Journal of Hydrology: Regional Studies 63 (2026) 103075

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Investigating the mechanisms of flood susceptibility with the use of multi-basin machine learning models in data-scarce environments in Cyprus

Constantinos F. Panagiotou [a,*], Giorgia Guerrisi [b], Davide De Santis [b], Fabio Del Frate [b], Marios Tzouvaras [a]

[a] Eratosthenes Centre of Excellence, 82 Franklin Roosevelt, Limassol 3012, Cyprus
[b] Department of Civil Engineering and Computer Science Engineering, Tor Vergata University of Rome, Rome 00133, Italy

ARTICLE INFO

Keywords:
Flood risk management
Machine learning
Shallow neural network
Generalized models
Simplified models
Data-scarce environments

ABSTRACT

Study region: The island of Cyprus is dominated by small-scale watersheds that favor the occurrence of flash floods. Climate projections indicate the increase in frequency and intensity of these events.

Study focus: The development of rapid flood screening tools is essential for better urban planning. This study uses four different machine learning algorithms, namely support vector machine (SVM), extreme gradient boosting (XGBoost), random forest (RF), and multilayer perceptron (MLP), to build models based on data collected from eight watersheds to enhance their within-region (Cyprus) generalization. Seven features were selected for tuning and testing the performance of these models. T-based confidence intervals were calculated to quantify uncertainty.

New hydrological insights for the region: All models achieved good agreement with the inventory database. RF model was selected to build multi-level susceptibility maps. Half of the Georskipou watershed is classified as highly susceptible to flooding, mostly urban and semi-urban regions, whereas 38 % of the test watershed is not expected to experience severe flood events. Simplified RF models were developed by selecting different combinations of the most important features, revealing that land-use, terrain slope, terrain elevation, and flow accumulation are sufficient to achieve good accuracy (95 %) with flood inventory data. The results highlight the ability of simple, computationally efficient data-driven models to provide rapid predictions, thus avoiding the compilation of fully detailed physically-based models.

1. Introduction

Flooding is the most disastrous natural hazard in terms of economic losses, causing annual mean damage that exceeds 12 billion euros at the European scale (EEA, 2020). For instance, 44 billion euros of damages are estimated because of the occurrence of extreme flood events in 2021 in Central Europe (Belgium, Germany, Netherlands), whereas the 2023 flood events in Slovenia have caused economic losses of approximately €10 billion. Moreover, these events affected at least 680,000 people across 24 European countries between 2014 and 2024 (EDJNet, 2025). At a global scale, the occurrence of more intense precipitation events is expected to increase

* Corresponding author.
E-mail address: constantinos.panagiotou@eratosthenes.org.cy (C.F. Panagiotou).

https://doi.org/10.1016/j.ejrh.2025.103075
Received 7 October 2025; Received in revised form 12 December 2025; Accepted 19 December 2025
Available online 6 January 2026

in the near future (Canadell and Monteiro, 2023) because of climate change. Moreover, economic exposure to flooding is expected to increase by approximately 300 % by 2050, mainly because of population growth and economic development in flood-prone areas (Jongman et al., 2012).

Consequently, the adverse consequences of flooding undermine global efforts to achieve the majority of sustainable development goals (SDGs) (UN, 2025). In particular, this phenomenon can limit the accessibility of financial resources which are needed to restore and develop business activities, including initiatives for alleviating poverty (SDGs 1, 8 and 11). These limitations can reduce citizens' capacity to cover their basic needs, such as access to clean water and safe food (SDG 2). Additional negative impacts that are directly linked to the SDGs include exacerbating imbalances among different groups of people (SDG 1,2,5, and 10); damaging infrastructure and delaying or postponing development initiatives (SDG 4 and 9); and worsening health quality (SDG 3). Therefore, effectively integrating the concept of disaster risk reduction into water and land development policies, including actions to support climate adaptation measures to reach sustainability goals, is essential.

To address the above issues, European authorities have allocated time and resources to establish a European flood directive for assessing and managing flood risk (EC, 2007) with respect to financial prosperity, human health, the sustainable development of cultural heritage and the environment. According to this directive (Article 5), all European Union (EU) member states are responsible for the identification of high-risk flood regions and the compilation of flood risk maps under different climatic scenarios. The resulting maps should serve as screening tools for the design of flood risk management strategies and action plans (e.g., optimum escape routes and shelters for citizens) that also include socioeconomic information (Article 4).

Nevertheless, proper adaptation of the EU framework is a difficult task because of the complex nature of this phenomenon, which is driven by the combined effects of several factors, including topography, hydrogeology, precipitation patterns, and land use development, among others. Hence, it is essential to identify the associations among these factors and quantify their influence on flood evolution before compiling effective policies for counteracting flood impacts in urban, semi-urban, and rural contexts (Sett et al., 2024). The identification of regions that are highly susceptible to flooding is an essential component of risk management, referring to the potential of an area to experience flood events on the basis of its physical characteristics. A plethora of modeling tools are used for monitoring and forecasting the spatiotemporal distribution of flood susceptibility, such as physically based models (de Sousa et al., 2022; Hasan Tanim et al., 2024; Kountouri et al., 2024; Sett et al., 2024; Suwannachai et al., 2024; Zhang et al., 2024), multicriteria decision analysis (MCDA) (Bathrellos et al., 2016; Chen et al., 2015; Hagos et al., 2022; Kazakis et al., 2015; Shadmehri Toosi et al., 2019), and machine learning models (MLMs) (Al-Sheriadeh and Daqdouq, 2024; Chen et al., 2021; Tsangaratos et al., 2023; Zhao et al., 2019).

Physically based models simulate the physical processes that determine the evolution of flood-related parameters using of topographical, meteorological, and hydrogeological information. In particular, these models provide spatiotemporal predictions of flood properties by solving unsteady partial differential flow equations, enabling practitioners to evaluate the response of the physical system to different climatic and non-climatic scenarios, and to estimate flood-related indicators that are difficult to measure during actual events (Mudashiru et al., 2021). As expected, many input parameters are required to properly emulate the physical processes, which are also sources of uncertainties and errors that propagate to the model outputs (Chen et al., 2016). Moreover, these models can become computationally demanding depending on the scale of the study area, the spatiotemporal resolutions, and the characteristics of the input data (Lyu et al., 2019). MCDA approaches, integrated in GIS-based environments, aim at balancing the perspectives of multiple decision makers to propose trade-offs in a transparent and logical framework, resulting in flood susceptibility maps on the basis of spatially distributed parameters (i.e., flood-related factors). These approaches are often preferred by policy-makers because they are flexible in handling multiple types of datasets over a wide range of scales and resolutions. Several MCDA algorithms are used in flood susceptibility assessment, such as fuzzy logic AHP (Bokhari et al., 2024; Mokhtari et al., 2024), the analytical hierarchical process (AHP) (Khouz et al., 2023; Panagiotou, 2025a; Panagiotou et al., 2025b), logistic regression Rahman et al., (2019), frequency ratio (Cao et al., 2016; Samanta et al., 2018), and weights of evidence (Gupta and Dixit, 2023). However, the main disadvantage of most of these approaches is that they are heavily influenced by experts' opinions and inputs (Danumah et al., 2016), leading to subjective results.

MLMs have been increasingly used in flood modeling during the last decade because of their ability to identify complex relationships among several attributes and provide data-driven, non-subjective outcomes. Several studies have focused on applying MLMs to address flood risk challenges. Tree-based algorithms are often used because of their proven ability to handle large high-dimensional datasets. For example, Abedi et al., (2022) applied tree-based algorithms to generate flash-flood maps of the Basca Chiojdului watershed (Romania). The authors selected 10 input variables that contained topographical, hydrogeological, and land-use land cover information, whereas 962 flash flood events were collected to compile the inventory database. The authors ranked random forest (RF) as the best algorithm on the basis of the area under the curve (AUC), followed by extreme gradient boosting (XGBoost), and the classification and regression tree (CART). The support vector machine (SVM) is an alternative algorithm that can provide reliable predictions in flood risk management. For example, Liu et al. (2022) used a SVM algorithm in conjunction with the logistic regression method to map flood susceptibility based on data collected from three continents (i.e., Asia, Europe, and Africa), allowing the identification of the most critical locations in a large area of interest. Chen et al. (2021) trained six MLMs based on data collected from the Pearl River Delta (China), involving 2000 entries and 12 flood-related indicators. Hyperparameter optimization was conducted to ensure fair comparison among the MLMs. The gradient boosting decision tree (GBDT) algorithm outperformed both shallow (i.e., SVM, RF, XGBoost) and deep-learning models (i.e., multilayer perceptron (MLP), convolutional neural network), suggesting that MLMs with shallow structures might be more suitable than computationally expensive deep-learning models for these applications.

There is a growing number of recent studies in the Mediterranean context that explored the use of MLMs for flood susceptibility management (Ilia et al., 2022; Kotaridis and Lazaridou, 2022; Rondinone et al., 2025; Tepetidis et al., 2025). For example, Tsangaratos

et al. (2023) compared the ability of several classification MLMs to predict flood susceptibility on the island of Euboea (Greece). The authors collected data from previous studies, cartographic products, and satellite images (Sentinel-2) to construct the flood inventory database. Several shallow and deep learning models were then trained and tested, whereas Shapley additive explanation (SHAP) was used to assess the importance of the selected features. Their results demonstrated the ability of one-dimensional convolutional neural networks to correctly capture the interrelationships among the driving factors and flood processes. Additionally, the authors highlighted the strong dependence of this algorithm on data availability and quality, and the need for sufficient computational resources for training and inference.

Ilia et al. (2022) presented a novel framework for predicting flash flood susceptibility on the basis of stacking ensemble modeling, particularly RF and artificial neural network (ANN). The performance of these models was compared to other models based on data collected from 54 flash flood locations and 14 flood-related parameters in the island of Rhodes (Greece). The results showed the ability of the ensemble models to effectively combine other predictive models to produce accurate predictions. A similar study was conducted by Tepetidis et al., (2025) to produce flood susceptibility maps for Thessaly (Greece) corresponding to 1000-year return period rainfall scenario. Thirteen flood-related factors were selected as explanatory features for several MLMs, and Synthetic Aperture Radar (SAR) image data from the Sentinel-1 satellite mission were used to construct the flood inventory database. In alignment with other studies, the results further validated the efficiency of tree-based models to provide accurate predictions. Rondinone et al. (2025) reported a predictive framework for multi-hazard susceptibility assessment in the Basento River Basin, located in southern Italy. The authors trained and tested an XGBoost algorithm using geological, topographical, and remote sensing data. The best models were then used to build flood and landslide susceptibility maps, while Explainable AI techniques (i.e., SHAP) were applied to clarify the influence of individual flood-related predictors on the target variable. The results showed a very good agreement with the inventory database for both hazards, with accuracy values exceeding 0.92.

Despite the existence of numerous studies reporting the successful application of MLMs to identify associations among flood parameters and processes, most of these studies have focused on developing site-specific models with unique characteristics. Consequently, the applicability of these models is expected to be limited to other, "unseen" regions that include different land contexts and physical characteristics. Moreover, most of these studies reported the collection of information from multiple databases, including satellite repositories, in situ monitoring systems, and detailed hydrogeological maps, which are not readily available for all regions. For example, flood events on the island of Cyprus are generated by short-duration precipitation events (flash floods), making it challenging to capture their dynamics using standard satellite missions (e.g., Sentinel, Landsat, Planet).

To address these gaps, this study focuses on a) merging data collected from eight different watersheds into a single dataset (called the training dataset) for developing multi-basin, within-region generalized MLMs which are then tested in an "unseen" watershed based on available data (called the test dataset), b) developing and comparing the performance of multiple MLMs to provide a comprehensive assessment of flood susceptibility by analyzing the influence of driving factors on flood emergence, and c) determining the minimum number of parameters (features) for the development of simplified models that can be useful for cost-effective and fast screening or monitoring in data-scarce environments. Consequently, the novelty of this study lies in explicit external-basin testing and

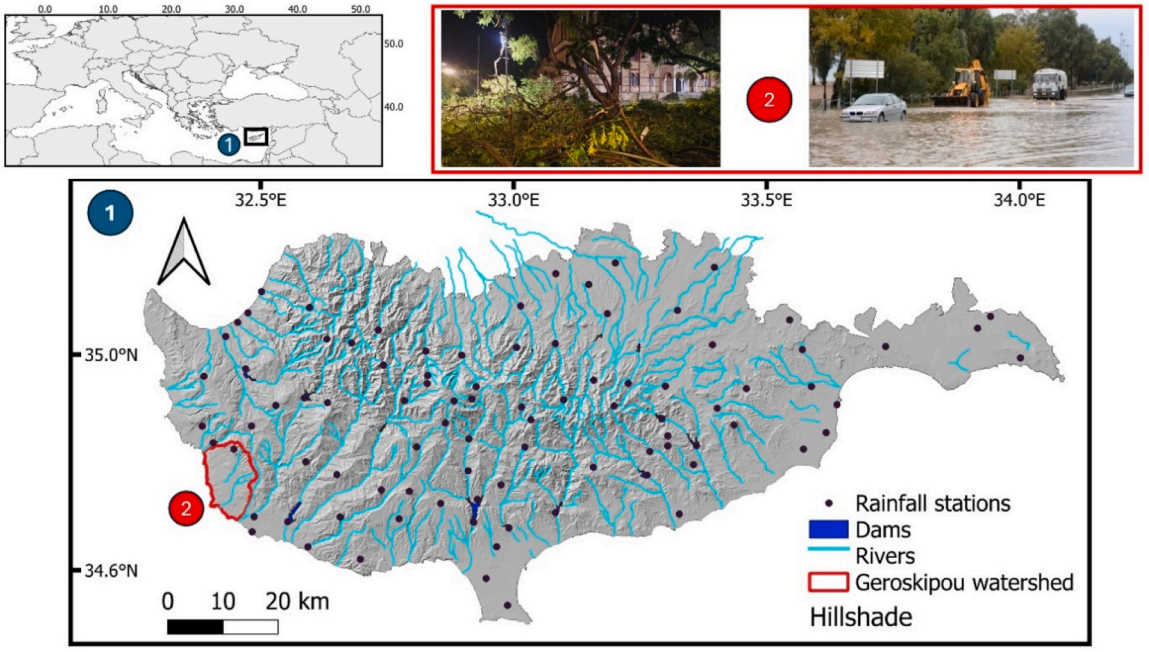

Fig. 1. 1) Geographical location of the entire study area, including the Geroskipou (test) watershed (enclosed by red-colored borders). 2) Photos of recent flood events in Paphos city (November 2024).

an operational, parsimonious recipe for flood susceptibility management. Additional contributions include a) the implementation of hyperparameter optimization for each MLM, unlike most existing studies, and b) the assessment of the functional relationships between major input variables and the flood susceptibility indicator, as well as the identification of threshold values for key flood indicators beyond which susceptibility varies substantially.

A brief description of the study area is provided in Section 2, whereas Section 3 presents the implementation phases of the current methodology, involving the development of the required datasets, model development, and comparison. Section 4 provides the major results of this study, whereas a discussion regarding the applicability of MLMs and current limitations is given in Section 5. Finally, the major conclusions of this study are provided in Section 6.

2. Description of the study area

The island of Cyprus is located in the southeastern region of the Mediterranean Sea (35.4°E, 33.5°N), covering approximately 9250 km^2 (Fig. 1). The climatic condition of the island is characterized as semiarid according to the UN aridity index (UNESCO, 1979), consisting of fluctuating wet periods that typically occur between mid-November and March, whereas the remaining hydrological period is dominated by dry conditions. The annual air temperature, which is averaged over daily data collected from 25 ground-based stations for the period 2011–2024, varies between 30 °C and 35 °C during the dry period, and from 14 °C to 17 °C during the wet period. According to daily rainfall values collected from 90 ground-based stations from 2011 to 2024, the mean annual rainfall rates fluctuate from 314 mm in the vicinity of the coastline, to 857 mm in the high-altitude mountainous regions that are present in the central part of the island (Panagiotou et al., 2024).

There are four major geomorphologic regions in Cyprus (GSD, 2025), particularly the Mesaoria plain, the Troodos Mountain range, the Mamonia Terrane, and the Pentadaktylos Mountain Range. The Mamonia terrane extends primarily in the southwestern part of the island, and comprises igneous, sedimentary, and metamorphic rocks that emerged between the Middle Triassic epoch and the Upper Cretaceous series. As described in Christofi et al. (2020), the combined influence of weathering (sedimentary rocks), fracturing (ophiolitic rock), and rapid uplift of mountainous layers significantly increased the secondary porosity of subsurface layers, leading to the emergence of aquifer systems in these areas. The Mamonia Complex is located in the southwestern region of Cyprus and comprises sedimentary, igneous, and metamorphic rocks that formed between the Middle Triassic and Upper Cretaceous periods.

According to the most recent governmental reports (WDD, 2025), the hydrological network of the entire island is partitioned into 70 major river basins and 387 subbasins (Figure SM1 in the "Supplementary material" file). It comprises primarily ephemeral rivers that are supplied with freshwater from mountainous, more rainy regions, whereas more than 100 surface water reservoirs capture approximately 30 % of the available surface runoff. In alignment with the EU Flood Directive, the national water authorities identified 38 areas as having significant potential for flooding, 37 of which are categorized as flash floods (WDD, 2025). This is attributed to the dominant presence of small-scale watersheds (< 100 km^2) which favor the occurrence of flash floods during the emergence of short-duration, high-intensity rainfall events. In this study, nine watersheds containing the majority of these high-risk regions were selected. Most of them cover the central and eastern parts of the island, encompassing different land contexts (i.e., urban, semi-urban, rural), as well as diverse topographical and meteorological characteristics.

The Geroskipou watershed was selected to test the performance of the MLMs, being classified as a high-risk region for flooding by the national water authorities. Recently, extreme flood events occurred in the coastal part of the watershed in November 2024, causing extensive damage and sediment transport in highly urbanized units (Fig. 1). This event highlighted the need for better preparedness by authorities to mitigate the negative impact of such phenomena, including the generation of high-fidelity flood maps that can support the development of action plans for citizen safety (e.g., optimized escape routes, public shelters). Consequently, multiple discussions have been held with representatives of the regional water authorities to address their needs.

3. Methodology

Fig. 2 presents an overview of the implemented methodology, comprising 3 distinct phases. Phase I focused on constructing the sampling dataset by collecting spatiotemporal information on flood-related factors and inventories from multiple sources. It also included the multicollinearity analysis, conducted to reduce data redundancy of the feature dataset. In Phase II, the sampling dataset was used to build the MLMs and included steps such as hyperparameter optimization and feature importance analysis. The last phase, Phase III, involved generating susceptibility maps using the best-performing machine learning algorithm, evaluating the performance of simplified models based on the most important features, and assessing the directionality of input variables using partial-dependence plots.

3.1. Data collection for flood susceptibility indicators

Similar to Panagiotou (2025a), seven flood-related factors were selected to predict the probability of flooding, which are shown in Table 1: rainfall intensity (RI), Euclidean, two-dimensional, distance from the drainage network (DD), terrain elevation (TE), terrain slope (TS), the combined effect of soil texture and depth (Soil), land use and land cover (LULC), and flow accumulation (FA). The preselection of these factors was jointly discussed and approved by key representatives of the national water authorities, who are responsible for the effective implementation of the EU Flood directive in the Republic of Cyprus.

The topographical characteristics of an area play a major role in flooding. Steep profiles of terrain elevation favor the acceleration of gravity-driven water streams, leading to high surface runoff and potentially reducing the effective time that is required for soil

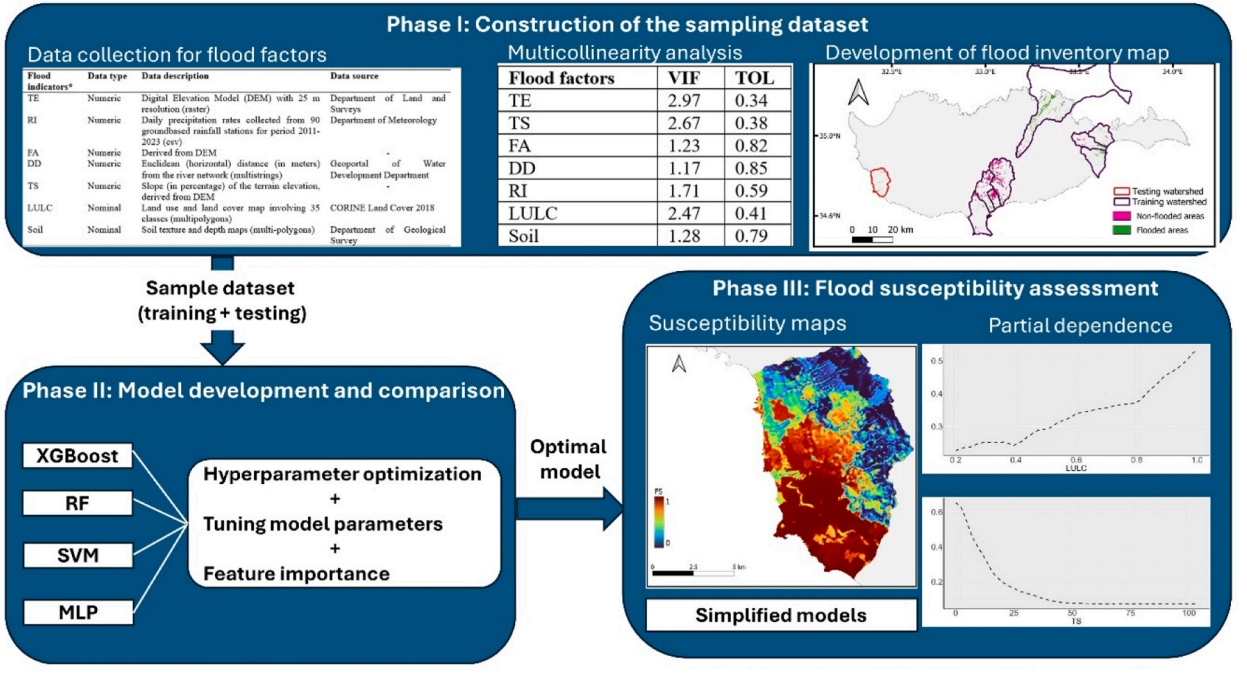

Fig. 2. Overview of the implemented methodology, which includes the following acronyms: Flood-related factors: TE (terrain elevation), RI (rainfall intensity), FA (flow accumulation), DD (distance from drainage network), TS (terrain slope), LULC (land use and land cover); Machine learning models: XGBoost (extreme gradient boosting), RF (random forest), SVM (support vector machine), and MLP (multilayer perceptron neural network).

Table 1
Details of the data collected from different sources to construct the feature dataset. CRS denotes the Coordinate Reference System.

Flood indicators*	Data type	Data description	CRS (EPSG)	Data source
TE	Numeric	Digital Elevation Model (DEM) with 25 m resolution (raster)	32636	Department of Land and Surveys **(version 2023)**
RI	Numeric	Daily rainfall data collected from 90 ground-based rainfall stations for period 2011–2023 (csv)	4326	Department of Meteorology
FA	Numeric	Derived from DEM (raster)	32636	Same as TE
DD	Numeric	Euclidean (horizontal) distance (in meters) from the river network (multi-linestrings)	3857	Geoportal of Water De- velopment Department
TS	Numeric	Slope (in percentage) of the terrain elevation, derived from DEM (raster)	32636	Same as TE
LULC	Nominal	Land use and land cover map involving 35 classes (multi-polygons)	4326	CORINE Land Cover 2018 **(version 2020_20u1)** (CLMS, 2018)
Soil	Nominal	Soil texture and depth maps (multi-polygons)	32636	Geological Survey Department

* TE: Terrain elevation (DEM); RI: rainfall intensity; FA: flow accumulation; DD: drainage distance; TS: terrain slope; LULC: land-use land cover; Soil: soil properties (texture and depth).

infiltration processes to become significant. In Cyprus, regions located at high altitudes are associated with steep terrain slopes and hence are less likely to suffer from flood phenomena. The drainage network, which comprises reservoirs (natural and artificial) and river bodies, is another major driver of flood mechanisms because it is responsible for the distribution of available water resources. Regions near these water bodies are more susceptible to flooding during the emergence of extreme rainfall events or other situations that can lead to high surface runoff.

Another important factor that affects infiltration processes is associated with the hydrogeological properties of the soil layer, which determine the permeability and connectivity of the soil zones, as well as the effective amount, and rates, of water that can infiltrate and be stored in porous subsurface systems. Rainfall intensity is also an important flood driver, especially in small-scale (i.e., areas typically less than 100 km^2) watersheds, that favor flash floods (Pfister et al., 2004). Precipitation events with short durations and high intensities favor the emergence of flooding, contributing to the deterioration of soil properties (e.g., soil erosion) and rapid saturation of the topsoil layers that favor flow accumulation on the land surface. Land use contributes to the occurrence of flooding as urban areas are dominated by impermeable ground surfaces (e.g., paved surfaces and dense infrastructure) that prevent subsurface infiltration, whereas forests and regions with dense vegetation tend to decelerate surface runoff while affecting infiltration and evapotranspiration processes. Fig. 3 shows the geographical distribution of the preselected flood-related factors in the study area. Moreover, the geographical distribution of these factors in the test watershed ("Geroskipou") is shown in Figure SM2.

Similar to Panagiotou (2025a), scores were assigned to the ordinal data collected for LULC and Soil, whereas the geographical distribution of the original (ordinal) information for LULC and Soil properties (i.e., texture and depth) is shown in the supplementary material (Figure SM3, Tables SM1 and SM2). With respect to RI, the daily modified Fournier index (dMFI) was calculated on the basis of precipitation values collected from 90 ground-based rainfall stations.

3.2. Development of flood inventory database

The flood inventory dataset was constructed by comprising the spatial distributions of flooded and non-flooded areas (Fig. 4). The target variable was treated as a Boolean factor with two possible labels: "flood" (positive class) and "non-flood" (negative class). Preliminary efforts to construct the inventory map focused on collecting information from different satellite missions (i.e., Sentinel, Landsat, and Planet); however, the short duration of past flash flood events made this impossible. Consequently, geospatial information was collected from the geoportal of the national water authorities (WDD, 2023), which represents the spatial extent of areas classified as high-risk for flood occurrence under the worst-case scenario that corresponds to a 500-year return period (Figure SM4 in the Supplementary Material). Additional information was collected from alternative governmental reports of the national water authorities. However, these data were not included in the dataset because of several deficiencies (e.g., no georeferenced data, missing dates).

In detail, regions of interest were identified and delineated on the basis of hydraulic and hydrological studies conducted in the most vulnerable watersheds for three different return periods (20, 100, and 500 years), following the guidelines of the EU Flood Directive. As introduced earlier, regions corresponding to the worst-case flood scenario (500-year return period) were classified as flooded in this study and inherently include regions subject to less extreme scenarios. This choice ensures that the entire spectrum of flood susceptibility is captured. For consistency, non-flooded areas were selected from the same nine watersheds that contain high risk regions.

Additional information was incorporated into the inventory database on the basis of the expected flood susceptibility map recently developed by Panagiotou (2025a). Particularly, locations with very high susceptibility values (> 0.8) were classified as flooded areas, while locations with very low susceptibility values (< 0.2) were classified as non-flooded areas.

The reasoning behind the collection of data from several watersheds is twofold: a) to compile a sufficient number of entries for training the data-demanding MLMs, and b) to generate a dataset that contains information from several watersheds with different characteristics for the compilation of multi-basin machine learning models that can capture a wide range of flood patterns. Indeed, flooded areas are compact regions that mainly cluster around hydrological networks and within urban units, so nearby points often

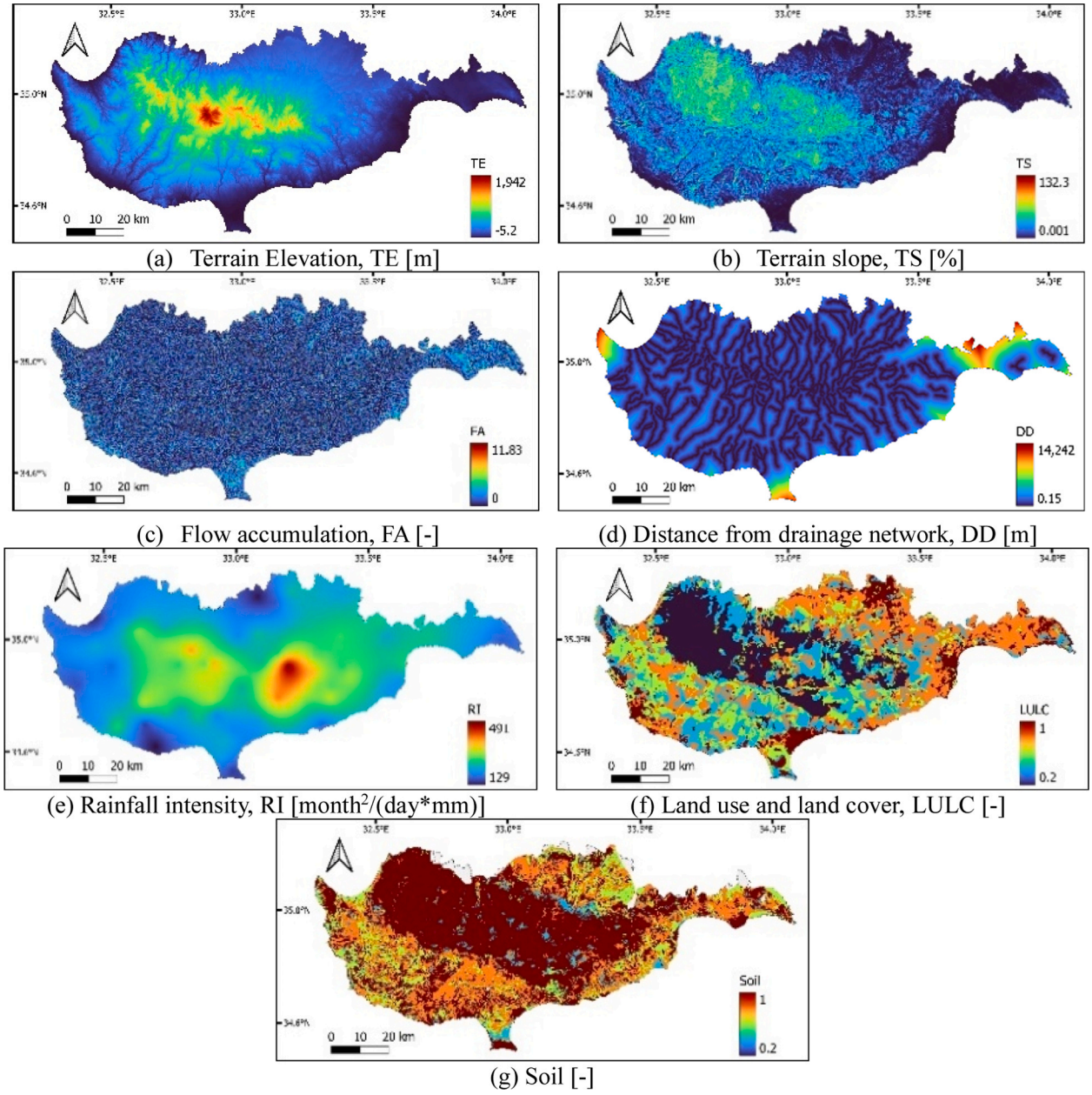

Fig. 3. Geographical distribution of the flood-related factors used to construct the sampling dataset. Actual units are used for the factors that contain numeric data (TE, TS, FA, DD, and RI), whereas ordinal data are converted to scores according to Panagiotou (2025a).

share similar characteristics. The non-uniform spatial distribution of the selected flooded areas provided diverse conditions for model training and allowed robust validation on independent sites.

In the end, eight watersheds were used to compile the training dataset, namely Aradippou, Agios Athanasios, Akrotiri, Garyllis, Germasogeia, Larnaka salt lakes, Pediaios, and Voroklini, whereas, as explained in Section 2, a single watershed (Geroskipou) was used to compile the test dataset. The final database comprised 46,299 data points for the training dataset (31,311 non-flooded areas and 14,988 flooded areas) and 898 data points for the test dataset (449 for each class).

3.3. Multicollinearity analysis and feature selection

Pearson and Spearman correlation coefficients were used to quantify the linear and monotonical associations among pairs of variables, respectively. The former coefficient is commonly applied in continuous data to measure linear correlations, whereas the latter coefficient is more suitable in the presence of outliers and nonlinear patterns in the dataset. Both coefficients range from −1 to

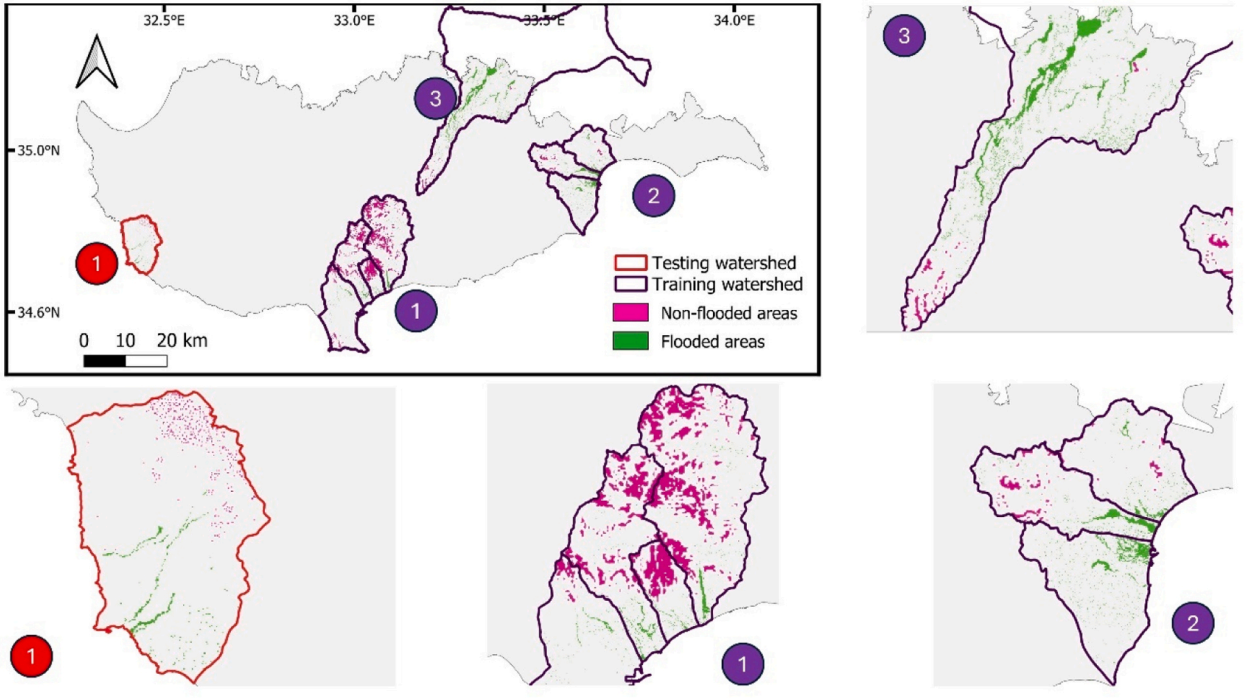

Fig. 4. Geographical distribution of flooded (green-colored points) and non-flooded (purple-colored points) areas used to construct the inventory database, with zoomed views of selected watersheds.

+1, where coefficient values close to zero indicate a weak correlation, values close to −1 indicate a strong negative correlation, and values close to 1 indicate a strong positive correlation.

A multicollinearity analysis was also conducted to identify the flood-related factors used as input variables for model development. Multicollinearity is a common phenomenon in complex environmental phenomena because of the dependencies that inherently exist among variables and are used to model environmental indicators in classification or regression optimization problems. It can prompt a number of deficiencies that make interpretability of the modeling expressions difficult, such as an increase in the standard errors and variance of each coefficient, which can lead to statistically insignificant variables (Chan et al., 2022).

Consequently, it is essential to identify variables that contribute to this issue and exclude them during the model development. Two popular metrics were used to conduct this analysis, namely the variance inflation factor (VIF) and the tolerance index (TOL). The former metric quantifies the inflation of the regression coefficient of a variable due to the existence of correlation among predictor variables in the model, calculated by the following equation:

$$VIF_j = \frac{1}{1 - R_j^2}, \tag{1}$$

where R_j^2 denotes the R-squared (i.e., coefficient of determination) value obtained by regressing the j-th predictor on the remaining predictors. As a rule of thumb, VIF values larger than 5 indicate the presence of strong multicollinearity (Tsangaratos et al., 2023). The latter metric is simply the reciprocal of the VIF and is often used because it provides values within the unit interval [0,1]. Values close to 0 (typically less than 0.1) indicate the presence of strong multicollinearity. More information regarding the multicollinearity analysis is given in the "Supplementary Material" file (see Section "Other").

3.4. Model development and feature importance

Four machine learning algorithms commonly used in environmental classification applications were selected for assessing flood susceptibility, namely support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost) and multilayer perceptron (MLP). Hyperparameter optimization was conducted prior to the training process to ensure a fair comparison among the different MLMs, a process that is often omitted in similar studies (Chen et al., 2021). The leave-one-cluster-out (LOCO) was incorporated in the k-fold cross-validation process to enforce strict spatial/cluster independence between training and validating datasets. In this study, k-1 clusters (watersheds) were used to train the selected models, which were validated in a separate, "unseen" cluster (i.e., the held-out watershed). Consequently, a grid search algorithm and an eight-fold LOCO cross-validation strategy were employed to control the training process for the RF, XGBoost, and SVM algorithms, without applying any early-stopping criteria, except for MLP.

Brief descriptions of these models are provided below.

3.4.1. Support vector machine (SVM)

For classification tasks, SVM aims to identify the optimal decision boundary that maximizes the margin between classes, thereby improving its ability to generalize to unseen data. Regularization terms are applied to penalize the data points that fall outside the margin, ensuring a balance between maximizing the margin and minimizing classification errors. For that purpose, kernel functions are used to implicitly map the input datasets into a higher-dimensional feature space, enabling the separation of classes using a linear decision boundary. SVMs have been proven to be effective in learning nonlinear patterns that are present in high-dimensional datasets, and they are widely used by practitioners for assessing flood risk (Tehrany et al., 2015).

3.4.2. Random forest (RF)

The RF algorithm is an ensemble machine learning technique (Breiman, 2001)) that generates multiple decision trees on the basis of bagging methods for solving both regression and classification problems. For classification problems, each decision tree predicts a class label (and/or probability). The final prediction is obtained by aggregating the outputs of all decision trees, typically through majority vote for class labels or averaging predicted probabilities. The RF algorithm is popular among practitioners for assessing flood risk because of its ability to a) handle high-dimensional datasets effectively, b) provide robust predictions in the presence of noisy information, and c) produce generalized models that achieve very good agreement with the actual data (Wang et al., 2015).

3.4.3. Extreme gradient boosting (XGBoost)

Similar to RF, XGBoost is another ensemble learning model used for both classification and regression tasks (Panagiotou et al., 2024). It adopts a gradient boosting framework to build trees in a sequential manner ("one-by-one"), where each new tree is trained to reduce the residual errors of the previous ones, thereby improving predictive accuracy. This is done by including regularization terms (i.e., L1 and L2-type) in its objective function to control model complexity and reduce overfitting during the training process, allowing it to capture complex nonlinear patterns while maintaining strong generalization performance.

3.4.4. Multilayer perceptron (MLP)

The multilayer perceptron (MLP) algorithm is a feed-forward artificial neural network which is widely used for both regression and classification problems. A MLP comprises a multilayer architecture that usually includes three types of layers: the input layer, one or more hidden layers, and the output layer (Petracca et al., 2022). The input layer contains the features of the sampling dataset, the hidden layers process the input information through a set of nonlinear activation functions (i.e., artificial neurons). The last layer

provides the predictions of the MLP, either as continuous values or discrete classes (Silva et al., 2022).

3.5. Performance and feature importance metrics

This subsection provides a brief introduction to the classification performance metrics used to evaluate the four MLMs, as well as the metrics employed to rank the selected features by their relative importance for the model responses. Additional information can be found in the "Supplementary Material" file.

The performance of the MLMs was evaluated on the basis of 4 classification metrics, which are: Balanced Accuracy (BA), BA t-based 95% confidence interval, F1 score, and Brier Score (BS). To account for class imbalance and prevent misleadingly high scores, BA was selected instead of Accuracy to evaluate model performance. This metric denotes the average Sensitivity and Specificity of model responses. High BA values indicate that the model is highly capable of distinguishing between the negative ("non-flood") and positive ("flood") classes. To account for the uncertainty of the cross-validation folds, the t-based 95% confidence interval of the BA is calculated based on the out-of-fold predictions. The F1 score (i.e., harmonic mean of Recall and Precision) was also adopted to determine the optimal probability cut-off, as it provides a balanced measure of classification performance by penalizing both false negatives (FN) and false positives (FP). High F1 scores indicate good classification performance with a balanced trade-off between Recall and Precision. Moreover, this metric was used to convert the predicted probabilities to binary values, therefore it was used to construct susceptibility binary maps. Particularly, the optimal probability threshold (FS_{thr}) was determined by maximizing the F1 score over all possible thresholds (i.e., within the unit interval [0,1]). BS measures the mean squared error between the model predictions and the observed data. Unlike threshold-dependent performance metrics (e.g., BA, F1 score), BS evaluates both calibration (i.e., how well predicted probabilities reflect the true likelihood of an event) and discrimination (i.e., the model's ability to separate positive and negative classes). Low BSs indicate better calibrated and more accurate predictions. These classification metrics were used to train and test the four MLMs under different hyperparameter configurations. The best-performing model from each algorithm was then used to construct binary susceptibility maps (4 maps in total). Subsequently, the best overall model was used to produce a five-level susceptibility map based on the optimal probability threshold and common practices in flood-susceptibility mapping.

Feature importance analysis was also conducted to identify the most significant features of the trained MLMs. With respect to the XGBoost algorithm, the relative significance among the input variables was quantified on the basis of the 'Gain' of a feature which measures the improvement in accuracy brought by a feature to the branches it is on. Gini impurity is adopted for RF algorithms to determine the best split across individual decision trees within the ensemble. The increase in node purity is calculated by the decrease in the total sum of squared errors when a variable is selected for splitting. The absolute value of the model coefficients was used to determine the feature significance in the SVM since this model becomes linear in the higher-dimensional variable space. For the MLP, the evaluation of feature importance is less direct because of the nonlinear transformations and multiple hidden layers. In this study, SHAP values (Abioye et al., 2025) were used to estimate the relative contribution of each input variable. SHAP is based on cooperative game theory and assigns an importance value to each feature by measuring its marginal contribution to the prediction across all possible combinations of features. This method is well suited for neural networks because it offers a clear and general way to explain their predictions. The resulting SHAP values allowed both the ranking of features by relevance and the visualization of their influence, offering insight into the internal decision process of the MLP.

Next, the most significant features of the best algorithm were used to develop simplified MLMs. This was achieved by sequentially excluding the less important features and repeating the model development process. Lastly, partial dependence plots were constructed to assess the directionality of the input variables in the predicted variable, enabling the determination of threshold values beyond which susceptibility values exhibit sharp variations. The marginal effect of the four most important features on the predicted outcome of the MLM was assessed either one-feature-at-a-time or in pairs.

The training and testing process for the three shallow MLMs (RF, XGBoost, and SVM) was conducted in R-Studio environment with the use of **caret** library. The **vip** (SVM), **varImp** (RF) and **xgb.Importance** (XGBoost) R libraries were used to rank the input variables, whereas **pdp** library was used to construct partial dependence plots. The modeling development of MLP was carried out in a Python environment with **scikit-learn** library for data standardization, **pytorch** package for neural network design and training, and the **SHAP** library for feature importance analysis.

4. Results

4.1. Multicollinearity analysis and feature selection process

Table 2 presents the results of the pairwise correlations among the seven flood-related factors on the basis of the training dataset with the use of Pearson and Spearman coefficients (correlation plot is also shown in Figure SM5). Both coefficients captured strong positive correlations (> 0.70) between terrain slope and terrain elevation. This observation is consistent with the topographical characteristics of the island of Cyprus: coastal regions are dominated by flat terrains and elevations close to sea level, whereas higher altitudes and steep terrain profiles are observed in the mountainous inland part. A strong negative correlation was observed between LULC and terrain elevation/slope (less than -0.67 for both correlation metrics), reflecting the presence of forests and dense vegetation in high-altitude mountainous regions, which do not favor flooding. A moderate link between rainfall intensity and terrain elevation was also observed, attributed to the more intense and frequent rainfall events that occur in high-altitude mountainous regions. Weak correlations were observed for the remaining pairwise comparisons, especially for pairs [DD, FA] and [DD, Soil] which are practically uncorrelated. With respect to the Boolean variable (1 represents "flood" class, whereas 0 represents "noflood" class), the LULC factor

exhibits the strongest positive correlation, with values 0.77 and 0.83 for the Spearman and Pearson coefficients, respectively, followed by flow accumulation. Moderate-to-strong negative correlations are found with respect to terrain slope and elevation, indicating that high values of these factors are associated with a lower probability of flood occurrence. In addition, the frequency distributions of the flood indicators show the presence of outliers and highly skewed distributions in the dataset (Figure SM6 in the Supplementary material).

Table 3 shows the VIF and TOL values for each flood indicator, revealing the absence of strong multicollinearity effects. As expected, the largest VIF values are associated with TE, TS, and LULC, because these variables are associated with the strongest pairwise correlation coefficients. Nevertheless, the VIF values are much lower than the threshold value (5) even in the presence of high pairwise correlations, indicating the absence of strong correlations between the response variable and the explanatory variables via multi-regression expressions (see Section 3.3). With respect to the TOL, all variables have values beyond the threshold value (0.1), whereas the smallest values are observed for TE, TS, and LULC. Consequently, all flood indicators were included in the feature dataset.

4.2. Comparison of the machine learning models

Table 4 displays the optimum hyperparameter values obtained during the optimization process for each MLM, whereas the entire set of hyperparameter values considered during the model development is given in Table SM3. In Table 4, for the MLP model, the optimal number of units in the hidden layer is either 15 or 20, depending on the LOCO fold considered. The selected hyperparameter values were then used to train and test the MLMs on the basis of the sampling dataset, as described in Section 3.2.

Table 5 reports the results achieved by the evaluated MLMs over a) the training dataset, considering LOCO cross-validation strategy, and b) the test dataset, which includes data collected from a single, "unseen" watershed (Geroskipou). The results are evaluated using the performance metrics described in Section 3.5, i.e., BA, BA t-based 95% confidence interval, F1 score, and BS. For the training dataset, the optimal probability threshold FS_{thr} is also reported, whereas the performance estimates from the out-of-fold predictions are used to estimate the mean (μ_{BA}) and 95 % confidence interval ($\boldsymbol{95\%\ CI_{BA}}$) of the BA. For the test dataset, BA, F1 score, and BS score are reported. For completeness, Table SM4 reports the elements of the confusion matrix for all MLMs. Overall, performance achieved by the four MLMs is aligned, and higher than 0.95 in terms of μ_{BA}, BA, and F1 score, for both the training and test datasets. The MLP model exhibits the best performance in terms of μ_{BA} for the training dataset; however, SVM is the best performing model for the test dataset, with a BA of 0.995 and an F1 score of 0.996. With respect to Balance accuracy, RF exhibits similar accuracy to MLP ($\mu_{BA} = 0.974$) for the training dataset, whereas it ranks second behind SVM.

Table 6 shows the feature importance scores of the trained MLMs. The consistency among the tuned MLMs in terms of the identification of the most important features (i.e., LULC, TS, FA, and TE) indicates that a subset of the feature dataset can be used to build simplified models, allowing to a) reduce model complexity, and b) collect high-fidelity information for a smaller number of features (i.e., dimensionality reduction). In alignment with the correlation analysis, all MLMs proposed LULC and TS as the two most influential flood-related factors, an observation that supports the robustness of the results. This observation is consistent with the fact that these parameters are well known for being strongly associated with susceptibility levels because of their direct influence on flood processes. Especially in urban context, the likelihood of flood occurrence is elevated due to the simultaneous presence of impermeable surfaces (pavements, asphalt) and dense infrastructure. This situation is further amplified in flat terrains, where water tends to accumulate. Such conditions characterize the coastal part of the Geroskipou watershed, where the city of Paphos is located. TE and FA are also important, particularly for RF, SVM, and MLP, while RI, DD, and soil properties have comparatively lower influence, although SVM assigns higher weight to RI. Overall, XGBoost shows slightly different feature rankings compared to RF, SVM, and MLP, most likely because of its sequential boosting process that emphasizes features that most reduce the residual error in early trees. All MLMs predicted negligible contribution of the soil factor, indicating that this factor contains trivial information and can be excluded from the modeling development. Another factor that is shown to play a minor role in the resulting models is the DD, especially in the case of RF, SVM, and MLP. These findings provide a different perspective than a previous study that used feedback from multiple stakeholders to select and rank the same set of flood factors to conduct flood susceptibility assessment (Panagiotou, 2025a). Compared with subjective weighting methods, data-driven approaches enable practitioners to reduce the influence of biased stakeholders' opinions in feature importance analysis and facilitate the decision-making process.

Table 2
Pearson (upper triangular) and Spearman (lower triangular) correlation coefficients for all pairs of flood factors. The strongest correlations are highlighted in bold text.

	TE	TS	FA	DD	RI	LULC	Soil	Boolean
TE	1	**0.70**	-0.32	-0.26	**0.63**	**-0.67**	0.40	**-0.59**
TS	**0.76**	1	-0.39	-0.17	0.38	**-0.75**	0.41	**-0.78**
FA	-0.37	-0.33	1	-0.07	-0.13	0.39	-0.19	0.53
DD	-0.15	-0.14	-0.03	1	-0.33	0.09	0.04	-0.05
RI	**0.57**	0.37	-0.14	-0.06	1	-0.33	0.26	-0.15
LULC	**-0.75**	**-0.75**	0.39	0.10	-0.45	1	-0.39	**0.83**
Soil	0.44	0.42	-0.21	0.11	0.37	-0.44	1	-0.34
Boolean	**-0.67**	**-0.79**	0.53	0	-0.15	**0.77**	-0.34	1

Table 3
Multicollinearity analysis on the basis of the variance inflation factor (VIF) and tolerance index (TOL).

Flood indicators	TE	TS	FA	DD	RI	LULC	Soil
VIF	2.97	2.67	1.23	1.17	1.71	2.47	1.28
TOL	0.34	0.38	0.82	0.85	0.59	0.41	0.79

Table 4
Selected values of the hyperparameters for the machine learning models. No early stopping criteria are applied (except for MLP), whereas the Leave-One-Cluster-Out (LOCO) cross-validation scheme is employed in all models.

Model	Hyperparameter	Optimum values
SVM	Cost parameter that controls the trade-off between maximizing the margin and minimizing classification error	0.7
RF	Number of predictors randomly sampled at each split	2
	Splitting rule	Extratrees
	Minimal node size for splitting	6
XGBoost	Number of boosting iterations	100
	Learning rate (contribution of each tree in the ensemble)	0.1
	Maximum depth of a decision tree	4
	Minimum loss reduction	0
MLP	Number of layers	3
	Hidden size	15/20
	Learning rate	0.001
	Number of epochs (maximum)	500
	Activation function	ReLU
	Optimizer	Adam
	Loss function	BCE With LogitsLoss

Table 5
Prediction performance of the MLMs with respect to the training and test dataset. LOCO denotes the Leave-One-Cluster-Out cross-validation strategy, μ_{BA} and 95% CI_{BA} denote the mean and the t-based 95 % confidence interval of BA across the out-of-fold performance estimates, respectively. FS_{thr} denotes the susceptibility threshold.

Model	Training dataset (LOCO Cross-validation strategy)			Test dataset (Geroskipou watershed)		
	μ_{BA}	95% CI_{BA}	FS_{thr}	BA	F1 score	Brier score
XGBoost	0.971	[0.939, 1]	0.21	0.952	0.949	0.033
RF	0.974	[0.946, 1]	0.48	0.991	0.991	0.015
SVM	0.973	[0.954, 0.992]	0.12	**0.995**	**0.996**	**0.003**
MLP	**0.984**	[0.965, 1]	**0.11**	0.990	0.991	0.005

4.3. Assessment of flood susceptibility

To predict the susceptibility values across the entire testing watershed, the performance of all MLMs was evaluated. The results showed that SVM and MLP achieved the best agreement with the flood inventory, closely followed by RF. Although SVM slightly outperforms RF on the test dataset, the latter model was selected for compiling flood susceptibility maps of the Geroskipou watershed because: a) RF can capture complex nonlinear relationships more effectively than SVM and is more robust to noise in the input data, making it better suited for retraining and testing in other Mediterranean regions with potentially different environmental conditions; and b) RF provides an optimal probability threshold close to 0.5 (specifically, 0.48), which facilitates the generation of multi-class susceptibility maps that are easier for policy-makers to interpret and use in decision-making. With respect to MLP, RF is: a) much more efficient in handling computational costs, and b) less demanding in terms of available data, making it easier to apply in other data-scarce environments.

Nevertheless, FS binary maps are also constructed with the use of the remaining MLMs (Figure SM7). Fig. 5 shows the spatial distribution of flood-susceptibility scores, which are used to partition the study area into distinct zones on the basis of five susceptibility intervals: 0–0.2 [“Very Low”], 0.2–0.4 [“Low”], 0.4–0.6 [“Moderate”], 0.6–0.8 [“High”], and [0.8–1.0] [“High”]. The selection of the intervals is justified by the fact that the optimal flood-susceptibility threshold (FS_{thr}) is close to 0.5 (0.48), indicating that regions with moderate likelihood of flood occurrence are associated with FS values around this cut-off. Moreover, partitioning the study area into multiple susceptibility zones is common practice in flood-susceptibility studies (Panagiotou, 2025a; Panagiotou et al., 2025b), and is particularly useful for policy-makers to prioritize areas when designing mitigation and management actions. To facilitate visual interpretation of the results, the mean and median values of the flood-related factors for each susceptibility zone are calculated and reported in Table 7.

The results show that regions less likely to experience flood events are located in the northern and east-central parts of the watershed, where the highest altitudes and terrain slopes are present (Figure SM2). These regions are mainly rural areas and are

Table 6
Feature importance scores (FISs) of the trained MLMs with respect to the test dataset.

Feature*	Importance Score (Rank)			
	XGBoost	RF	SVM	MLP
TS	**0.65 (1)**	75.12 (2)	**100 (1)**	**0.21** (1)
LULC	0.18 (2)	**100 (1)**	97 (2)	0.13 (2)
RI	0.07 (3)	4.97 (5)	59 (5)	0.04 (5)
FA	0.03 (5)	11.86 (4)	83 (4)	0.05 (4)
TE	0.03 (5)	20.90 (3)	91 (3)	0.12 (3)
DD	0.04 (4)	1.03 (6)	0.5 (6)	0.01 (6)
Soil	0 (7)	0 (7)	0.69 (7)	0.01 (7)

* TE: Terrain elevation (DEM); RI: rainfall intensity; FA: flow accumulation; DD: drainage distance; TS: terrain slope; LULC: land-use land cover; Soil: soil properties (texture and depth).

covered by dense vegetation and forests, which are associated with intense agricultural and economical activities. The major branches of two ephemeral rivers, namely Limnaria and Kochinas, pass across the majority of the central part of the study area. As a result, the central and southern parts of the watershed are characterized by low and moderate DD scores, whereas the highest scores are observed in the western coastal part. The results indicate that half of the study area is expected to experience flooding. Specifically, 36.9 % of the study area falls into the "Very-High" susceptibility class, whereas 13 % is classified as "High". Small spatial variations in rainfall intensity are observed across the entire watershed, except in the northeastern part, where the lowest rainfall scores are present. The smooth spatial patterns of this factor are mainly attributed to the interpolation of RI values based on a sparse network of rainfall stations (Figure SM2).

With respect to the soil aspect, moderate FS scores are present in the majority of the northeastern part which is characterized mainly by a medium soil texture. The southern-central part of the watershed is covered mostly by dense urban units and construction sites, which are associated with impermeable soil surfaces and buildings, whereas the southern areas are covered by arable land.

The sensitivity of the classification to the susceptibility cut-off (FS_{thr}) was evaluated by comparing the RF binary map obtained using the F1-maximizing probability threshold, with binary maps generated using three different susceptibility percentiles (i.e., 25th, 40th, and 60th), thereby assessing the robustness of model predictions to threshold selection (Figure SM8). This comparison reveals that the southern and western parts of the test dataset, particularly near the coastline, are consistently classified as "flood" regardless of threshold selection. This observation supports the robustness of the model predictions, indicating that these areas are expected to be highly prone to flood events.

Next, different combinations of features were selected to build RF models using subsets of the available information, aiming to provide models with lower computational complexity than the optimum model without significantly compromising performance. This is done by ranking the features in descending order according to the feature importance analysis that was conducted with the use of the RF algorithm (see Table 6) and repeating the modeling development process, each time by excluding the less important features one-by-one (Table 8). As in the previous analysis, the LOCO cross-validation strategy was employed again to evaluate the performance of the resulting models. The relative contribution of each feature is better illustrated on the basis of the cumulative feature importance

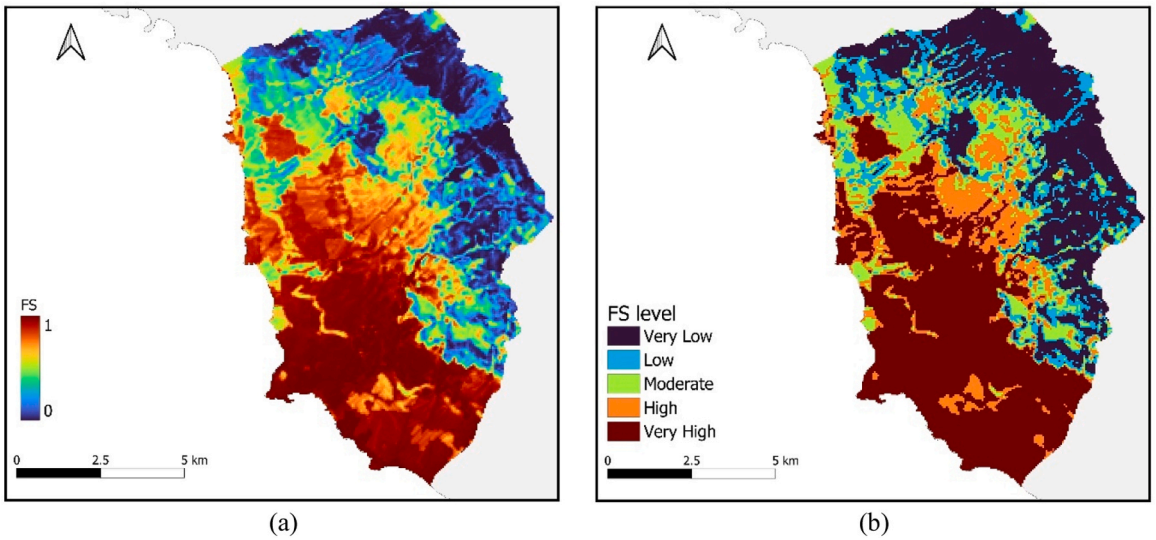

Fig. 5. Geographical distribution of flood-susceptibility (a) scores and (b) levels. FS denotes the flood susceptibility indicator. Risk levels are determined with respect to FS as: "Very Low" [0–0.2], "Low" [0.2–0.4], "Moderate" [0.4–0.6], 0.6–0.8 ["High"], and [0.8–1.0] ["High"].

Table 7
Median/mean values of the features for each susceptibility class in the Geroskipou watershed.

Feature*	Median/Mean feature values at each susceptibility zone				
	Very Low [0–0.2]	Low [0.2–0.4]	Moderate [0.4–0.6]	High [0.6–0.8]	Very High [0.8–1.0]
TE	358.70/360.12	223.37/240.20	171.25/177.86	158.76/145.32	61.61/61.53
TS	20.48/22.30	10.55/11.99	6.78/7.71	5.55/6.36	3.18/3.70
FA	0.86/1.11	1.28/1.52	1.56/1.77	1.65/1.88	1.83/2.00
DD	899.08/931.54	1160.20/1287.34	1273.30/1477.92	832.14/1184.68	1109.16/1182.77
RI	202.09/198.23	198.52/193.79	193.88/192.32	195.99/194.57	197.88/197.30
LULC	0.60/0.63	0.67/0.74	0.74/0.75	0.85/0.86	1.00/0.94
Soil	0.80/0.79	0.75/0.76	0.75/0.77	0.80/0.77	0.80/0.77

* TE: Terrain elevation (DEM); RI: rainfall intensity; FA: flow accumulation; DD: drainage distance; TS: terrain slope; LULC: land-use land cover; Soil: soil properties (texture and depth).

percentage (CFIP) index, expressed by

$$CFIP_{n_{sub}} = 100\% \times \frac{\sum_{j=1}^{n_{sub}} FIS_j}{\sum_{i=1}^{n_{cr}} FIS_i}, \tag{7}$$

where n_{sub} denotes the number of features in the subset, and n_{cr} denotes the total number of criteria (7 in this study). In principle, the performance of the model deteriorates each time a new feature is excluded, with the largest performance reduction associated with the subtraction of the RI. As expected, excluding the Soil factor had a negligible impact on model performance, as this parameter contributes practically no information to the CFIP (see Table 6). The BA values decrease with a reducing number of features, ranging from 94 % to 97 % across all feature combinations. This observation is supported by the fact that a large portion of the feature-importance information (CFIP=91 %) is contained in three flood-related factors, namely LULC, TS, and TE. This finding is also consistent with the fact that these features exhibit the strongest correlations with the Boolean target variable (see Table 2). The optimal probability cut-off ranges from 0.41 to 0.55 across all feature subsets, further supporting the robustness of the model predictions according to the threshold selection. With respect to the test dataset, a similar behavior is observed for the BA values, while predictive errors increase as the number of features is reduced, reaching a Brier score of 0.054 for [LULC, TS, TE] combination. Overall, all feature combinations yield accurate results, demonstrating that simplified models can be effectively used for flood susceptibility assessment.

To further assess the influence and directionality of the features on the target variable, partial dependence plots (PDPs) are constructed on the basis of the predictions of the best algorithm (RF). These plots offer supplementary insights with respect to feature importance analysis, allowing the evaluation of the model sensitivity to a single feature while keeping the values of the remaining input variables of the training dataset fixed. Fig. 6 presents the partial dependence of the susceptibility scores on the four most significant features, particularly LULC, TS, TE, and FA.

With respect to LULC, the highest variations in FS are observed for LULC values larger than 0.8, whereas a sharp change in the monotonic trend (from decreasing to increasing) occurs at 0.4. This observation suggests that the replacement of LULC types with scores equal to 0.8 (e.g., arable crops) with less flood-prone LULC types (e.g., vineyards, olives) can have a significant impact on highly susceptible regions. A monotonic trend is observed for the terrain slope, characterized by sharp reductions in the susceptibility values for low TS values, whereas the impact of this feature on the predicted outcomes becomes trivial beyond 50 %, after which a plateau is reached. This observation illustrates the fact that the rapid motion of gravity-driven flows in steep-slope terrains limit water accumulation and therefore reduces flood potential. A monotonic trend is also observed with respect to FA, where the largest increasing trend of FS is observed for FA values of approximately 5.8, supplemented by a plateau profile that occurs for values larger than 9. This observation is consistent with the fact that regions where large amounts of water are expected to accumulate during extreme precipitation events, are more likely to face flood issues. This is exactly the situation in the urban units which are located in the coastal part of the Geroskipou watershed, where high FA values and flat terrains are simultaneously present. With respect to the TE, a sharp

Table 8
Performance scores for RF models built based on different subsets of the feature dataset.

n_{sub}	Subset of features*	CFIP (%)	LOCO Cross-validation			Test dataset		
			μ_{BA}	95% CI_{BA}	FS_{thr}	BA	F1 score	Brier score
7	LULC+TS+TE+FA+RI+DD+Soil	100	0.974	[0.946, 1.0]	0.47	0.991	0.991	0.016
6	LULC+TS+TE+FA+RI+DD	100	0.975	[0.951, 1.0]	0.48	0.989	0.989	0.016
5	LULC+TS+TE+FA+RI	99	0.969	[0.936, 1.0]	0.47	0.986	0.986	0.019
4	LULC+TS+TE+FA	97	0.956	[0.908, 1.0]	0.41	0.951	0.953	0.033
3	LULC+TS+TE	91	0.943	[0.886, 0.99]	0.55	0.921	0.924	0.054

* TE: Terrain elevation (DEM); RI: rainfall intensity; FA: flow accumulation; DD: drainage distance; TS: terrain slope; LULC: land-use land cover; Soil: soil properties (texture and depth).

reduction in the decreasing trend occurs close to 350 m above sea level (m.a.s.l), followed by a smoother variation at TE values approximately 700 m.a.s.l. Fig. 6e-f displays the partial dependence of the two most significant pairs of input variables, particularly [TS, LULC] and [TE, LULC], which exhibit distinct discrepancies with respect to the variability of the susceptibility score. High susceptibility scores are observed with increasing LULC values, especially those above 0.4, as long as small TS values (less than 10 %) are present. This pattern is particularly evident in the coastal, urban part of the watershed, which is characterized by flat terrains (i.e., low TS values) and the highest LULC scores. With respect to the [TE, LULC] pair, high susceptibility values occur in the simultaneous presence of high TE and LULC. Compared with the [TS, LULC] pair, higher FS values are observed for LULC scores above 0.5 regardless of the TE values. This observation indicates the stronger dependence of FS to the land cover compared to terrain elevation, as shown in Table 6. Moreover, it is consistent with the fact that the FS reaches a plateau earlier for the TS than for the TE (Fig. 6b-c).

5. Discussion

This study explored the applicability of four machine learning algorithms, particularly the SVM, RF, XGBoost, and MLP, to assess flooding in the Geroskipou watershed. The major flood drivers were identified through multicollinearity analysis, and used to generate susceptibility maps, which achieved very good agreement with the flood inventory data. Consequently, the findings of this study can assist future research and planning, and need to be further discussed.

5.1. Guidance for flood susceptibility management

Although the highest susceptibility is expected in urban areas, moderate-to-high probabilities are predicted in semi-urban and rural areas where important economic activities and land use developments are currently present. Particularly, rapid land use changes have occurred due to the expansion of urban areas and the adaptation of water consumption patterns to climate conditions. Moreover, the extensive drought conditions that have intensified during recent years have led to a reduction in irrigated land because of reduced water availability and the development of economic sectors with lower water demands, such as domestic and tourism.

The emergence of extreme wildfires has further contributed to the acceleration of land cover changes, with most notably, a very recent (June 2025) fire event that completely burnt approximately 124 km^2 (1.3 % of the entire area of Cyprus) (EFFIS, 2025). These events are expected to increase soil erosion and favor surface runoff, hence increasing the probability for flood events occurring and leading to the transition of some regions toward higher susceptibility zones. Consequently, additional scenarios should be conducted to assess the impact of these changes on susceptibility levels, whereas revised mitigation action plans should be compiled in rapidly developing rural regions on the basis of scientific evidence. In alignment with the EU flood directive (Article 7), the resulting maps and action plans should be integrated with early warning systems and forecasting tools to enhance the effective coordination among multiple key actors, including researchers, first responders, local authorities, and policy-makers.

Moreover, near-real time satellite data should be combined with model outputs and in situ information to support reliable products for flood monitoring and forecasting, such as Google Flood Hub[1] and Copernicus Global Flood Awareness System.[2] In this study, different satellite missions (i.e., Sentinel, Planet) were unsuccessfully explored to identify historical flood events within the entire domain of Cyprus. Nevertheless, with the rapid technological advancement of the satellite industry, and the recent installation of the first Earth Observation Data Acquisition Station in Cyprus,[3] this information is expected to become more easily accessible for use in forthcoming research works and studies.

LULC, TS, and TE are found to be the most influential drivers with respect to the feature importance analysis, in alignment with the findings of the correlation analysis in Section 4.1. These factors constitute major drivers of surface run-off and water accumulation. For example, high LULC values are observed in the coastal city of Paphos, where a dense network of impervious surfaces and buildings reduces infiltration capacity and increases surface run-off, thereby amplifying flood peaks. Moreover, flat terrain dominates the urban context, further enhancing the accumulation potential. These results are consistent with existing studies that highlight the essential role of topographical and land-cover characteristics in controlling hydrological processes. Among these factors, it is suggested that the revised plans should focus on land use development for better flood risk management. For example, urban units can be constructed at relatively high altitudes to decrease the likelihood of persistent waterlogging (Zarekarizi et al., 2020), promote green infrastructure activities that include reforestation, especially in burnt areas, wetland restoration, and the installation of nature-based solutions, for example managed aquifer recharge (Dillon et al., 2019; Panagiotou, Eisenreich, et al., 2024), which are able to capture water and store it in subsurface systems.

5.2. Limitations of the current approach and recommendations

Flooding is a complex and highly dynamic physical phenomenon that evolves according to the interactions among climatic and non-climatic processes. This study focused on a limited number of physical factors (seven in total) which inevitably cannot fully reflect the characteristics of flood susceptibility. Additional factors that contain socioeconomic information (e.g., population age, income and density, land and infrastructure value) are needed to quantify the influence of the remaining risk components (i.e., vulnerability,

[1] https://sites.research.google/floods/l/0/0/3
[2] https://global-flood.emergency.copernicus.eu/
[3] https://cyprus-mail.com/2025/07/02/new-earth-observation-data-antenna-arrives-in-cyprus

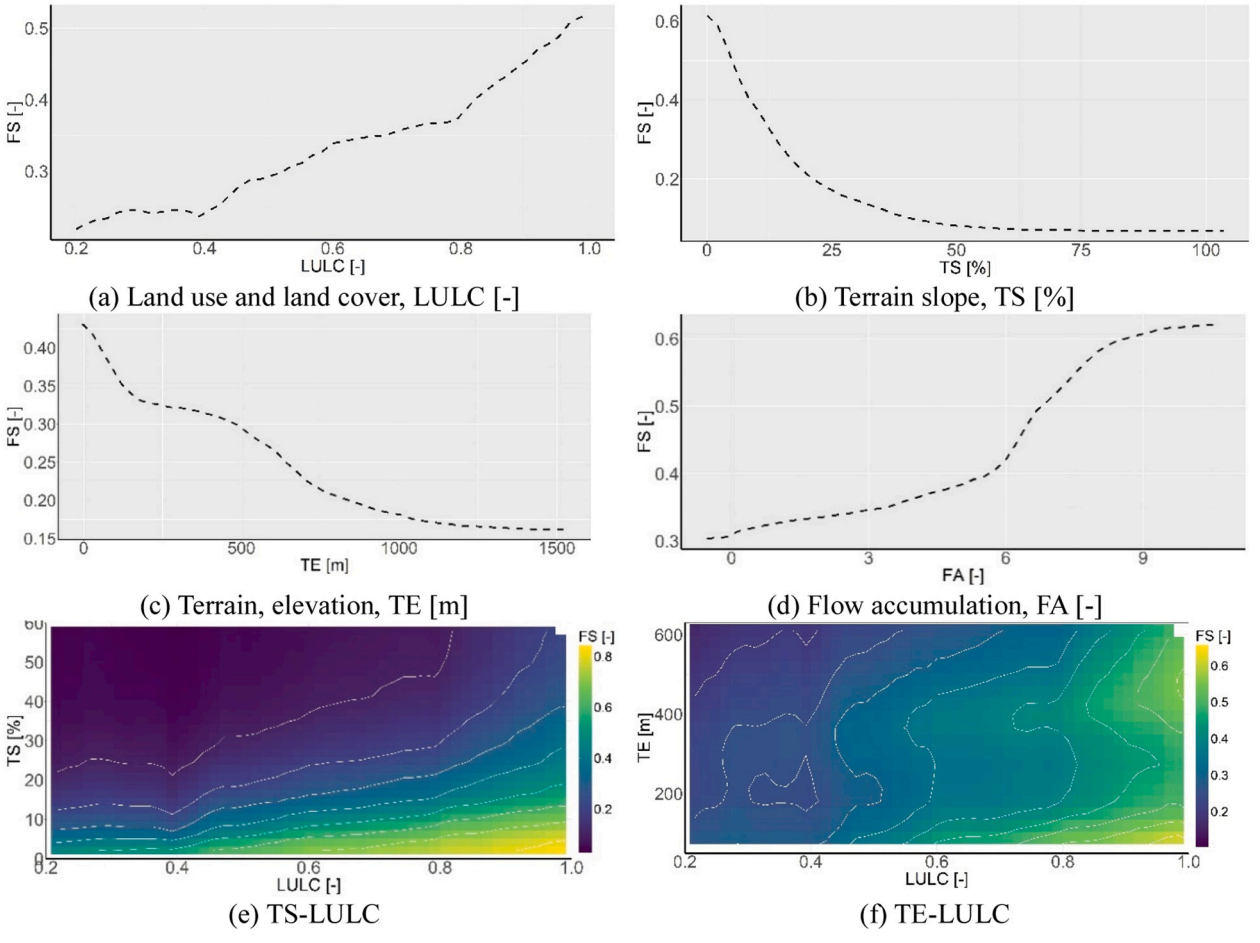

Fig. 6. (a) Partial dependence plots for the major input variables. Variations are conducted to (a)-(d) single features and (e)-(f) pairs of features. The vertical axis (a-d) and the contour values (e-f) denote the susceptibility values.

exposure), enabling policy-makers to assess flood risk in a holistic manner. This study focused on developing a multi-basin model with general applicability within Cyprus, producing flood susceptibility maps that can serve as screening tools for the Cypriot authorities because of the difficulty of collecting high-fidelity socioeconomic information. Consequently, future studies should aim to include this information with sufficient spatiotemporal resolution and coverage. Although the MLMs were trained and tested using data collected from multiple watersheds (nine in total), all algorithms achieved high agreement with the inventory database. To ensure the reliability of these results, several diagnostic checks were conducted to identify potential deficiencies in the model development process, including data linkage errors, duplicated records, and excessively high correlations between the Boolean variable and individual predictors. These analyses confirmed that none of these issues is responsible for the high performance observed in the resulting models. Therefore, additional training and testing should be conducted using data collected from other Mediterranean regions to enhance the general applicability of these models, particularly in areas with topographical, meteorological, and geological characteristics that differ significantly from those of Cyprus.

Another limitation of this study is that it does not account for the uncertainties that inherently exist in the sampling dataset and propagates through the modelling expressions to the target outputs. A plethora of sources of uncertainty are present in the input dataset, which are associated with the spatiotemporal resolution (e.g., interpolation of rainfall data), three-dimensional effects (e.g. Euclidean distance from the drainage network), estimations of subsurface parameters (e.g., soil properties), the dynamic nature of land use and land cover, etc. With respect to the most important factor, LULC, the Corine Land Cover inventory correctly predicts more than 85 % of the land use categories in all biogeographical areas, except for the Black Sea territory (Büttner et al., 2021). Therefore, the errors associated with this factor are not expected to have a significant effect on the model outputs. With respect to the topographical factors (i.e., TS, TE, and FA), multiple correction schemes are applied to the original DEM, such as filling holes, pits, and depressions, via the **whitebox** library in R prior to constructing the sampling dataset. These corrections have led to an improved, more connected representation of the terrain elevation. With respect to rainfall intensity, deterministic interpolation methods are used to estimate daily rainfall values on the basis of a limited number of ground-based stations. Future studies should focus on acquiring rainfall data with higher resolution in terms of spatial coverage to reduce the uncertainty of the predictions far from the stations, whereas subdaily information should be considered to capture the dynamics of the short duration of flash flood events. Moreover, additional information should be included in the drainage network, such as roads and closed channels, since these infrastructures can also affect the evolution of surface runoff.

With respect to the cross-validation process, this study adopted a Leave-One-Cluster-Out (LOCO) approach to account for spatial autocorrelation by ensuring that the trained model is evaluated on an unseen dataset (i.e., a held-out watershed). This approach is more appropriate than other spatially sensitive approaches, such as spatial or block cross-validation, for problems involving multiple disconnected geographical areas. Nevertheless, alternative sampling strategies can be employed to better characterize inherent uncertainty. For example, bootstrap resampling can generate multiple realizations of spatial folds, which can be used to derive distributions of performance metrics. While this study provides estimations of confidence intervals for performance metrics and assesses the partial-dependence of model responses to major flood-related factors, incorporating spatial bootstrap framework into the modeling process could provide an additional layer of uncertainty quantification. Moreover, mapping predictive entropy and standard deviation could help the identification of regions where model predictions are less reliable.

Another challenge in this study was the compilation of a reliable flood inventory database due to the lack of satellite data and sufficient details on historical data (e.g., lack of geographical coordinates of the recorded events). As already mentioned in Section 2, Cyprus consists mainly of small-scale water basins (< 100 km^2) which favor the occurrence of short-duration extreme rainfall events (flash floods), making it difficult for satellite missions to sufficiently capture the dynamics of such events. Therefore, sub-daily meteorological information is required to capture the temporal patterns of the rainfall intensity. A limited number of ground-based stations equipped with hourly rainfall gauges is currently installed at a national scale, hindering their use for producing reliable maps of the rainfall intensity. With respect to other sources of information, there are currently two radar stations in Cyprus which are monitored by the Department of Meteorology. As part of future work, a meticulous processing of the radar data will be conducted and used to generate RI maps. Consequently, data should be collected from different satellite repositories and fused with in situ information to achieve the necessary spatiotemporal resolutions for monitoring these events.

Additionally, only a limited range of model hyperparameter combinations was considered in this study due to computational constraints, leading to the determination of local optimum values within the grid search space rather than global optima. As a result, this shortcoming might lead to biased comparisons among the different models. In this study, all models selected the same pair of features (LULC, TS) as the most important flood-related factors on the susceptibility scores, indicating that the tuned models captured the general patterns of the sampling dataset.

Another factor that needs to be considered is the impact of the sampling size on the performance of the MLMs. A wide range of sampling sizes is found in similar studies (Chen et al., 2021; Tsangaratos et al., 2023) for several MLMs, leading to different conclusions regarding the selection of the optimal algorithm. Consequently, additional computations are performed in this study at different proportions of the training dataset (32.4 %, 10.8 %, and 3.2 %) to quantify the impact of the sampling size on the performance of the MLMs (Figure SM9, Table SM5). The results showed a weak influence of the sampling size on the performance of the MLMs, further supporting the robustness of the predictions.

6. Conclusions

This study explored the applicability of four multi-basin MLMs to assess flood susceptibility which are trained on the basis of data collected from eight watersheds and tested on “unseen” data collected from the Geroskipou watershed on the basis of governmental

reports and existing studies. Furthermore, this study is the first to apply machine learning algorithms to assess flood susceptibility with the use of MLMs in Cyprus. The main conclusions are summarized as follows:

- MLP neural network and SVM exhibited the highest performance among the MLMs, whereas RF achieved similar results across several classification performance metrics. RF was selected for compiling susceptibility maps because it enables the partitioning of predicted probabilities into multiple susceptibility classes, providing a fast-screening tool for policy-makers when prioritizing among different areas.
- Feature importance analysis revealed that LULC and TE are the most significant flood factors for all algorithms, supporting the fidelity of the resulting models.
- Half of the regions within the test watershed are highly prone to flood events, whereas the majority of the remaining regions are less likely to experience flooding. Particularly, the regions that are least prone to flooding are observed in the northern and eastcentral parts of the watershed. TE and TS exhibit their lowest median values in the high-risk zone, whereas the opposite is observed for LULC, providing valuable information to policy-makers for future planning.
- The simplified RF models on the basis of the four most important features, particularly LULC, TS, TE, and FA, achieved very good agreement (BA=95 %) with the flood inventory data, whereas the mean squared error (Brier score) is 0.033. Consequently, these models can serve as computationally effective flood screening tools in data-scarce environments.
- The directionality assessment of the model predictions revealed threshold values for major flood indicators (LULC, TS, TE, FA) that are associated with large changes in the susceptibility levels. With respect to LULC, these values can be used by the authorities to design land use development policies that take into account flood mitigation planning.

This study demonstrates that shallow MLMs can be used instead of advanced, deep learning algorithms, as screening tools to classify susceptibility levels in terms of a limited number of major flood indicators. Particularly, simplified MLMs such as RF and SVM can assist policymakers as rapid flood screening tools, being also suitable for near-real time applications due to their computational efficiency, and ability to provide quick predictions. In contrast, fully detailed hydrodynamic models are associated with high computational burden and require vast amounts of information and time. This is particularly useful in data-scarce environments, which may not provide sufficient data to train a deep learning model. Moreover, these data-driven models can provide subjective predictions without incorporating the biased and often conflicting perspectives of different policy-makers, which are frequently required in decision making processes.

Funding

This work was supported by the European Union's Horizon Europe Framework Programme HORIZON-WIDERA-2021-ACCESS-03 (101079468) and the EXCELSIOR H2020 Widespread Teaming project funded from the European Union's Horizon 2020 research and innovation programme (857510).

CRediT authorship contribution statement

Davide De Santis: Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Panagiotou Constantinos:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Giorgia Guerrisi:** Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Fabio Del Frate:** Writing – review & editing, Funding acquisition. **Marios Tzouvaras:** Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The present work was conducted in the framework of AI-OBSERVER project (https://ai-observer.eu/) titled "Enhancing Earth Observation capabilities of the Eratosthenes Centre of Excellence on Disaster Risk Reduction through Artificial Intelligence", that has received funding from the European Union's Horizon Europe Framework Programme HORIZON-WIDERA-2021-ACCESS-03 (Twinning) under the Grant Agreement No. 101079468. The authors also acknowledge the 'EXCELSIOR': ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment H2020 Widespread Teaming project (www.excelsior2020.eu). The 'EXCELSIOR' project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857510, from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination and Development and the Cyprus University of Technology.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2025.103075.

Data availability

Data will be made available on request.

References

Abedi, R., Costache, R., Shafizadeh-Moghadam, H., Pham, Q.B., 2022. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto Int. 37 (19), 5479–5496. https://doi.org/10.1080/10106049.2021.1920636.

Abioye, S.O., Babatunde, Y.O., Abikoye, O.A., Shaibu, A.N., Bankole, B.J., 2025. Optimized machine learning algorithms with SHAP analysis for predicting compressive strength in high-performance concrete. AI Civ. Eng. 4 (1), 16. https://doi.org/10.1007/s43503-025-00061-x.

Al-Sheriadeh, M.S., Daqdouq, M.A., 2024. Robustness of machine learning algorithms to generate flood susceptibility maps for watersheds in Jordan. Geomat. Nat. Haz. Risk 15 (1). https://doi.org/10.1080/19475705.2024.2378991.

Bathrellos, G.D., Karymbalis, E., Skilodimou, H.D., Gaki-Papanastassiou, K., Baltas, E.A., 2016. Urban flood hazard assessment in the basin of Athens Metropolitan city, Greece. Environ. Earth Sci. 75 (4), 319. https://doi.org/10.1007/s12665-015-5157-1.

Bokhari, B.F., Tawabini, B., Baalousha, H.M., 2024. A fuzzy analytical hierarchy process -GIS approach to flood susceptibility mapping in NEOM, Saudi Arabia. Front. Water *6*. https://doi.org/10.3389/frwa.2024.1388003.

Breiman, L., 2001. Random forests. Mach. Learn 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Büttner, G., Kosztra, B., Gergely, M., Pataki, R., Kleeschulte, S., Hazeu, G., Vittek, M., Schröder, C., Littkopf, A., 2021. Copernic. Land Monit. Serv. CORINE Land Cover Use Man. ⟨https://land.copernicus.eu/en/technical-library/clc-product-user-manual⟩.

Canadell, J.G., Monteiro, P.M.S., 2023. Global Carbon and Other Biogeochemical Cycles and Feedbacks. In Climate Change 2021 – The Physical Science Basis. Cambridge University Press, pp. 673–816. https://doi.org/10.1017/9781009157896.007.

Cao, C., Xu, P., Wang, Y., Chen, J., Zheng, L., Niu, C., 2016. Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. Sustainability 8 (9), 948. https://doi.org/10.3390/su8090948.

Chan, J.Y.-L., Leow, S.M.H., Bea, K.T., Cheng, W.K., Phoong, S.W., Hong, Z.-W., Chen, Y.-L., 2022. Mitigating the multicollinearity problem and its machine learning approach: a review. Mathematics 10 (8), 1283. https://doi.org/10.3390/math10081283.

Chen, J., Huang, G., Chen, W., 2021. Towards better flood risk management: assessing flood risk and investigating the potential mechanism based on machine learning models. J. Environ. Manag. 293, 112810. https://doi.org/10.1016/j.jenvman.2021.112810.

Chen, H., Ito, Y., Sawamukai, M., Tokunaga, T., 2015. Flood hazard assessment in the Kujukuri Plain of Chiba Prefecture, Japan, based on GIS and multicriteria decision analysis. Nat. Hazards 78 (1), 105–120. https://doi.org/10.1007/s11069-015-1699-5.

Chen, Y., Li, J., Xu, H., 2016. Improving flood forecasting capability of physically based distributed hydrological models by parameter optimization. Hydrol. Earth Syst. Sc. 20 (1), 375–392. https://doi.org/10.5194/hess-20-375-2016.

Christofi, C., Bruggeman, A., Kuells, C., Constantinou, C., 2020. Hydrochemical evolution of groundwater in gabbro of the Troodos Fractured Aquifer. A comprehensive approach. Appl. Geochem. 114, 104524. https://doi.org/10.1016/j.apgeochem.2020.104524.

CLMS. (2018, April 24). *Copernicus Land Monitoring Service*. ⟨https://land.copernicus.eu/En/Products/Corine-Land-Cover⟩.

Dillon, P., Stuyfzand, P., Grischek, T., Lluria, M., Pyne, R.D.G., Jain, R.C., Bear, J., Schwarz, J., Wang, W., Fernandez, E., Stefan, C., Pettenati, M., van der Gun, J., Sprenger, C., Massmann, G., Scanlon, B.R., Xanke, J., Jokela, P., Zheng, Y., Sapiano, M., 2019. Sixty years of global progress in managed aquifer recharge. Hydrogeol. J. 27 (1), 1–30. https://doi.org/10.1007/s10040-018-1841-z.

Danumah, J.H., Odai, S.N., Saley, B.M., Szarzynski, J., Thiel, M., Kwaku, A., Kouame, F.K., Akpa, L.Y., 2016. Flood risk assessment and mapping in Abidjan district using multi-criteria analysis (AHP) model and geoinformation techniques, (Cote d' Ivoire). Geoenviron. Disasters 3 (10). https://doi.org/10.1186/s40677-016-0044-y.

EC, 2007. Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. Off. J. Eur. Union 288, 27–34.

EDJNet. (2025). *Troubled Waters*. European Data Journalism Network ⟨https://www.europeandatajournalism.eu/troubled-waters/⟩.

EEA. (2020). *The European environment – State and outlook 2020*. ⟨https://www.eea.europa.eu/en/analysis/publications/soer-2020⟩.

EFFIS. (2025, September 9). *European Forest Fire Information System*. European Commission Joint Research Centre. ⟨https://forest-fire.emergency.copernicus.eu/⟩.

GSD. (2025, September 9). ⟨https://www.moa.gov.cy/moa/gsd/gsd.nsf/page45_en/page45_en?OpenDocument⟩. Department of Geology Survey.

Gupta, L., Dixit, J., 2023. Assessment of urban flood susceptibility and role of urban green space (UGS) on flooding susceptibility using GIS-based probabilistic models. Environ. Monit. Assess. 195 (12), 1518. https://doi.org/10.1007/s10661-023-12061-4.

Hagos, Y.G., Andualem, T.G., Yibeltal, M., Mengie, M.A., 2022. Flood hazard assessment and mapping using GIS integrated with multi-criteria decision analysis in upper Awash River basin, Ethiopia. Appl. Water Sci. 12 (7), 148. https://doi.org/10.1007/s13201-022-01674-8.

Hasan Tanim, A., Warren McKinnie, F., Goharian, E., 2024. Coastal compound flood simulation through coupled multidimensional modeling framework. J. Hydrol. 630, 130691. https://doi.org/10.1016/j.jhydrol.2024.130691.

Ilia, I., Tsangaratos, P., Tzampoglou, P., Chen, W., Hong, H., 2022. Flash flood susceptibility mapping using stacking ensemble machine learning models. Geocarto Int. 37 (27), 15010–15036. https://doi.org/10.1080/10106049.2022.2093990.

Jongman, B., Ward, P.J., Aerts, J.C.J.H., 2012. Global exposure to river and coastal flooding: Long term trends and changes. Glob. Environ. Change 22 (4), 823–835. https://doi.org/10.1016/j.gloenvcha.2012.07.004.

Kazakis, N., Kougias, I., Patsialis, T., 2015. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: application in Rhodope–Evros region, Greece. Sci. Total Environ. 538, 555–563. https://doi.org/10.1016/j.scitotenv.2015.08.055.

Khouz, A., Trindade, J., Santos, P.P., Oliveira, S.C., El Bchari, F., Bougadir, B., Garcia, R.A.C., Reis, E., Jadoud, M., Saouabe, T., Rachidi, S., 2023. Flood susceptibility assessment through statistical models and HEC-RAS analysis for sustainable management in Essaouira Province, Morocco. Geosciences 13 (12), 382. https://doi.org/10.3390/geosciences13120382.

Kotaridis, I., Lazaridou, M., 2022. Integration of convolutional neural networks for flood risk mapping in Tuscany, Italy. Nat. Hazards 114 (3), 3409–3424. https://doi.org/10.1007/s11069-022-05525-2.

Kountouri, J., Panagiotou, C.F., Tsouni, A., Sigourou, S., Pagana, V., Loulli, E., Evagorou, E., Mettas, C., Kontoes, C., Hadjimitsis, D., 2024. Flood Hazard Assessment and Vulnerability Analysis in Garyllis River Basin, Cyprus. IGARSS 2024 2024 IEEE Int. Geosci. Remote Sens. Symp. 3194–3198. https://doi.org/10.1109/IGARSS53475.2024.10641237.

Liu, Y., Liu, R., Shang, R., 2022. GLOBMAP SWF: a global annual surface water cover frequency dataset during 2000–2020. Earth Syst. Sci. Data 14 (10), 4505–4523. https://doi.org/10.5194/essd-14-4505-2022.

Lyu, H.-M., Shen, S.-L., Yang, J., Yin, Z.-Y., 2019. Inundation analysis of metro systems with the storm water management model incorporated into a geographical information system: a case study in Shanghai. Hydrol. Earth Syst. Sc. 23 (10), 4293–4307. https://doi.org/10.5194/hess-23-4293-2019.

Mokhtari, E., Abdelkebir, B., Djenaoui, A., Hamdani, N.E.H., 2024. Integrated analytic hierarchy process and fuzzy analytic hierarchy process for Sahel watershed flood susceptibility assessment, Algeria. Water Pract. Technol. 19 (2), 453–475. https://doi.org/10.2166/wpt.2024.012.

Mudashiru, R.B., Sabtu, N., Abustan, I., Balogun, W., 2021. Flood hazard mapping methods: a review. J. Hydrol. 603, 126846. https://doi.org/10.1016/j.jhydrol.2021.126846.

Panagiotou, C.F., 2025a. Copula-based assessment of flood susceptibility in the island of Cyprus via stochastic multicriteria decision analysis. Sci. Total Environ. 979, 179469. https://doi.org/10.1016/j.scitotenv.2025.179469.

Panagiotou, C.F., Eisenreich, S., Barouta, O.T., Chekirbane, A., Martins, T., Neophytides, S., Khemiri, K., Stefan, C., 2024. Identification of feasible regions for managed aquifer recharge in the Republic of Cyprus using a co-participative multi-criteria decision analysis. Groundw. Sustain. Dev. 27, 101323. https://doi.org/10.1016/j.gsd.2024.101323.

Panagiotou, C.F., Feloni, E., Aristidou, K., Eliades, M., 2025b. Probabilistic assessment of flood susceptibility via a coparticipative multicriteria decision analysis. Environ. Process 12 (2), 22. https://doi.org/10.1007/s40710-025-00766-2.

Panagiotou, C.F., Konstantinou, C., Chekirbane, A., 2024. A generalized machine learning approach for cost-effective monitoring of irrigation suitability: a demonstration case in El Fahs aquifer (Tunisia. Groundw. Sustain. Dev. 27, 101324. https://doi.org/10.1016/j.gsd.2024.101324.

Petracca, I., De Santis, D., Picchiani, M., Corradini, S., Guerrieri, L., Prata, F., Merucci, L., Stelitano, D., Del Frate, F., Salvucci, G., Schiavon, G., 2022. Volcanic cloud detection using Sentinel-3 satellite data by means of neural networks: the Raikoke 2019 eruption test case. Atmos. Meas. Tech. 15 (24), 7195–7210. https://doi.org/10.5194/amt-15-7195-2022.

Pfister, L., Kwadijk, J., Musy, A., Bronstert, A., Hoffmann, L., 2004. Climate change, land use change and runoff prediction in the Rhine–Meuse basins. River Res. Applic 20 (3), 229–241. https://doi.org/10.1002/rra.775.

Rahman, M., Ningsheng, C., Islam, M.M., Dewan, A., Iqbal, J., Washakh, R.M.A., Shufeng, T., 2019. Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. Earth Syst. Environ. 3 (3), 585–601. https://doi.org/10.1007/s41748-019-00123-y.

Rondinone, M., Dal Sasso, S.F., Aung, H.H., Contillo, L., Dimola, G., Schiattarella, M., Fiorentino, M., Telesca, V., 2025. Assessing Flood and Landslide Susceptibility Using XGBoost: case study of the Basento River in Southern Italy. Appl. Sci. 15 (10), 5290. https://doi.org/10.3390/app15105290.

Samanta, R.K., Bhunia, G.S., Shit, P.K., Pourghasemi, H.R., 2018. Flood susceptibility mapping using geospatial frequency ratio technique: a case study of Subarnarekha River Basin, India. Model. Earth Syst. Environ. 4 (1), 395–408. https://doi.org/10.1007/s40808-018-0427-z.

Sett, D., Trinh, T.P., Wasim, T., Ortiz-Vargas, A., Nguyen, D.G.C., Büche, K., Assmann, A., Nguyen, H.K.L., Walz, Y., Souvignet, M., Bachofer, F., Vu, T.B., Garschagen, M., Hagenlocher, M., 2024. Advancing understanding of the complex nature of flood risks to inform comprehensive risk management: findings from an urban region in Central Vietnam. Int. J. Disast. Risk Re. 110, 104652. https://doi.org/10.1016/j.ijdrr.2024.104652.

Shadmehri Toosi, A., Calbimonte, G.H., Nouri, H., Alaghmand, S., 2019. River basin-scale flood hazard assessment using a modified multi-criteria decision analysis approach: a case study. J. Hydrol. 574, 660–671. https://doi.org/10.1016/j.jhydrol.2019.04.072.

Silva, C.A., Guerrisi, G., Del Frate, F., Sano, E.E., 2022. Near-real time deforestation detection in the Brazilian Amazon with Sentinel-1 and neural networks. Eur. J. Remote Sens 55 (1), 129–149. https://doi.org/10.1080/22797254.2021.2025154.

de Sousa, M.M., de Oliveira, A.K.B., Rezende, O.M., de Magalhães, P.M.C., Pitzer Jacob, A.C., de Magalhães, P.C., Miguez, M.G., 2022. Highlighting the role of the model user and physical interpretation in urban flooding simulation. J. Hydroinform 24 (5), 976–991. https://doi.org/10.2166/hydro.2022.174.

Suwannachai, L., Sriworamas, K., Sivanpheng, O., Kangrang, A., 2024. Application of SWAT model for assessment of surface runoff in flash flood areas. Water 16 (3), 495. https://doi.org/10.3390/w16030495.

Tehrany, M.S., Pradhan, B., Mansor, S., Ahmad, N., 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. CATENA 125, 91–101. https://doi.org/10.1016/j.catena.2014.10.017.

Tepetidis, N., Benekos, I., Iliopoulou, T., Dimitriadis, P., Koutsoyiannis, D., 2025. Combining machine learning models and satellite data of an extreme flood event for flood susceptibility mapping. Water 17 (18), 2678. https://doi.org/10.3390/w17182678.

Tsangaratos, P., Ilia, I., Chrysafi, A.-A., Matiatos, I., Chen, W., Hong, H., 2023. Applying a 1D convolutional neural network in flood susceptibility assessments—the case of the Island of Euboea, Greece. Remote Sens 15 (14), 3471. https://doi.org/10.3390/rs15143471.

UN. (2025, September 9). *The Sustainable Development Goals.* ⟨https://www.un.org/Sustainabledevelopment/Development-Goals/⟩.

UNESCO. (1979). *Map of the World Distribution of Arid Regions, MAB Technical Note 7.*

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X., 2015. Flood hazard risk assessment model based on random forest. J. Hydrol. 527, 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008.

WDD. (2023, October 1). *Geoportal of Water Development Department (WDD).* ⟨https://geoportal-wdd.hub.arcgis.com/Search?Collection=Dataset⟩.

WDD. (2025, February 1). *Water Development Department Geoportal.* ⟨https://geoportal-wdd.hub.arcgis.com/⟩.

Zarekarizi, M., Srikrishnan, V., Keller, K., 2020. Neglecting uncertainties biases house-elevation decisions to manage riverine flood risks. Nat. Commun. 11 (1), 5361. https://doi.org/10.1038/s41467-020-19188-9.

Zhang, J., Yang, X., Fan, G., Li, H., Zhou, J., 2024. Physical and numerical modeling of a landslide dam breach and flood routing process. J. Hydrol. 628, 130552. https://doi.org/10.1016/j.jhydrol.2023.130552.

Zhao, G., Pang, B., Xu, Z., Peng, D., Xu, L., 2019. Assessment of urban flood susceptibility using semi-supervised machine learning model. Sci. Total Environ. 659, 940–949. https://doi.org/10.1016/j.scitotenv.2018.12.217.