Prediction of Flash Flood Susceptible Zone using Machine Learning and Advanced Geospatial Techniques in parts of Teesta Basin, India

Prepared & Presented by

Moslem Miah

Roll: VU/PG/22/34/02-IVS-0014

Registration No. VP221000161 of 2022-2023

Under the Supervision of

Dr. Dipanwita K. Dutta

Assistant Professor

Department of Remote Sensing and GIS

Vidyasagar University

Paschim Mednipur, 721102

Date – 09/07/2024

Introduction

- Flash floods are fast and unexpected flood events that can happen within an hour of heavy rainfall, dam breaks, or other causes.
- Understanding and predicting areas prone to flash floods is crucial for minimizing the adverse effects on communities and ecosystems.
- Remote sensing is a valuable method for monitoring and analyzing environmental factors across large areas, offering wide coverage, high-resolution data, and multi-spectral information.
- Machine learning (ML) techniques offer powerful tools for analyzing complex datasets and improving predictive accuracy.

Aim & Objectives

The aim of the study is to prediction of flash flood susceptibility zone using ML and advanced geospatial techniques in Teesta basin, India.

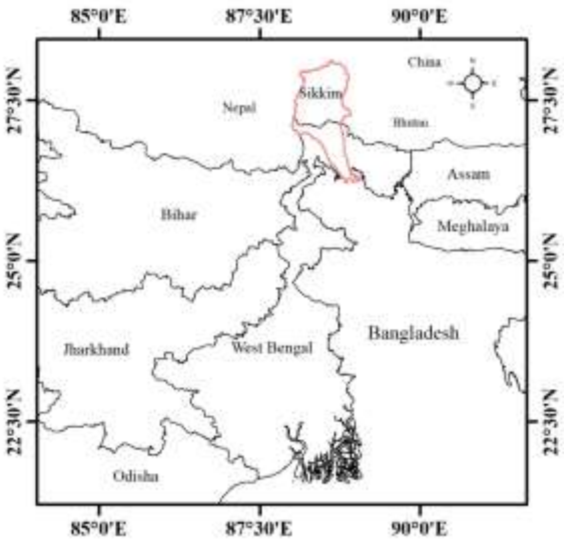
The major objectives of the study are –

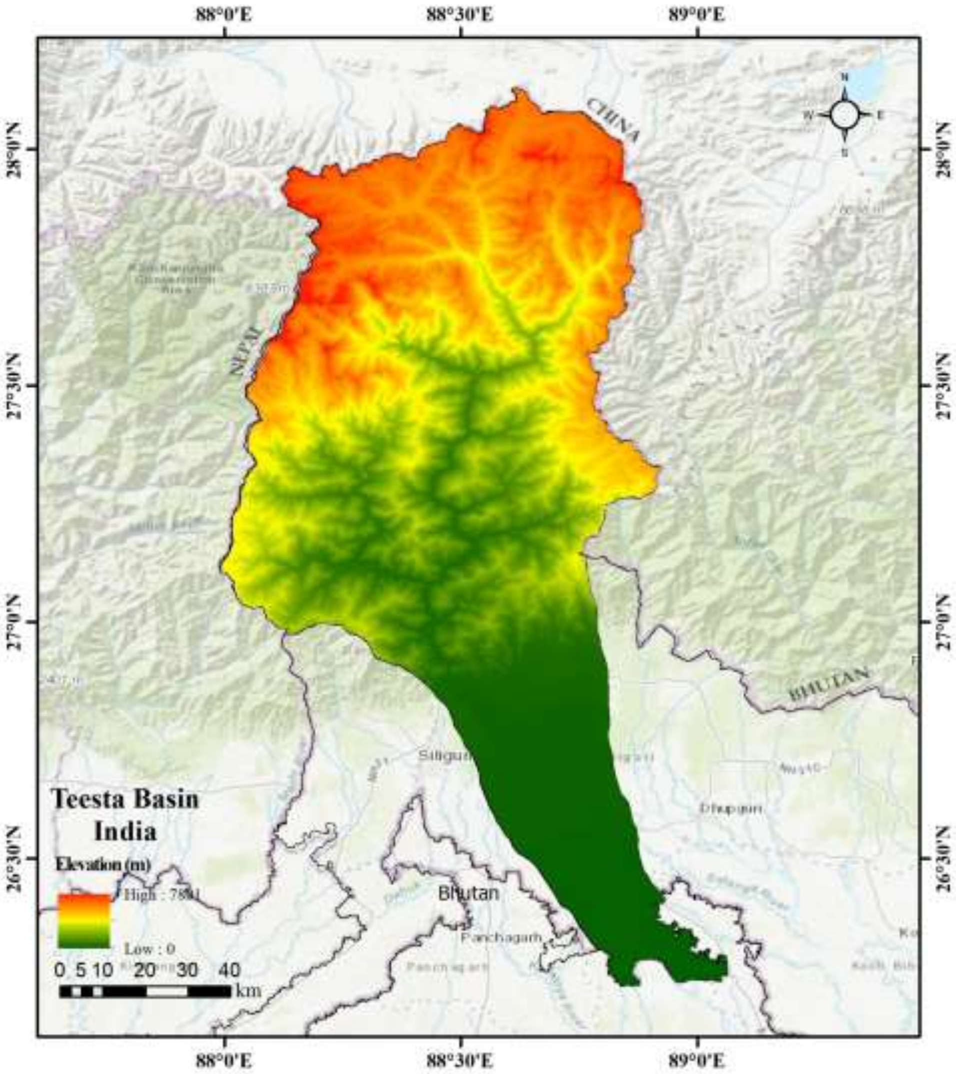
- **Objective 1:** To identify and analyze the key physical, environmental, and anthropogenic factors that contribute to the occurrence of flash floods in the Teesta Basin, India.

- **Objective 2:** To develop and validate a machine learning-based flash flood susceptibility model using the identified contributing factors and advanced geospatial techniques

Study Area

- This study area is situated between latitudes 26°13'49.97“ N and 28°7'39.93“ N, and longitudes 87°59'30.85“ E and 89°3'48.08" E.
- **Area coverage**: 9,855 sq. km.
- **Elevation:** 0 – 7800 m
- **Major river:** Teesta



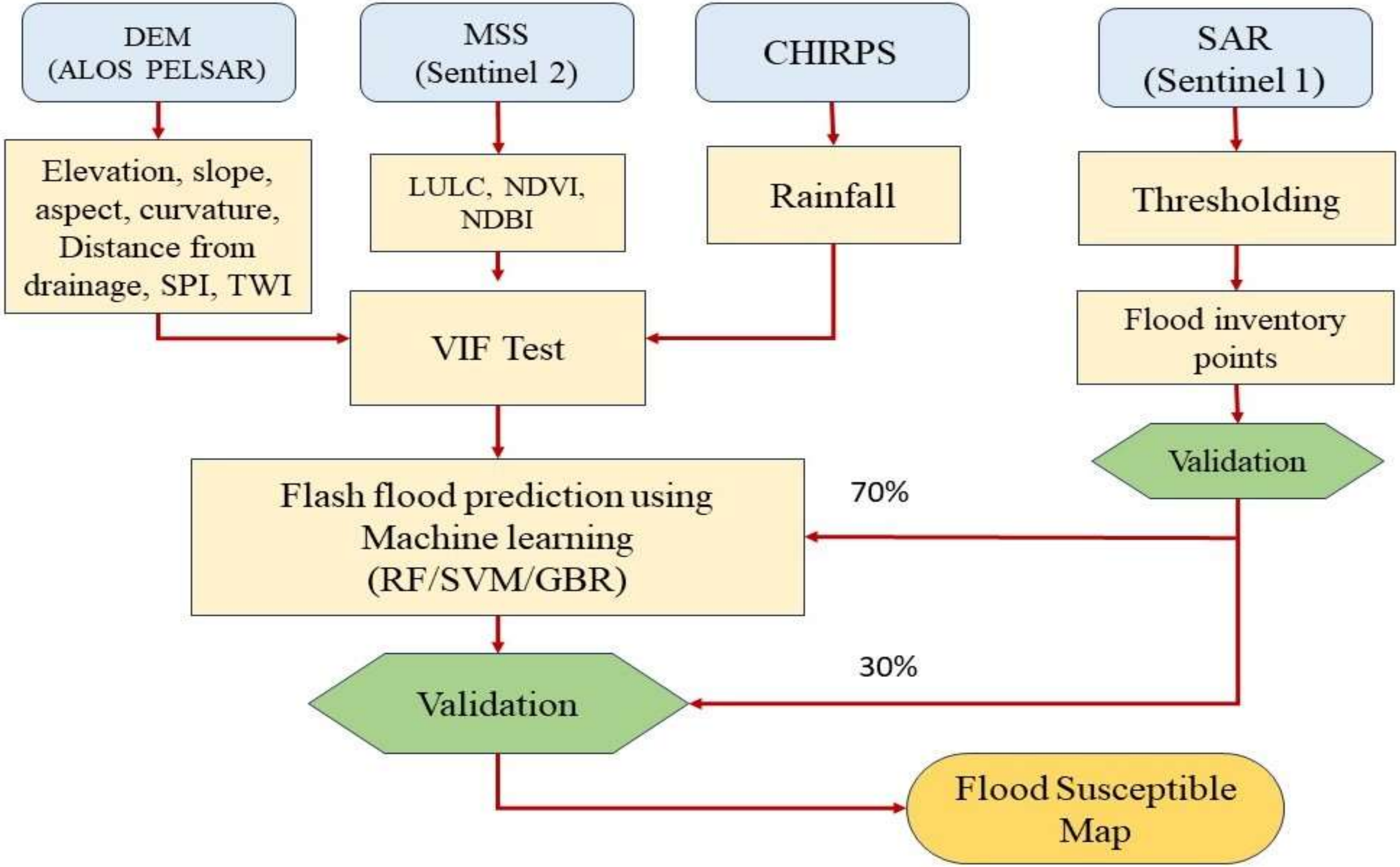


Datasets

Data	Resolution	Source	Acquisition Date
Sentinel – 1	10 m	GEE Data Catalog	Before flood: 27/09/2023-03/10/2023 After flood: 04/10/2023-11/10/2023
Sentinel – 2	10	GEE Data Catalog	01/09/2023-30/09/2023
ALOS PALSAR DEM	12.5 m	Alaska Satellite Facility	2006 - 2011
CHIRPS	5.5 km.	GEE Data Catalog	27/09/2024-04/10/2024

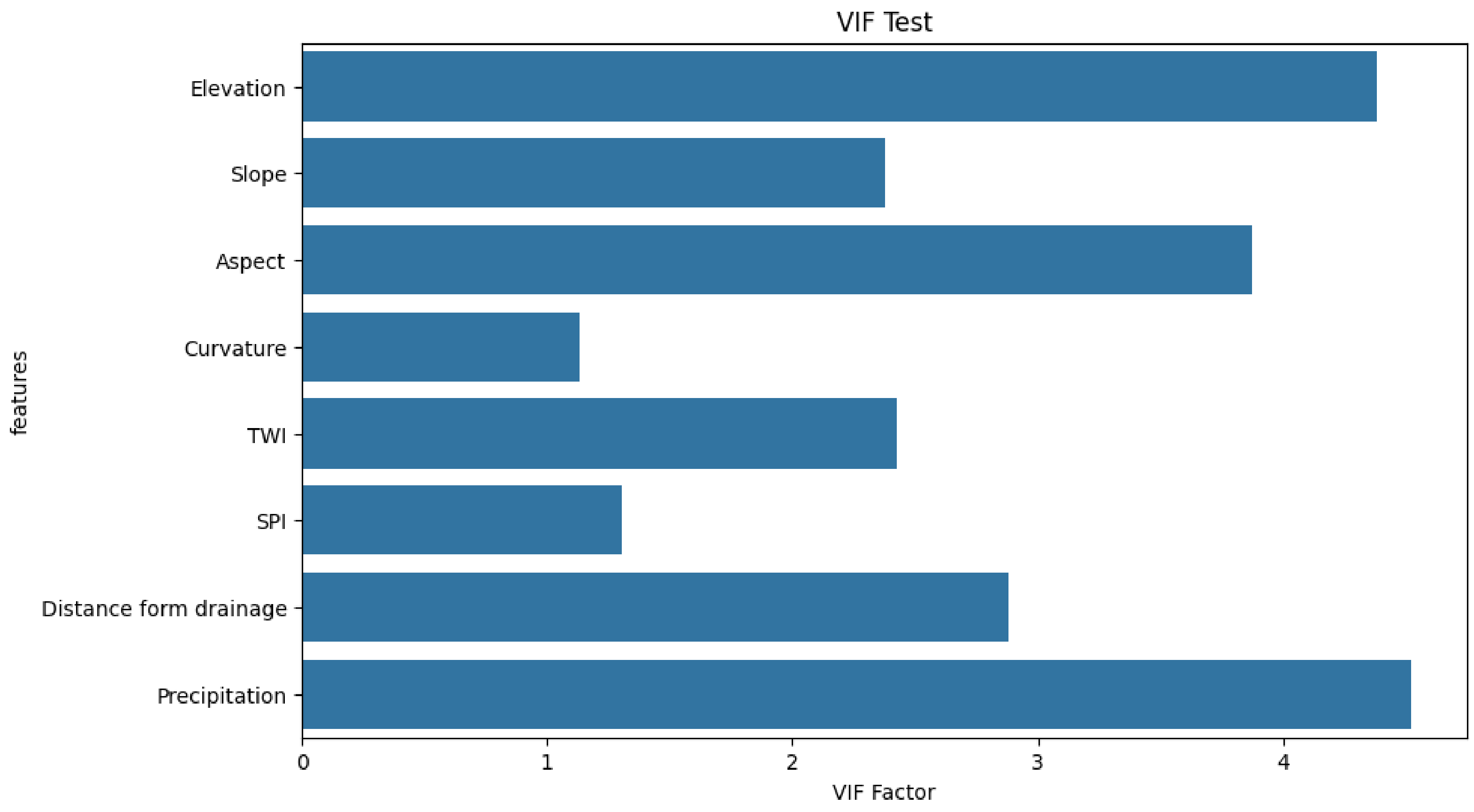
Software & Platform used:
ArcGIS, ArcGIS Pro, QGIS, Google Earth Pro, Google Earth Engine, Jupyter Notebook (Python)

Methodology

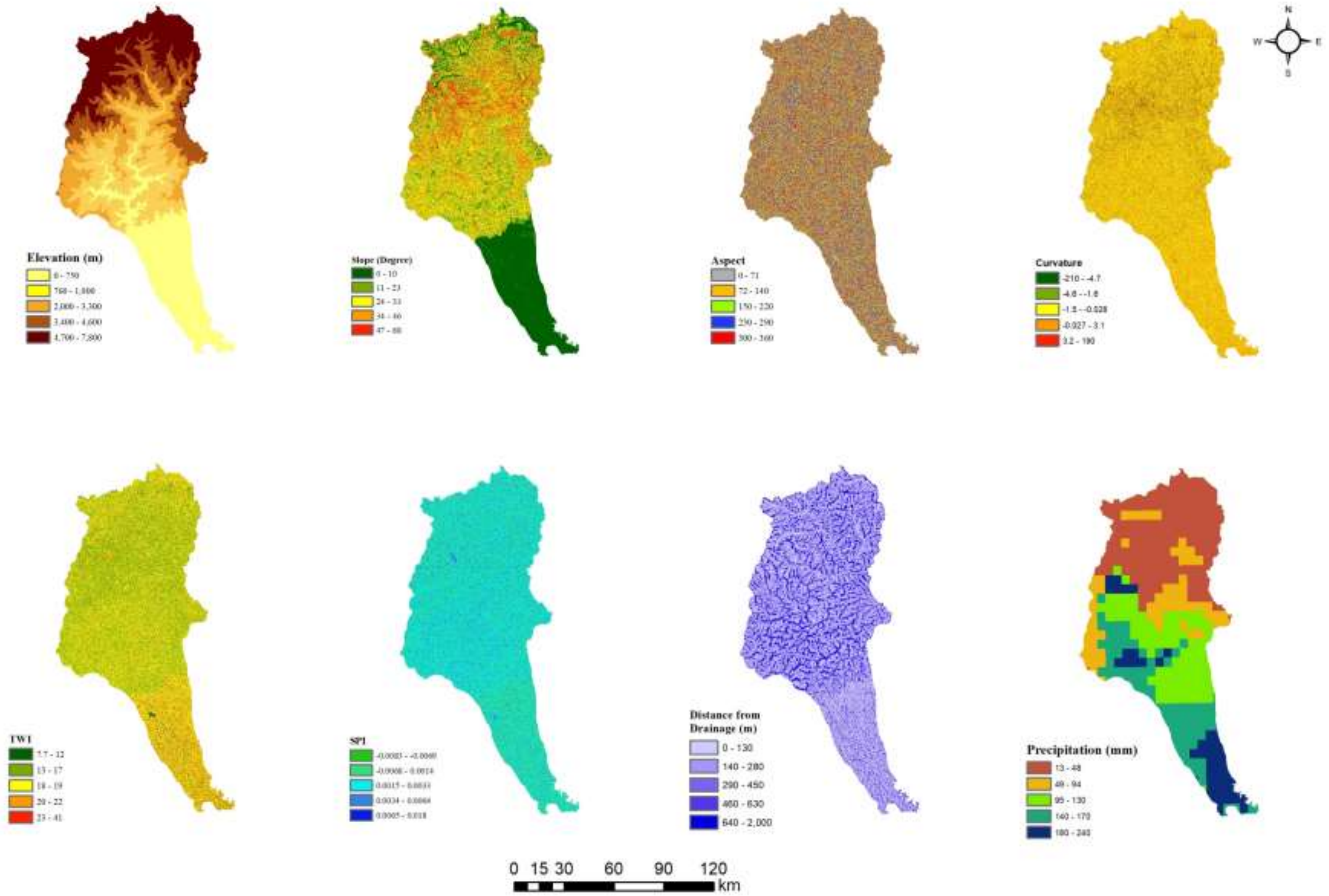


Influencing Factors

- The study began by selecting numerous factors based on a review of previous literature and data availability.
- After conducting a multicollinearity test (VIF), eight factors were chosen for the study: Elevation, Slope, Aspect, Curvature, distance from river, Stream Power Index (SPI), Topographic Wetness Index (TWI), and Precipitation.

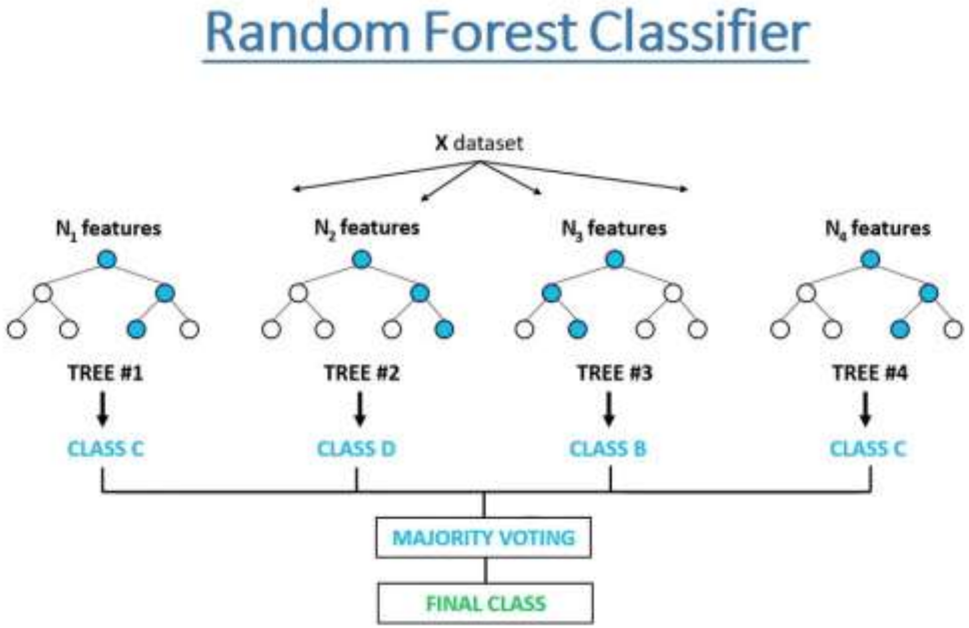


Influencing Factors

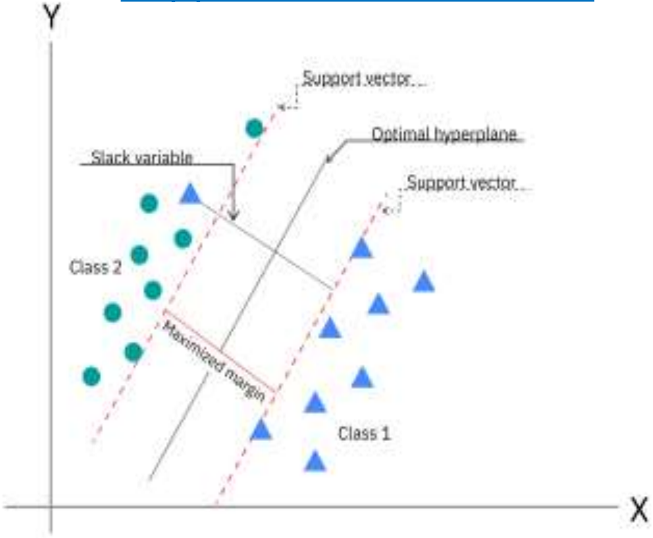


ML Model

Random Forest (RF)	Support Vector Machine (SVM)	Gradient Boosting Regression (GBR)

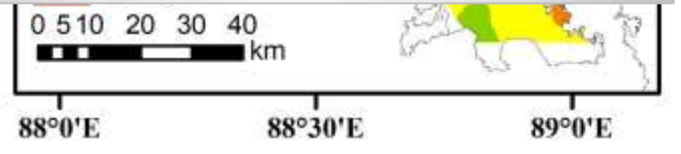


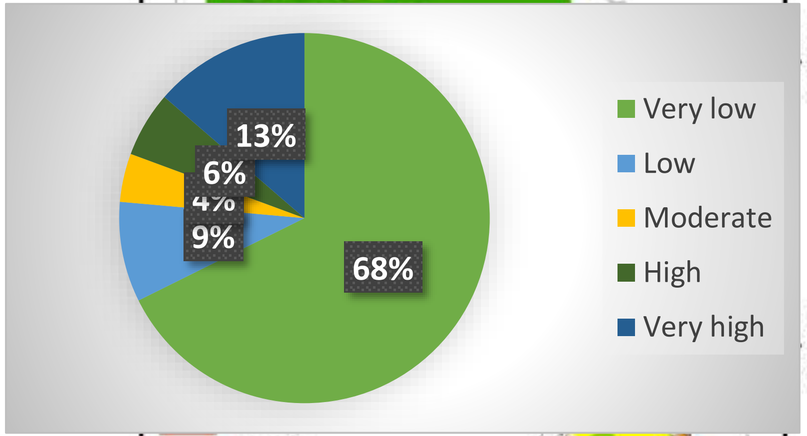
Support Vector Machine



SVM

Zone	Area (Sq.Km.)	Area (%)
Very low	6699.63	67.98
Low	853.07	8.66
Moderate	413.17	4.19
High	557.33	5.66
Very high	1331.79	13.51



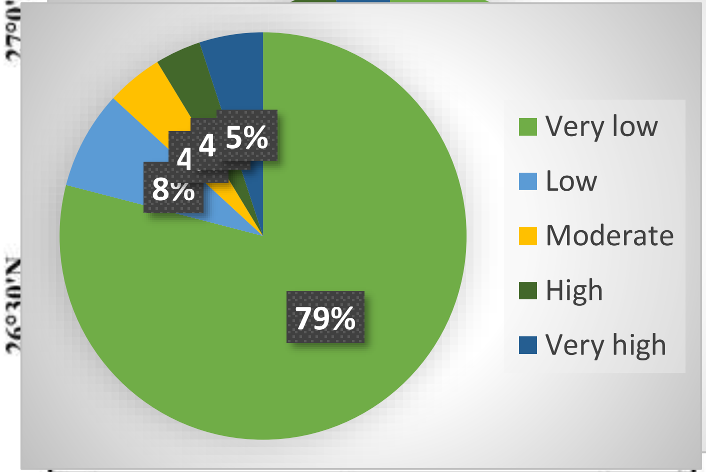


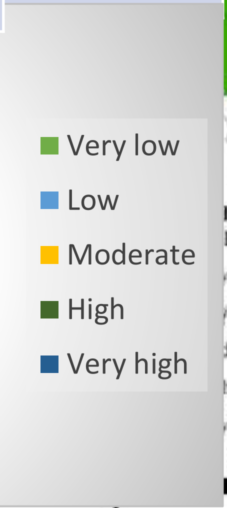
RF

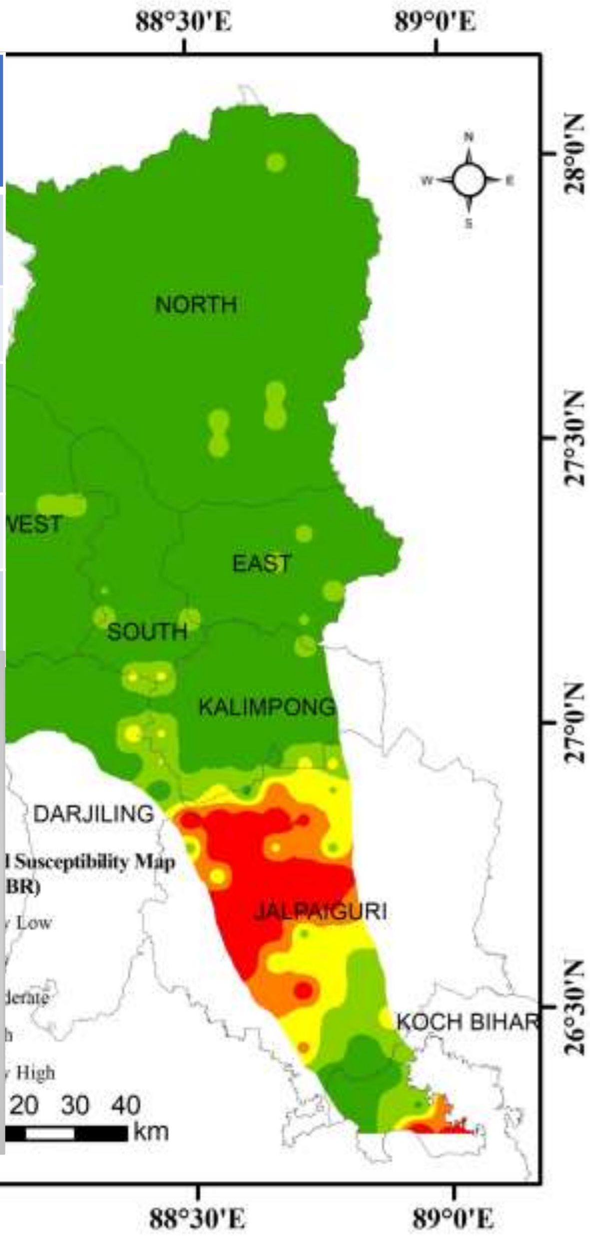
Zone	Area	Area (%)
		.7
		.7
		.9
		2
		.8

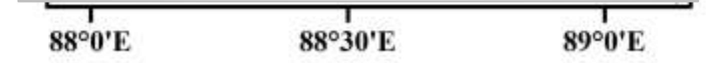
GBR

Zone	Area (Sq.Km.)	Area (%)
Very low	7787.77	79.02
Low	768.42	7.80
Moderate	440.65	4.47
High	359.06	3.64
Very high	499.11	5.06









Comparison result of RF, SVM, & GBR model

RF

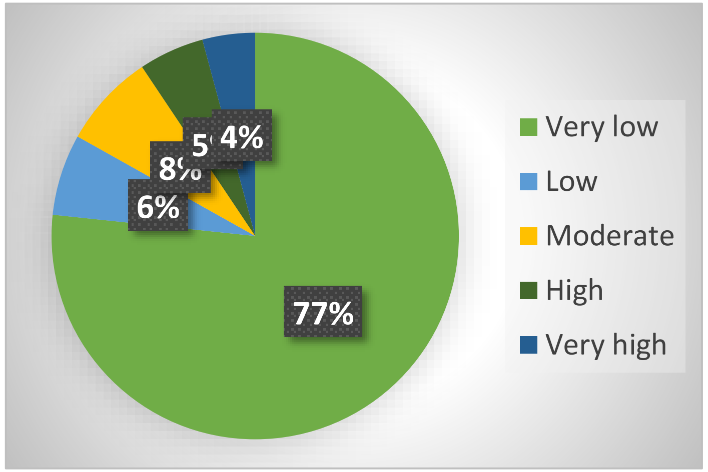
Zone	Area (Sq.Km.)	Area (%)
Very low	7555.11	76.7
Low	637.30	6.47
Moderate	737.92	7.49
High	512.37	5.2
Very high	412.31	4.18

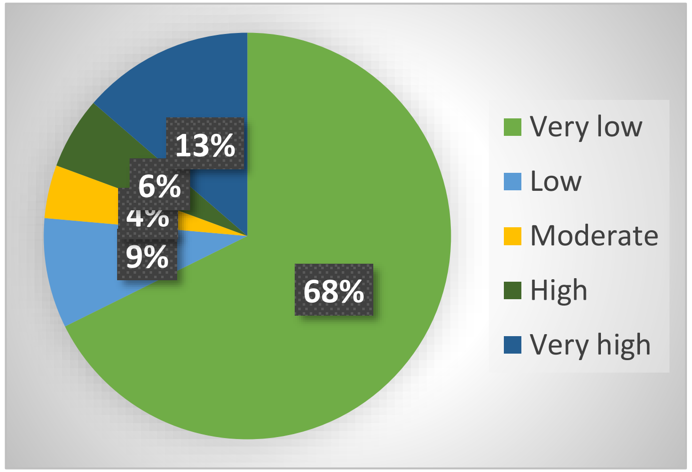
SVM

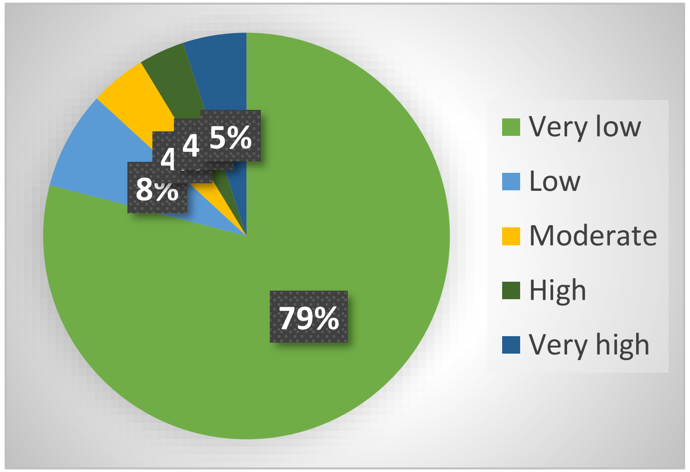
Zone	Area (Sq.Km.)	Area (%)
Very low	6699.63	67.98
Low	853.07	8.66
Moderate	413.17	4.19
High	557.33	5.66
Very high	1331.79	13.51

GBR

Zone	Area (Sq.Km.)	Area (%)
Very low	7787.77	79.02
Low	768.42	7.80
Moderate	440.65	4.47
High	359.06	3.64
Very high	499.11	5.06

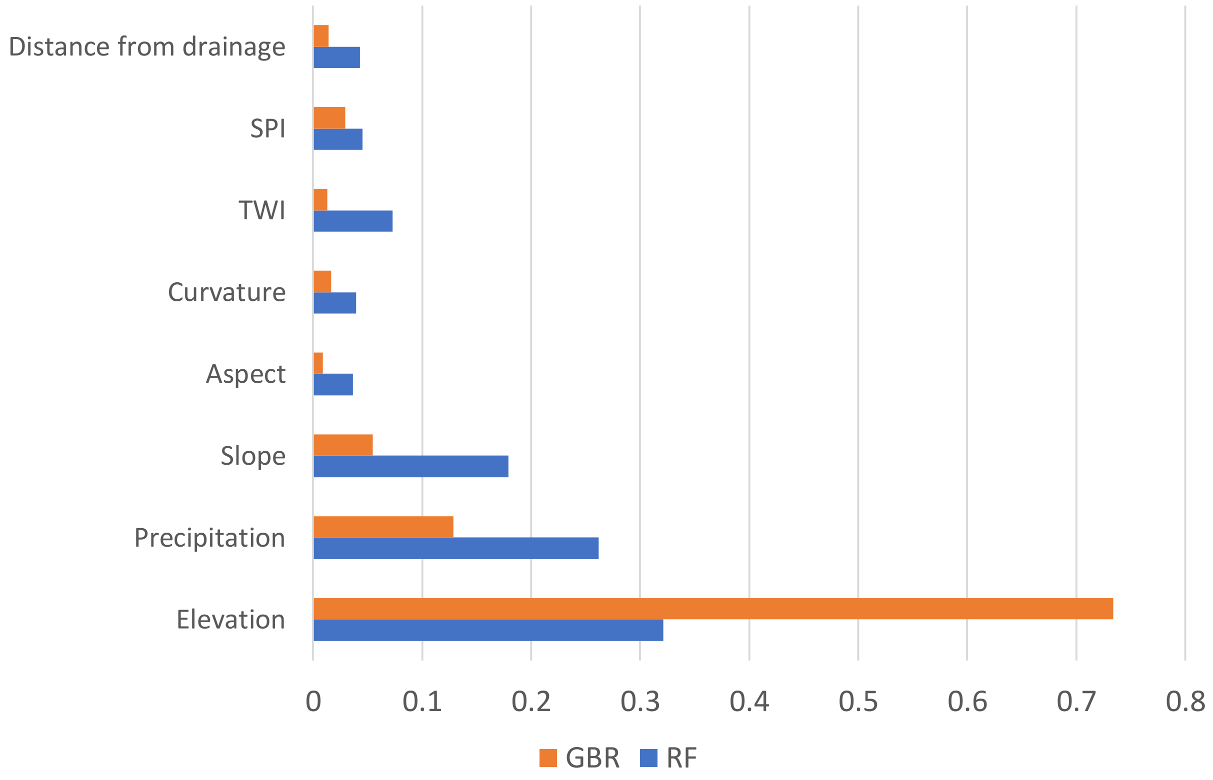






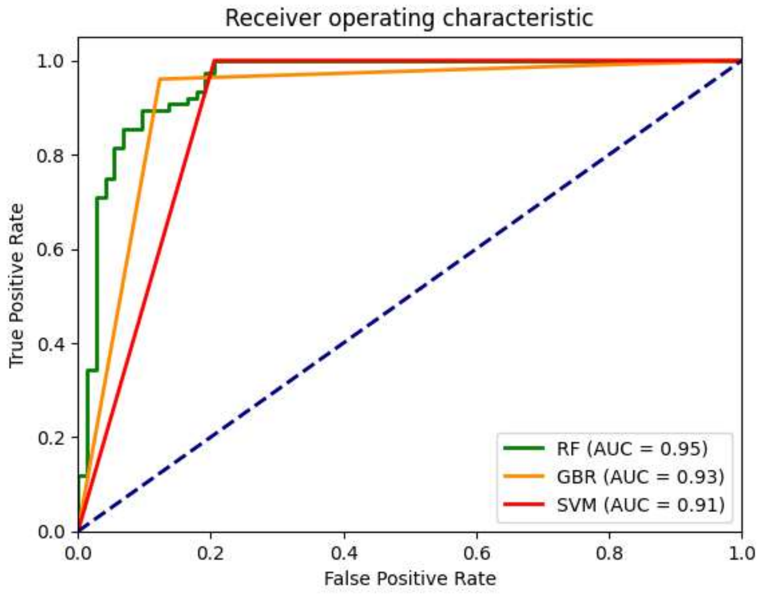
Feature Importance

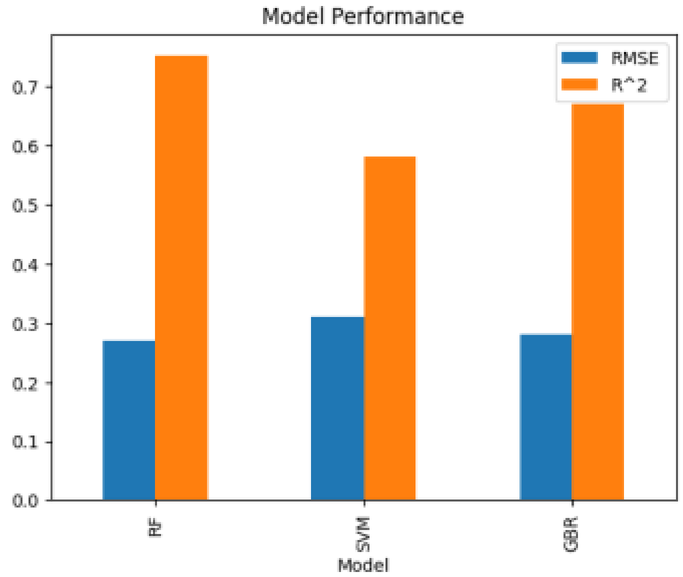
Features	Importances (RF)	Importances (GBR)
Elevation	0.321499	0.734011
Precipitation	0.262130	0.128763
Slope	0.179428	0.054906
TWI	0.072786	0.013103
SPI	0.045124	0.029252
Aspect	0.036723	0.009039
Curvature	0.039202	0.016795
Distance from drainage	0.043108	0.014132



Model Performance

Model	Root Mean Squared Error (RMSE)	R^2	Area Under Cover (AUC)
Random Forest (RF)	0.27	0.75	0.95
Support Vector Machine (SVM)	0.31	0.58	0.91
Gradient Boosting Regression	0.28	0.67	0.93





Conclusion

- The study successfully identified and analyzed the key physical, environmental, and anthropogenic factors contributing to flash floods in the Teesta basin.
- The study successfully developed and validated a machine learning-based flash flood susceptibility model for the Teesta basin using the identified contributing factors and advanced geospatial techniques.
- The model demonstrated high predictive accuracy and reliability in identifying areas prone to flash flood events, as validated through cross-validation and independent testing.
- The flash flood susceptibility maps generated from the models can serve as a valuable tool for local authorities and decision-makers to prioritize flood-prone areas, implement targeted mitigation measures, and enhance disaster preparedness in the Teesta basin.

Field Photograph

Thank You