






Article

Combining Machine Learning Models and Satellite Data of an Extreme Flood Event for Flood Susceptibility Mapping

Nikos Tepetidis ¹, Ioannis Benekos ², Theano Iliopoulou ¹, Panayiotis Dimitriadis ¹
and Demetris Koutsoyiannis ^{1,*}

¹ Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, 15780 Zographou, Greece; nikostepe191201@gmail.com (N.T.); tiliopoulou@hydro.ntua.gr (T.I.); pandim@itia.ntua.gr (P.D.)

² Laboratory of Risk Management and Resilience, Hellenic Institute of Transport, Centre of Research and Technology Hellas, 34 Ethnarchou Makariou, 16341 Ilioupoli, Greece; ibenekos@certh.gr

* Correspondence: dk@itia.ntua.gr

Abstract

Machine learning techniques have been increasingly used in flood management worldwide to enhance the effectiveness of traditional methods for flood susceptibility mapping. Although these models have achieved higher accuracy than traditional ones, their application has not yet reached full maturity. We focus on applying machine learning models to create flood susceptibility maps (FSMs) for Thessaly, Greece, a flood-prone region with extreme flood events recorded in recent years. This study utilizes 13 explanatory variables derived from topographical, hydrological, hydraulic, environmental and infrastructure data to train the models, using Storm Daniel—one of the most severe recent events in the region—as the primary reference for model training. The most significant of these variables were obtained from satellite data of the affected areas. Four machine learning algorithms were employed in the analysis, i.e., Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF) and eXtreme Gradient Boosting (XGBoost). Accuracy evaluation revealed that tree-based models (RF, XGBoost) outperformed other classifiers. Specifically, the RF model achieved Area Under the Curve (AUC) values of 96.9%, followed by XGBoost, SVM and LR, with 96.8%, 94.0% and 90.7%, respectively. A flood susceptibility map corresponding to a 1000-year return period rainfall scenario at 24 h scale was developed, aiming to support long-term flood risk assessment and planning. The analysis revealed that approximately 20% of the basin is highly prone to flooding. The results demonstrate the potential of machine learning in providing accurate and practical flood risk information to enhance flood management and support decision making for disaster preparedness in the region.

Keywords: flood susceptibility mapping (FSM); machine learning (ML); LR; SVM; RF; XGBoost; Greece; Thessaly; Storm Daniel



Academic Editor: Álvaro Sordo-Ward

Received: 21 May 2025

Revised: 14 August 2025

Accepted: 7 September 2025

Published: 10 September 2025

Citation: Tepetidis, N.; Benekos, I.; Iliopoulou, T.; Dimitriadis, P.; Koutsoyiannis, D. Combining Machine Learning Models and Satellite Data of an Extreme Flood Event for Flood Susceptibility Mapping. *Water* **2025**, *17*, 2678. <https://doi.org/10.3390/w17182678>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. General and Literature Review

Flood events are natural phenomena in which normally dry land becomes inundated due to excessive water accumulation. These result in devastating consequences worldwide, causing significant economic, environmental, and social impacts [1,2]. They originate from various factors, including intense or prolonged rainfall, river overflow, storm surges, and dam failures. Floods are a global challenge, affecting regions regardless of geographical

boundaries and endangering millions of lives. Their impact highlights the need for continuous research and strategic planning to mitigate associated risks. To this end, in the framework of the implementation of the European Union Flood Directive [3], the assessment and management of flood risk have been prioritized. Floods can be categorized according to their origin, root cause and occurrence, as river floods, flash floods, coastal floods, urban floods and floods induced by the failure of respective flood protection structures [4,5]. Effective flood management is crucial for minimizing damage to communities, infrastructure and the environment.

As suggested by the current legal framework (i.e., Floods Directive 2007/60/EC [3]), flood susceptibility and hazard maps are essential at the river basin level. By the term Flood Susceptibility Map (FSM), we refer to a spatial representation of a region or river basin that identifies areas that are prone to flooding. These maps help to highlight regions with a higher likelihood of flooding, considering various factors, e.g., topography, rainfall, land use and historical floods. On the other hand, Flood Hazard Maps (FHMs) measure the water depth and extent across a flooded area. They play a crucial role in risk assessment, land-use planning and the development of effective flood management strategies.

Traditionally, flood susceptibility mapping has been conducted using a variety of qualitative and quantitative approaches. Qualitative methods, such as Multi-Criteria Decision Analysis (MCDA) [6], integrate expert knowledge and spatial criteria to evaluate flood-prone areas. Statistical approaches, including linear regression and frequency ratio analysis [7], are utilized to assess the relationships between historical flood occurrences and influencing factors. Hydrological-based models, including HyMOD [8] and the Soil and Water Assessment Tool (SWAT) [9], simulate the hydrological cycle to estimate surface runoff and flood potential. Additionally, hydraulic models, such as the Hydrologic Engineering Center's River Analysis System (HEC-RAS) [10], are employed to simulate water flow dynamics and floodplain extents based on channel geometry and discharge data. These conventional methods, while widely used, often require large amounts of data, which must be weighted by selecting appropriate weights, and often require time-consuming flood simulations that are based on numerical models, which, in turn, also depend on the discretization of the physical domain.

To address above limitations, alternative approaches such as Machine Learning (ML) and Deep Learning (DL) models have been explored in flood susceptibility mapping, offering potential benefits in terms of efficiency and accuracy. ML models can handle large and complex datasets, and identify non-linear relationships and complicated patterns. Extensive research has been conducted on the application of these models across various river basins worldwide. A comprehensive literature review of 58 recent publications is presented in [4], covering various aspects of flood mapping, including flood inundation, flood susceptibility and flood hazard, with a focus on DL approaches. Various ML algorithms have been employed for flood susceptibility mapping. In particular, Artificial Neural Networks (ANNs), Multi-Layer Perceptron (MLP) and Least absolute shrinkage and selection operator (Lasso) have been used in [11], as well as Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network (CNN) in [12]. Additionally, various boosting algorithms, such as AdaBoost, Gradient Boosting, XGBoost, CatBoost, and Stochastic Gradient Boosting (SGB), have been utilized [13]. Recent advances in flood susceptibility modeling include the use of the Kolmogorov–Arnold Network (KAN), which leverages mathematical decomposition techniques to improve prediction accuracy and has been successfully applied to river basins in Iran [14]. Table 1 presents a brief summary of selected studies in this field. For each algorithm listed in the table, the best-performing model is highlighted in bold. Table 2 lists the flood conditioning factors used in the respective studies, which serve as inputs for the corresponding models. Although machine

learning models are pretty powerful tools, they require a significant amount of high-quality and labeled data for training. If the data is limited, biased or unreliable, their accuracy can suffer—making them less effective, especially in areas with little historical flood data. However, unlike traditional methods, they can often work with smaller, more focused datasets or utilize techniques like data augmentation and transfer learning to improve accuracy.

Machine learning models can be very useful when combined with satellite data, offering state-of-the-art, data-driven methods for analyzing large-scale patterns. Satellite imagery offers high-resolution, real-time data on some of the most important flood-influencing variables such as topography, land cover, soil moisture and precipitation, as well as the flooded area. ML algorithms can process and analyze these large datasets to detect patterns and flood-prone areas, identify the most influencing features and improve predictive accuracy. Flood conditioning factors serve as input to the model (independent variables), while the output (dependent variable) determines whether an area is classified as flooded or not. This process can be framed as either a classification or regression problem, depending on the nature of the output values. By integrating ML with satellite observations, researchers can enhance flood prediction capabilities, even in regions with limited ground-based hydrological data [15]. However, challenges such as data preprocessing, cloud cover interference, and model interpretability remain key considerations in optimizing the use of ML for satellite-based flood susceptibility mapping.

Despite the widespread use and numerous applications of ML models for FSM globally, their adoption in Greek basins remains limited. This is largely due to challenges such as a lack of technical expertise and the relatively early stage of ML integration into flood management practices. An attempt was made in [16] at the national scale for flood hazard mapping using multi-criteria analysis and ANN. The full potential of ML in FSM remains underexplored. Similar barriers to ML adoption exist in other regions as well, where limited resources and expertise block broader implementation. Smaller regions and basins should also be considered for flood susceptibility mapping to improve accuracy and obtain more region-specific results. Thessaly, the focus of this study, is a flood-prone region that has faced severe flood events in recent years, highlighting the importance of flood susceptibility mapping. One of the most recent extreme events was Storm “Daniel” in September 2023, which resulted in devastating damage [17]. Storm Daniel, also known as Cyclone Daniel, took place on 4–7 September 2023. It affected not only Greece but also Bulgaria, Turkey and Libya, causing extensive flooding. A post-event analysis of Storm Daniel in Thessaly [18] revealed that the Peneus river basin received an average rainfall of approximately 360 mm, which triggered extensive flooding. The disaster resulted in the loss of human lives and caused severe environmental and economic damage. The study calculated that, although the station-based return period point estimates exhibited substantial variability, the overall area averaged return period for the 72 h timescale reached up to 150 years, reflecting the remarkable intensity of the event. Given its unprecedented scale and basin-wide impact, Storm Daniel presented a unique opportunity to study flood susceptibility under extreme hydrological conditions.

Table 1. Summary of selected machine learning applications for FSM. The best-performing model from each study is highlighted in bold.

Area of Interest	Algorithms	Evaluation	Data Split Train/Test (%)	Total Points	Resolution (m)	Ref.
Ibaraki, Japan	ANN-MLP SVR GBR Lasso	MAE MSE RMSE R ² AUC/ROC	70/30	224	30 × 30	[11]
Berlin, Germany	CNN ANN RF SVM	AUC/Kappa	80/20	3934	30 × 30 10 × 10 5 × 5 2 × 2	[12]
Idukki, Kerala India	AdaBoost Gradient Boosting XGBoost CatBoost SGB	AUC Precision Recall F ₁ score NP	70/30	1500	30 × 30	[13]
Periyar River, India	LR SVM Naive Bayes RF Ada Boosting Gradient Boosting XGBoost	AUC/ROC	30/70	188	30 × 30	[19]
Fujairah, UAE	xDeepFM DNN SVM RF	recall F ₁ score precision accuracy	75/25	2400	30 × 30	[20]
Salzburg, Austria	MCDA (AHP, ANP) ML (RF , SVM)	AUC/ROC	70/30		30 × 30	[21]
Metlili, Morocco	RF CART SVM XGBoost	AUC	70/30	204	30 × 30	[22]
Karun, Iran Gorganrud, Iran	Deep Forest CFM Multi-gained scanning	AUC/ROC OA KC	27/73	4160 1278	30 × 30	[23]
Wilayat As-Suwayq, Oman	XGBoost RF CatBoost	AUC	70/30	446	5 × 5	[24]
Haraz, Iran	ANN CART FDA GLM GAM BRT MARS MaxEnt	AUC/ROC	–	201	20 × 20	[25]

Notes: ANN: Artificial Neural Network; MLP: Multilayer Perceptron; SVR: Support Vector Regression; GBR: Gradient Boosting Regressor; CNN: Convolutional Neural Network; RF: Random Forest; XGBoost: eXtreme Gradient Boosting; SVM: Support Vector Machine; SGB: Stochastic Gradient Boosting; LR: Logistic Regression; xDeepFM: eXtreme Deep Factorization Machine; DNN: Deep Neural Network; AHP: Analytical Hierarchical Process; ANP: Analytical Network Process; CART: Classification And Regression Trees; CFM: Cascade Forest Model; FDA: Flexible Discriminant Analysis; GLM: Generalized Linear Model; GAM: Generalized Additive Model; BRT: Boosted Regression Trees; MARS: Multivariate Adaptive Regression Splines; MaxEnt: Maximum Entropy.

Table 2. Flood conditioning factors used in each of the selected machine learning applications for FSM.

Ref.	Flood Conditioning Factors
[11]	Elevation, Slope, Aspect, Plane Curvature, Profile Curvature, TWI, SPI, DTStreams, DTRiver, DTRoads, Land Cover
[12]	Elevation, Slope, Aspect, Curvature, TWI, DTRiver, DTRoads, DTDrainage, CN, AP, FP
[13]	Elevation, Slope, Aspect, Curvature, STI, TRI, TWI, SPI, DTRoads, DTStreams, Soil, Geology, Geomorphology, LULC, NDVI, Rainfall
[19]	Elevation, Slope, Aspect, Flow Direction, Drainage Density, SPI, STI, TPI, NDWI, Rainfall
[20]	Elevation, Slope, Curvature, Drainage Density, SPI, TWI, STI, TRI, NDVI, DTDrainage, Rainfall, Land Use, Geology
[21]	Elevation, Slope, Aspect, TWI, SPI, DTRoads, DTDrainage, NDVI, Geology, Rainfall, Land Cover
[22]	Elevation, Slope, Aspect, Plan Curvature, TWI, SPI, DTStreams, DTRoads, Lithology, Rainfall, LULC, NDVI
[23]	Elevation, Slope, Aspect, Curvature, Plan Curvature, Profile Curvature, TPI, TRI, TWI, SPI, Convergence Index, LULC, NDVI, Valley Depth, LS Factor, Flow Accumulation, MCA, HOFD, VOFD, CN, MFI
[24]	Elevation, Slope, Curvature, TRI, TWI, SPI, DTDrainage, Drainage Density, DTRoads, NDVI, Geology, Soil Type, Rainfall
[25]	Elevation, Slope, Curvature, SPI, TWI, River Density, DTRiver, NDVI, Land Cover, Lithology, Rainfall

Notes: TWI: Topographic Wetness Index; SPI: Stream Power Index; STI: Sediment Transport Index; TPI: Topographic Position Index; TRI: Topographic; DTStreams: Distance from Streams; DTRiver: Distance from Rivers; DTRoads: Distance from Roads; DTDrainage: Distance from Drainage; CN: Curve Number; AP: Annual Precipitation; FP: Frequency Precipitation; NDVI: Normalized Difference Vegetation Index; LULC: Land Use Land Cover; NDWI: Normalized Difference Water Index; MCA: Modified Catchment Area; HOFD: Horizontal Overland Flow Distance; VOFD: Vertical Overland Flow Distance; MFI: Modified Fournier Index.

1.2. Objectives and Structure of the Study

This study focuses on performing flood susceptibility mapping on the Peneus River Basin (PRB) with machine learning algorithms, contributing to the advancement of data-driven approaches in flood risk assessment. The first objective is to evaluate and compare the performance of various machine learning models—Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF) and eXtreme Gradient Boosting (XGBoost)—during an extreme flood event, specifically Storm Daniel. Thereafter, the objective is to further investigate the predictive capability and generalizability of the most effective model; the best model is applied under different initial training conditions. Finally, a flood susceptibility map corresponding to a 1000-year return period rainfall scenario is generated.

The contributions of the present paper can be outlined as follows:

- Supporting the development of an early warning system for flood susceptibility through the exploitation of satellite-derived rainfall data.
- Application and comparative analysis of machine learning models (LR, SVM, RF, XGBoost) to produce accurate flood susceptibility maps.
- Calculation of feature importance scores to evaluate the influence of each input variable on model predictions.
- Investigation of the impact of different initial training conditions on model performance.
- Development of an FSM based on a 1000-year return period rainfall scenario at the 24 h scale.
- Provision of recommendations for flood risk management and planning in the case study, based on the results obtained.

The novelty of this study lies in its integration of state-of-the-art machine learning models with region-specific data to evaluate flood susceptibility in the Peneus watershed—

a previously underexplored yet flood-prone area. Unlike previous studies that apply ML models in more generic or well-documented settings, this work addresses a recent extreme event (i.e., Storm Daniel) in a topographically diverse and vulnerable region, characterized by extensive flooding. By focusing on Thessaly's unique combination of lowlands and hydrological complexity, we not only validate the models in a challenging environment but also provide actionable insights tailored to local risk management. By applying these advanced ML techniques to the Peneus watershed, this study contributes to the growing maturity of ML in FSM, addressing gaps in previous research and providing insights that have yet to be explored in similar contexts. The integration of regional data with a real-world extreme event enhances the practical relevance and transferability of the proposed methodology, offering valuable insights to authorities for informed decision making.

The paper is structured as follows: The case study and data used are described in Section 2, while the methodological framework is detailed in Section 3. The results, including the simulation and experimentation outcomes, and comparisons between different models are presented in Section 4, followed by a broader discussion in Section 5. Finally, Section 6 summarizes the key findings and conclusions of this study.

2. Study Area and Data

2.1. Study Area

The selected study region is the Peneus river basin (Figure 1), situated in Thessaly, Greece. The Peneus basin occupies approximately 85% of the Thessaly Water District, covering a total area of 11,062 km². Thessaly's topography is defined by four surrounding mountain ranges: Olympus–Kamvounia to the north, Pindus to the west, Othrys to the south, and Pelion–Ossa to the east. Originating in the Pindus Mountains, the Peneus river traverses the Thessalian Plain and discharges into the Aegean Sea. The region exhibits a transition from a Mediterranean climate to an eastern coastal climate. The climatic conditions vary geographically, with a continental climate prevailing in the central lowlands and a mountainous climate in the western highlands. Annual precipitation is highest in the western highlands and decreases toward the lowlands, before increasing again in the eastern highlands, and it is strongly seasonal [26]. The average annual temperature ranges between 16 and 17 °C. The region holds significant importance due to its extensive agricultural activity, encompassing approximately 450,000 hectares (ha) of agricultural land, of which 250,000 hectares (ha) is irrigated. We chose to apply the aforementioned methodology in the Peneus river basin due to its regional significance and the recurring flood-related challenges it faces, including extreme events [27]. The area has experienced numerous flood events over time [28]. Two of the most recent and significant flood events were caused by Mediterranean cyclone Ianos (18–20 September 2020) [29] and Storm Daniel (4–7 September 2023) [18]. Both events resulted in loss of life and extensive damage to infrastructure and agricultural land. Among them, Storm Daniel was the most intense, causing widespread inundation and catastrophic impacts, which is why we selected it as the main event in this study. After each similar event, the case for preventive measures is revived, yet no substantial intervention follows from the responsible authorities. All the maps (e.g., Figure 1) are generated using the Greek Geodetic Reference System 1987 (GGRS87)/Greek Grid (EPSG:2100), a geodetic coordinate system specifically designed for the Greek mainland. Coordinates represent eastings and northings (X, Y) in meters.

The correspondence between the grid coordinates and their respective latitude (ϕ) and longitude (λ) in degrees, in the World Geodetic System 1984 (WGS84), is provided in Table 3.

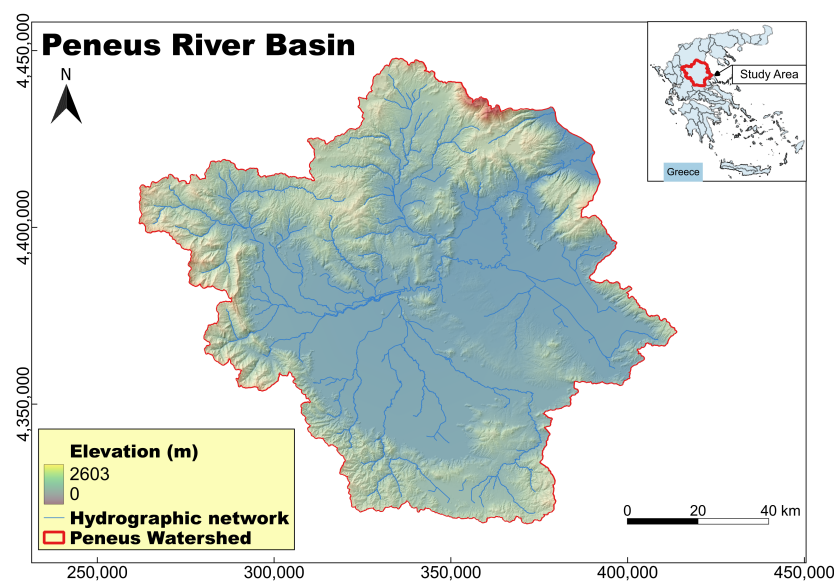


Figure 1. Geomorphological map of the study area—Peneus river basin—along with its main hydrographic network. Coordinate System: GGRS87/Greek Grid (EPSG:2100).

Table 3. Correspondence between coordinates in the Greek Geodetic Reference System 1987 (GGRS87)/Greek Grid (EPSG:2100) and the World Geodetic System 1984 (WGS84) for the extent of the PRB, showing the conversion between eastings and northings (X, Y) in meters and the corresponding latitude (ϕ) and longitude (λ) in degrees.

GGRS87 (m)		WGS84 (°)	
X	262,000	Long.	38.95 N
	414,000		40.18 N
Y	4,315,000	Lat.	21.25 E
	4,448,000		22.99 E

2.2. Flood Inventory Map

Flood maps are fundamental for identifying and extracting samples of flooded and non-flooded areas. The accuracy of this inventory dataset directly influences the potential reliability of the models. In this context, we created a flood inventory map using the most recent significant flood event, i.e., the Storm Daniel flood event in September 2023. Synthetic Aperture Radar (SAR) image data was used to identify the area of the catchment flooded by this event, and more specifically, Sentinel-1 mission data was used (<https://browser.dataspace.copernicus.eu/> (accessed 10 February 2025)). The flood extent was derived by processing satellite images captured before and after the flood event—on 30 August 2023 and 7 September 2023, respectively (<https://rapidmapping.emergency.copernicus.eu/EMSR692/download> (accessed 10 February 2025)). The flood coverage is highlighted in Figure 2. The flooded and non-flooded areas were assigned values of “1” (positive class) and “0” (negative class), respectively. From the derived inventory map of the watershed, 3950 data points were randomly selected to create our dataset for the modeling, as shown in Figure 3. The selection process ensured that the dataset remained balanced, maintaining an equal ratio of flooded and non-flooded points. While this balanced dataset does not reflect the true spatial distribution, where non-flooded pixels vastly outnumber flooded ones, it is a common and necessary approach in classification tasks to mitigate model bias toward the majority class and to improve the model’s ability to learn meaningful patterns for both classes [30]. Direct training on such an imbalanced dataset would likely result in a model that performs poorly on the minority (flooded) class.

This initial dataset was divided into 80% for training and validation purposes, and the remaining 20% for testing the models. Table 4 summarizes the dataset split, showing the number of samples allocated to the training and testing sets, along with the distribution of the instances of each class.

Table 4. Distribution of samples and classes in training and testing splits.

Description	Split Percentage (%)	Number of Samples	Flooded	Non-Flooded
Training	80	3160	1595	1565
Testing	20	790	379	411
Total	100	3950	1974	1976

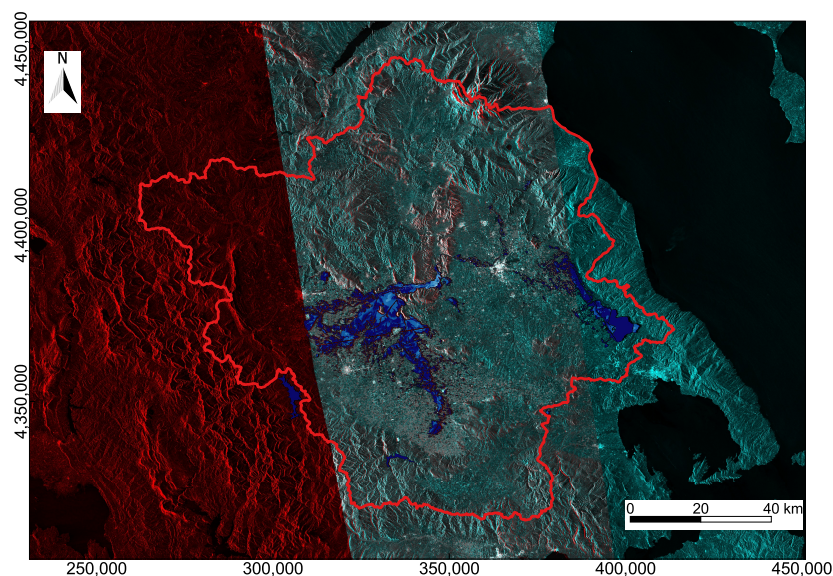


Figure 2. Satellite image showing the area flooded by Storm Daniel. Flood extent is depicted in blue. Left red part indicates no satellite coverage. Source: Sentinel-1. Coordinate System: GGRS87/Greek Grid (EPSG:2100).

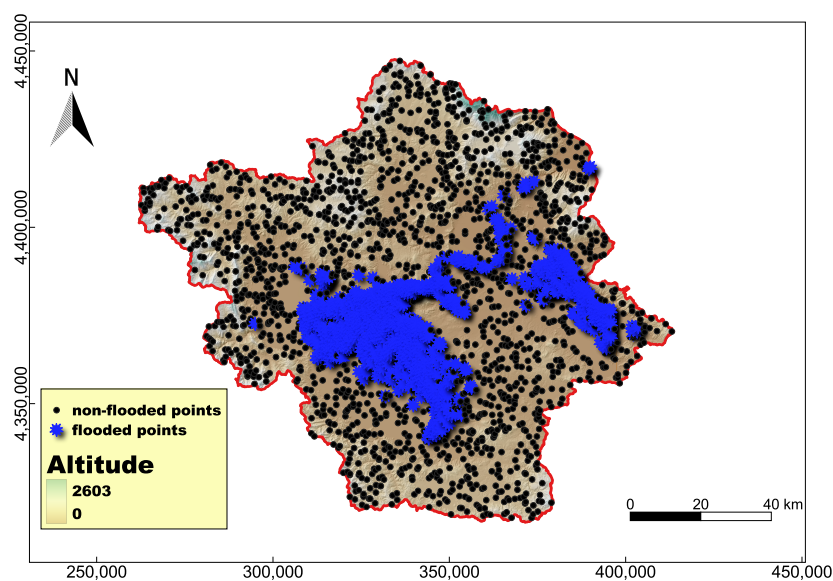


Figure 3. Illustration of flooded and non-flooded samples used in the study. Coordinate System: GGRS87/Greek Grid (EPSG:2100).

2.3. Flood Conditioning Factors

This section outlines and explains each of the factors selected and their relevance to floods. The selection of conditioning variables for flood susceptibility mapping was based on a combination of data availability, the relevance to flood-generating processes, and a brief literature review of previous studies in the field. This approach guarantees that the selected variables are both publicly accessible and scientifically supported. The following analysis, processing of input variables, needed calculations and mapping were conducted using tools available in QGIS 3.34.13 software. We ended up with 13 factors, i.e., elevation, slope, aspect, curvature, distance from roads, distance from rivers, drainage density, topographic wetness index (TWI), stream power index (SPI), curve number (CN), land use/land cover (LULC), normalized difference vegetation index (NDVI) and rainfall. Figure 4 presents the maps of the derived factors, while Table 5 provides a summary of their sources. All the data were converted into a 30×30 m grid (the grid size of the Digital Elevation Model) for analysis if their original format differed.

Elevation, obtained from the digital elevation model (DEM), is one of the most critical factors in flood susceptibility, as lower elevation areas tend to accumulate runoff and are more likely to flood during heavy rainfall events. On the other hand, high altitudes facilitate faster drainage, reducing the likelihood of flood occurrence. The source provides data with a 30×30 m size resolution, collected from [31] (Figure 4a).

Slope affects flood dynamics by controlling the speed and direction of surface runoff. Steeper slopes typically promote faster water flow, reducing the potential for accumulation. Flat flows increase the likelihood of flooding. Slope angle is calculated directly from the DEM and is expressed in degrees (Figure 4b).

Aspect, which refers to the direction of the slope, can indirectly influence flood susceptibility by affecting variables such as frontal precipitation direction and evapotranspiration, and as a result, it also affects vegetation. Aspect is calculated from the DEM and takes values from 0 to 360 that express the slope direction, starting from North (0°) and continuing clockwise (Figure 4c).

Curvature describes the shape of the land surface and influences how water flows across it. Curvature is calculated from the DEM in the QGIS environment. Positive curvature refers to concave areas, zero values to flat areas, and negative values to convex areas. Thus, terrain curvature plays a key role in determining water flow patterns and flood-prone zones (Figure 4d).

Distance from roads (DTRoads) is a significant factor influenced by infrastructure development, as roads often disrupt natural drainage patterns and can contribute to flooding. Euclidean distance was used to determine this feature from the road network (Figure 4e).

Distance from rivers (DTRiver) is crucial, as areas located closer to rivers are more likely to experience flooding. Euclidean distance was also used for this feature (Figure 4f).

Drainage density refers to the total length of streams and rivers per unit area, and it is expressed in km/km^2 (Figure 4g).

The topographic wetness index (TWI) is a widely used hydrological indicator that estimates the spatial distribution of moisture in a catchment. High TWI values indicate areas that are likely to be wetter, such as valleys or flat regions—often prone to flooding. The TWI was calculated using Equation (1) (Figure 4h).

The topographic wetness index (TWI) is a widely used hydrological indicator that estimates the spatial distribution of soil moisture and potential water accumulation in a landscape. It is a quantitative measure of the potential for water accumulation in a

landscape, based on the local slope and upstream contributing area, and can identify areas where water is likely to collect

$$TWI = \ln\left(\frac{\alpha}{\tan(\beta)}\right) \quad (1)$$

where α is the specific catchment area (the flow accumulation at the grid cell multiplied by the grid cell area and then divided by the contour width, in m^2/m) and β is the slope at the corresponding grid cell (in radians).

The stream power index (SPI) is a hydrological factor that quantifies the erosion from flowing water based on both the slope and upstream contributing area, helping us to identify areas at risk of concentrated runoff and flooding. It is calculated with Equation (2) (Figure 4i).

$$SPI = \alpha \tan(\beta) \quad (2)$$

where α and β are as previously defined.

The curve number (CN) is an empirical hydrologic parameter for predicting direct runoff from rainfall excess. It was developed by the USDA Natural Resources Conservation Service, with the method of the Soil Conservation Service (SCS) [32]. It reflects how easily rainfall turns into surface runoff in a particular area (Figure 4j).

A land use/land cover (LULC) map represents the physical and human-defined characteristics of the watershed surface, which significantly influence flood behavior. Five classes were identified, taking the first level of the given subdivision, i.e., Artificial Surfaces, Agricultural Areas, Forest/Semi Natural Areas, Wetlands and Water Bodies (Figure 4k).

The normalized difference vegetation index (NDVI) is a satellite-derived index that measures the density and health of vegetation by comparing the difference between near-infrared (NRI) light (which vegetation strongly reflects) and red light (which vegetation absorbs). Images from the Sentinel 2 mission were used, and Equation (3) was applied. Four classes were created (Figure 4l) depending on their values; that is,

- −1–0: dead plant;
- 0–0.33: diseased plant;
- 0.33–0.66: moderate healthy plant;
- 0.66–1: very healthy plant.

$$NDVI = \frac{NRI - Red}{NRI + Red} \quad (3)$$

Rainfall is one of the most critical factors in floods, as it directly contributes to the volume and intensity of the surface runoff. In this study, we used the rainfall observed in the IMERG (Integrated Multi-satellite Retrievals for GPM) obtained from Giovanni. This is a widely used dataset that combines multiple satellite observations to estimate precipitation at fine spatial and temporal scales. Since the flooded areas used in this study arose from the extreme Storm Daniel event, it was essential to incorporate the corresponding rainfall that caused these floods. Therefore, we obtained the rainfall data at 30-min intervals for the period of 4–7 September 2020, from IMERG, and aggregated it to compute the total rainfall of the event. The maximum rainfall observed by the satellite inside the river basin was 250 mm (Figure 4m).

Table 5. Data sources of the selected flood conditioning factors.

Data	Source
DEM	From Copernicus GLO 30 https://dataspace.copernicus.eu/ (accessed on 15 December 2024) [31]
Slope	Calculated from DEM

Table 5. Cont.

Data	Source
Aspect	Calculated from DEM
Curvature	Calculated from DEM
DTRoads	From OSM https://www.openstreetmap.org/ (accessed on 10 December 2024)
DTRiver	From GeoData http://geodata.gov.gr/ (accessed on 8 January 2025)
Drainage Density	From hydrographic network—GeoData http://geodata.gov.gr/ (accessed on 8 January 2025)
TWI	Calculated from DEM
SPI	Calculated from DEM
CN	Calculated from slope, soil, land use [33]
LULC	From National Cadastre CORINE 2018 https://ktimatologio.gr/ (accessed on 5 January 2025)
NDVI	From Copernicus Sentinel 2 https://browser.dataspace.copernicus.eu/ (accessed on 5 February 2025)
Rainfall	From Giovanni https://giovanni.gsfc.nasa.gov/giovanni/ (accessed on 7 March 2025)
Flooded Area	From Copernicus Sentinel 1 https://browser.dataspace.copernicus.eu/ (accessed 10 February 2025)

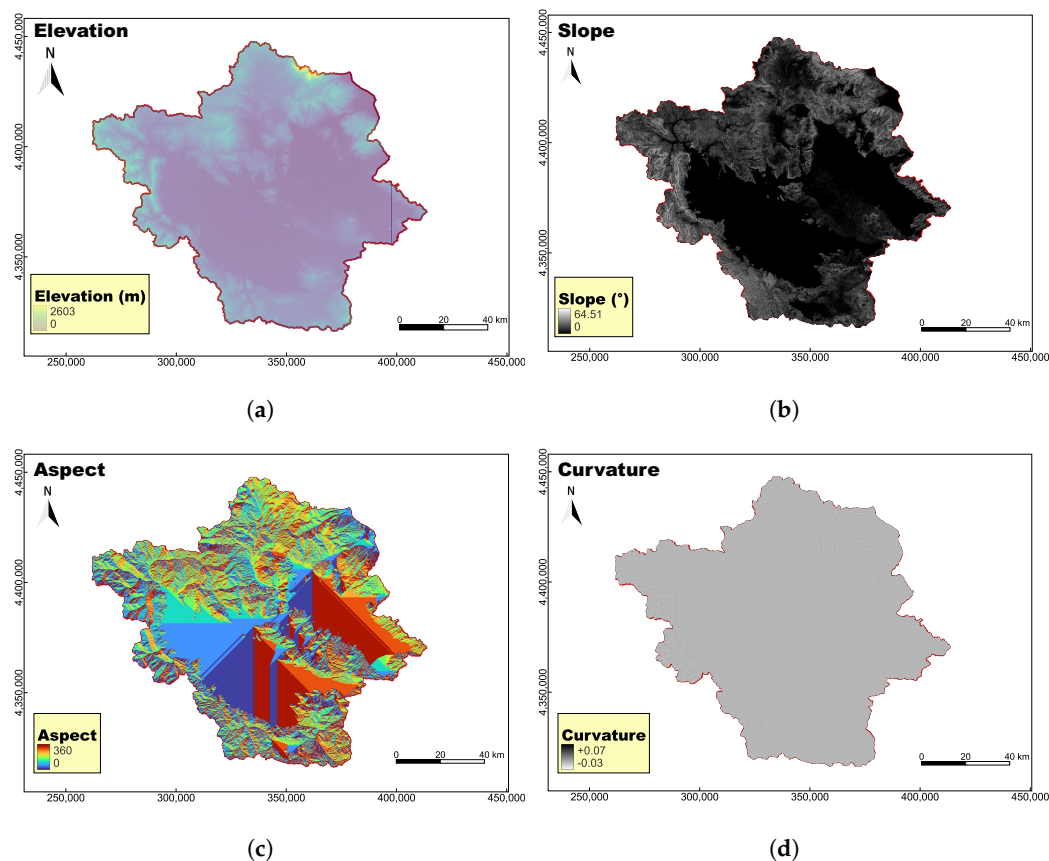


Figure 4. Cont.

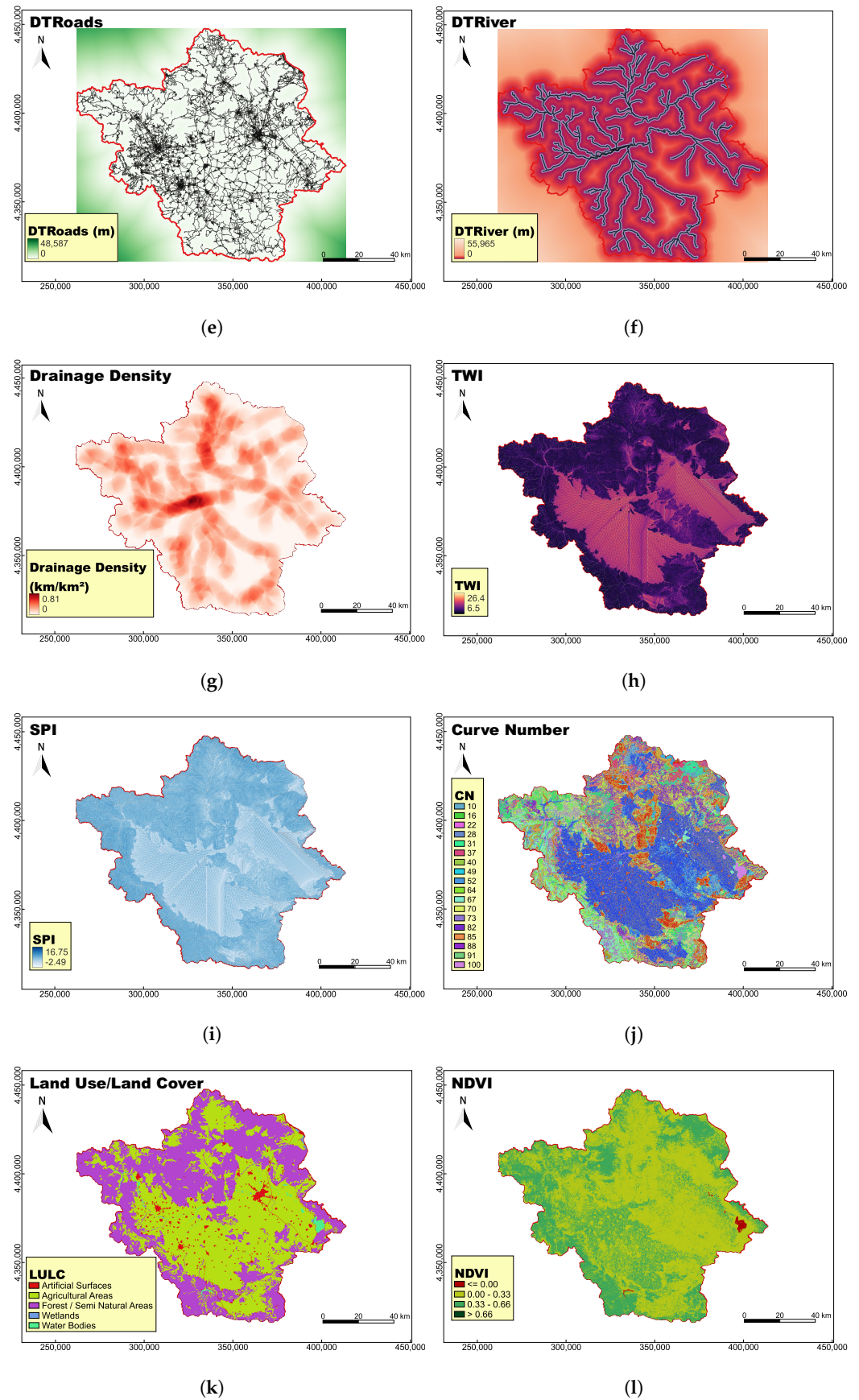
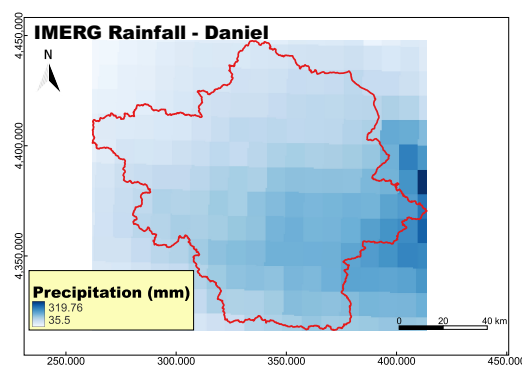


Figure 4. Cont.



(m)

Figure 4. Flood conditioning factors used. (a) Elevation. (b) Slope. (c) Aspect. (d) Curvature. (e) DTRoads. (f) DTRiver. (g) Drainage Density. (h) SPI. (i) TWI. (j) CN. (k) LULC. (l) NDVI. (m) Total rainfall during Storm Daniel. Coordinate System: GGRS87/Greek Grid (EPSG:2100).

3. Materials and Methods

There are several ways to develop a model for flood susceptibility analysis. One approach treats it as a regression problem, predicting water depth as a continuous variable. Another frames it as a classification task, distinguishing between flooded and non-flooded areas. In this study, we adopt a classification approach, where the objective is to predict the likelihood of a location being classified as either flooded (1) or non-flooded (0), based on a set of input variables. Those explanatory variables are some factors, also referred to as flood conditioning factors, which significantly influence the probability of flooding. The selected factors are described in Section 2.3. In general, we create a dataset comprising some independent variables (flood conditioning factors) and a single dependent variable (classified as flood-prone or not flood-prone). This setup formulates a binary classification task suitable for machine learning applications. The overview of the methodology is depicted in Figure 5. Four well-established machine learning models are chosen to support this approach (i.e., Logistic Regression, Support Vector Machine, Random Forest and Extreme Gradient Boosting). Their underlying mechanisms and limitations are discussed in the following sections.

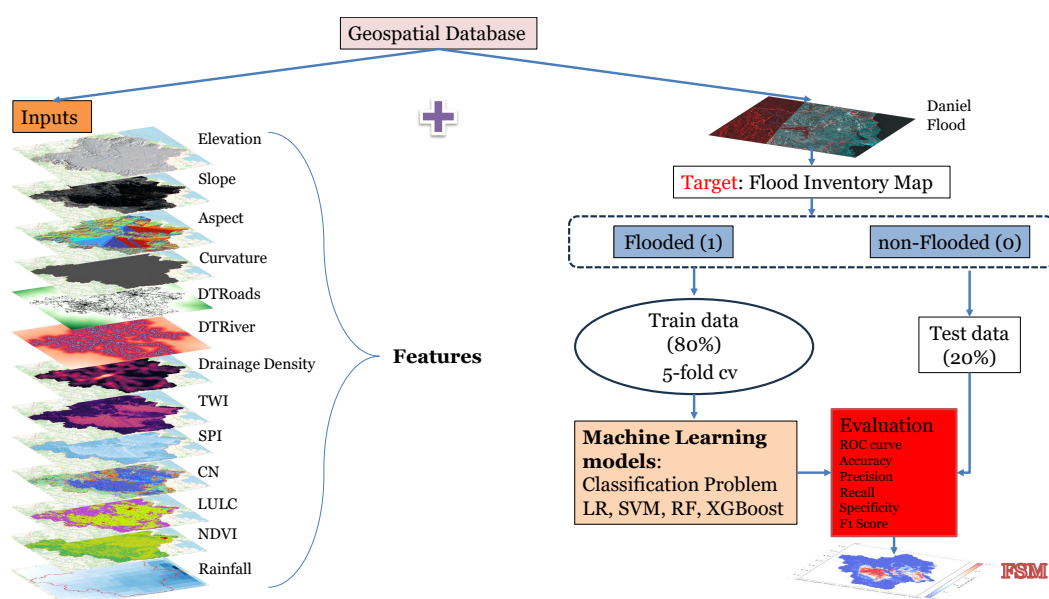


Figure 5. Flowchart of the methodology used for Flood Susceptibility Mapping.

3.1. Logistic Regression (LR)

Logistic Regression (LR) [34] is a supervised machine learning algorithm commonly used to predict the probability of a categorical outcome, particularly in binary classification tasks, where the goal is to distinguish between two classes. The key characteristic of logistic regression is that it models the relationship between the input features and the output probability using an S-shaped curve (Figure 6), known as the logistic or sigmoid function (Equation (4)). This ensures that the predicted values are always restricted between zero and one, allowing them to be interpreted as probabilities.

$$f(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

$$z = \mathbf{w}^T \mathbf{x} + w_0 \quad (5)$$

where z is a linear combination of the input features and their corresponding weights; \mathbf{x} is the feature vector; \mathbf{w} is the weight vector; and w_0 is the bias term. The weights are determined during the training process with the maximum-likelihood estimation algorithm [35]. LR is simple, interpretable and efficient for binary classification tasks, especially when the relationship between the input and output is approximately linear. However, it struggles with more complex relationships and non-linear decision boundaries.

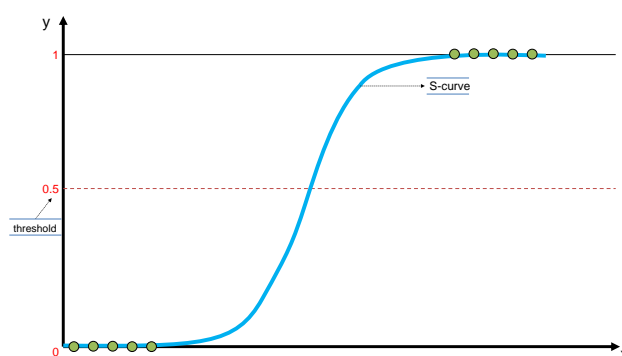


Figure 6. Logistic Regression model.

3.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) [36] is a supervised machine learning algorithm commonly used for binary classification, though it can also be applied to regression tasks. SVM's goal is to find the optimal decision boundary or hyperplane (for >3 features) that separates the data into two or more classes, while maximizing the margin between them. An example of the model's key characteristics is depicted in Figure 7.

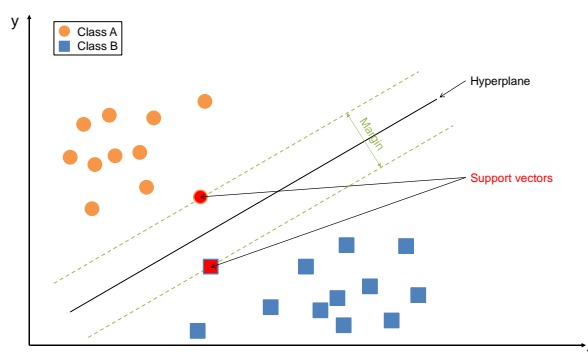


Figure 7. Support Vector Machine model.

Mathematically, the decision boundary (hyperplane) in linear SVM can be described by Equation (6):

$$w^T x + w_0 = 0 \quad (6)$$

where x is the input vector, w is the weight vector and w_0 is the bias term. Parameters are calculated through solving an optimization problem that aims to find the hyperplane which maximizes the margin. When data points are not linearly separable in their original space, SVM uses the kernel trick [37] to map them into a higher-dimensional space where a linear separation is possible. Common kernels include the linear, polynomial, radial basis function (RBF), and sigmoid kernels, which measure similarity between data points to enable effective separation. SVM is effective in high-dimensional spaces and remains memory-efficient by relying only on a subset of training points (support vectors) to determine the decision boundary, but careful normalization is required. At the same time, it is computationally intensive and slow to train on large datasets, which can negatively impact scalability and performance.

3.3. Random Forest (RF)

Random Forest (RF) [38] is an ensemble learning algorithm widely used for classification and regression tasks. It operates by constructing a large number of decision trees [39] during training and outputting the class that is the majority vote of the individual trees. Each tree is trained on a random subset of the data (using a bootstrapping sampling), and at each split, it considers a random subset of features, which helps reduce overfitting and improves generalization. This randomness ensures that the trees are independent, making the overall model more robust than a single decision tree. The final prediction for a classification problem is typically the class with the most votes among all trees. Figure 8 summarizes the architecture of the RF model. Random Forests are particularly effective in handling high-dimensional data, capturing non-linear relationships, and working well with both numerical and categorical features, in tabular form. However, they can be computationally expensive for large datasets and are less interpretable than simpler models like logistic regression. Despite this, their accuracy, resistance to overfitting and minimal need for parameter tuning make them a powerful tool in prediction assessment.

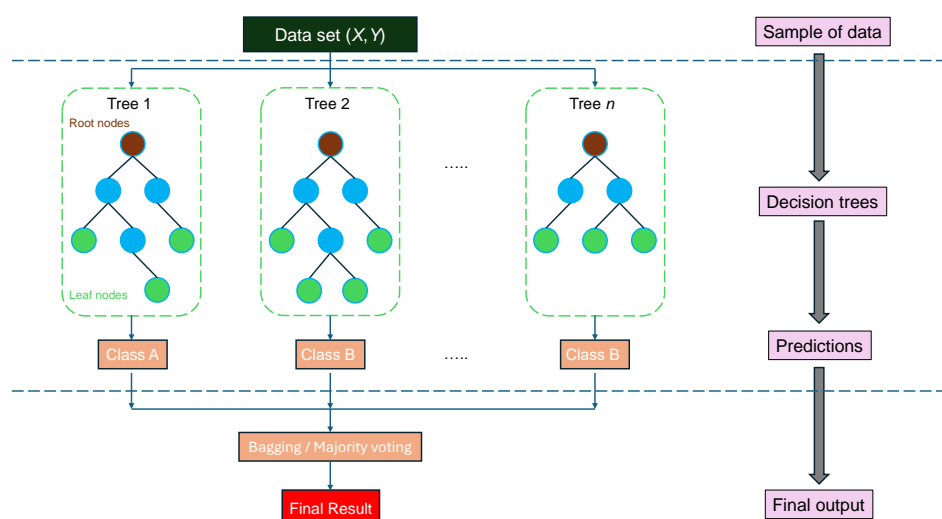


Figure 8. The architecture of the Random Forest algorithm.

3.4. Extreme Gradient Boosting (XGBoost)

XGBoost (eXtreme Gradient Boosting) [40] is a high-performance machine learning algorithm built upon the gradient boosting framework and has gained widespread

recognition due to its success in numerous Kaggle competitions [41]. It operates as an ensemble method, meaning it combines multiple base learners—typically decision trees [39]—through a boosting process. In this process, models are trained sequentially, with each one aiming to reduce the errors made by its predecessors. This iterative optimization leads to continuous performance improvement by minimizing a defined objective function. Boosting techniques like XGBoost are particularly effective at balancing bias and variance. In the case of XGBoost, each decision tree assigns a numerical weight to its leaf nodes, and samples are directed to these leaves based on their feature values, with the leaf's weight representing the prediction. The term “gradient” in gradient boosting refers to the use of gradient descent to minimize errors as more models are added. XGBoost distinguishes itself from other boosting algorithms through the inclusion of regularization within its objective function to prevent overfitting and its capability to build trees in parallel, greatly enhancing computational speed. These features contribute to its reputation as a highly efficient, flexible and scalable algorithm. The architecture and training process of XGBoost are illustrated in Figure 9.

Following [40], the objective function that needs to be minimized at the t -th iteration is

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (7)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (8)$$

where y_i is the target value, \hat{y}_i is the predicted value, l is a differentiable loss function and $f_t(x_i)$ is the output from the new model at iteration t , with n being the number of training samples. The first component of Equation (7) quantifies the model's performance on training data, while the second component, Ω , acts as a regularization term that controls the complexity of the individual learners (i.e., trees), to prevent overfitting. In Equation (8), T refers to the number of leaves in the tree, with leaf weights denoted by w , and γ and λ being regularization parameters. The task of optimizing this objective function can be simplified to minimizing a quadratic function, through the application of the second-order Taylor expansion [42].

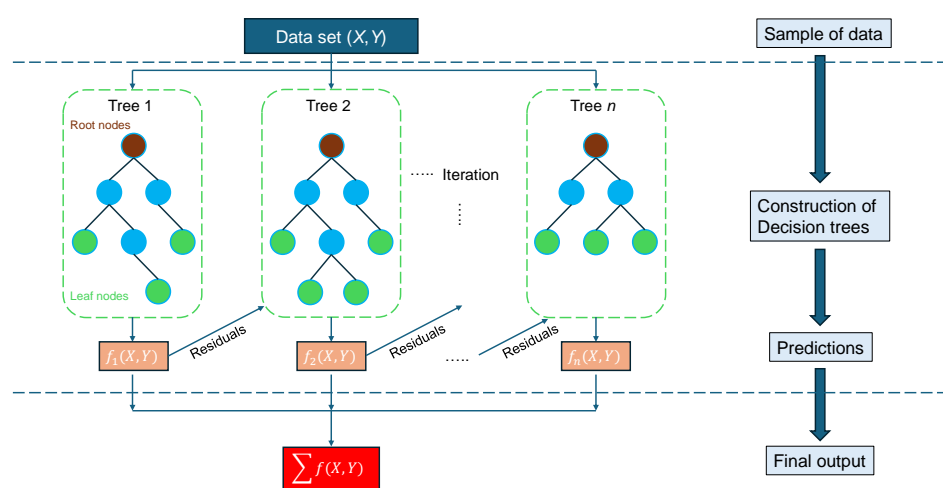


Figure 9. The architecture of the XGBoost algorithm.

The main difference between RF and XGBoost lies in how they build and combine decision trees. RF, a bagging-based method, creates multiple trees and computes their predictions independently and in parallel. In contrast, XGBoost is a sequential model, where each tree is influenced by the predictions of the previous one. XGBoost requires

careful hyperparameter tuning to perform optimally, and this fine-tuning can be challenging and time-consuming, especially when working with large datasets. Additionally, XGBoost is designed to handle only numerical data. Its more complex structure also makes it less interpretable compared to simpler algorithms.

3.5. Feature Importance

With the expanding use of machine learning in decision making, it is becoming ever more important to ensure transparency and to understand the factors contributing to a model's predictions [43]. Feature Importance (FI) in ML refers to techniques used to quantify the contribution of each input feature to the model's prediction. Understanding which features are most influential helps to improve model transparency, supports better feature selection, and can reveal insights into the underlying data. There are several ways to calculate feature importance, such as [44]:

1. Coefficient-based feature importance;
2. Permutation-based feature importance;
3. Tree-based feature importance;
4. SHapley Additive exPlanations (SHAP).

Coefficient-based FI is used in linear models (like linear or logistic regression), where the magnitude of the model coefficients indicates the relative importance of each feature. Permutation-based FI measures the impact of each feature on model performance by randomly shuffling its values and observing the decrease in metrics. In a tree-based model (like RF, XGBoost, etc.), FI is typically calculated based on how much each feature reduces impurity (e.g., Gini impurity or entropy) [45]. SHAP [46] is a model-agnostic method based on cooperative game theory. It provides a clear way to see how much each feature contributes to a model's prediction by looking at how the prediction changes, when we include or exclude that feature in different combinations. This makes SHAP a valuable tool for understanding model behavior. SHAP feature importance serves as an alternative to permutation-based feature importance. While permutation importance evaluates the impact of a feature by measuring the decrease in model performance when the feature values are randomly shuffled, SHAP quantifies importance based on the magnitude of feature attributions derived from their contributions to individual predictions. By identifying key drivers of a model's output, feature importance aids in building more interpretable and efficient models.

3.6. Evaluation Metrics

Selecting the appropriate evaluation metrics is essential during the model assessment phase, as they offer a clear understanding of the model's effectiveness in practical applications. In binary classification tasks, which is our case, the choice of metrics significantly impacts how performance is interpreted [47]. The most popular and used metric to evaluate the performance of classification models is accuracy. Accuracy simply measures the overall proportion of correct predictions with respect to the total number of data. While accuracy gives an overall idea of a model's performance, it can be a misleading metric in some cases like imbalanced datasets. To overcome the drawbacks, we usually use more metrics like precision, recall, specificity and F_1 score, which provide a more detailed evaluation, especially when dealing with imbalanced data or situations where different types of errors have different costs. Precision indicates the proportion of true positives among predicted positives, and recall or sensitivity indicates the proportion of true positives among actual positives. Specificity provides the number of negative records correctly predicted. The F_1 score combines precision and recall into one metric, by calculating the harmonic mean.

Additionally, Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) are widely used for binary classification algorithms to evaluate a model's ability to rank positive instances higher than negatives across various threshold settings. The ROC curve is produced by calculating and plotting the True Positive Rate (TPR) on the y-axis, against the False Positive Rate (FPR) on the x-axis, for a classifier at a variety of thresholds. A good model will have a large ROC-AUC, while a poor model will be positioned near the diagonal $x = y$ line, which represents random performance. The ROC-AUC metric is also very useful for comparing different models against each other. Selecting appropriate metrics depends on the specific goals of the task, particularly in cases of class imbalance where accuracy alone can be misleading.

Equations (9)–(13) present the formulas for the aforementioned metrics.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (9)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (12)$$

$$F_1\text{score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

$$\text{TPR} = \text{Sensitivity} = \text{Recall} \quad (14)$$

$$\text{FPR} = 1 - \text{Specificity} = 1 - \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (15)$$

where TP, TN, FP, and FN stand for True Positives, True Negatives, False Positives, and False Negatives, respectively. To elaborate, each term means the following:

- *True Positives*: The number of instances where the model correctly predicted the positive class;
- *True Negatives*: The number of instances where the model correctly predicted the negative class;
- *False Positives*: The number of instances where the model incorrectly predicted the negative class for a positive case;
- *False Negatives*: The number of instances where the model incorrectly predicted the positive class for a negative case.

4. Analysis and Results

4.1. Problem and Model Setup

As stated in Section 2.2, the overall dataset that is used for the training and testing of the algorithms contains 3950 grid cell points, equally divided between flooded and non-flooded instances. Each point includes the 13 independent variables that affect the flood along with a dependent variable indicating whether the location is flooded or non-flooded. We then split the data into 80% (i.e., 3160) for the training purposes and the remaining 20% (i.e., 790) for testing the models, ensuring that the ratio of flooded and non-flooded samples remained consistent across both subsets.

In order to explore and visualize the linear relationships among the flood conditioning variables in the dataset, a Pearson correlation matrix was computed. The resulting matrix, visualized through a heatmap in Figure 10, reveals both the strength and direction of pairwise correlations. The values range from -1 , indicating a perfect negative

linear relationship, to +1, indicating a perfect positive linear relationship. Elevation and slope exhibited the strongest positive correlation (+0.7), while slope and TWI showed the strongest negative correlation (−0.7). Due to the fact that the highest absolute correlation coefficient is 0.7, which is below the common multicollinearity threshold of 0.8 [48], we proceeded to include all features in the training set. We also experimented with removing the variables showing the highest correlations, but this had no noticeable impact on the models' performance, so all the features were retained.

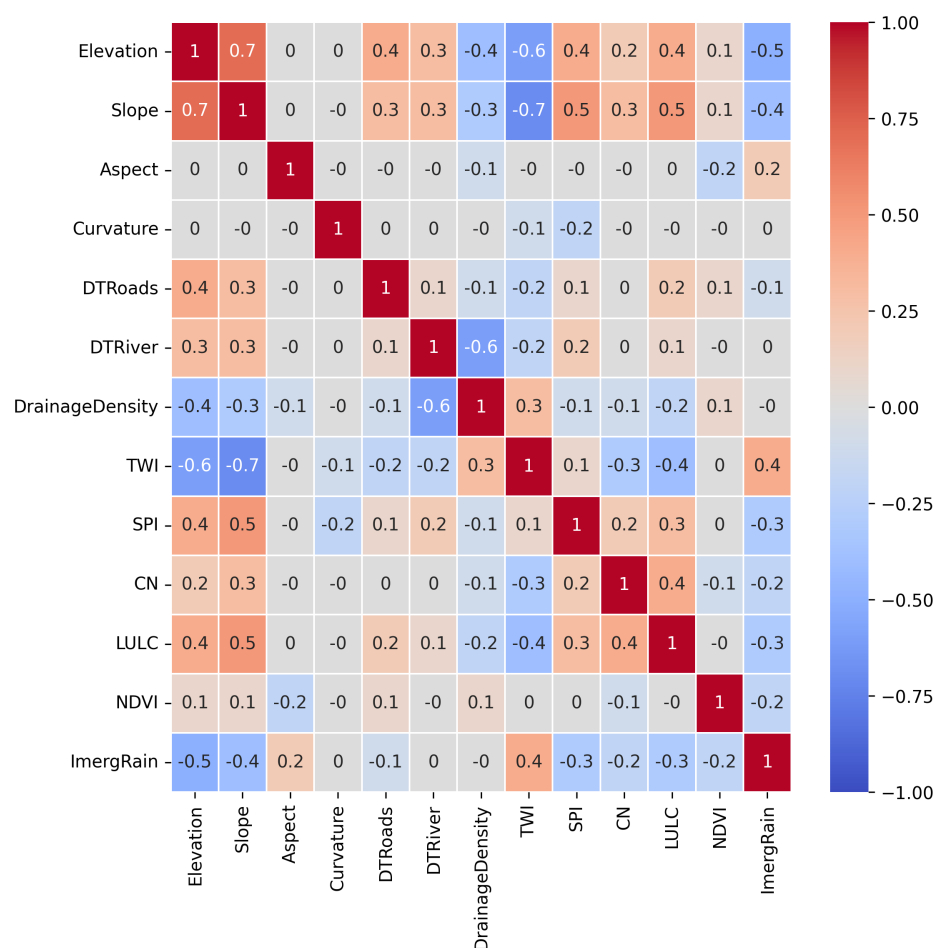


Figure 10. Pearson correlation matrix of all input (independent) variables.

To ensure robust model performance, cross-validation was employed during the evaluation process. Each model used in this study was trained using five-fold cross-validation to mitigate overfitting, and accuracy was used as the scoring metric. In this approach, the dataset is divided into five equal parts; in each iteration, four parts are used for training and the remaining one for validation, rotating until every subset has been used for validation once. Following hyperparameter optimization using Grid Search, the finalized parameter values of the models are as presented in Table 6. All the analyses and code implementation were carried out using Python 3.11.12 with relevant scientific libraries, e.g., the XGBoost Python package 2.1.4 and scikit-learn version 1.6.1. Model training was performed on Google Colab, utilizing an Intel Xeon CPU with 2 vCPUs (virtual CPUs) and 13 GiB of RAM, to provide a suitable computational environment for the machine learning tasks.

Table 6. Values derived from hyperparameter optimization task for the Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF) and eXtreme Gradient Boosting (XGBoost).

Model	Hyperparameter Values
LR	max iter = 1000; scoring = ‘accuracy’; multi class = ‘auto’; solver = ‘lbfgs’
SVM	C = 2; gamma = ‘scale’; kernel = ‘radial basis function (rbf)’; scaler = ‘StandardScaler’
RF	no. estimators = 200; max depth = 20; min samples split = 2; min samples leaf = 1; max features = ‘sqrt’
XGBoost	no. estimators = 100; max depth = 7; learning rate = 0.1; subsample = 1.0; colsample by tree = 0.8

4.2. Results and Comparison

With the models being trained, we evaluate their performance on the testing set. A comparison of confusion matrices for the different models is demonstrated in Figure 11. The format of the illustrated matrix is as follows:

$$\begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix}$$

Both Logistic Regression (LR) and Support Vector Machine (SVM) yield a high value of classification errors (i.e., 121 and 107), which means that they predict that the point is either non-flooded or flooded when it is actually not. Almost all of them are false positive points, meaning that the model incorrectly flagged them as positive cases (flood) despite them actually belonging to the negative class (non-flood). Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) minimize the false positive and negative values, misclassifying only 55 and 58 samples, respectively, out of 790 total testing samples (cells).

Table 7 summarizes the values of the five metrics, i.e., accuracy, precision, recall, specificity and F₁score. For each evaluation metric, the highest scores are emphasized in bold. As shown in the table, the tree-based models (i.e., Random Forest and Extreme Gradient Boosting) demonstrate superior performance in general, achieving consistently high results across all metrics. Both Random Forest (RF) and XGBoost achieved the highest accuracy of 0.93, reflecting their superior predictive capability. These two models also led in precision (0.89) and F₁score (0.93), indicating a good balance between precision and recall. While Logistic Regression (LR) achieved a perfect recall of 1.00, it did so at the cost of lower specificity (0.71) and precision (0.76), suggesting more false positives. Support Vector Machine (SVM) showed moderate improvements over LR but was still outperformed by RF and XGBoost across most metrics. Notably, XGBoost slightly surpassed RF in specificity (0.90 compared to 0.89), emphasizing its effectiveness in correctly identifying non-flood-prone areas.

Furthermore, Figure 12 presents the Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC). The ROC curve shows the trade-off between sensitivity (or TPR) and 1 – specificity (or FPR). The dotted lines represent the ideal classifier (top-left curve) and random classifier (the diagonal where FPR = TPR). Curves that approach the perfect classifier indicate better performance, whereas on the contrary, those closer to the random line suggest weaker or poor performance. Aside from visual inspection, which can be critical in some cases, a measure that summarizes the performance of each classifier is needed in order to compare different models. The AUC score measures the area under the ROC curve and gives a general predictive score. From Figure 12, it can be observed that LR performs worse than the other models (90.7%). SVM performs slightly better (94%), while the tree-based models (RF and XGBoost) show the best performance with similar ROC curves—occasionally, one slightly surpasses the other. Quantitatively, RF achieves the best AUC score of 96.9%, slightly outperforming XGBoost (AUC = 96.8%). This improved performance can be attributed to their ability to capture

complex, non-linear relationships and interactions within the data, which simpler models may fail to detect.

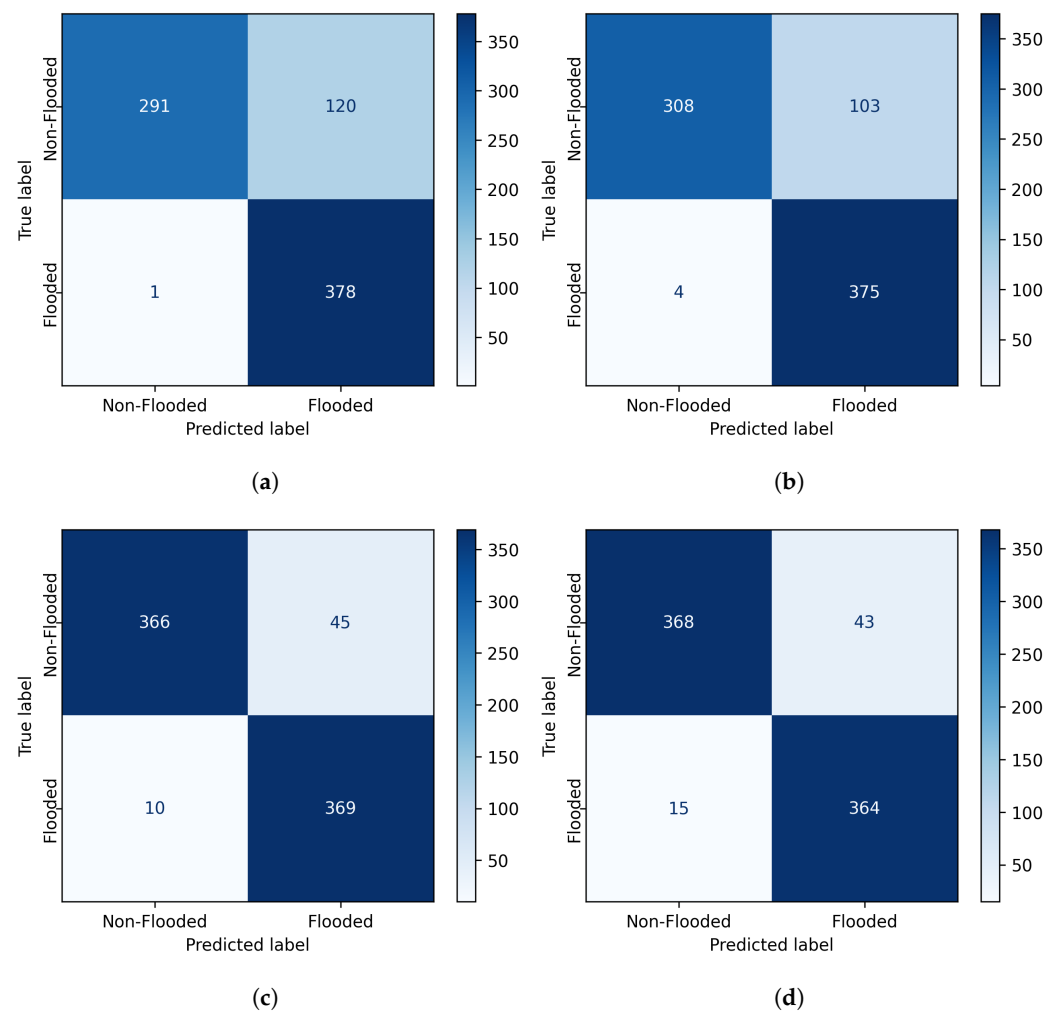


Figure 11. Confusion matrices showing model performance on the test dataset. (a) LR. (b) SVM. (c) RF. (d) XGBoost.

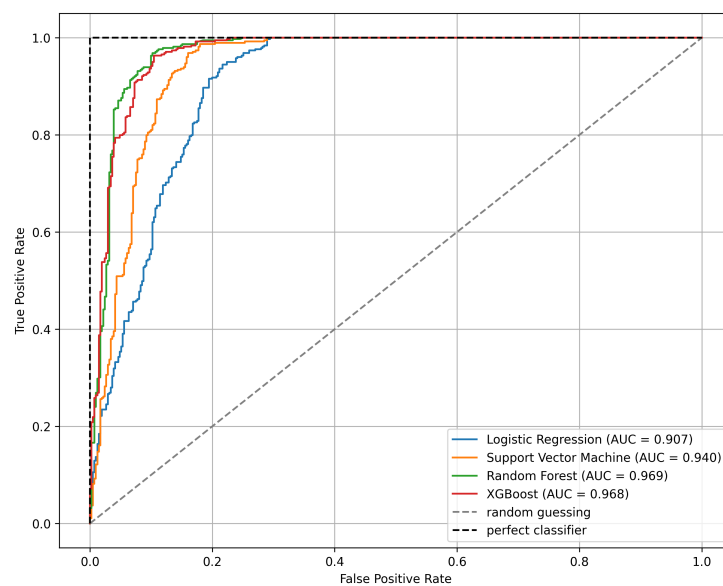


Figure 12. Comparison analysis of models' performance through ROC curve and AUC score, evaluated on test dataset.

Table 7. Models' results for each evaluation metric on the test set. The best-performing model for each metric is highlighted in bold.

Model	Accuracy	Precision	Recall/Sensitivity	Specificity	F ₁ Score
LR	0.85	0.76	1.00	0.71	0.86
SVM	0.86	0.78	0.99	0.75	0.88
RF	0.93	0.89	0.97	0.89	0.93
XGBoost	0.93	0.89	0.96	0.90	0.93

4.3. FSM Maps

It is important to have a flood susceptibility map for the complete watershed, when an extreme event like Storm Daniel is taking place. Hence, Figure 13 depicts the flood susceptibility maps generated by each of the models applied, indicating the probability of flooding for each cell. Higher values (close to one) indicate greater susceptibility to flooding, while lower values (close to zero) indicate lower susceptibility. LR and SVM predict significantly larger flood-prone areas compared to RF and XGBoost, which identify more limited regions as susceptible to flooding. According to tree-based models, which demonstrated superior performance, the most flood-prone areas are located in the eastern region and the central-to-southern region, as highlighted in red.

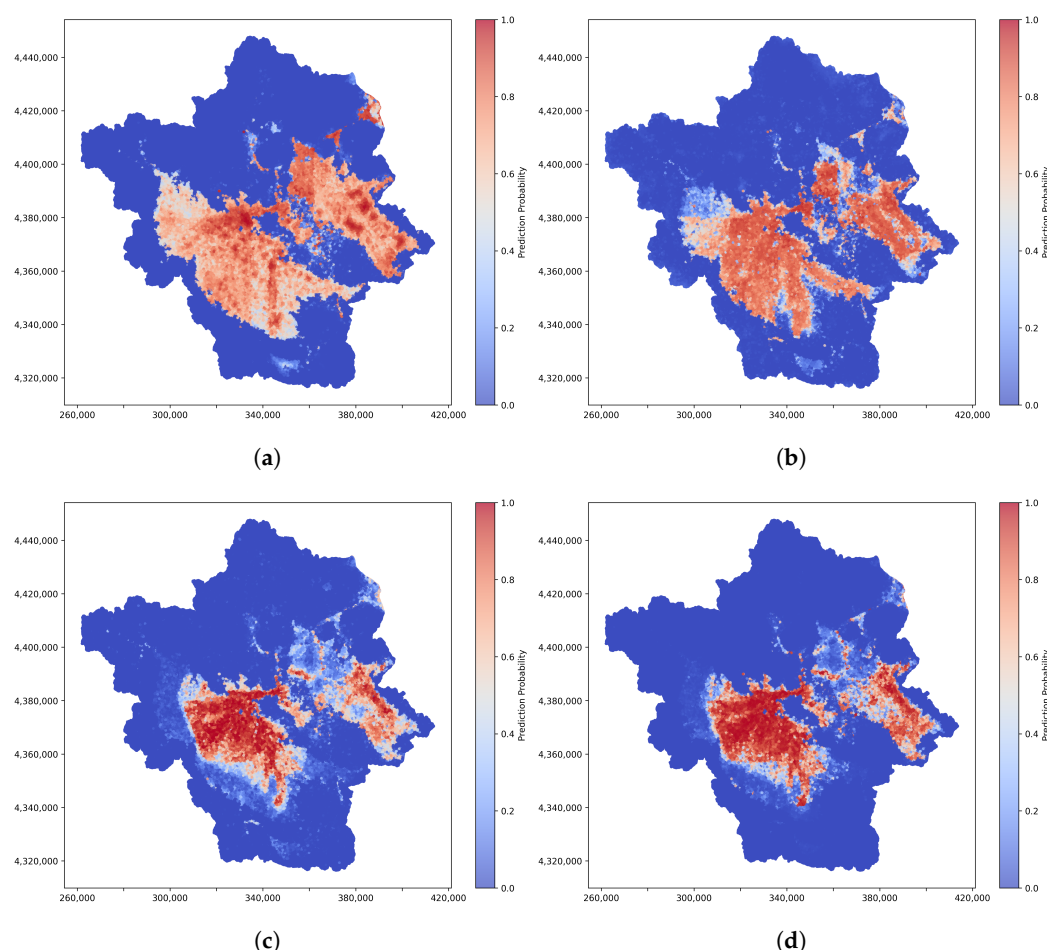


Figure 13. Flood susceptibility maps of Peneus river basin for each of the trained models. (a) LR. (b) SVM. (c) RF. (d) XGBoost.

4.4. Feature Importance

As mentioned in Section 3.5, feature importance allows us to understand the relationship between the features and the target output. In that manner, which feature is the

most effective for the model is distinguished. As Random Forest performed the best in the evaluation assessment, with respect to the ROC curve and AUC score, it was selected as the primary model for further analysis and investigation. To measure how each feature influences the output in RF, we used Mean Decrease in Impurity (MDI) (Figure 14a) and a permutation-based method that involved measuring the Mean Decrease in Accuracy (MDA) (Figure 14b). MDI evaluates the importance of a feature by summing the total decrease in impurity caused by that feature over all trees in the model. MDA, in contrast, quantifies the drop in predictive accuracy when a feature's values are randomly permuted, providing insight into how much the model depends on that variable. Both methods suggest that the most influential factors in the RF model are elevation, slope, rainfall and TWI. The elevation, slope and TWI variables are critical as they directly influence the topographic characteristics of the landscape, which are crucial in flood modeling. Additionally, rainfall provides important environmental context, representing precipitation patterns.

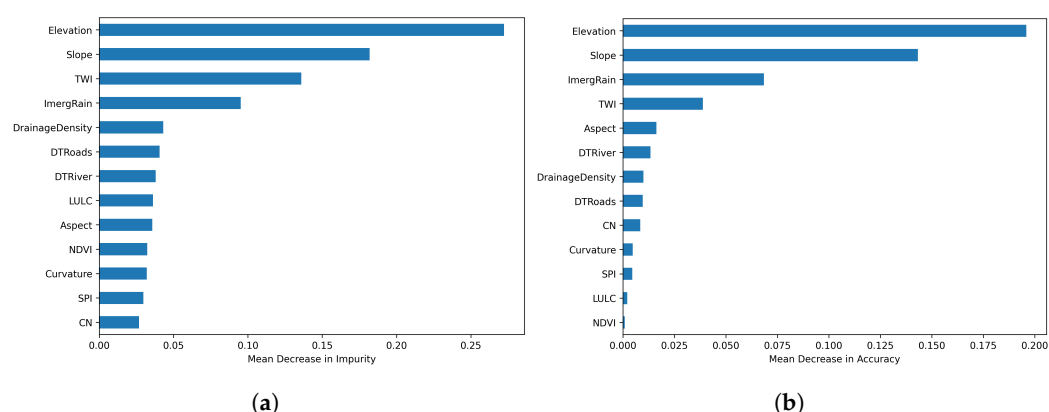


Figure 14. Feature importance for the RF model across different methods. (a) Mean Decrease Impurity. (b) Mean Decrease in Accuracy.

4.5. Initial Data Experimentation

With the intention of verifying the generalizability of the model and evaluating the potential impact of the initial training data selection, we conducted two additional experiments. In each experiment, the Peneus watershed was divided into two halves. For the first experiment, 3950 training data points were sampled exclusively from the western half, while in the second, the same number of points was selected from the eastern half. After training the RF model on both datasets, two trained models were developed. FSMs for the entire basin were generated, and the results are presented in Figure 15. As we can observe, the training data have some influence on the resulting FSM. To quantify and compare the results, we produced the ROC and AUPRC (Area Under Precision–Recall) curve in Figure 16 for the three models, i.e., trained with points from the whole basin, trained with points from the right-half basin and trained from the left-half basin. Not unexpectedly, the model trained using data points from the entire basin—thus capturing the spatial variability across the full study area—exhibited better performance compared to models trained on spatially limited subsets. Nevertheless, the accuracy of the remaining models remained at a high level.

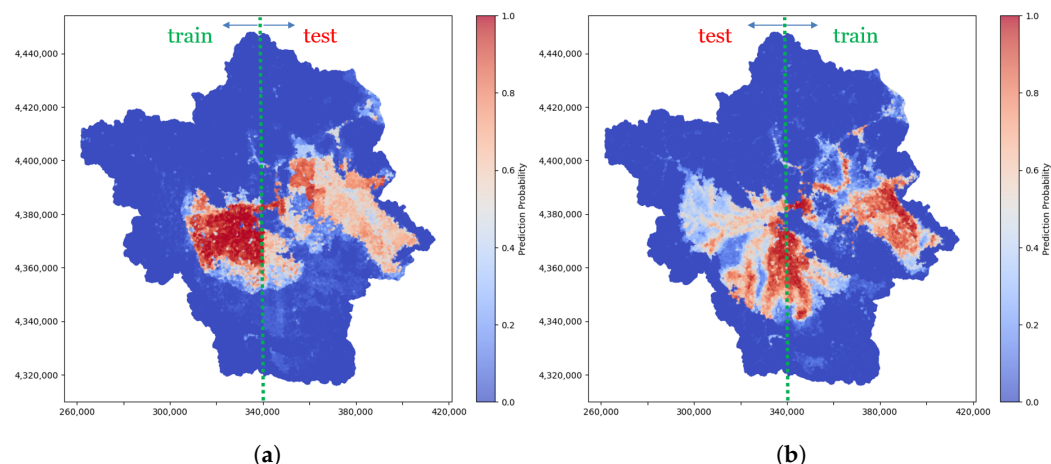


Figure 15. Flood susceptibility maps of Peneus river basin with Random Forest for different initial training data. (a) Training with left-half catchment. (b) Training with right-half catchment.

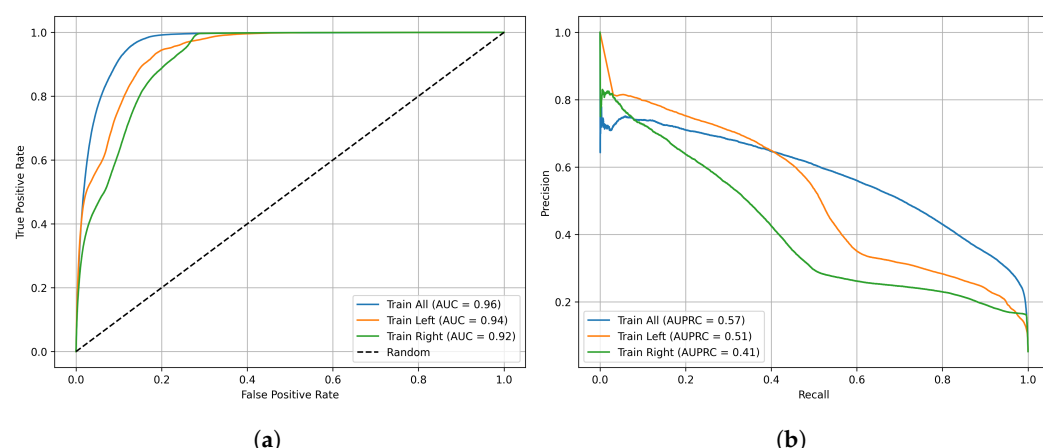


Figure 16. Comparison analysis of Random Forest models trained on different initial training data, through ROC and AUPRC curve (Train All: trained using data from the entire basin; Train Left: trained using data from the left side of the basin; Train Right: trained using data from the right side of the basin). (a) ROC curves. (b) AUPRC curves.

4.6. FSM for $T = 1000$ -Year Rainfall Scenario

Flood susceptibility maps are essential for identifying high-risk areas and guiding flood mitigation and land-use planning efforts in the area. Inference in machine learning refers to the process of applying a trained model to new data in order to generate predictions. Under these circumstances, we constructed an FSM referring to a 1000-year return period rainfall scenario at a 24 h scale. The new methodological stochastic framework for the construction of rainfall intensity–time scale–return period relationships [49] (referred as ombrian curves), was adopted. This approach was applied recently within Greek territory [50,51]. The relationship between the rainfall intensity x for any timescale k and return period T is given by Equation (16):

$$x = \lambda \frac{(T/\beta)^{\xi} - 1}{(1 + k/\alpha)^{\eta}} \quad (16)$$

where five parameters are essential, i.e., λ , an intensity scale parameter in units of x (e.g., mm/h); β , a timescale parameter in the units of the return period (e.g., years); α , a timescale parameter in the units of the timescale (e.g., h) with $\alpha > 0$; η , a dimensionless parameter with $0 < \eta < 1$; and ξ , the tail index of the process. The total precipitation level can be determined by multiplying the rainfall intensity x by the corresponding timescale (i.e., 24 h). Figure 17 shows the rainfall distribution across the study area. The corresponding

FSM is depicted in Figure 18, highlighting three main classes: low-, moderate- and high-flood-susceptibility zones. Pixels with prediction probability values between 0 and 0.33 are classified as low-susceptibility, 0.33–0.66 as moderate, and values greater than 0.66 as high-susceptibility.

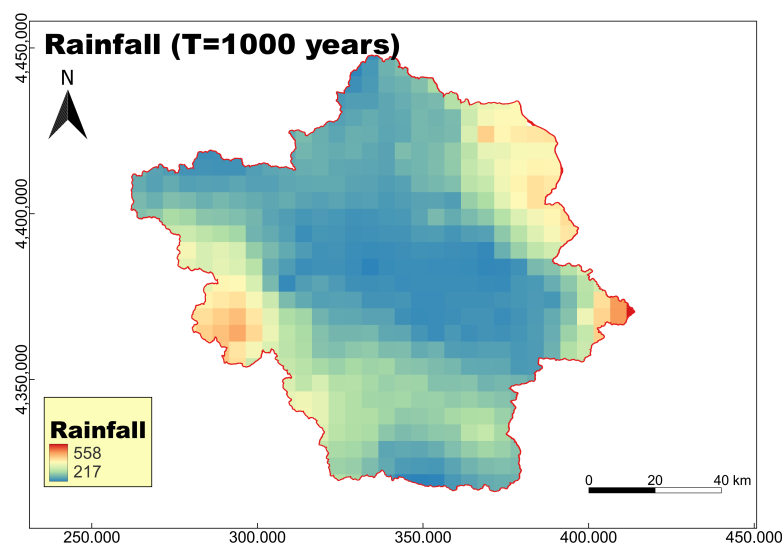


Figure 17. Rainfall for $T = 1000$ years and $k = 24$ h.

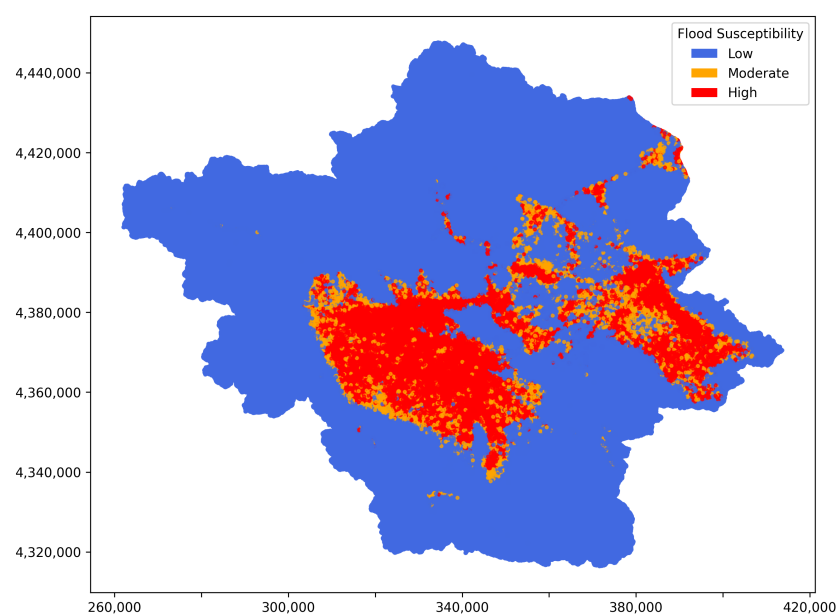


Figure 18. FSM for Peneus basin based on a rainfall scenario for $T = 1000$ years and $k = 24$ h.

5. Discussion and Future Research Directions

In recent years, machine learning techniques have been used to map flood susceptibility across many regions globally. These are highly effective in this task because of their ability to leverage historical flood data to uncover hidden patterns in environmental data. Through our work, we highlight how these models can be leveraged to make more accurate, data-driven predictions. However, despite their growing popularity, the application of these techniques has not yet reached its full potential, and there is still significant room for improvement. In the present study, we applied machine learning algorithms for the creation of flood susceptibility maps in the region of Thessaly, aiming to enhance the flood protection of the area and provide more insights into the applied methodology. We utilized

satellite-derived data to assess flooded areas from a massive recent extreme event, i.e., Storm Daniel, along with the corresponding satellite-based rainfall of the event.

We created a dataset of 3 950 data points consisting of 13 flood conditioning factors as independent variables and corresponding binary flood labels. The factors included elevation, slope, aspect, curvature, distance from roads, distance from rivers, drainage density, topographic wetness index, stream power index, curve number, land use, normalized difference vegetation index and rainfall. As we divided the balanced dataset into 80% for training and the remaining 20% for testing, we compared the performance of LR, SVM, RF and XGBoost in a classification procedure. In this performance assessment, according to various metrics, like accuracy, precision, recall, specificity, F_1 score and ROC curve, the tree-based algorithms performed better, achieving an Area Under the Curve (AUC) score of 0.97, with 1.0 representing the highest possible score. More specifically RF, XGBoost, SVM and LR achieved AUC values of 0.969, 0.968, 0.940 and 0.907, respectively. The feature importance of RF revealed that terrain variables, such as elevation, slope and TWI, along with the event's rainfall, are the most significant contributors. Elevation and slope achieved the highest scores in both MDI (0.27 and 0.18) and MDA (0.20 and 0.15), respectively.

In comparison with previous studies conducted under similar contexts but different geographic regions, the AUC values achieved in this work ($\approx 97\%$ with RF and XGBoost) surpass those reported for RF in Periyar (94%) [19], Salzburg (88%) [21], and Metlili (81%) [22], and are comparable to the higher performance observed for RF in Berlin (96%) [12] and XGBoost in Wilayat As-Suwayq (98%) [24], among others. These findings indicate that the proposed modelling framework achieves accuracy levels at the upper range of those reported in the literature, highlighting the strong potential demonstrated by all the evaluated ML models to effectively capture complex flood-related patterns in extreme event conditions.

The choice to use data from the Storm Daniel event was driven by both practical and methodological reasons. As an extreme flood event that impacted the entire river basin, Storm Daniel offered a rare opportunity to capture the full spatial extent of flood dynamics across the catchment. Using data from such an event enabled us to derive a rich, representative sample of flooded and non-flooded areas based on satellite images. The availability of data for such events is not always present. In fact, other historical flood events lacked similarly extensive and high-quality satellite data coverage.

To investigate the generalizability of the model and assess the impact of the initial training data, we conducted additional experiments. As we divided the catchment in half, we used data from one side each time to train a random forest model. The results highlight the model's sensitivity to the training data, suggesting that generalizability may be influenced by spatial characteristics. In particular, the models trained using data from the entire basin, the half-left side and the half-right side achieved AUC scores of 0.96, 0.94 and 0.92, respectively, and AUPRC values of 0.57, 0.51 and 0.41. This finding highlights the need for caution when applying the model to new areas with different spatial features. To improve generalizability, it may be necessary to incorporate diverse data sources or ensure a more representative distribution of training data that covers a broader range of spatial conditions.

Early flood warning systems are critical tools for disaster risk reduction, enabling timely alerts and proactive measures to minimize damage and protect communities at risk [52]. The proposed approach can also serve as an early warning system, wherein a pre-trained model operates for a real-time rainfall forecast to identify areas likely to be most affected by the event. This method could complement or even replace time-consuming simulations traditionally conducted with conventional hydraulic models. Physically based flood models offer detailed process simulation but demand extensive data and computational resources, which can limit their applicability in data-poor or urgent scenarios. By

contrast, our machine learning–satellite data approach achieves comparable or superior accuracy with lower data requirements and faster processing, making it highly suitable for large-scale or rapid flood susceptibility mapping. Such a scenario was included in the study, in which a map based on a 1000-year return period rainfall event was generated, and the susceptible areas were identified. This scenario was created using the updated parameters of the ombrian curves in Greece, within the implementation of the EU Directive 2007/60/EC [51]. A trained model required less than 5 min to produce the outputs. The resulted map exhibits approximately 20% of the basin as a highly flood-prone area, showing two major zones of high flood susceptibility inside the basin, specifically in the eastern and central-to-southern regions, highlighting the need for targeted measures and actions.

It is important to provide decision makers with information about solutions for every situation, regardless of the challenges involved, and also to not attribute failure to unforeseen events and unrealistic theories. Anticipating and mitigating potential risks can greatly reduce the aftermath of the resulting consequences; all civil protection guidance highlights the importance and added value of anticipatory planning and design in disaster management. Our case study area, Thessaly, has been suffering from both flooding and a lack of water for many years, yet the responses have proven insufficient in many cases. Solutions can only be achieved through civil engineering works and infrastructure designed to mitigate these challenges; still, the responsible authorities and sometimes the wider community often resist the implementation of these projects. A robust strategic plan could explore the potential benefits of multipurpose dams all around the PRB, evaluating whether they could serve dual functions, such as providing water for irrigation, industrial use, and human consumption, while also supporting flood control and hydroelectric power generation. These infrastructure projects, properly studied, are among the most effective ways to prevent flooding, and their landscape design is beneficial for landscape quality perception as well as touristic development [53].

Although the integration of satellite data with machine learning shows great potential for flood susceptibility analysis, several challenges remain to be addressed. Firstly, there can be limited spatial and temporal resolution in satellite imagery, and high-quality satellite data may not be available for the specific period. SAR images often contain noise, which can reduce the accuracy of flood detection and necessitates complex preprocessing [54]. Moreover, the rainfall data derived from satellite observations may not always be accurate and can sometimes diverge from the ground truth measurements [55]. Accurate ground truth data labeling for training a machine learning model can be challenging in some regions. ML models trained on a specific region may not generalize well to other areas. In addition, processing high-resolution satellite data may require significant computational resources.

This study can serve as a foundation for future research that can build upon it to explore potential improvements to the methods used in the current work. To begin with, not only point-based models (SVM, RF, etc.), but also image-based models like Convolutional Neural Networks (CNNs) can be adopted. The main benefit from such models is that they can capture spatial patterns and information, as pixels of the image are not treated independently. On top of that, comparing the proposed approach with a hydrological model, along with considering the socio-economic aspects, could offer valuable insights. Additionally, this study could serve as a key reference for evaluating differences in model performance using future extreme flood events; however, in that case the feature of rainfall should be adopted by accounting for all precipitation that contributed to those events. A key aspect is the model's generalizability, and future work could explore this by testing it on different extreme flood events and watersheds to better assess its performance across varied conditions. It is also worth noting that a key question to address is how many data samples are necessary to consider the ML model accurate. Another important aspect

is the reliable quantification of the uncertainty of each model. One way to accomplish this is via the BlueCat [56] approach, which operates by transforming a point prediction provided by a model to a corresponding stochastic formulation. While conventional ML models ignore any underlying physical law, certain types of neural networks, such as Physics-Informed Neural Networks (PINNs) [57] and Neural Operators [58], which fall under the broader domain of Scientific Machine Learning, are specifically designed to incorporate them into the modeling process. PINNs solve partial differential equations (PDEs) by embedding physical laws into neural networks, while neural operators learn mappings between function spaces to rapidly approximate solutions to complex systems. These could be useful, as flood modeling typically employs the shallow water equations or the Navier–Stokes equations. Last, but equally important, is the growing effort to interpret how some machine learning models function (Interpretable AI [59]) and provide clear explanations for their outputs and help decision-making processes (Explainable AI [60]). It is important to understand the reasoning behind the prediction of each model, regardless of its complexity.

6. Summary and Conclusions

Our research focuses on the contribution of machine learning algorithms to flood susceptibility mapping. Using 13 flood conditioning factors and satellite data, we trained four models on a classification task, focusing on the Peneus Basin case study and with an extreme flood event serving as a key reference. The performance of Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) was compared using standard evaluation indicators. Feature importance was analyzed for the best-performing model (RF). To assess the generalizability, the RF model was trained using different subsets of data by splitting the basin in half. Finally, a rainfall scenario for a 1000-year return period with a 24-hour scale was developed to support long-term flood risk assessment and planning. Based on the analysis and evaluation of the applications, the key findings are outlined below:

- Tree-based models, particularly RF and XGBoost, outperformed the other algorithms, achieving the highest Area Under the Curve AUC scores—RF: 0.969; XGBoost: 0.968; SVM: 0.940; LR: 0.907.
- Feature importance analysis revealed that the most influential factors contributing to flood susceptibility, in decreasing order, are elevation, slope, rainfall and TWI, providing valuable insights for model interpretation and decision making.
- The choice of initial training data impacts model performance and generalizability, highlighting the need for careful dataset selection in spatial prediction tasks.
- The identified areas with high flood susceptibility are the eastern and the central-to-southern regions, offering useful information for targeted risk mitigation and planning efforts.
- Machine learning algorithms offer a promising approach for flood susceptibility mapping and can be extended as an early warning system, requiring significantly less time compared to traditional models.

Floods continue to be a major socio-environmental challenge, causing widespread damage and disruption. While accurate FSMs are essential for effective flood risk management and preparedness, this remains a complex task due to the variability of climatic, hydrological and geographical factors. Machine learning offers a promising avenue for improving predictive capabilities by leveraging data-driven insights. However, ongoing research is needed to enhance model robustness, generalizability and integration into real-time early warning systems. Ultimately, the integration of domain knowledge with advanced technologies offers powerful solutions for developing effective flood control strategies.

Author Contributions: Conceptualization, N.T., I.B. and D.K.; methodology, N.T.; software, N.T.; validation, N.T.; formal analysis, N.T.; investigation, N.T., I.B., D.K., T.I. and P.D.; resources, N.T.; data curation, N.T. and I.B.; writing—original draft preparation, N.T.; writing—review and editing, N.T., I.B. and D.K.; visualization, N.T.; supervision, N.T., I.B., D.K., T.I. and P.D.; project administration, N.T., I.B., D.K., T.I. and P.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. N.T. received partial support through a scholarship from the Special Account for Research Funding (E.L.K.E) of the National Technical University of Athens.

Data Availability Statement: The data were collected from various sources. All the data used in this study are freely available from public sources. Detailed descriptions of the datasets and their respective sources are provided in the corresponding sections of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Jonkman, S.N. Global Perspectives on Loss of Human Life Caused by Floods. *Nat. Hazards* **2005**, *34*, 151–175. [\[CrossRef\]](#)
- Yu, Q.; Wang, Y.; Li, N. Extreme Flood Disasters: Comprehensive Impact and Assessment. *Water* **2022**, *14*, 1211. [\[CrossRef\]](#)
- European Commission. *Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the Assessment and Management of Flood Risks*; L 288, 6.11.2007; Official Journal of the European Communities: Aberdeen, UK, 2007; p. 27.
- Bentivoglio, R.; Isufi, E.; Jonkman, S.N.; Taormina, R. Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 4345–4378. [\[CrossRef\]](#)
- Avila-Aceves, E.; Plata-Rocha, W.; Monjardin-Armenta, S.A.; Rangel-Peraza, J.G. Geospatial modelling of floods: A literature review. *Stoch. Environ. Res. Risk Assess.* **2023**, *37*, 4109–4128. [\[CrossRef\]](#)
- Abdullah, M.F.; Siraj, S.; Hodgett, R.E. An Overview of Multi-Criteria Decision Analysis (MCDA) Application in Managing Water-Related Disaster Events: Analyzing 20 Years of Literature for Flood and Drought Events. *Water* **2021**, *13*, 1358. [\[CrossRef\]](#)
- Liuzzo, L.; Sammartano, V.; Freni, G. Comparison between Different Distributed Methods for Flood Susceptibility Mapping. *Water Resour. Manag.* **2019**, *33*, 3155–3173. [\[CrossRef\]](#)
- Chen, H.; Yang, D.; Hong, Y.; Gourley, J.J.; Zhang, Y. Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting. *Adv. Water Resour.* **2013**, *59*, 209–220. [\[CrossRef\]](#)
- Soomro, S.E.H.; Hu, C.; Boota, M.W.; Ahmed, Z.; Chengshuai, L.; Zhenyue, H.; Xiang, L.; Soomro, M.H.A.A. River Flood Susceptibility and Basin Maturity Analyzed Using a Coupled Approach of Geo-morphometric Parameters and SWAT Model. *Water Resour. Manag.* **2022**, *36*, 2131–2160. [\[CrossRef\]](#)
- Stoleriu, C.C.; Urzica, A.; Miha-Pintilie, A. Improving flood risk map accuracy using high-density LiDAR data and the HEC-RAS river analysis system: A case study from north-eastern Romania. *J. Flood Risk Manag.* **2020**, *13*, e12572. [\[CrossRef\]](#)
- Wahba, M.; Sharaan, M.; Elsadek, W.M.; Kanae, S.; Hassan, H.S. Examination of the efficacy of machine learning approaches in the generation of flood susceptibility maps. *Environ. Earth Sci.* **2024**, *83*, 429. [\[CrossRef\]](#)
- Seleem, O.; Ayzel, G.; de Souza, A.C.T.; Bronstert, A.; Heistermann, M. Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. *Geomat. Nat. Hazards Risk* **2022**, *13*, 1640–1662. [\[CrossRef\]](#)
- Saravanan, S.; Abijith, D.; Reddy, N.M.; Kss, P.; Janardhanam, N.; Sathiyamurthi, S.; Sivakumar, V. Flood susceptibility mapping using machine learning boosting algorithms techniques in Idukki district of Kerala India. *Urban Clim.* **2023**, *49*, 101503. [\[CrossRef\]](#)
- Seydi, S.T.; Sadegh, M. Novel Kolmogorov-Arnold network architectures for accurate flood susceptibility mapping: A comparative study. *J. Hydrol.* **2025**, *661*, 133553. [\[CrossRef\]](#)
- Wedajo, G.K.; Lemma, T.D.; Fufa, T.; Gamba, P. Integrating Satellite Images and Machine Learning for Flood Prediction and Susceptibility Mapping for the Case of Amibara, Awash Basin, Ethiopia. *Remote Sens.* **2024**, *16*, 2163. [\[CrossRef\]](#)
- Kourgialis, N.N.; Karatzas, G.P. A national scale flood hazard mapping methodology: The case of Greece—Protection and adaptation policy approaches. *Sci. Total Environ.* **2017**, *601–602*, 441–452. [\[CrossRef\]](#)
- Tsinidis, G.; Koutas, L. Geotechnical and Structural Damage to the Built Environment of Thessaly Region, Greece, Caused by the 2023 Storm Daniel. *Geotechnics* **2025**, *5*, 16. [\[CrossRef\]](#)
- Dimitriou, E.; Efstratiadis, A.; Zotou, I.; Papadopoulos, A.; Iliopoulou, T.; Sakki, G.-K.; Mazi, K.; Rozos, E.; Koukouvinos, A.; Koussis, A.D.; et al. Post-Analysis of Daniel Extreme Flood Event in Thessaly, Central Greece: Practical Lessons and the Value of State-of-the-Art Water-Monitoring Networks. *Water* **2024**, *16*, 980. [\[CrossRef\]](#)
- Sreekala, S.; Geetha, P.; Madhu, D. Flood Susceptibility Map of Periyar River Basin Using Geo-Spatial Technology and Machine Learning Approach. *Remote. Sens. Earth Syst. Sci.* **2024**, *8*, 1–21. [\[CrossRef\]](#)

20. Al-Ruzouq, R.; Shanableh, A.; Jena, R.; Gibril, M.B.A.; Hammouri, N.A.; Lamghari, F. Flood susceptibility mapping using a novel integration of multi-temporal sentinel-1 data and eXtreme deep learning model. *Geosci. Front.* **2024**, *15*, 101780. [CrossRef]
21. Nachappa, T.G.; Piralilou, S.T.; Gholamnia, K.; Ghorbanzadeh, O.; Rahmati, O.; Blaschke, T. Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *J. Hydrol.* **2020**, *590*, 125275. [CrossRef]
22. Hitouri, S.; Mohajane, M.; Lahsaini, M.; Ali, S.A.; Setargie, T.A.; Tripathi, G.; D'antonio, P.; Singh, S.K.; Varasano, A. Flood Susceptibility Mapping Using SAR Data and Machine Learning Algorithms in a Small Watershed in Northwestern Morocco. *Remote Sens.* **2024**, *16*, 858. [CrossRef]
23. Seydi, S.T.; Kanani-Sadat, Y.; Hasanlou, M.; Sahraei, R.; Chanussot, J.; Amani, M. Comparison of Machine Learning Algorithms for Flood Susceptibility Mapping. *Remote Sens.* **2023**, *15*, 192. [CrossRef]
24. Al-Kindi, K.M.; Alabri, Z. Investigating the Role of the Key Conditioning Factors in Flood Susceptibility Mapping through Machine Learning Approaches. *Earth Syst. Environ.* **2024**, *8*, 63–81. [CrossRef]
25. Shafizadeh-Moghadam, H.; Valavi, R.; Shahabi, H.; Chapi, K.; Shirzadi, A. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *J. Environ. Manag.* **2018**, *217*, 1–11. [CrossRef]
26. Iliopoulou, T.; Malamos, N.; Koutsoyiannis, D. Regional Ombrian Curves: Design Rainfall Estimation for a Spatially Diverse Rainfall Regime. *Hydrology* **2022**, *9*, 67. [CrossRef]
27. Mimikou, M.; Koutsoyiannis, D. Extreme floods in Greece: The case of 1994. In Proceedings of the US-ITALY Research Workshop on the Hydrometeorology, Impacts, and Management of Extreme Floods, Perugia, Italy, 13–17 November 1995.
28. Bathrellos, G.D.; Skilodimou, H.D.; Soukis, K.; Koskeridou, E. Temporal and Spatial Analysis of Flood Occurrences in the Drainage Basin of Pinios River (Thessaly, Central Greece). *Land* **2018**, *7*, 106. [CrossRef]
29. Valkaniotis, S.; Papathanassiou, G.; Marinos, V.; Saroglou, C.; Zekkos, D.; Kallimogiannis, V.; Karantanellis, E.; Farmakis, I.; Zalachoris, G.; Manousakis, J.; et al. Landslides Triggered by Medicane Ianos in Greece, September 2020: Rapid Satellite Mapping and Field Survey. *Appl. Sci.* **2022**, *12*, 12443. [CrossRef]
30. Kumar, P.; Bhatnagar, R.; Gaur, K.; Bhatnagar, A. Classification of Imbalanced Data: Review of Methods and Applications. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1099*, 012077. [CrossRef]
31. Copernicus DEM. 2022. Available online: <https://doi.org/10.5270/esa-c5d3d65> (accessed on 15 December 2024).
32. USDA. Hydrology. In *National Engineering Handbook*; United States Department of Agriculture, Soil Conservation Service, US Government Printing Office: Washington, DC, USA, 1972.
33. Efstratiadis, A.; Koukouvinos, A.; Michaelidi, E.; Galiouna, E.; Tzouka, K.; Koussis, A.D.; Mamassis, N.; Koutsoyiannis, D. *Description of Regional Approaches for the Estimation of Characteristic Hydrological Quantities, DEUCALION—Assessment of Flood Flows in Greece under Conditions of Hydroclimatic Variability: Development of Physically-Established Conceptual-Probabilistic*; Department of Water Resources and Environmental Engineering—National Technical University of Athens, National Observatory of Athens: Athens, Greece, 2014.
34. Cox, D.R. The Regression Analysis of Binary Sequences. *J. R. Stat. Soc. Ser. (Methodol.)* **1958**, *20*, 215–232. [CrossRef]
35. Myung, I.J. Tutorial on Maximum Likelihood Estimation. *J. Math. Psychol.* **2003**, *47*, 90–100. [CrossRef]
36. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992. [CrossRef]
37. Schölkopf, B. The Kernel Trick for Distances. In Proceedings of the 13th International Conference on Neural Information Processing Systems (NIPS'00), Hong Kong, China, 3–6 October 2006; MIT Press: Cambridge, MA, USA, 2000; pp. 283–289.
38. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
39. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
40. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 13–17 August 2016; pp. 784–794.
41. Kaggle Competitions. Available online: <https://www.kaggle.com/competitions> (accessed on 24 September 2024).
42. Amjad, M.; Ahmad, I.; Ahmad, M.; Wróblewski, P.; Kamiński, P.; Amjad, U. Prediction of Pile Bearing Capacity Using XGBoost Algorithm: Modeling and Performance Evaluation. *Appl. Sci.* **2022**, *12*, 2126. [CrossRef]
43. Rengasamy, D.; Mase, J.M.; Kumar, A.; Rothwell, B.; Torres, M.T.; Alexander, M.R.; Winkler, D.A.; Figueredo, G.P. Feature importance in machine learning models: A fuzzy information fusion approach. *Neurocomputing* **2022**, *511*, 163–174. [CrossRef]
44. Vatsal. Feature Importance in Machine Learning, Explained. Towards Data Science. Available online: <https://towardsdatascience.com/feature-importance-in-machine-learning-explained-443e35b1b284/> (accessed on 20 April 2025).
45. Tangirala, S. Evaluating the Impact of GINI Index and Information Gain on Classification Using Decision Tree Classifier Algorithm. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 612–619. [CrossRef]
46. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
47. D'Agostino, A. The Explanation You Need on Binary Classification Metrics. Medium. Available online: <https://medium.com/data-science/the-explanation-you-need-on-binary-classification-metrics-321d280b590f> (accessed on 25 April 2025).

48. Grewal, R.; Cote, J.A.; Baumgartner, H. Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Mark. Sci.* **2004**, *23*, 519–529. [[CrossRef](#)]
49. Koutsoyiannis, D.; Iliopoulou, T.; Koukouvinos, A.; Malamos, N. A stochastic framework for rainfall intensity-time scale-return period relationships. Part I: Theory and estimation strategies. *Hydrol. Sci. J.* **2024**, *69*, 1082–1091. [[CrossRef](#)]
50. Iliopoulou, T.; Koutsoyiannis, D.; Malamos, N.; Koukouvinos, A.; Dimitriadis, P.; Mamassis, N.; Tepetidis, N.; Markantonis, D. A stochastic framework for rainfall intensity-time scale-return period relationships. Part II: Point modelling and regionalization over Greece. *Hydrol. Sci. J.* **2024**, *69*, 1092–1112. [[CrossRef](#)]
51. Koutsoyiannis, D.; Iliopoulou, T.; Koukouvinos, A.; Malamos, N.; Mamassis, N.; Dimitriadis, P.; Tepetidis, N.; Markantonis, D. *Production of Maps with Updated Parameters of the Ombrian Curves at Country Level (Implementation of the EU Directive 2007/60/EC in Greece)*; Technical Report; Department of Water Resources and Environmental Engineering—National Technical University of Athens: Athens, Greece, 2023. Available online: <https://www.itia.ntua.gr/2273/> (accessed on 20 April 2025).
52. Perera, D.; Seidou, O.; Agnihotri, J.; Rasmy, M.; Smakhtin, V.; Coulibaly, P.; Mehmood, H. *Flood Early Warning Systems: A Review of Benefits, Challenges and Prospects*; UNU-INWEH Report Series; Issue 08; United Nations University-Institute for Water, Environment and Health: Hamilton, ON, Canada, 2019; ISBN 978-92-808-6096-2. Available online: <https://inweh.unu.edu/flood-early-warning-systems-a-review-of-benefits-challenges-and-prospects/> (accessed on 15 April 2025).
53. Ioannidis, R.; Sargentis, G.-F.; Koutsoyiannis, D. Landscape design in infrastructure projects - is it an extravagance? A cost-benefit investigation of practices in dams. *Landsc. Res.* **2022**, *47*, 370–387. [[CrossRef](#)]
54. Singh, P.; Diwakar, M.; Shankar, A.; Shree, R.; Kumar, M. A review on SAR image and its despeckling. *Arch. Comput. Methods Eng.* **2021**, *28*, 4633–4653. [[CrossRef](#)]
55. Yan, J.; Gebremichael, M. Estimating actual rainfall from satellite rainfall products. *Atmos. Res.* **2009**, *92*, 481–488. [[CrossRef](#)]
56. Koutsoyiannis, D.; Montanari, A. Bluecat: A Local Uncertainty Estimator for Deterministic Simulations and Predictions. *Water Resour. Res.* **2022**, *58*, e2021WR031215. [[CrossRef](#)]
57. Cuomo, S.; Di Cola, V.S.; Giampaolo, F.; Rozza, G.; Raissi, M.; Piccialli, F. Scientific Machine Learning through Physics-Informed Neural Networks: Where We Are and What's Next. *J. Sci. Comput.* **2022**, *92*, 88. [[CrossRef](#)]
58. Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Fourier Neural Operator for Parametric Partial Differential Equations. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual Event, 3–7 May 2021.
59. Dikshit, A.; Pradhan, B. Interpretable and explainable AI (XAI) model for spatial drought prediction. *Sci. Total Environ.* **2021**, *801*, 149797. [[CrossRef](#)]
60. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2020**, *23*, 18. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.