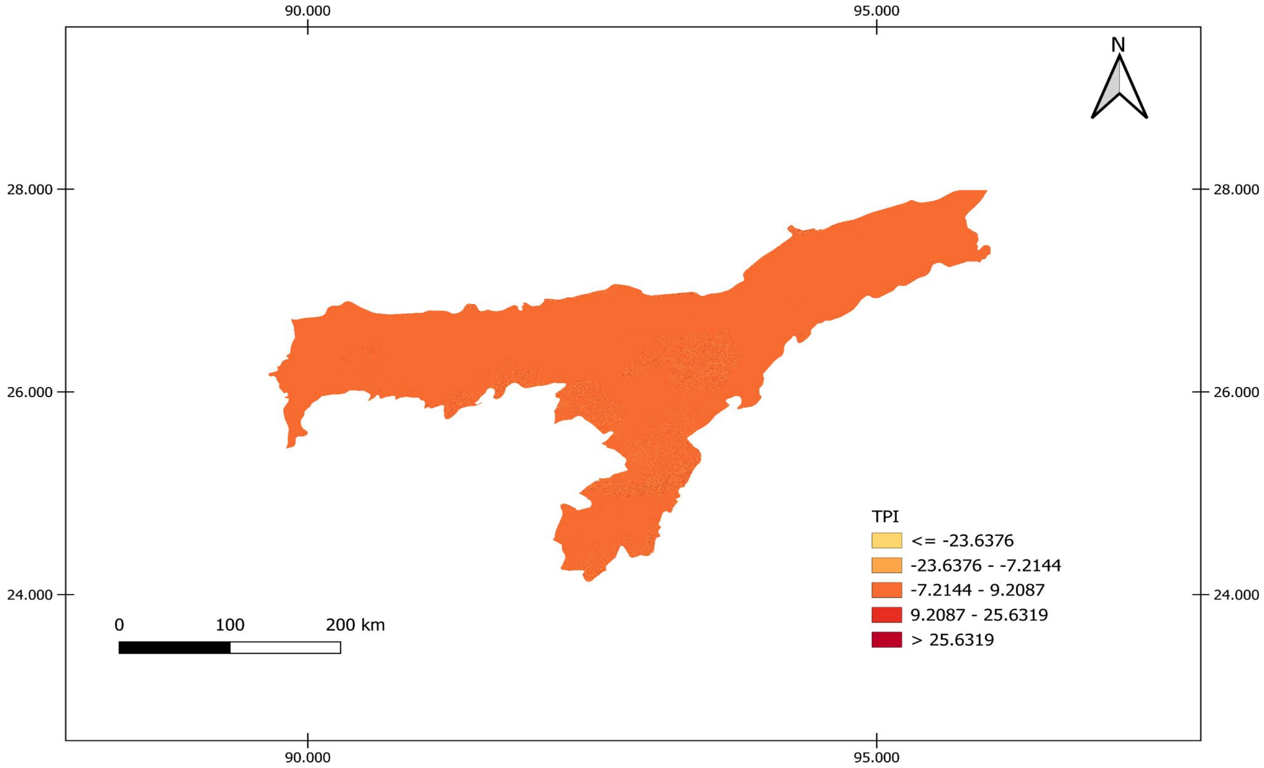
Importance of TPI in Flood Analysis:

Positive TPI (Peaks/Ridges): Indicates higher elevation areas that promote faster water runoff, reducing flood risk.

Negative TPI (Valleys/Depressions): Represents lower elevation areas prone to water pooling and increased flood risk.

Zero TPI (Flat Areas): Shows areas with similar elevation to surrounding cells, which may lead to slow water movement and potential flooding.

Topographic Position Index (TPI)



Distance From River

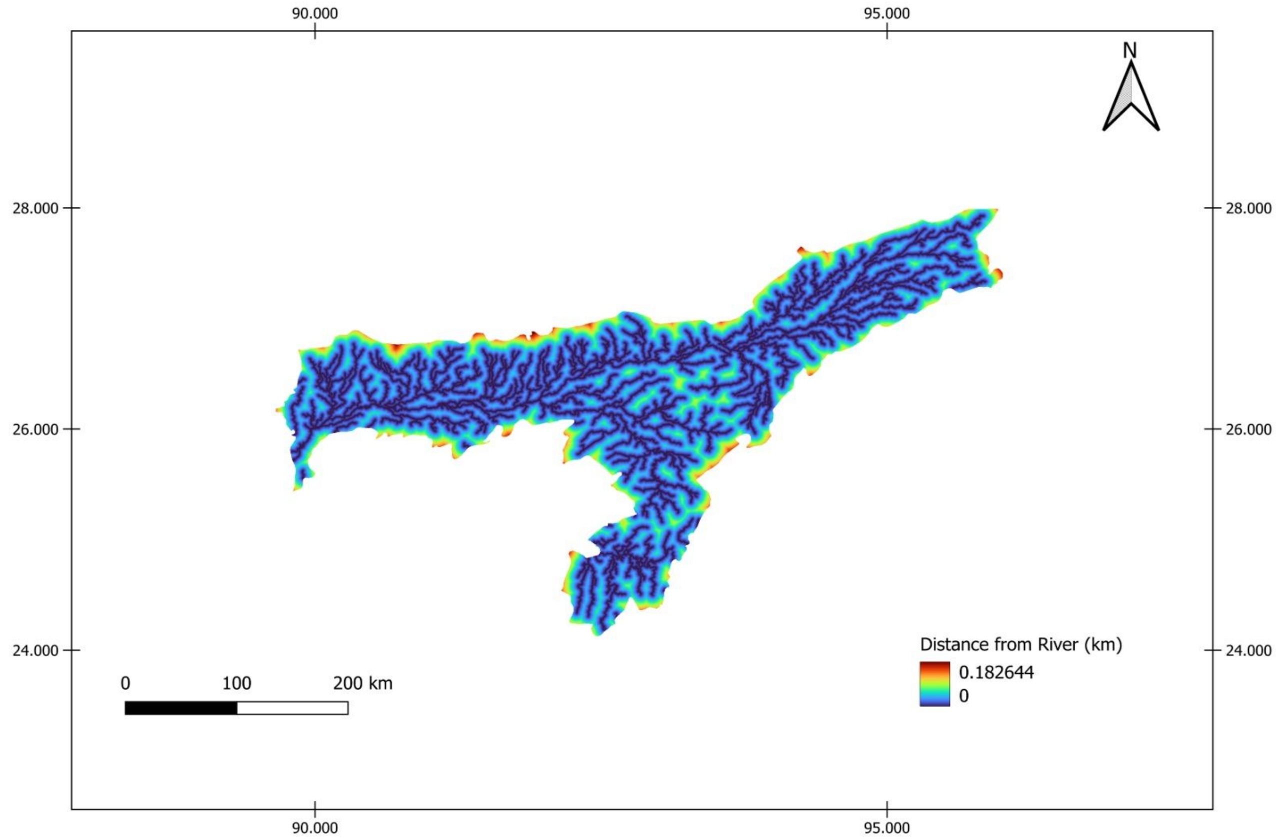
- The distance from a river represents the proximity of a location to the nearest watercourse, calculated as a straight-line (Euclidean) distance.

Importance of "Distance From River" in Flood Analysis:

Closer to River: Areas near rivers are more vulnerable to flooding due to overflow during high rainfall or river surges.

Farther from River: Locations at greater distances generally have a lower risk of direct riverine flooding.

Distance From River



Rainfall

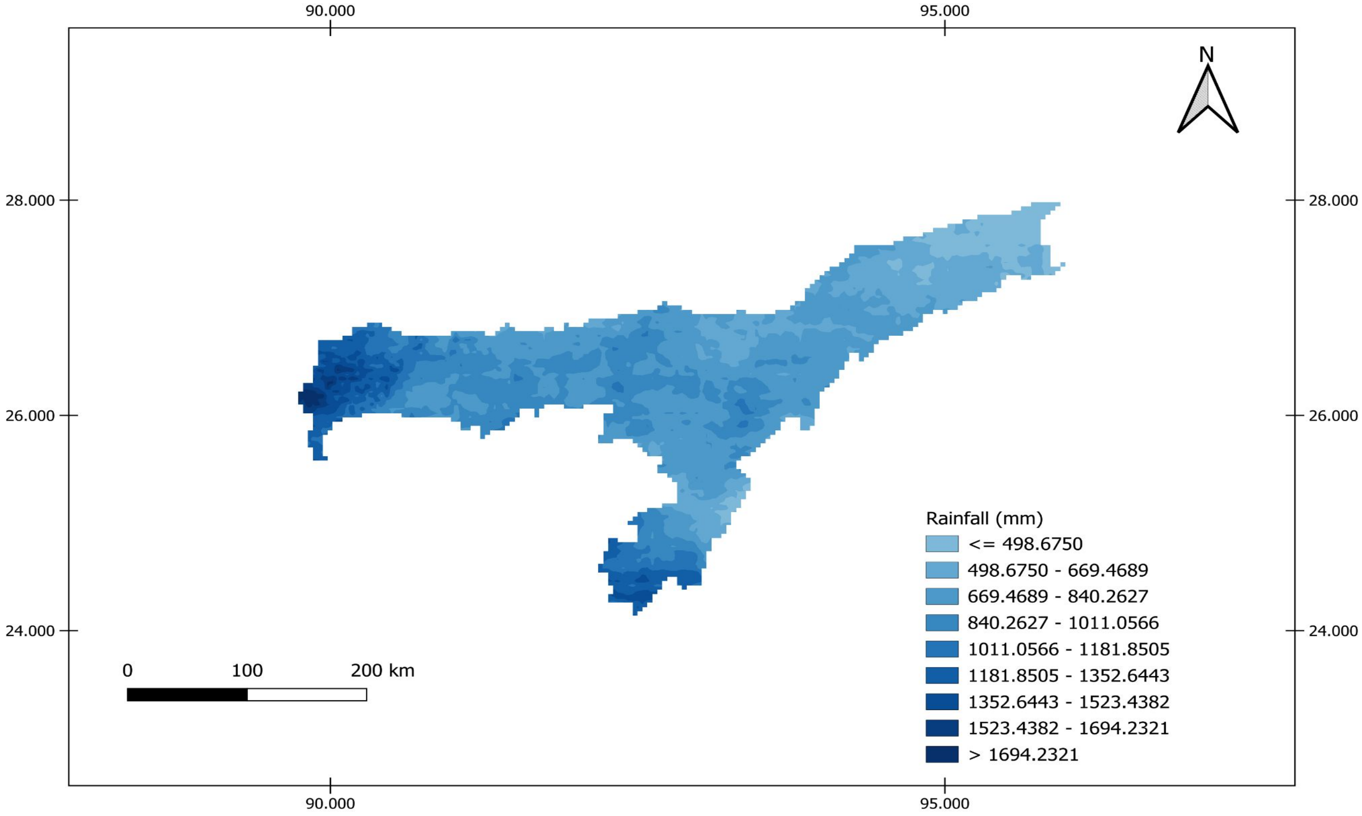
- Rainfall refers to the amount of precipitation that falls over a specific area within a given time period.

Importance of Rainfall in Flood Analysis:

Heavy Rainfall: Higher yearly average rainfall increases the potential for flooding, as more precipitation leads to greater surface runoff and higher river levels.

Low Rainfall: Lower yearly average rainfall decreases the potential for flooding, as there is less water available to contribute to runoff and river overflow.

Rainfall


90.000
95.000
N
28.000
26.000
24.000
Rainfall (mm)
<= 498.6750
498.6750 - 669.4689
669.4689 - 840.2627
840.2627 - 1011.0566
1011.0566 - 1181.8505
1181.8505 - 1352.6443
1352.6443 - 1523.4382
1523.4382 - 1694.2321
> 1694.2321
0
100
200 km

Normalized Difference Vegetation Index (NDVI)

- NDVI is a remote sensing index used to measure and monitor vegetation health and density by analyzing the difference between near-infrared (NIR) and visible red light reflectance from the Earth's surface.

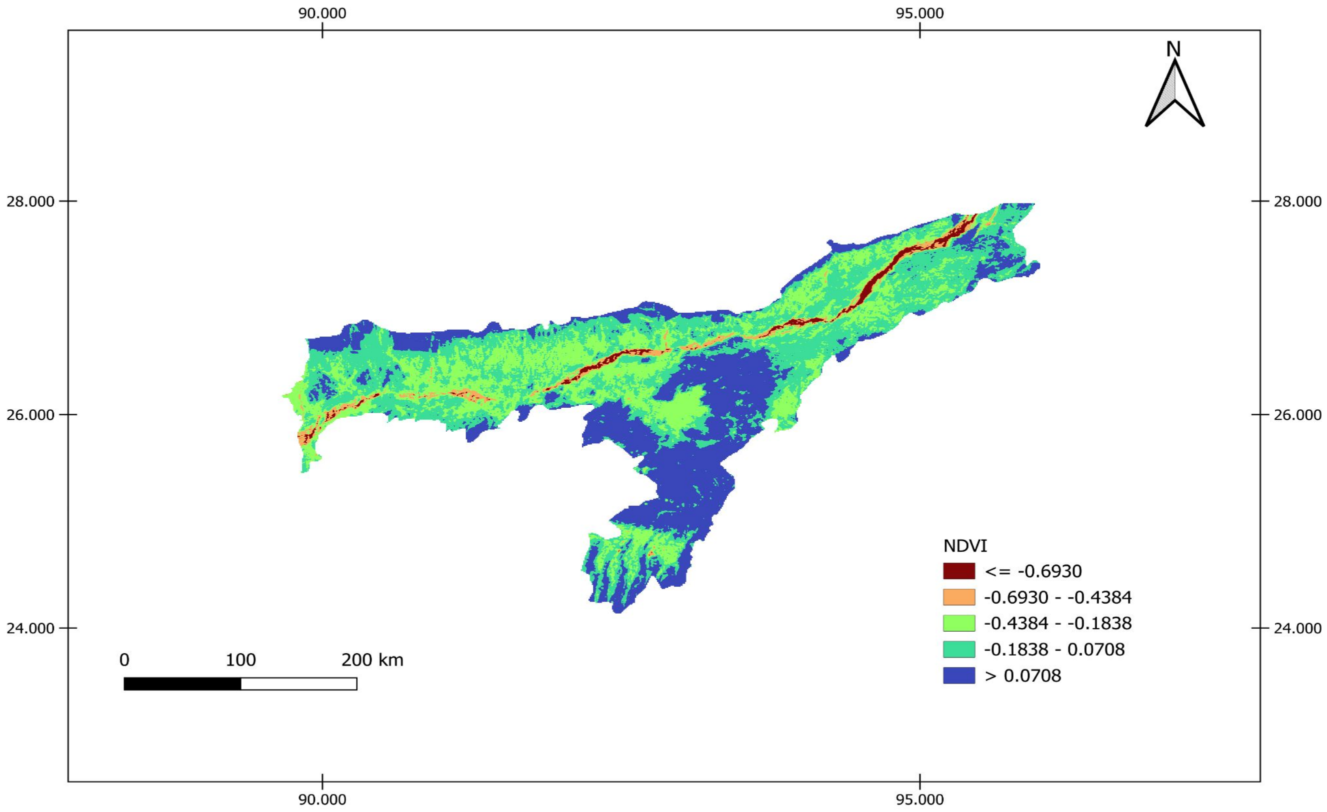
$$NDVI = \frac{NIR - Red}{NIR + Red}$$

Importance of NDVI in Flood Analysis:

High NDVI: Higher NDVI values indicate areas with dense vegetation that can absorb rainfall and reduce surface runoff, potentially lowering flood risk.

Low NDVI: Areas with low NDVI values, indicating sparse or no vegetation, are more susceptible to higher runoff rates and increased flooding because of reduced water absorption.

Normalized Difference Vegetation Index (NDVI)



Land Use Land Cover (LULC)

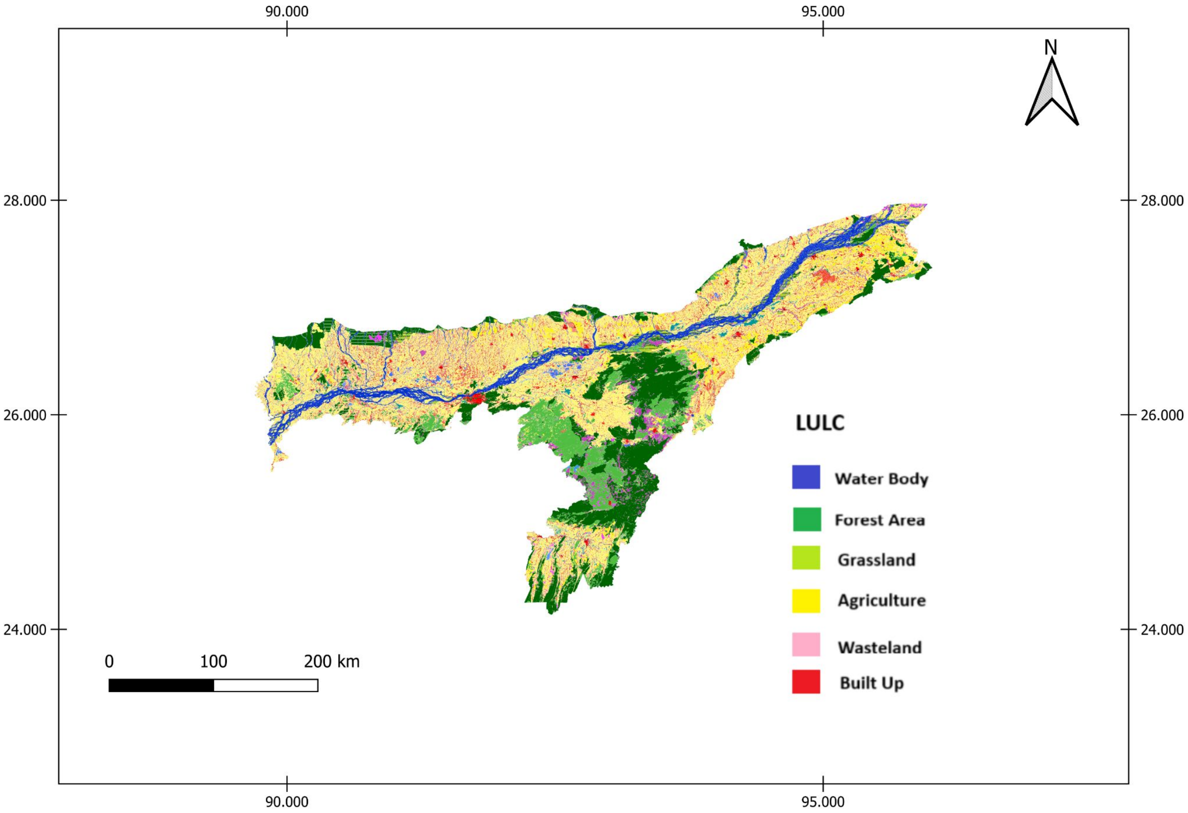
- Land Use/Land Cover (LULC) refers to the classification of land areas based on their current use (e.g., urban, agricultural, forested) or natural cover type (e.g., water bodies, grasslands).

Importance of LULC in Flood Analysis:

Urban Areas: Urban areas with high proportions of impervious surfaces (e.g., roads, buildings) reduce water infiltration, increase runoff, and contribute to a higher risk of flooding.

Vegetation Areas: Forested and vegetated areas help absorb water, slow down runoff, and reduce flood potential. Higher coverage of such land types generally leads to a lower flood risk.

Land Use Land Cover (LULC)



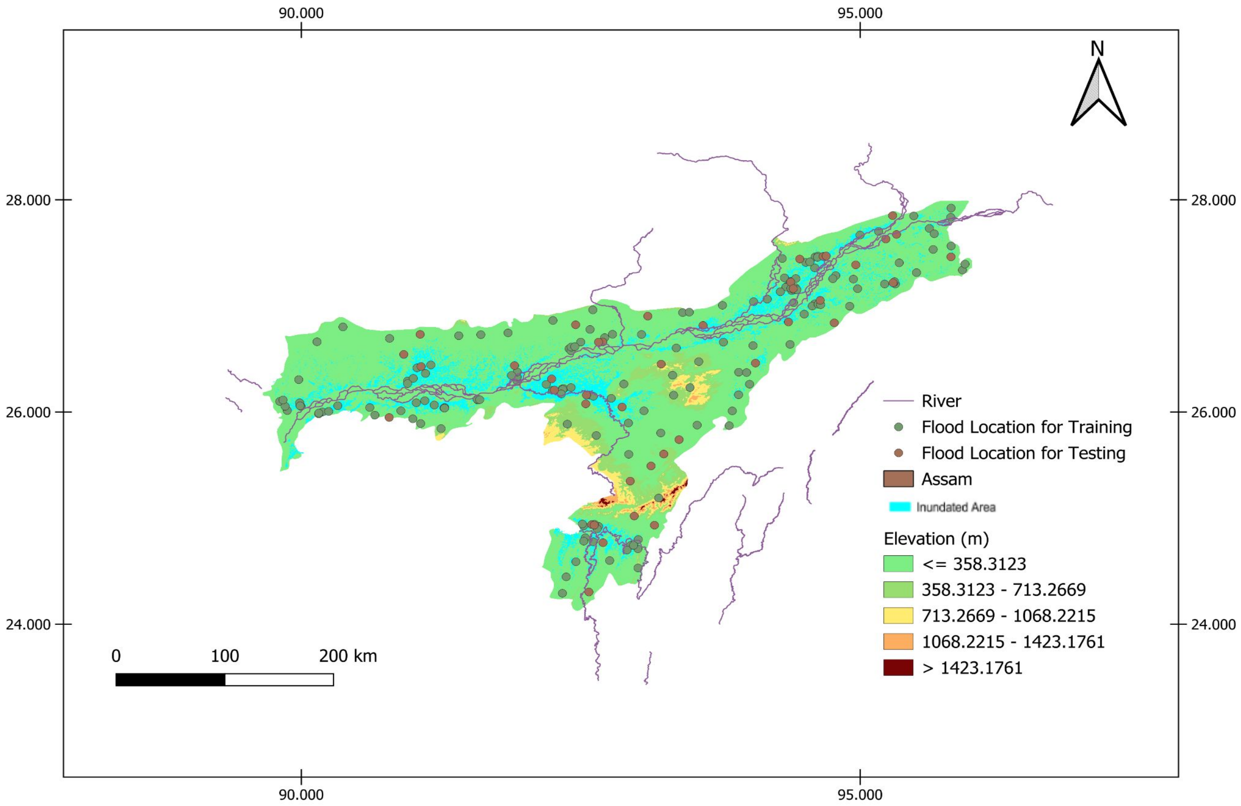
Model Training, Testing, and Accuracy Evaluation for Flood Susceptibility Mapping

Flood Inventory Map

A flood inventory map is a detailed depiction showing the locations, extent, and frequency of past floods in a specific area, aiding in the identification of flood-prone regions and guiding flood management efforts.

For analysis, 200 points were used, equally divided between flooded and unflooded areas. Models were trained using 159 points and tested on 41 points to evaluate accuracy.

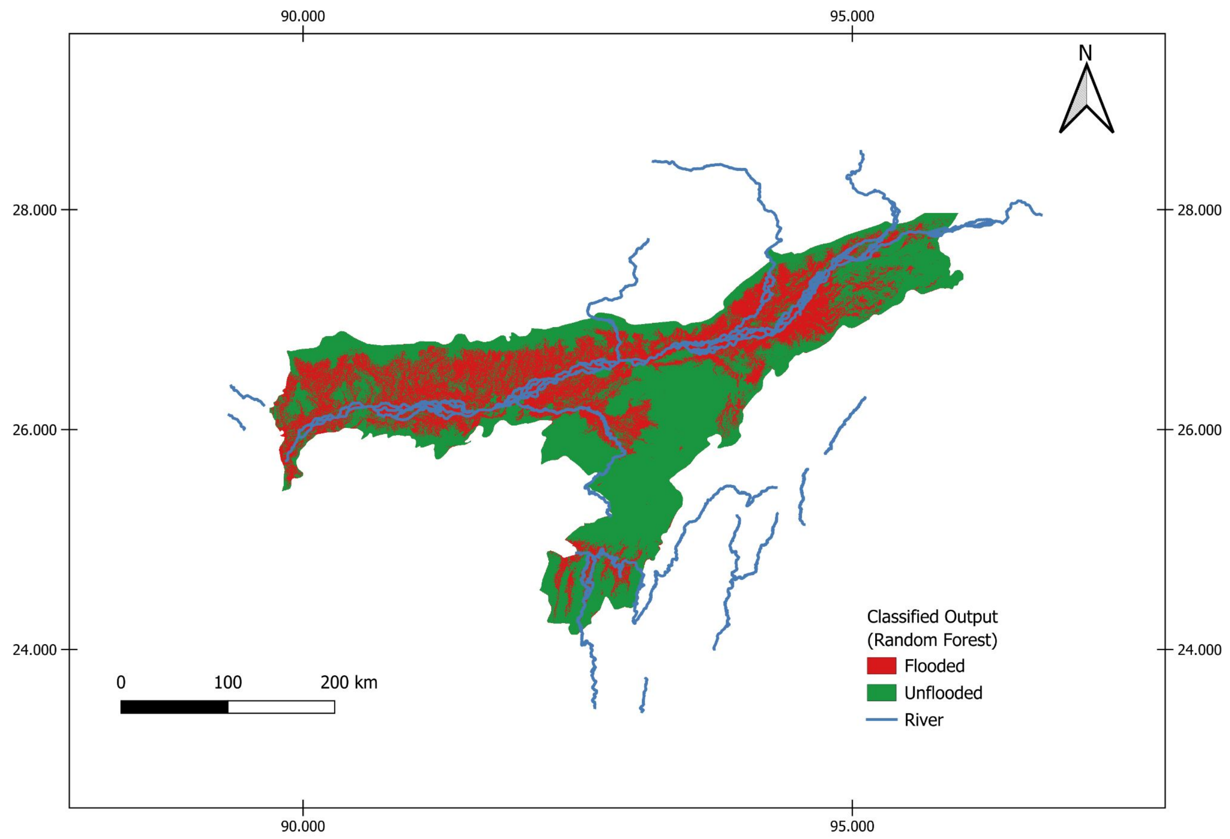
Flood Inventory Map



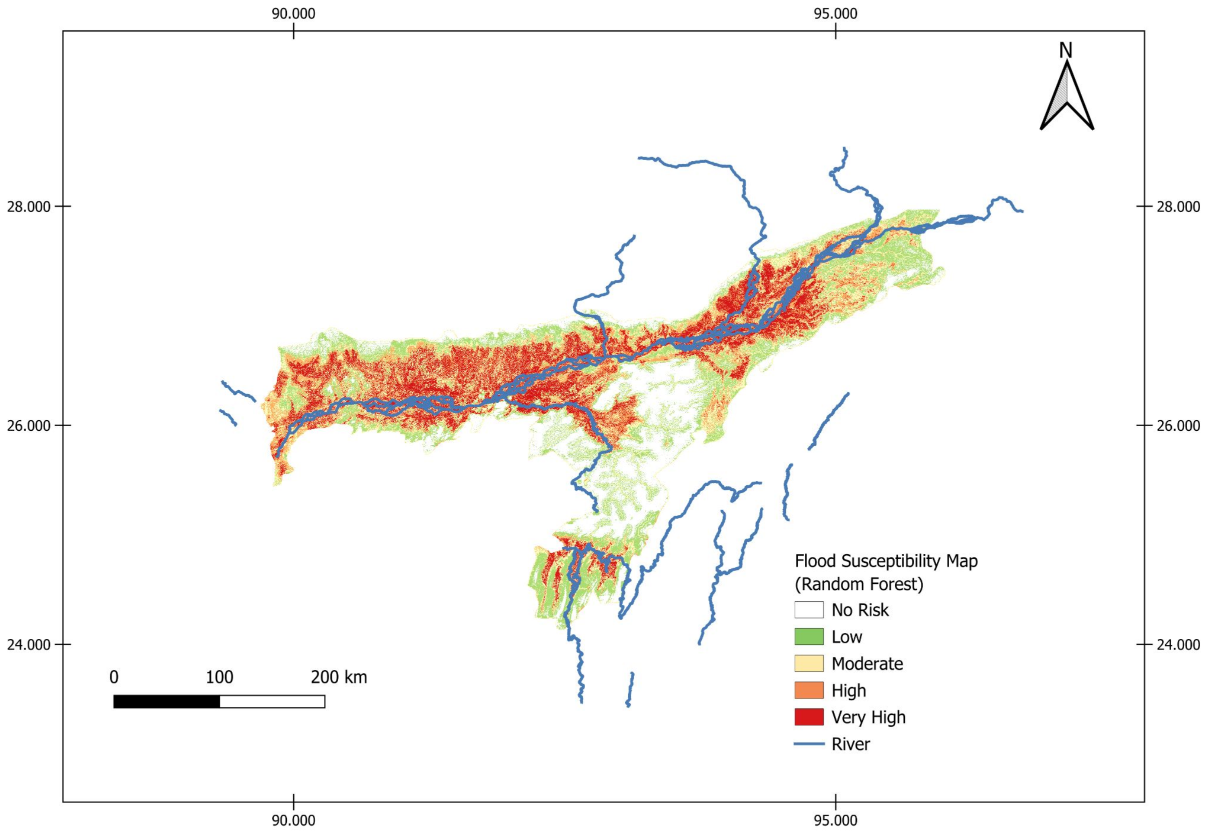
Models Used:

- Random Forest
- Support Vector Machines (SVM)
- MultiLayer Perceptron (MLP - Neural Networks)

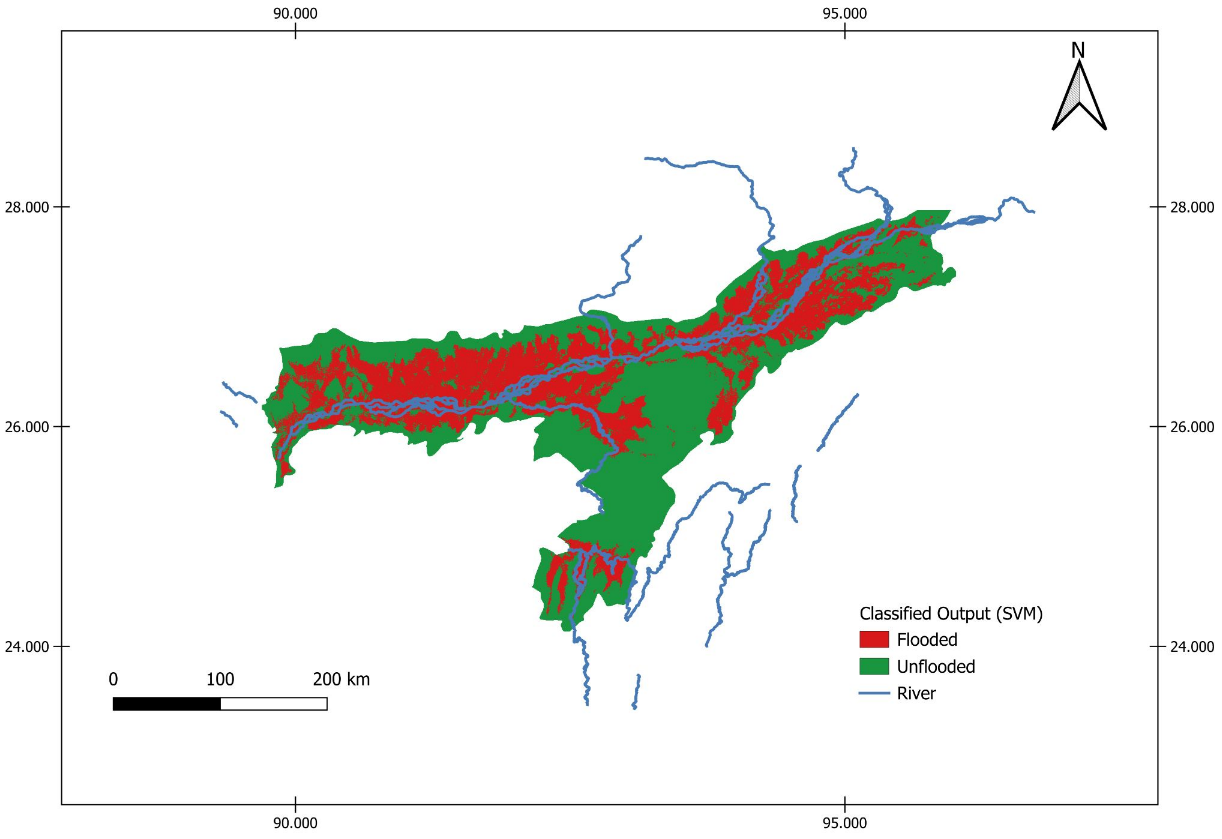
Random Forest



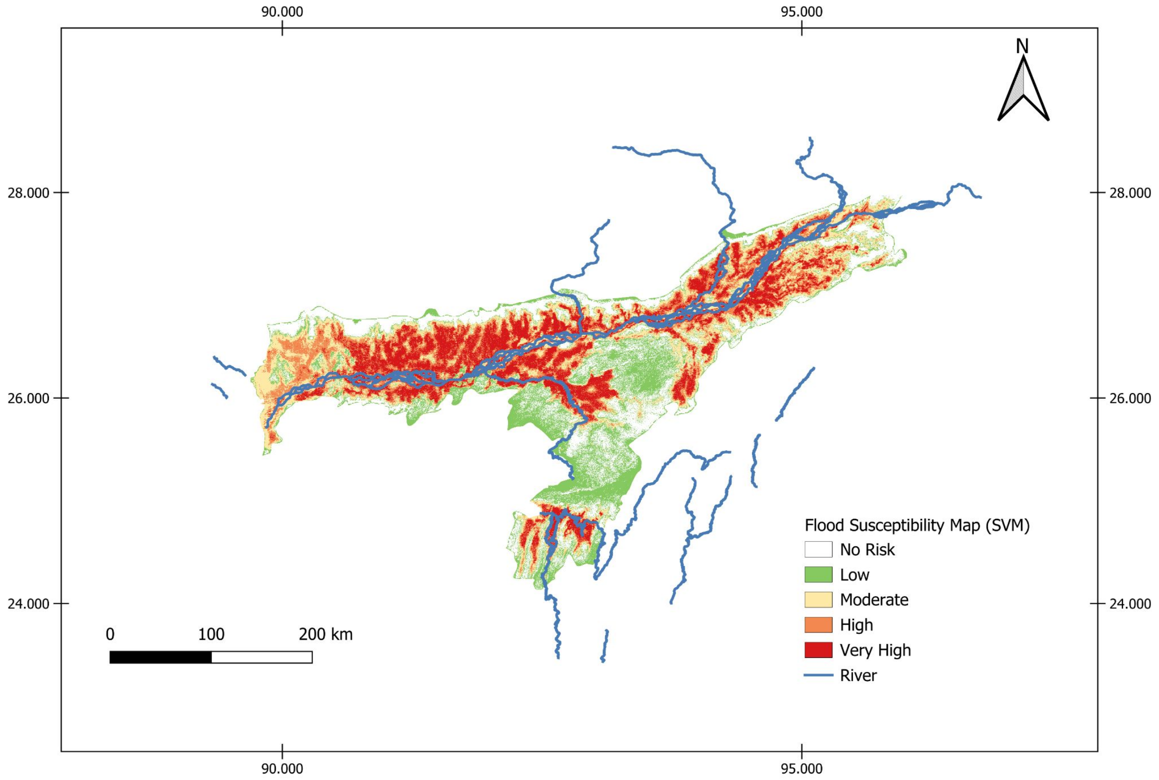
Random Forest



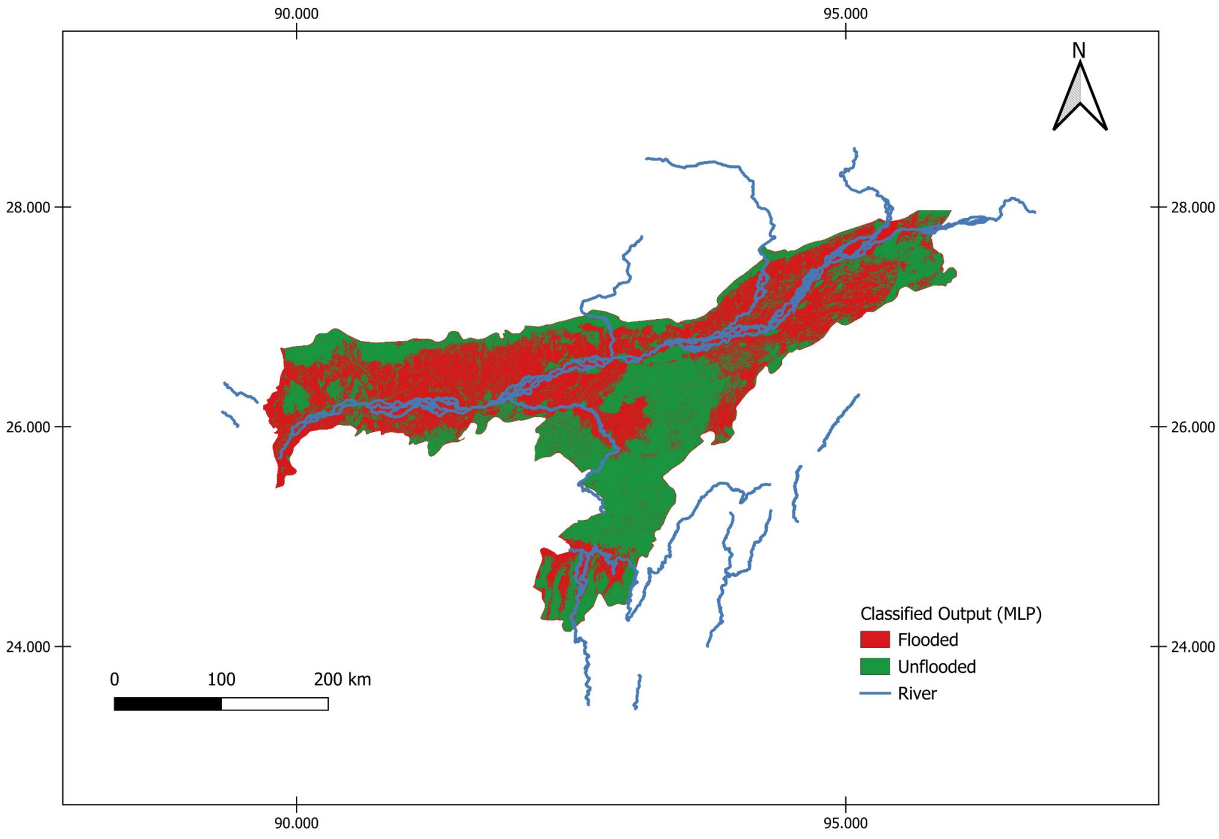
SVM


90.000
95.000
28.000
26.000
24.000
N
0
100
200 km
Classified Output (SVM)
Flooded
Unflooded
River

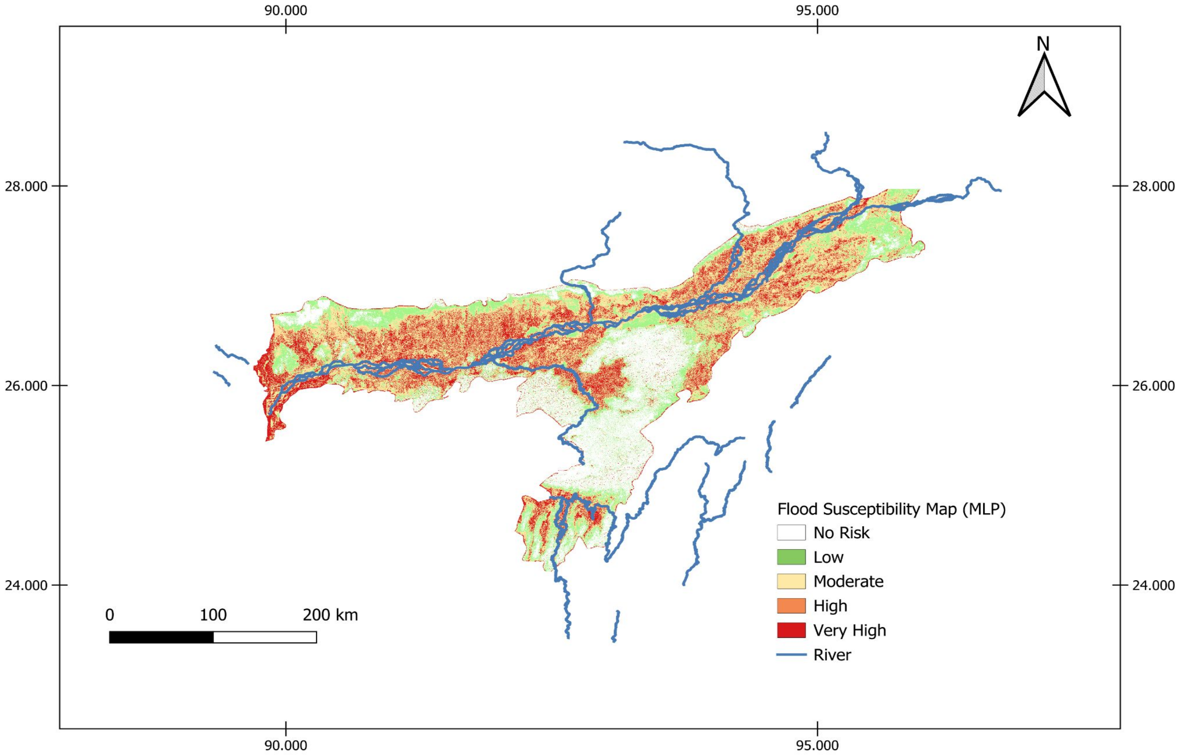
Support Vector Machines (SVM)



MultiLayer Perceptron (MLP - Neural Networks)



MultiLayer Perceptron (MLP - Neural Networks)



Accuracy Results and Model Comparison for Flood Prediction

Random Forest

Reference/Classified	Flooded	Unflooded	Total
Flooded	17	0	17
Unflooded	1	23	24
Total	18	23	41
PA[%]	94.14	100	
UA[%]	100	95.83	

Accuracy[%] = 97.50

SVM

Reference/Classified	Flooded	Unflooded	Total
Flooded	16	3	19
Unflooded	2	20	22
Total	18	23	41
PA[%]	87.00	88.84	
UA[%]	84.21	90.90	

Accuracy[%] = 88.09

MLP (Neural Networks)

Reference/ Classified	Flooded	Unflooded	Total
Flooded	18	7	25
Unflooded	0	16	16
Total	18	23	41
PA[%]	100	76.96	
UA[%]	72	100	

Accuracy[%] = 85.53

Among all tested models, Random Forest achieved the best results in terms of accuracy and reliability, making it the top-performing model for flood prediction. It was followed by SVM (Support Vector Machine) and MLP (Neural Networks), which also showed strong but slightly lower performance.

Data Sources

- Bhuvan
- Bhoonidhi
- CHRS
- Open Government Data

Access Our Data, Referenced Papers and Results Here

OneDrive Link

Q & A

THANK YOU