



Optimizing flood susceptibility assessment in semi-arid regions using ensemble algorithms: a case study of Moroccan High Atlas

Youssef Bammou¹ · Brahim Benzougagh² · Brahim Igmoullan¹ ·
Abdessalam Ouallali³ · Shuraik Kader^{4,5} · Velibor Spalevic⁶ · Paul Sestras^{7,8} ·
Paolo Billi^{9,10} · Slobodan B. Marković¹¹

Received: 14 November 2023 / Accepted: 4 March 2024
© The Author(s) 2024

Abstract

This study explores and compares the predictive capabilities of various ensemble algorithms, including SVM, KNN, RF, XGBoost, ANN, DT, and LR, for assessing flood susceptibility (FS) in the Houz plain of the Moroccan High Atlas. The inventory map of past flooding was prepared using binary data from 2012 events, where “1” indicates a flood-prone area and “0” a non-flood-prone or extremely low area, with 762 indicating flood-prone areas. 15 different categorical factors were determined and selected based on importance and multicollinearity tests, including slope, elevation, Normalized Difference Vegetation Index, Terrain Ruggedness Index, Stream Power Index, Land Use and Land Cover, curvature plane, curvature profile, aspect, flow accumulation, Topographic Position Index, soil type, Hydrologic Soil Group, distance from river and rainfall. Predicted FS maps for the Tensift watershed show that, only 10.75% of the mean surface area was predicted as very high risk, and 19% and 38% were estimated as low and very low risk, respectively. Similarly, the Haouz plain, exhibited an average surface area of 21.76% for very-high-risk zones, and 18.88% and 18.18% for low- and very-low-risk zones respectively. The applied algorithms met validation standards, with an average area under the curve of 0.93 and 0.91 for the learning and validation stages, respectively. Model performance analysis identified the XGBoost model as the best algorithm for flood zone mapping. This study provides effective decision-support tools for land-use planning and flood risk reduction, across globe at semi-arid regions.

Keywords Flood susceptibility · GIS · Machine learning · Factor importance · Tensift watershed

Abbreviations

ABHT	Tensift Basin Hydraulic Agency
ANN	Artificial neural network
AUC-ROC	Area under curve of the receiver operating characteristic
CM	Correlation matrix
DEM	Digital elevation model

Extended author information available on the last page of the article

DT	Decision tree
FAO	Food and Agriculture Organisation
FR	Frequency ratio
FS	Flood susceptibility
GIS	Geographic informatic system
HSG	Hydrologic Soil Group
IDW	Inverse distance weighted
KNN	K-nearest neighbor
LR	Logistic regression
LULC	Land use and land cover
MAE	Mean absolute error
MI	Mutual information
ML	Machine learning
NFR	Normalized frequency ratio
RMSE	Root mean square error
RF	Random forests
SPI	Stream power index
SVM	Support vector machines
TOL	Tolerance
TPI	Topographic position index
TRI	Topographic roughness index
TWI	Topographic wetness index
USGS	United States Geological Survey
VIF	Variance inflation factor
XGBoost	EXtreme Gradient Boosting contraction

1 Introduction

The current technological advancements are dealing hard to cope with the recovery from various natural disasters across the world because of their unpredictability, and the enormous damages result as consequences. Among these natural occurrences, flooding stands out as destructive phenomenon that directly impact about 200 million people annually in terms of infrastructural, economic detriments, and loss of lives (Abdulrazzak et al. 2019; Sivakumar 2005). In the twentieth century, more than 100,000 lives were killed by floods and affected over 1.4 billion people in terms of infrastructure and economic destructions (El Alfy 2016; Han and Sharif 2021). Systematic risk reduction strategies with the integration of sophisticated scientific technology to forecast flooding scenarios in advance, that would incorporate all the associated parameters in particular regions related to climatic and geographic features are necessary to mitigate the adverse impacts of floods.

In contrary to popular beliefs, floods are more damaging in semi-arid and desert areas than in wetlands. The main reason for these drastic losses is the absence of a systematic prediction tool for runoffs in these areas because of the lack of hydrometeorological data and gauging of the flash floods (Aryal et al. 2020). After earthquakes, floods are the most frequent natural catastrophe in Morocco in terms of fatalities and injuries. The Tensift watershed and the Haouz plain are very vulnerable to flooding. In fact, the sub-catchment regions of the Tensift have witnessed disastrous floods because of torrential rains in 1995

and 2002, which not only caused the deaths of more than 200 people but also enormous material damage (Youssef et al. 2023a, b).

The Haouz plain and Tensift watershed are categorized as belonging to semi-arid zones on the Kopper-Geiger world climate map (Bouramtane et al. 2020), where it is exceedingly challenging and complex to investigate these phenomena. Due to climatic changes in rainfall, urbanization, and variables relating to the sub-watersheds of Ourika, R'dat, Zat, Rheraya, Assif El Mal, N'fis and Seksaou, the flood disaster in the Tensift watershed and the Haouz plain has gotten worse. Consequently, comprehensive research and investigations on flood susceptibility (FS) have become imperative in accurately identifying high-risk areas and implementing effective preventative measures (Echogdali et al. 2022).

Apart from climate-related influences, various other factors play a crucial role in determining flood dynamics, necessitating their consideration for accurate flood risk evaluation. Among these factors, slope characteristics directly affect hydrological processes, influencing the direction of precipitation. A higher permeability of the topsoil layer on the land enhances the infiltration capacity and reduces runoff (Bennani et al. 2019; Benssaou et al. 2003). Additionally, factors such as geological formations and land use changes, such as transitions from forested areas to agricultural fields or from agricultural zones to urbanized regions, play significant roles in flood management (Kader et al. 2023). Additionally, there are climatic factors that affect flood risk, including precipitation rate and volume and melting (Anjos et al. 2015; Bennani et al. 2019; Soulaïmani and Bouabdelli 2005).

Recent advances in ML algorithms have been combined with GIS and remote sensing methods, greatly enhancing the mapping of flood risk and spatial variability. The commonly used ML techniques for predicting flood risk are the ANN—Artificial Neural Network (Dahri et al. 2022), SVM—Support Vector Machine (Choubin et al. 2019a, b), RF—Random Forests (Billah et al. 2023), LR—logistic Regression (Ali et al. 2020), adaptive neuro-fuzzy inference systems (Ahmadlou et al. 2019), and long-term memory (Apyadin et al. 2020; Dazzi et al. 2021). In order to predict the spatial and temporal variation of flood risk by relating runoff to precipitation and which require field observation data, several studies on the likelihood of flood occurrence have been conducted using various models and techniques, such as rainfall runoff like GSSHA (Kirker and Toran 2023), MIKE DHI (Beden and Ulke Keskin 2021), HEC-HMS (El Alfy 2016), and SWAT (Tan et al. 2020). The regions vulnerable to flooding in this model are divided into two categories (1) not floodable (2) floodable: on the basis of historical and geo-environmental data (Benkirane et al. 2020), which do not consider the field observation data. To forecast floods by incorporating all three data types including the field observation data, further models based on GIS and remote sensing have been developed (Costache et al. 2022; Mosavi et al. 2018).

Despite the publication of considerable studies, the use of machine learning (ML) techniques to analyze floods in the Tensift watershed and Haouz plain has lacked the required attention. These methodologies must be incorporated to map and forecast FS in the locations, and it is an utmost important to identify the best feasible model that is suitable for the study area in terms of relevance and serviceability. To resolve this identified research gap, this study was designed based on the following primary objectives: were firstly to evaluate the ability of the SVM, KNN, RF, XGBoost, ANN, DT, and LR methods to predict flood susceptibility areas, secondly to study the role that each flood conditioning factor plays in the development of flood susceptibility maps, Thirdly to create flood susceptibility maps using an ensemble modelling strategy for each model and finally to identify the areas at risk of flooding in the Haouz plain and the Tensift catchment.)

The novelty of this study lies in the comprehensive evaluation of the multiple advanced algorithms on flood prediction, offering a diverse toolkit for tackling this complex issue.

Unlike the published research sources available within this field in scientific literature, this research is structured on a systematic comparison of various ensembled algorithms, including RF, SVM, k-nearest neighbor (KNN), and eXtreme Gradient Boosting contraction (XGBoost). This study is first-of-its-kind to the North African region, providing a comprehensive FS assessment crucial for risk management in the study region.

This research study offers a sophisticated methodology for scientific community to assess floods that integrates a Multifactor analysis approach by integrating various influential factors, such as slope, elevation, vegetation index, terrain ruggedness, land use and land cover. By analyzing the binary data from past flooding events, the study skillfully produced a FS map that delineates risk levels across the landscape. This research is especially significant since it can provide extremely precise flood prediction models, which are useful for reducing the likelihood of future floods and lessening their negative effects. These findings hold valuable implications for flood management and risk reduction strategies and offers insights for policymakers, enabling informed decisions for land use planning and disaster preparedness. This research's outcome furnishes decision-support tools crucial for informed land-use planning and mitigation efforts.

2 Material and methods

2.1 Study area

The Tensift watershed, located in the central part of Morocco near Marrakech, has an area of 20,000 km² (Fig. 1). It is made up of two separate hydrological zones with opposing tendencies. The southern slopes of the Atlas Mountains, which stand above 4000 m in height, get significant precipitation and snowfall, amounting to up to 600 mm every year. These mountains provide an important source of water for the large Haouz plain downstream, which is semi-arid with an annual precipitation of 250 mm. Particularly, irrigation operations benefit a substantial area of this plain, notably the 2000 km² irrigated Haouz plain.

Geologically, there are three primary geological formations that make up the watershed of the High Atlas near Marrakech (Duclaux 2005): (1) The Permo-Triassic is the dominant formation in the east. It frequently coexists with Precambrian and Ordovician schistose

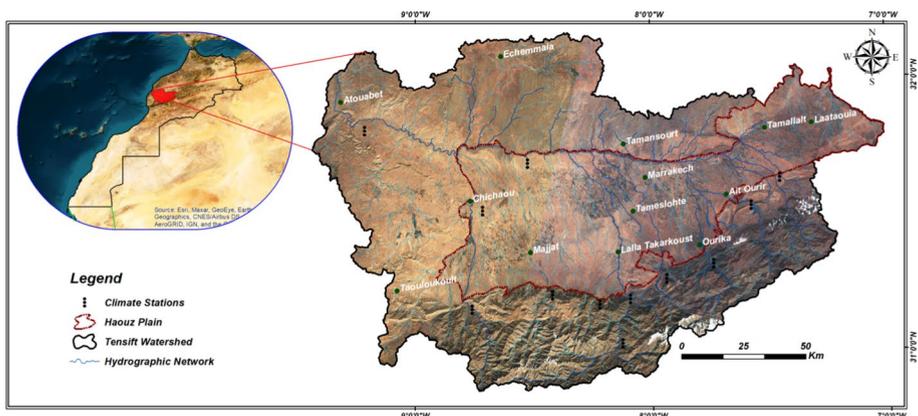


Fig. 1 Study area of the Tensift watershed and the Haouz plain [Source – Study authors]

rocks; (2) Precambrian eruptive and metamorphic rocks may be found in the central region, which is home to the highest peaks of the Atlas; and (3) primary and secondary limestone formations can be found in the western region. Since most of these geological formations have limited permeability, continuous surface runoff and, eventually, the development of considerable runoff after heavy rains, are encouraged.

In addition, this area has a varied and unpredictable hydrological behavior that is influenced by its geomorphological and climate factors. It gives rise to long-lasting storms that frequently result in significant damage, such as the inundations that occurred in Ourika on August 17, 1995, and October 28, 1999, which resulted in 200 fatalities, the destruction of 142 buildings, and the inundation of more than 300 ha of agricultural land. These events were reported by the Tensift Basin Hydraulic Agency (ABHT).

2.2 Predictors and conditioning factors of flooding

Based on the research area and data available, the flood-conditioning parameters, including topography, elevation, slope, aspect, lithology, land use, precipitation, and habitation, were taken into consideration. The four categories of these flood conditioning elements are hydrology, geography, environment, and ethnography. In this study, the following factors were taken into account from the Digital Elevation Model (DEM) at a resolution of 30 m, downloaded from Were downloaded from the website of United States Geological Survey (USGS); slope, aspect, elevation, flow accumulation, topographic position index (TPI), topographic roughness index (TRI), stream power index (SPI), plane curvature, topographic wetness index (TWI), profile curvature, distance to river, for precipitation, this factor is calculated by interpolating data from 9 climate stations supplied by ABHT using IDW method, for land use and land cover (LULC) is extracted by supervised classification using the Maximum Likelihood method from a Sentinel 2A satellite scene supplied by Scihub Copernicus, the soil type factor was developed using the database provided by the FAO, the Hydrologic Soil Group (HSG) factor was developed using data from EARTH-DATA, illustrated by Fig. 2 and Table 1.

Precipitation is an important climatic factor that impacts the likelihood of floods. As a result, it is critical to underline that the annual mean rainfall acts as a core component in most flood vulnerability studies. Hydrometeorological aspect has been proven as an indispensable flood predictor due to its substantial relationship with soil moisture changes (Ighile et al. 2022; Verma et al. 2022). The aspect grid values were used to construct a single flat region as well as the ninth north, northeast, east, southeast, southwest, west, and northwest divisions. The TPI which makes it possible to identify one of these most recent cells, considers the height of the surrounding cells (Jenness, 2000). The following classes have been defined for the TPI card in this study: $(-198)-(-123)$; $(-123)-(-40)$; $(-40)-(-10)$; $(-10)-40$; $40-140$; and $140-247$.

SPI is another morphometric parameter that will be utilized for flood forecasting. Erosive force and transport capacity are included while calculating the water values. The SPI maps have been divided into the following categories based on professional opinion: <1 ; $1-2$; $2-6$; $6-8$ and $8-15$.

Similarly, slope is a major morphometric feature that has a significant impact on flooding (Ganie et al. 2022). It is well known that an area with a steep slope that causes significant surface runoff is more prone to flooding than a flat area. The slope was separated into the following ranges: $0-5$; $5-12$; $12-22$; $22-33$; and $33-71$ to create the slope map. Due to the permeability of the rock, the soil types primarily control the amount of water

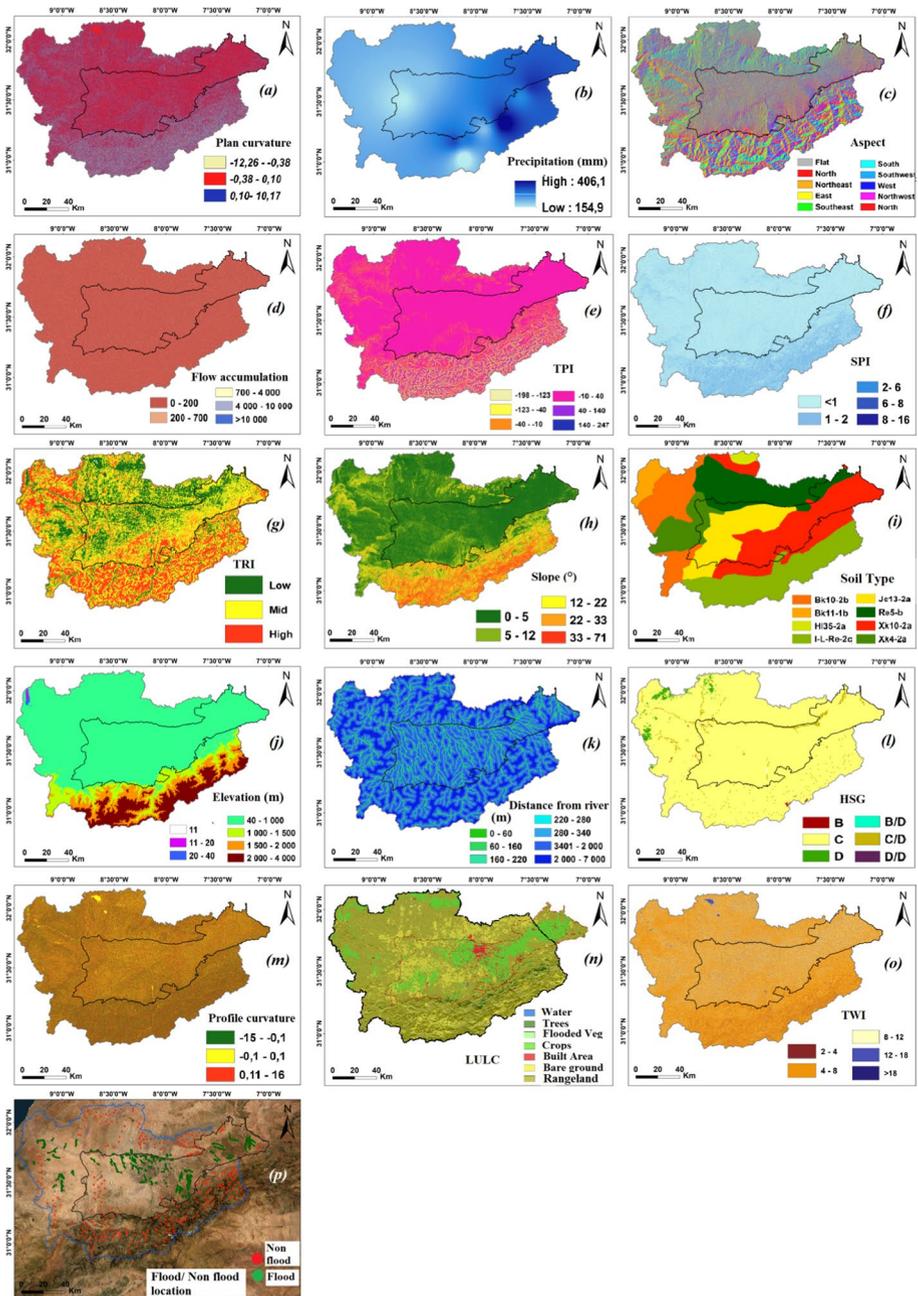


Fig. 2 The fifteen flood factors developed for the study area **a** plan curvature, **b** precipitation, **c** Aspect, **d**: Flow accumulation, **e** TPI, **f** SPI, **g** TRI, **h** Slope, **i** Soil type, **j** Elevation, **k** Distance from river, **l** HSG, **m** Profile-curvature, **n** LULC, **o** TWI, and **p** inventory of flooded and non-flooded areas). [Source – Study authors]

Table 1 Sources, equations and data processing used in this study

Factors	Layers and units	Equation & Processing	Source
Topographic factors	Elevation (m)	Fill tool for correction in ArcGIS	SRTM-DEM (Digital Elevation Model) were downloaded from the website of United States Geological Survey (USGS) https://earthexplorer.usgs.gov/
	Slope (°)	Processing in ArcGIS	
	Stream Power Index (SPI)	$SPI = As \times \tan\beta$ As is the upstream drainage area and β is the slope degree	
	Topographic Position Index (TPI)	Processing in ArcGIS	
	Aspect	Processing in ArcGIS	
	Plan Curvature	Processing in ArcGIS	
	Profile Curvature	Processing in ArcGIS	
Geomorphological factors	Topographic Wetness Index (TWI)	$TWI = \ln(As/\tan\beta)$ As is the upstream drainage area and β is the slope degree	
	Topographic Roughness Index (TRI)	Processing in ArcGIS	
	Distance to rivers (m)	Processing in SAGA GIS	
	Flow Accumulation		

Table 1 (continued)

Factors	Layers and units	Equation & Processing	Source
Geological factors	Soil Type	<p>Xk10-2a: (48.7% sand, 29.9% silt, 21.6% caly and 0.64 orgC)</p> <p>Bk10-2b: (81.6% sand, 6.8% silt, 11.7% caly and 0.44 orgC)</p> <p>Re5-b: (68.3% sand, 15.1% silt, 16.6% caly and 0.5 orgC)</p> <p>Xk4-2a: (47.7% sand, 30.9% silt, 22.6% caly and 0.65 orgC)</p> <p>Je13-2a: (39.6% sand, 39.9% silt, 20.6% caly and 0.65 orgC)</p> <p>IL-Re-2c: (58.9% sand, 16.2% silt, 24.9% caly and 0.97 orgC)</p>	<p>https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/faounesco-soil-map-of-the-world/en</p>
Climatic factors	Rainfall (mm)	<p>B: moderately low runoff potential (50–90% sand and 10–20% clay)</p> <p>C: moderately high runoff potential (<50% sand and 20–40% clay)</p> <p>D: high runoff potential (<50% sand and >40% clay)</p> <p>B/D: high runoff potential unless drained (50–90% sand and 10–20% clay)</p> <p>C/D: high runoff potential unless drained (<50% sand and 20–40% clay)</p> <p>D/D: high runoff potential unless drained (<50% sand and >40% clay)</p> <p>Interpolation of monthly rainfall using the IDW method</p>	<p>https://daac.ornl.gov/SOILS/guides/Global_Hydro_logic_Soil_Group.html</p> <p>Tensift Hydraulic Basin Agency (ABHT)</p>
LAND cover factors	LandUse-LandCover (LULC)	Supervised classification using the Maximum Likelihood method	<p>Sentinel-2 Images</p> <p>https://scihub.copernicus.eu/dhus/#/home</p>

penetration, which in turn affects the flooding occurrence (Tariq et al. 2022). Eight classes were discovered in this research region: (Bk10-2b; Bk11-1b; HI35-2a; I-L-Re-2c; Jc13-2a; Re5-b Xk10-2a and Xk4-2a). Because flooding is more common in low-lying areas, elevation is a second critical morphometric component in determining flood vulnerability. The elevation map for the case study was created using the following seven elevation classes: 11 m, 11–20 m, 20–40 m, 40–1000 m, 1000–1500 m, 1500–2000 m, and 2000–4000 m. The flood control parameter known as the hydrologic soil group has a big influence on how much water infiltration occurs.

Soil texture has a direct impact on infiltration due to its influence on hydraulic conductivity. There are six main hydrological soil types in the Tensift watershed: B, C, D, B/D, C/D and D/D. The direction of the slope in the vertical plane influences the curvature of the soil profile. Positive numbers indicate that surface runoff is decreasing, while negative values suggest that surface runoff is increasing. Three classes were used to create the profile curvature map: -15 to -0.1 , -0.1 to 0.1 , and 0.1 to 16 . LULC, a factor in flood forecasting, has a considerable impact on processes involving surface runoff and water storage. There are seven main types of land use in the Tensift watershed: water bodies, forest areas, flooded areas, vegetation, crops, built-up areas, bare lands, and rangeland. The slope angle and catchment area measurements are used to calculate the TWI. This metric highlights the role that geography plays in the phenomena of water accumulation. The classes described below were developed using the natural breaks approach to produce the topographic moisture index map: 1–4; 4–8; 8–12; 12–18 and > 18 .

2.3 Flood Locations Inventory

For management and mitigation strategies to be effective, it is crucial to comprehend and analyze the flood inventory. The past flood occurrences serve as crucial input variables for calculating FS. The origin of the flood inventory varies by geographic region, but commonly include previous engineering and scientific work, the hydraulic agency, field surveys, or recently developing technology. Variables that affect flood vulnerability have been employed as points for flood placement in several published studies including (Chapi et al. 2017; Costache et al. 2022; Wang et al. 2018). Based on information from scholarly journals, historical data, Google Earth, and satellite image analysis, the current study created a flood inventory map. As a result, the flooding that occurred in the Tensift watershed and the Haouz plain utilized a total of 890 flood event sites (Figs. 2 and 3). These selected flood plain locations were considered relevant in the study area, as they shed light on the complex issues associated with flooding.

2.4 Flood factor classification analysis

In the current study, seven prediction models were used to enhance the (ML) prediction of flood hazards. Various statistical tests were performed on these models to discover strong linear connections between different components. These tests, which included correlation matrix (CM) analysis, variance inflation factor (VIF) (Eq. 1), tolerance (TOL) (Eq. 2), and mutual information (MI) (Eq. 3), aided in identifying and eliminating non-significant components. VIF values larger than 10 and TOL values of 0.1, in particular, revealed significant multicollinearity amongst components (Miao et al. 2023). If two variables were significantly correlated and satisfied the multicollinearity criterion, the one with the higher VIF was eliminated based on the CM analysis. The MI

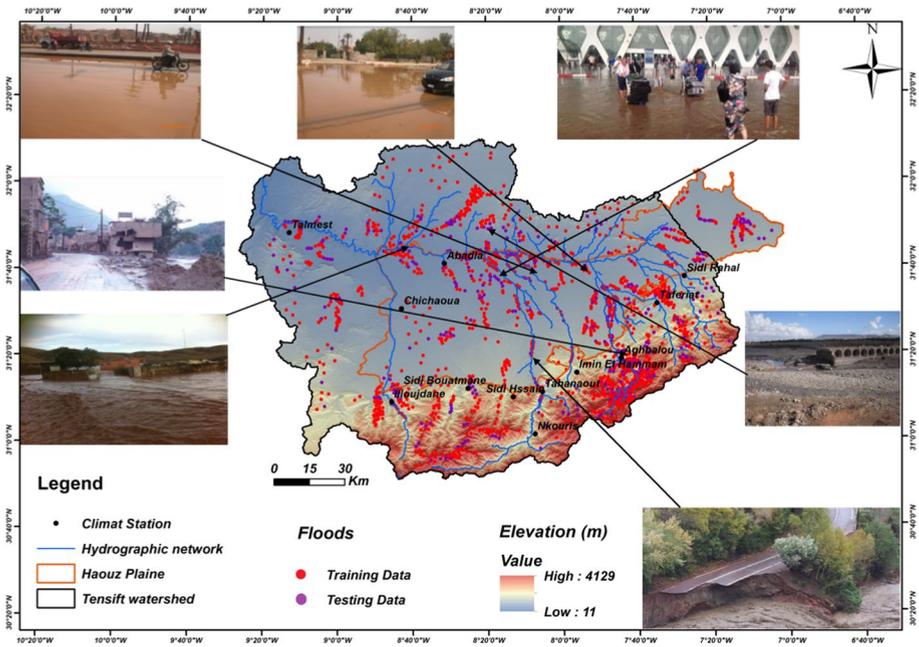


Fig. 3 Inventory map and examples of floods in the Tensift watershed and the Haouz plain [Source – Study authors]

analysis revealed the significance of factors causing floods, with low MI values suggesting little influence and leading to their removal.

$$VIF_j = \left[\frac{1}{Tol_j} \right] \tag{1}$$

$$Tol = 1 - R_j^2 \tag{2}$$

$$MI(n, j) = H(n) - H\left(\frac{n}{j}\right) \tag{3}$$

where *j* is the FS influence factor, *n* is the subclass of FS influence factors, Tol *i* is the tolerance of *j*, VIF *j* is the variance inflation factor of *j*, MI (*n*; *j*) is the mutual information for *n* and *j*, *R* is the determining coefficient of the regression for the predisposition of *j*, on all other predisposition factors, *H*(*n*) is the entropy of *n*, and *H* (*n*/*j*) is the conditional entropy for *n* given the flooded area state factor *j*.

The optimized selection analysis of FS influencing factors and model application was based on the determination of the normalized frequency ratio (NFR) (Eq. 4), which has recently been a recommended step to unify the importance of the type of input data for the different factors (Mao et al. 2022; Namous et al. 2021). Consequently, the frequency ratio (FR) (Eq. 5) was assigned to the subclass of factors influencing the FS in the sense of defining the relationship between the flooded locations and the factors influencing the

FS (Masoud et al. 2022). The results were then normalized using Eq. (5). As a result, all the maps used were converted to an NFR between 0 for low FS, and 1 for high FS.

$$NFR_n = \frac{FR_n - \text{Max}(FR_n)}{\text{Max}(FR_n) - \text{Min}(FR_n)} * (0.99 - 0.01) + 0.01 \tag{4}$$

$$FR_n = \frac{\frac{W_n}{W_t}}{\frac{P_n}{P_t}} \tag{5}$$

where n represents the subclass of factors influencing FS, FRn is the frequency ratio of n, NFRn is the normalized frequency ratio of n, Wn is the number of water sampling points located in n, Wt is the total number of water sampling points, Pn is the number of pixels in n, and Pt is the total number of all pixels.

The subclasses of FS influencing factors were determined by classifying maps produced using Jenks’ natural break technique (Sarker 2021); the exceptions are aspect, LULC, and lithology, which have been classified according to directional units, supervised classification, and lithological units, respectively.

2.5 Methodology flowchart

This research suggests the use of seven different ML models as well as GIS and remote sensing to monitor flooding. As our model uses a binary categorization (0: no flooding; 1: flooded), the location of non-flooded areas is included in the susceptibility mapping process. The methodological strategy used to draw up the susceptibility maps for this study involves the following main stages, described in Fig. 4.

The construction of the initial GIS database, which includes two types of data, namely the historical flooded areas extracted through different and the development of different maps of flood conditioning factors.

The transformation of this database into numerical mode used the FR method to indicate the relationships between the influencing factors and the FS. After applying multicollinearity analyses such as correlation matrix analysis (CM), VIFs, Tol and the mutual information test (MI), the factors influencing the FS according to their importance were selected. The second step consists of evaluating the performance and effectiveness of the seven algorithms applied, namely DT, ANN, LR, KNN, SVM, RF, XGBoost for training and test data, according to validation criteria such as specificity, sensitivity, false positive rate, precision, F1 score, accuracy, mean absolute error, root-mean-square error and the area under curve of the receiver operating characteristic (AUC-ROC). In the final stage, the database was divided into training and test data, with 70% and 30% of the total data sample being flooded and non-flooded respectively for the generation of the FS maps. External validation of the random sample for each site was carried out using ArcGIS 10.5.1 software to ensure a non-objective sampling process.

2.6 Description of the learning algorithms

In this work, seven algorithms were used to estimate flooding susceptibility: SVM, RF, K-NN, LR, ANN, DT, and XGBoost. Table 2 contains descriptions of the algorithms that were

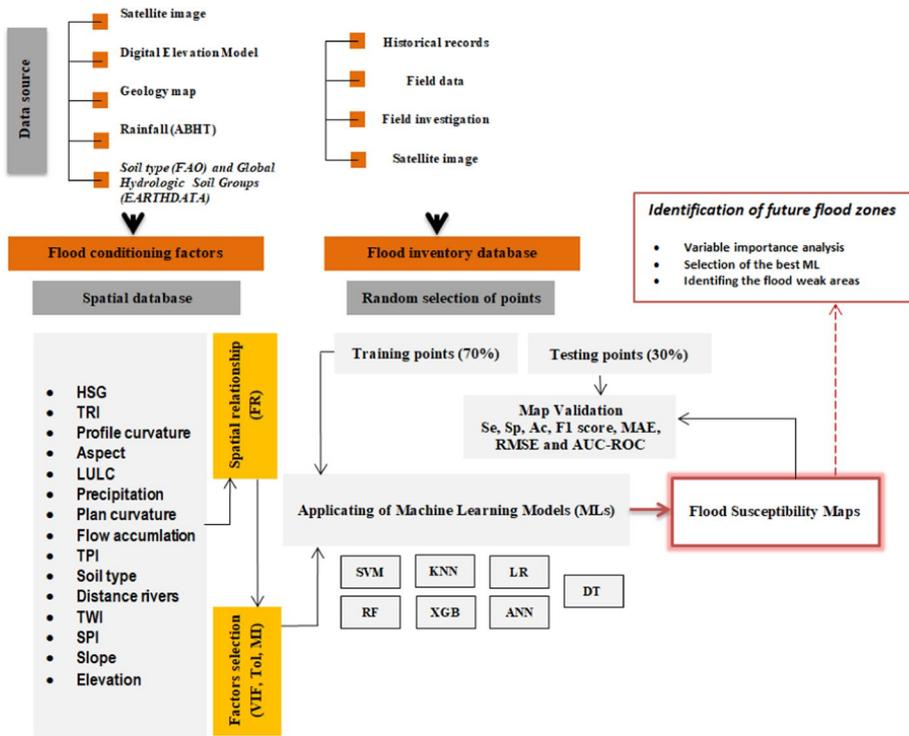


Fig. 4 Outline of the methodological workflow adopted in this study

chosen. Furthermore, there are surveys with scientific establishments that discuss the parameters, classification, and functioning of these systems (Liu and Lang 2019; Sarker 2021).

2.7 Validation techniques

The results of the proposed technique were validated for the seven models generated from various performance measures including specificity, precision, sensitivity, and accuracy. If there is a geographical connection between the measured floodable and non-floodable regions and the anticipated floodable areas, the performance indices are deemed significant, according to the outcomes of (Costache 2019a; Costache and Bui 2020).

The following parameters TP (true positives), TN (true negatives), FP (false positives) and FN (false negatives) were used to find Specificity, Sensitivity, Accuracy, Precision, recall and F1 score (Eqs. 6–10). The analysis also used another popular measure known as the receiver operating characteristic (ROC) curve. The most used ROC curve analyzes the area under the curve to determine the accuracy of prediction models (AUC) illustrated by Eq. 11. FS mapping has also used root mean square error (RMSE) and mean absolute error (MAE) as of Eqs. 12 and 13. Both types of indices have been used in several previous research studies.

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{6}$$

Table 2 Description of the four algorithms applied in this study

Model	Developed by	Description	Applied by
KNN	Fix and Hodges (1952) and Cover and Hart (1967)	One of the ML algorithms, the K-NN algorithm is based on the supervised learning method. It places the new case in the category that is most like the existing categories by assuming that there is a likelihood that the new case/data will be like the current cases. That is, flooded and non-flooded points	Costache (2019b), Sami and Ibrahim (2021), Shah et al. (2020), Thi Thuy Linh et al. (2022)
SVM	Bottou and Vapnik (1992) and Cortes and Vápnik (1995)	Each datum is represented by a point in an n-dimensional space (n—the number of features) in the SVM method, with each feature's value denoting a specific position. The best hyperplane that effectively distinguishes the two classes is then found during the classification. The prediction of FS is done using the radial basis function	Dazzi et al. (2021), Tehrani et al. (2014)
RF	Breiman (2001)	Based on various groups of the FS-influencing variables identified using the bootstrap aggregation (also known as bagging) and random feature selection procedures, a meta-estimator called random forest updates a specific number of DT classifiers. It enhances the model's propensity to predict outcomes	Chapi et al. (2017), Hosseini et al. (2020), Mosavi et al. (2018)
XGBoost	Chen and Guestrin (2016)	An enhanced Gradient Boosting second-order tree boosting is known as eXtreme Gradient Boosting second-order derivatives are used, which lower the loss function and provide more precise trees	Thi Thuy Linh et al. (2022)
ANN	Hopfield (1982)	(ANN) has been developed from the multi-layer perceptron with three neural layers: a primary layer and an intermediate or hidden layer. In the first stage of learning, the network predicts the label of everyone (flooded or non-flooded) using its weights and bias values. A cost function determines the difference between the calculated value and the observed value. In the backpropagation step, each weight is updated according to the gradient of the cost function. This process continues until convergence is reached or a maximum number of learning epochs is reached	Melitho et al. (2022), Parvin et al. (2022)

Table 2 (continued)

Model	Developed by	Description	Applied by
DT	Breiman (2001)	This algorithm selects the most suitable features (the factor influencing flooding) as the root and generates child nodes. Based on the top-down observations and the processing results at each level, predictions are obtained, from the root to the child nodes corresponding to the new data (flooded and/or non-flooded)	Choubin et al. (2019b), Seydi et al. (2022), Wang et al. (2015)
LR	Cox (1958)	This generalized linear statistical model uses a logistic function as the link function. It is used to study the relationships between a set of qualitative and qualitative variables. This model can be used to predict the probability of a flood occurring (value of 1) or not (value of 0), by optimizing the regression coefficients. The values are in a range from 0 to 1	Zhao et al. (2019)

$$\text{Sensitivity} = \frac{TP}{FN + TP} \tag{7}$$

$$\text{Accuracy} = \frac{TN + TP}{FP + TP + FN + TN} \tag{8}$$

$$\text{Precision} = \frac{TP}{FP + TP} \tag{9}$$

$$\text{F1score} = \frac{2}{\frac{1}{Pr} + \frac{1}{Recal}} \text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

$$\text{AUC} = \frac{(\sum TP + \sum TN)}{(P + N)} \tag{11}$$

where P and N are the total number of pixels with and without torrential events respectively, TP represent the true positive and TN represent the true negative.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{predicted} - X_{actual})^2} \tag{12}$$

$$MAE = \sqrt{\frac{1}{n} \sum_{i=1}^n |X_{predicted} - X_{actual}|} \tag{13}$$

where n is the total number of samples in the learning or testing phase, X predicted is the projected value from the FS model, and X actual is the observed value.

3 Results

3.1 Multicollinearity and factor selection

In Fig. 5, the correlation matrix (CM) represents Pearson’s association analysis between fifteen influencing variables that are briefly elaborated in Fig. 2. As the results show, the highest positive correlation value (0.75) was found between elevation and slope, and a strong linear correlation between the following factors: TWI and SPI, rainfall and slope, SPI and Profile curvature, TRI and distance to river, TRI, and elevation.

The results of the tolerance and VIF applied to check the multicollinearity of the feed influence factors in this study show a Tol value between 0.12 and 0.97 for TWI and HSG, respectively, as well as a maximum VIF value of 8.02 for TWI and a minimum value of 1.02 for HSG (Fig. 6). In accordance with the Tol and VIF requirements among the fifteen factors used in this study, TWI and SPI were removed in the following analysis. Next, the MI of the other thirteen factors (Fig. 7) shows positive values ranging from 0.382 (distance to river) to 0.013 (HSG). Consequently, distance from the river is ranked as the most important factor, followed by elevation (MI=0.245), Accumulation flow (MI=0.183) and slope (MI=0.159).

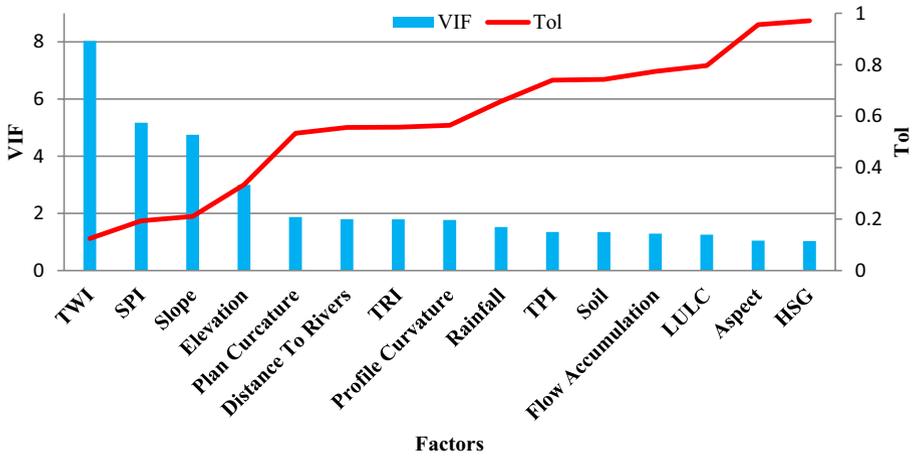


Fig. 5 Multicollinearity analysis of conditioning factors using the correlation matrix

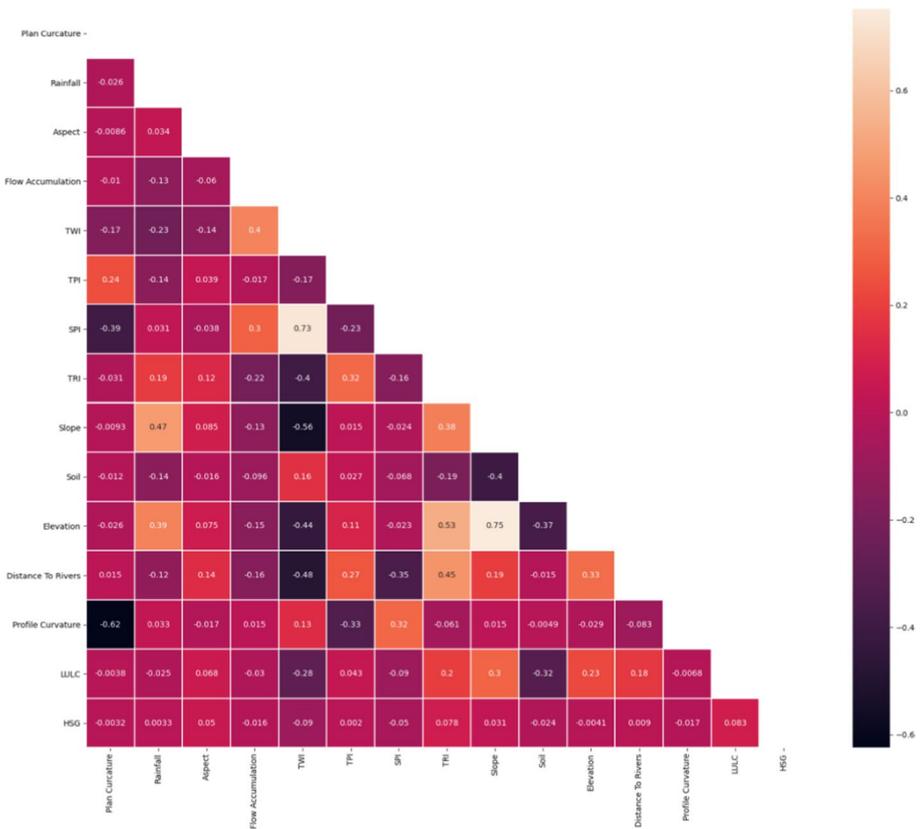


Fig. 6 Multicollinearity analysis of conditioning factors by variance inflation factor (VIF) and tolerance (TOL)

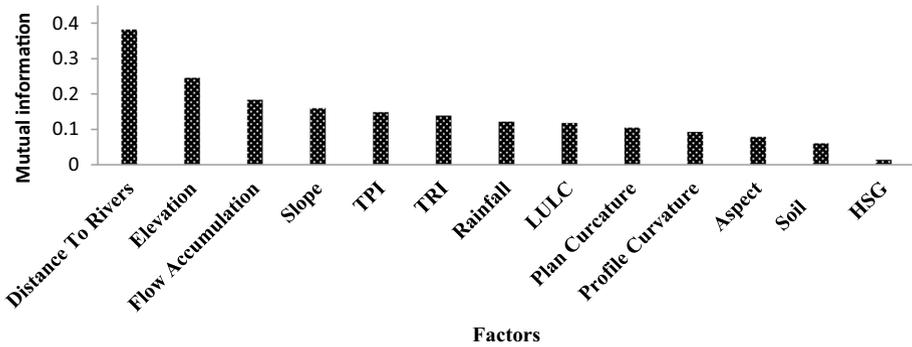


Fig. 7 Flash flood conditioning factors in relation to predictive strength for all models

3.2 Flood susceptibility maps

The FS development of the FS model was based on the application of the seven algorithms. The results were presented as a probability prediction ranging from 0 to 1, corresponding to the lowest and highest FS values respectively. Subsequently, the generated maps were then classified into the following five different zones using Jenks’ natural break classification: very low, low, moderate, high, and very high.

Initial visual analysis of the seven maps produced by the SVM (Fig. 8a), RF (Fig. 8b), LR (Fig. 8c), KNN (Fig. 8d), DT (Fig. 8e), ANN (Fig. 8f) and XGBoost (Fig. 8g) models shows that very high FS values are concentrated in the downstream part of the Tensift watershed, particularly in the lowland area (Haouz), and are slightly localized in the western part and in the Jbilet. On the other hand, very low FS values are found in the upstream part of the High Atlas chain.

The spatial distribution of the degree of sensitivity to flooding is quite remarkable in the very strong and very weak degrees especially in the Haouz plain, all the models used show a very weak sensitivity and the same differing degrees of sensitivity in the high altitudes of the High Atlas and the Ourika area except for the XGBoost model which also shows a weak sensitivity in the Jebilet area.

For the Haouz plain, the SVM model shows 23% of very weak sensitivity surfaces in the Majjat and Ait Ourir area and 20% of very strong sensitivity surfaces located in the centre of the area, the RF model shows 11% of very weak sensitivity surfaces scattered in several areas in the Mjjat and Takarkoust and 20% of very strong sensitivity surfaces generally in the Laataouia area, the LR model shows 14% of very low sensitivity surfaces in the middle altitudes of certain areas of Mjjat, Takarkoust, Ait Ourir and Lataouia and 19% of very high sensitivity surfaces located slightly in the centre of the area, the KNN model shows 11% of very low sensitivity surfaces and 18% of very high sensitivity surfaces, with the same geospatial distribution of the RF model, the DT model shows 13% of very weak sensitivity surface located in the region of Ait Ourir, Mejjat and 23% of very strong sensitivity surface observed over a large region of Laataouia and the urban area of Marrakech, the ANN model shows 33% of very weak sensitivity surface over a large area of Majjat, Takarkoust and Ait ourir and 23% of very strong sensitivity surfaces generally observed in Marrakech and Ourika and finally the XGBoost model, which shows 23% of very low sensitivity surfaces over large areas of

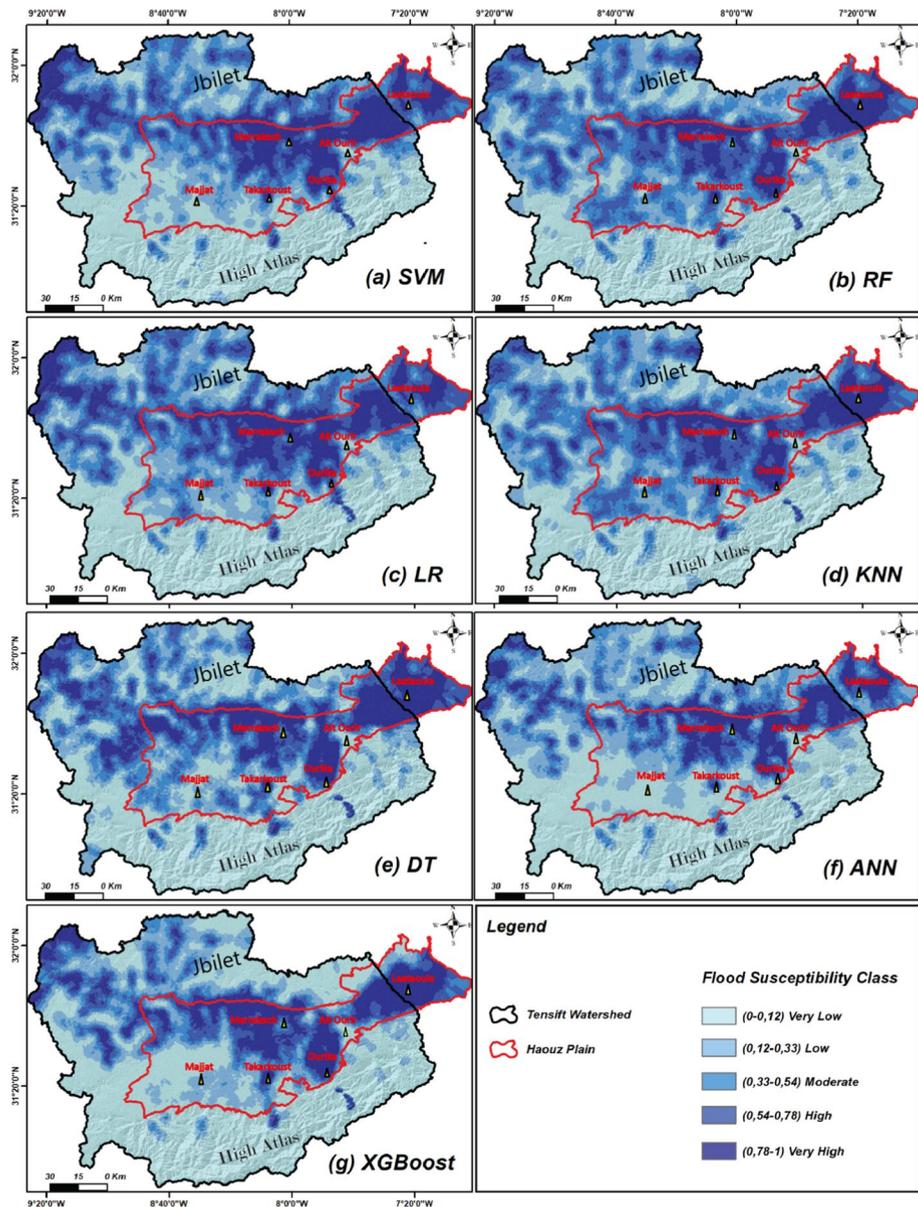


Fig. 8 FS maps Tensift watershed predicted by the SVM, RF, LR, KNN, DT, ANN and XGBoost models [Source Study authors]

Majjat and the surrounding Laataouia region and 30% of very high sensitivity surfaces concentrated in Ourika, Laataouia and Marrakech.

In the Tensift watershed, most areas are predicted to have very low (39.89%) and low (23.97%) FS, and the remaining areas are associated with moderate (15.72%), high (13.86%) and very high (10.75%) FS values provided by Fig. 9. In the Haouz plain zone,

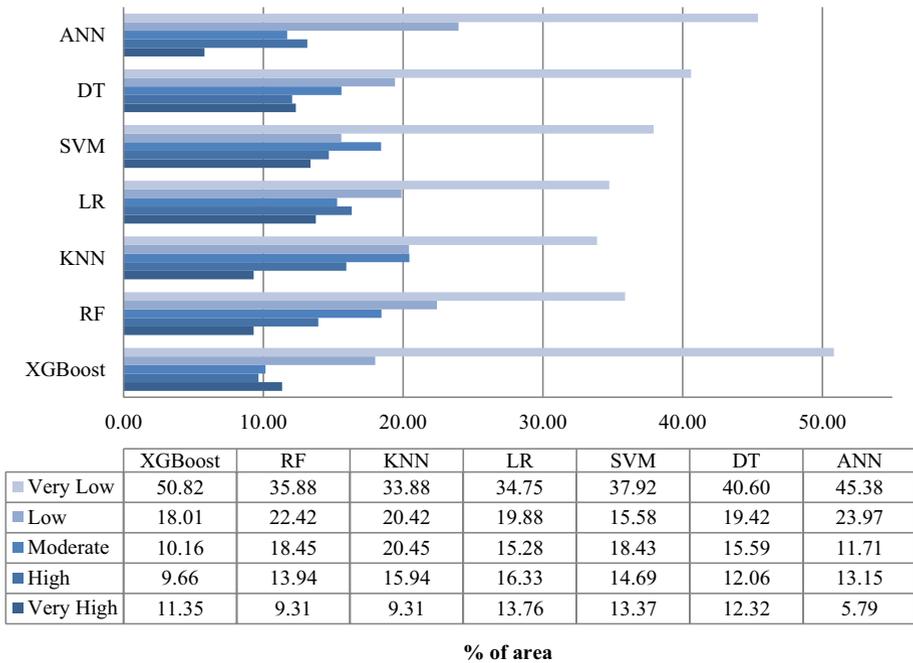


Fig. 9 Percentage of FS area in the Tensift watershed of the models used

on the other hand, there is an increase in the areas associated with high (22.14%) and very high (21.76%) FS values while the remaining areas show moderate (19.02%), low (18.88%) and very low (18.18%) percentages (Fig. 10).

Overall, Fig. 9 illustrates that the low and very low FS classes are predominant in the Tensift watershed, encompassing more than 59.85% of the total surface area. Conversely, Fig. 10 demonstrates that the high and very high FS classes dominate the Haouz plain, covering over 43.91% of the total surface area.

3.3 Performance of the seven ML models used

The results of this study were constructed using seven ML models to predict FS, namely SVM, RF, KNN, DT, ANN, LR and XGBoost, and evaluated the effectiveness of the training (70%) and test (30%) data used, and the performance indicators. Precision, Sensitivity, specificity, (FPR) false positive rate, accuracy, recall, F1 score, MAE, RMSE and AUC-ROC. The results of the training and validation phases were recorded as shown in Tables 3, 4, and Fig. 11.

In this research, for the training data set, the XGBoost model shows excellent performance, presented by scores (Precision = 0.95), (Sensitivity = 0.97), (Specificity = 0.95), (Accuracy = 0.97), (Recall = 0.96), (F1 score = 0.95), (FPR = 0.05), (MAE = 0.04) and AUC-ROC (values = 96.21%). The other six models RF, KNN, LR, SVM, DT, and ANN show high performance, indicated by Sensitivity, Specificity, Precision, Accuracy, AUC-ROC and F1 score values above 0.88 and negligible FPR and MAE scores. For the test data set, all models applied show high to very high performance. Sensitivity values

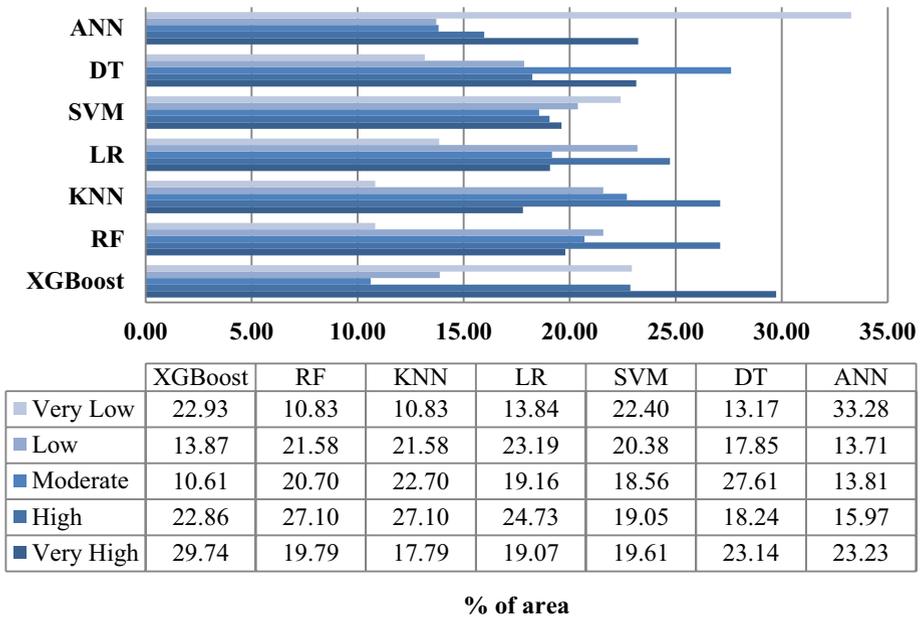


Fig. 10 Percentage of FS surface in the Haouz plain of the models used

Table 3 Performance of the four models using training data

Performance Indicators	Models						
	XGBoost	RF	KNN	LR	SVM	DT	ANN
<i>Training data</i>							
Precision	0.95	0.92	0.88	0.88	0.86	0.92	0.84
Sensitivity	0.97	0.96	0.96	0.95	0.94	0.95	0.98
Specificity	0.95	0.92	0.92	0.88	0.86	0.90	0.84
False positive rate (FPR)	0.05	0.08	0.08	0.12	0.14	0.10	0.16
Accuracy	0.97	0.94	0.94	0.92	0.91	0.93	0.91
recall	0.96	0.93	0.88	0.92	0.90	0.90	0.97
F1 score	0.95	0.93	0.92	0.90	0.88	0.92	0.90
MAE	0.04	0.06	0.06	0.08	0.09	0.07	0.09
RMSE	0.19	0.23	0.24	0.28	0.31	0.26	0.30
AUC-ROC	96.21	94.48	94.31	91.96	90.69	93.38	91.01

ranged from a minimum of 0.90 to a maximum of 0.98 for DT and ANN respectively, parameter values for Precision, Accuracy, F1-score, FPR, MAE varied between 0.78 and 0.93, 0.88 and 0.94, 0.87 and 0.92, 0.07 and 0.22, 0.06 and 0.12 respectively, maximum AUC-ROC values were all marked for the XGBoost model (0.9378), and minimum values were all marked for the ANN model (0.8756).

Table 4 Performance of the four models using testing data

Performance Indicators	Models						
	XGBoost	RF	KNN	LR	SVM	DT	ANN
<i>Testing data</i>							
Precision	0.93	0.93	0.89	0.90	0.91	0.91	0.78
Sensitivity	0.94	0.94	0.94	0.94	0.93	0.90	0.98
False positive rate (FPR)	0.07	0.07	0.14	0.10	0.09	0.09	0.22
Specificity	0.93	0.93	0.86	0.90	0.91	0.91	0.78
Accuracy	0.94	0.94	0.91	0.92	0.92	0.91	0.88
recall	0.92	0.91	0.93	0.91	0.89	0.86	0.98
F1 score	0.92	0.92	0.91	0.91	0.90	0.88	0.87
MAE	0.06	0.06	0.09	0.08	0.08	0.08	0.12
RMSE	0.25	0.25	0.31	0.28	0.28	0.31	0.38
AUC-ROC	93.78	93.71	90.53	92.29	92.02	90.67	87.56

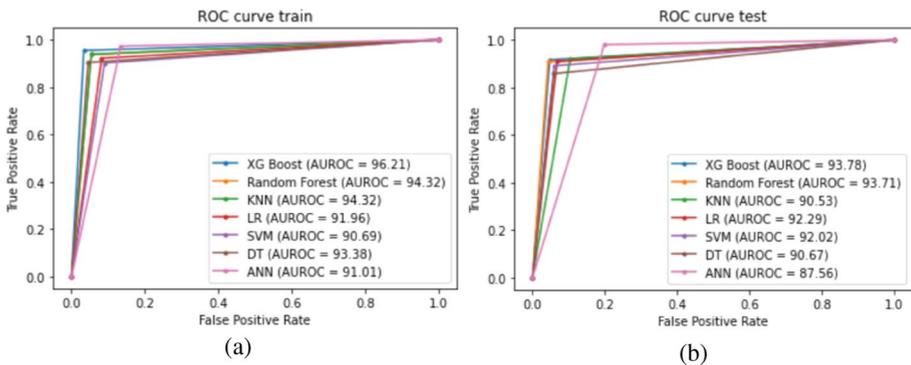


Fig. 11 ROC curve analysis of different flooding models using **a** training data, **b** validation data

Nonetheless, XGBoost outperforms the other four models in terms of accuracy, mostly because to its integration of all base learners’ prediction outcomes, which boosts the model’s recognition rate and generalization capacity. To address missing values on various nodes, distinct techniques will be applied while determining and storing the best course of action. In the meanwhile, XGBoost increases the learning rate by adding a regular term to the objective function and supporting custom loss functions, which prevent overfitting and simplify the learning model. Therefore, the XGBoost model eventually produces better simulation results, and the flood susceptibility mapping approach based on XGBoost is effective and workable.

In addition to the sensitivity and specificity criteria, statistical evaluations were conducted to make the comparison using RMSE, as shown in Tables 3 and 4. The overall RMSE ranking is shown in Fig. 12. The RMSE values for XGBoost indicate an optimal match with the observed and generated values; thus, the expected susceptibility probability has been achieved.

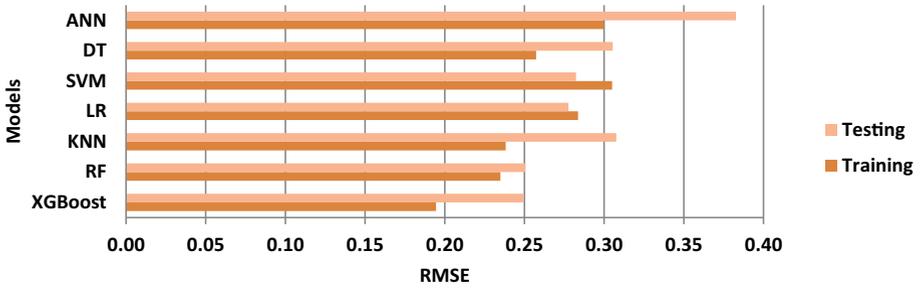


Fig. 12 The performance of different flooding models using RMSE values

3.4 Prioritization of the seven models used

The performance of each model in this study suggests that the XGBoost and RF models had the greatest accuracy in predicting FS for both the training and test datasets. For the training dataset, they are followed by LR, SVM, DT, KNN, and ANN. The following algorithms were ranked for the test dataset: KNN, DT, LR, ANN, and SVM mentioned by Fig. 13.

4 Discussion

In a semi-arid flood-prone area such as the Tensift catchment and the Haouz plain, it is important to categorize and develop FS maps for success and failure rates that ensure flood risk planning and management. In most research and studies, test data sets have been used to construct the reliability performance of flood models. According to (Janizadeh et al.

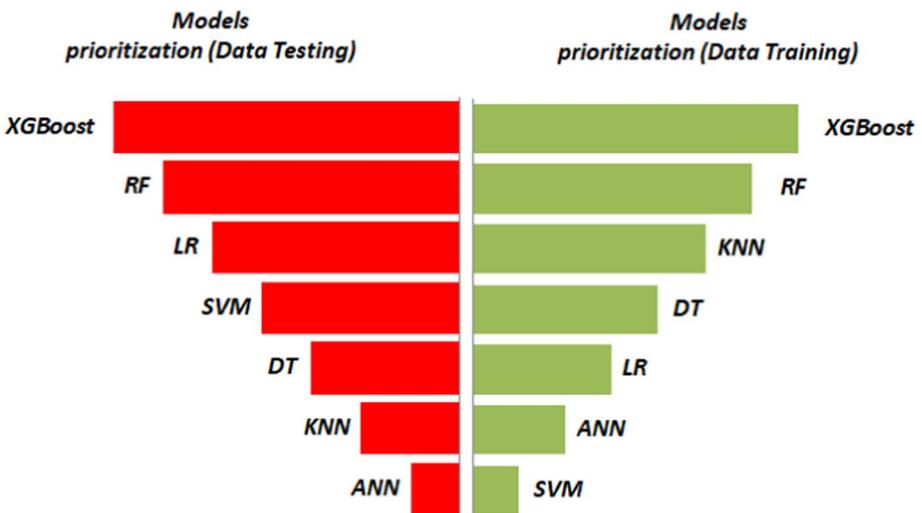


Fig. 13 Prioritization of models in the training and testing phases

2021; Sestras et al. 2023), the spatial prediction of flood risk areas has been attributed to the probability of different thematic flood risk maps. In addition, research and studies have investigated the impact of influential geo-environmental factors and the benefits of several ML algorithms in the mapping of FS in different regions of the world affected by this phenomenon.

The use of the ensemble modelling technique, which combines the seven models to provide more accurate forecasts than each of them in isolation, has increased the accuracy of the results. As a result, the ensemble models gained in importance and were able to anticipate future events more accurately than the individual models.

The importance of the variables affecting vulnerability to flooding, the effectiveness and periodicity of the models used, and the reliability of the technique chosen, are discussed in this section.

4.1 Factors influencing susceptibility to flooding

After the preparation of the inventory maps through field surveys, the collection of data from the processing of satellite scenes, and the analysis of the ABHT historical flood database, the first step in the spatial prediction of FS was the preparation of the set of influencing factors. Of the fifteen factors prepared, TWI and SPI were eliminated because of their collinearity with other factors, which limits the performance of the prediction. Furthermore, according to MI, the most important factor is the distance to the river, which is not consistent with the results of (Meliho et al. 2022), applied in the Ourika catchment which is a totally manganese area. But they are consistent with the results of (Al-Areeq et al. 2022), applied in a mixed area between low and high altitude, followed by elevation. The least important factor in our study is the HSG. In addition, Flow accumulation, Slope, TPI, TRI, Rainfall, LULC, Curvature Plane, Curvature Profile, Aspect, and soil type had a considerable impact on FS, respectively.

The results indicate that the most important locations at a distance from the river, particularly those of Tensift and Ourika along the Haouz plain at low altitude, were highly susceptible to flooding with a high density. The results obtained indicate that high susceptibility is associated with elevation, slope, TPI and TRI along the roads and sub-watersheds of Tensift, the sub-watersheds of Ourika, Zat and Rheraya. This assessment is confirmed by our results. The results and actions of the flood predictors also demonstrate that frequent flooding is expected due to the morphology of the Tensift catchment. In the Kingdom of Morocco, a great deal of research has been carried out to map regions and areas vulnerable to flooding to take targeted measures to reduce the incidence of these risks in the future. Information on the characteristics and impacts of floods is important for administrations and decision-makers to help them formulate policies, as well as flood management strategies, such as the construction of flood-resistant structures to improve emergency response plans in the event of flooding.

4.2 Performance of the models and effectiveness of the methodology

The 70/30% as of Fig. 4 split was used in this study for the choice of training and validation datasets shows high performance. The seven algorithms applied in this study have a high to very high efficiency for the prediction of the FS after the evaluation of the models, despite a slight weakness of the ANN which shows a (AUC-ROC=87.56% in the test stage), is

which has been reported by previous studies in the same theme (Mia et al. 2023; Towfiqul Islam et al. 2021; Youssef et al. 2023a, b).

According to the results, the XGBoost model obtained excellent results in both phases, as it is the case in other studies as well. Methodology adopted in this research, and the results are consistent with recent research studies of Parvin et al. (2022), Seydi et al. (2022), which have highlighted the high and rapid efficiency of ML-DL, RS and GIS in geo-spatial prediction in semi-arid and arid areas characterized by irregular and dangerous flood periods. It is based on a set of geo-environmental factors.

The XGBoost algorithm predicts the FS map effectively, making it a viable alternative for flood risk management in both the Tensift catchment and the Haouz plain. It can specifically help in detecting and classifying regions at high risk of flooding, allowing for the establishment of a monitoring system.

The effective result of this method reflected by the targeted choice of maximum available spatial data of the study area, also to the set of input data, as well as to the application and validation of the algorithm and finally by an extensive field survey, the historical database of the ABHT and satellite scenes of the inventory, the application and evaluation of seven models applied based on the same set of data and parameterized allowed a more logical prioritization of models. Optimizing the performance of the seven models is an essential objective that can be achieved in two significant ways. The first approach consists of enriching all these models by subjecting them to intensive training with a considerable mass of observed data. The second approach relies on the power of the synergy generated by combining these models.

5 Conclusion

This research examined some ML models RF, SVM, KNN, DT, ANN, LR, and XGBoost to map and predict FS in the Tensift watershed and Haouz plain. To this end, mutual information (MI) analysis was used for factor selection and classification. Of these, distance to river and elevation factors were identified as the main factors influencing FS, while TWI and SPI were eliminated from the analysis based on the multicollinearity test and the importance of MI. In addition, the performance measures (Precision, Sensitivity, specificity, FPR, accuracy, recall, F1 score, MAE, RMSE and AUC-ROC) were tested simultaneously with the model evaluation to validate the FS maps. The results for the FS prediction training and validation tests showed that all the models applied met the validation standards. The results of the analysis of the priorities of the models for the prediction of flood risk phenomena demonstrated the superiority of the XGBoost model (AUC-ROC training = 96.21%, AUC-ROC test = 93.78%) and, consequently, the effectiveness of the FS map as predicted by this model. The approach followed in this research has generated an essential ML-RS-GIS-based tool for flood vulnerability mapping, designed to implement prevention and protection plans in a semi-arid context.

It is inevitable to mention certain limitations encountered throughout this research study. The research uses binary flood data from 2012 to create the inventory map of past flooding. This data is from single year, which might not capture the full spectrum of flooding patterns. Future studies should consider using longer time series data to better understand FS, especially in regions with variable climate conditions. In terms of research context, a key concern is that the FS can change over time due to factors like urbanization,

climate change, and land-use modifications. Addressing these limitations will enhance the accuracy and applicability of FS models, improving land-use planning and risk efforts.

This study and its methodology were designed and developed such a way to be more applicable and generalizable across globe. Outcomes of this research highlights the role of advanced algorithms in enhancing precision for flood risk assessments, applicable globally. Methodology and insights can influence policies and practices in regions facing similar environmental challenges. Findings on the efficacy of XGBoost compared to other ML models have broader applications for FS mapping in similar semi-arid regions across globe.

Future research works should extend the study with temporal data to capture changes in FS patterns over time. Also, the integration of dynamic factors like urbanization trends, socio-economic factors, real-time remote sensing data would provide a more holistic assessment for real-time monitoring and would provide a broader understanding regarding the flood impacts on communities and infrastructure. It would be more comprehensive if future works conducted based on comparative studies across different regions to validate the effectiveness of the developed algorithm in varied contexts.

Authors contributions 1. Conceived and designed the studies—Bammou Youssef, Benzougagh Brahim, Igmoullan Brahim, Shuraik Kader, Velibor Spalevic, Paolo Billi, and Slobodan B. Marković. 2. Performed the analysis—Bammou Youssef, Benzougagh Brahim, Igmoullan Brahim, Ouallali Abdessalam, Shuraik Kader, Velibor Spalevic, Paul Sestras, Paolo Billi, and Slobodan B. Marković. 3. Analyzed and interpreted the data—Bammou Youssef, Benzougagh Brahim, Igmoullan Brahim, Ouallali Abdessalam, Shuraik Kader, Velibor Spalevic, Paolo Billi, and Slobodan B. Marković. 4. Contributed materials, analysis tools or data—Bammou Youssef, Benzougagh Brahim, Igmoullan Brahim, Ouallali Abdessalam, Shuraik Kader. 5. Preparation of draft—Bammou Youssef, Benzougagh Brahim, Igmoullan Brahim, Ouallali Abdessalam, Shuraik Kader, Velibor Spalevic, Paul Sestras, Paolo Billi, and Slobodan B. Marković. 6. Internal reviewers – Velibor Spalevic, Paolo Billi, and Slobodan B. Marković. 7. Project administration – Benzougagh Brahim, Shuraik Kader, Velibor Spalevic, Paolo Billi, and Slobodan B. Marković. All authors have read and agreed to the submitted version.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. No fundings were received for this research.

Availability of data and materials The data and materials will be available on request.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

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

References

- Abdulrazzak M, Elfeki A, Kamis A, Kassab M, Alamri N, Chaabani A, Noor K (2019) Flash flood risk assessment in urban arid environment: case study of Taibah and Islamic universities' campuses, Medina, Kingdom of Saudi Arabia. *Geomat Nat Hazards Risk* 10(1):780–796. <https://doi.org/10.1080/19475705.2018.1545705>
- Ahmadlou M, Karimi M, Alizadeh S, Shirzadi A, Parvinnejhad D, Shahabi H, Panahi M (2019) Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto Int* 34(11):1252–1272. <https://doi.org/10.1080/10106049.2018.1474276>
- Al-Areeq AM, Abba S, Yassin MA, Benaafi M, Ghaleb M, Aljundi IH (2022) Computational machine learning approach for flood susceptibility assessment integrated with remote sensing and GIS techniques from Jeddah. *Saudi Arabia Remote Sens* 14(21):5515. <https://doi.org/10.3390/rs14215515>
- Ali SA, Parvin F, Pham QB, Vojtek M, Vojteková J, Costache R et al (2020) GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: a case of Topľa basin. *Slovakia Ecol Indic* 117:106620. <https://doi.org/10.1016/j.ecolind.2020.106620>
- Anjos L, Gaistardo CC, Deckers J, Dondeyne S, Eberhardt E, Gerasimova M et al (2015) World reference base for soil resources 2014 International soil classification system for naming soils and creating legends for soil maps
- Apaydin H, Feizi H, Sattari MT, Colak MS, Shamshirband S, Chau K-W (2020) Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. *Water* 12(5):1500. <https://doi.org/10.3390/w12051500>
- Aryal SK, Silberstein RP, Fu G, Hodgson G, Charles SP, McFarlane D (2020) Understanding spatio-temporal rainfall-runoff changes in a semi-arid region. *Hydrol Process* 34(11):2510–2530. <https://doi.org/10.1002/hyp.13744>
- Beden N, Ulke Keskin A (2021) Flood map production and evaluation of flood risks in situations of insufficient flow data. *Nat Hazards* 105(3):2381–2408. <https://doi.org/10.1007/s11069-020-04404-y>
- Benkirane M, Laftouhi N-E, El Mansouri B, Salik I, Snineh M, El Ghazali FE et al (2020) An approach for flood assessment by numerical modeling of extreme hydrological events in the Zat watershed (High Atlas, Morocco). *Urban Water J* 17(5):381–389. <https://doi.org/10.1080/1573062X.2020.1734946>
- Bennani O, Druon E, Leone F, Trambly Y, Saidi MEM (2019) A spatial and integrated flood risk diagnosis: relevance for disaster prevention at Ourika valley (High Atlas-Morocco). *Disaster Prev Manag Int J* 28(5):548–564. <https://doi.org/10.1108/DPM-12-2018-0379>
- Benssaou M, Hamoumi NM (2003) Le graben de l'Anti-Atlas occidental (Maroc): contrôle tectonique de la paléogéographie et des séquences au Cambrien inférieur. *CR Geosci* 335(3):297–305. [https://doi.org/10.1016/S1631-0713\(03\)00033-6](https://doi.org/10.1016/S1631-0713(03)00033-6)
- Billah M, Islam AKMS, Mamooun WB, Rahman MR (2023) Random forest classifications for landuse mapping to assess rapid flood damage using Sentinel-1 and Sentinel-2 data. *Remote Sens Appl Soc Environ* 30:100947. <https://doi.org/10.1016/j.rsase.2023.100947>
- Bottou L, Vapnik V (1992) Local learning algorithms. *Neural Comput* 4(6):888–900. <https://doi.org/10.1162/neco.1992.4.6.888>
- Bouramtane T, Yameogo S, Touzani M, Tiouiouine A, Elanati MH, Ouadi J et al (2020) Statistical approach of factors controlling drainage network patterns in arid areas (Morocco). Application to the Eastern Anti Atlas Morocco. *J Afr Earth Sci* 162:103707. <https://doi.org/10.1016/j.jafrearsci.2019.103707>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ Model Softw* 95:229–245. <https://doi.org/10.1016/j.envsoft.2017.06.012>
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining
- Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019a) An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci Total Environ* 651:2087–2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>
- Choubin B, Rahmati O, Soleimani F, Alilou H, Moradi E, Alamdari N (2019) Regional groundwater potential analysis using classification and regression trees. In: *Spatial modeling in GIS and R for earth and environmental sciences*, Elsevier, pp 485–498

- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297. <https://doi.org/10.1007/BF00994018>
- Costache R (2019a) Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models. *Sci Total Environ* 659:1115–1134. <https://doi.org/10.1016/j.scitotenv.2018.12.397>
- Costache R (2019b) Flood susceptibility assessment by using bivariate statistics and machine learning models—a useful tool for flood risk management. *Water Resour Manag* 33(9):3239–3256. <https://doi.org/10.1007/s11269-019-02301-z>
- Costache R, Bui DT (2020) Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Sci Total Environ* 712:136492. <https://doi.org/10.1016/j.scitotenv.2019.136492>
- Costache R, Pham QB, Arabameri A, Diaconu DC, Costache I, Crăciun A et al (2022) Flash-flood propagation susceptibility estimation using weights of evidence and their novel ensembles with multicriteria decision making and machine learning. *Geocarto Int* 37(25):8361–8393. <https://doi.org/10.1080/10106049.2021.2001580>
- Cover T, Hart P (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27
- Cox DR (1958) The regression analysis of binary sequences. *J R Stat Soc Ser B Stat Methodol* 20(2):215–232
- Dahri N, Yousfi R, Bouamrane A, Abida H, Pham QB, Derdous O (2022) Comparison of analytic network process and artificial neural network models for flash flood susceptibility assessment. *J Afr Earth Sc* 193:104576. <https://doi.org/10.1016/j.jafrearsci.2022.104576>
- Dazzi S, Vacondio R, Mignosa P (2021) Flood stage forecasting using machine-learning methods: a case study on the Parma River (Italy). *Water* 13(12):1612. <https://doi.org/10.3390/w13121612>
- Duclaux A (2005) Modélisation hydrologique de 5 Bassins Versants du Haut-Atlas Marocain avec SWAT (Soil and Water Assessment Tool). Mémoire du diplôme d'Ingénieur Agronome de l'Institut National Agronomique de Paris-Grignon.
- Echogdali FZ, Kpan RB, Ouchchen M, Id-Belqas M, Dadi B, Ikirri M et al (2022) Spatial prediction of flood frequency analysis in a semi-arid zone: a case study from the Seyad Basin (Guelmim Region, Morocco). *Geospatial Technol Landsc Environ Manag Sustain Assess Plann*. https://doi.org/10.1007/978-981-16-7373-3_3
- El Alfy M (2016) Assessing the impact of arid area urbanization on flash floods using GIS, remote sensing, and HEC-HMS rainfall–runoff modeling. *Hydrol Res* 47(6):1142–1160. <https://doi.org/10.2166/nh.2016.133>
- Fix E, Hodges JL (1952) Discriminatory analysis: nonparametric discrimination: Small sample performance
- Ganie PA, Posti R, Kunal K, Kunal G, Sarma D, Pandey PK (2022) Insights into the morphometric characteristics of the Himalayan River using remote sensing and GIS techniques: a case study of Saryu basin, Uttarakhand, India. *Appl Geomat* 14(4):707–730. <https://doi.org/10.1007/s12518-022-00461-z>
- Han Z, Sharif HO (2021) Analysis of flood fatalities in the United States, 1959–2019. *Water* 13(13):1871. <https://doi.org/10.3390/w13131871>
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* 79(8):2554–2558. <https://doi.org/10.1073/pnas.79.8.2554>
- Hosseini FS, Choubin B, Mosavi A, Nabipour N, Shamsirband S, Darabi H, Haghghi AT (2020) Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. *Sci Total Environ* 711:135161. <https://doi.org/10.1016/j.scitotenv.2019.135161>
- Ighile EH, Shirakawa H, Tanikawa H (2022) Application of GIS and machine learning to predict flood areas in Nigeria. *Sustainability* 14(9):5039. <https://doi.org/10.3390/su14095039>
- Janizadeh S, Chandra Pal S, Saha A, Chowdhuri I, Ahmadi K, Mirzaei S et al (2021) Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future. *J Environ Manag* 298:113551. <https://doi.org/10.1016/j.jenvman.2021.113551>
- Kader S, Raimi MO, Spalevic V, Iyingiala AA, Bukola RW, Jaufer L, Butt TE (2023) A concise study on essential parameters for the sustainability of Lagoon waters in terms of scientific literature. *Turk J Agric* for 47(3):288–307. <https://doi.org/10.55730/1300-011X.3087>
- Kirker AN, Toran L (2023) When impervious cover doesn't predict urban runoff: Lessons from distributed overland flow modeling. *J Hydrol* 621:129539. <https://doi.org/10.1016/j.jhydrol.2023.129539>
- Liu H, Lang B (2019) Machine learning and deep learning methods for intrusion detection systems: a survey. *Appl Sci* 9(20):4396. <https://doi.org/10.3390/app9204396>
- Mao W, Xu C, Yang Y (2022) Investigation on strength degradation of sandy soil subjected to concentrated particle erosion. *Environ Earth Sci* 81:1–10. <https://doi.org/10.1007/s12665-021-10123-9>

- Masoud AM, Pham QB, Alezabawy AK, El-Magd SAA (2022) Efficiency of geospatial technology and multi-criteria decision analysis for groundwater potential mapping in a Semi-Arid region. *Water* 14(6):882. <https://doi.org/10.3390/w14060882>
- Meliho M, Khattabi A, Driss Z, Orlando CA (2022) Spatial prediction of flood-susceptible zones in the Ourika watershed of Morocco using machine learning algorithms. *Appl Comput Inf*. <https://doi.org/10.1108/ACI-09-2021-0264>
- Mia MU, Chowdhury TN, Chakraborty R, Pal SC, Al-Sadoon MK, Costache R, Islam ARMT (2023) Flood susceptibility modeling using an advanced deep learning-based iterative classifier optimizer. *Land* 12(4):810. <https://doi.org/10.3390/land12040810>
- Miao F, Zhao F, Wu Y, Li L, Török Á (2023) Landslide susceptibility mapping in Three Gorges Reservoir area based on GIS and boosting decision tree model. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-023-02394-4>
- Mosavi A, Ozturk P, Chau K-W (2018) Flood prediction using machine learning models: literature review. *Water* 10(11):1536. <https://doi.org/10.3390/w10111536>
- Namous M, Hssaisoune M, Pradhan B, Lee C-W, Alamri A, Elaloui A et al (2021) Spatial prediction of groundwater potentiality in large semi-arid and karstic mountainous region using machine learning models. *Water* 13(16):2273. <https://doi.org/10.3390/w13162273>
- Parvin F, Ali SA, Calka B, Bielecka E, Linh NTT, Pham QB (2022) Urban flood vulnerability assessment in a densely urbanized city using multi-factor analysis and machine learning algorithms. *Theor Appl Climatol* 149(1–2):639–659. <https://doi.org/10.1007/s00704-022-04068-7>
- Sami NA, Ibrahim DS (2021) Forecasting multiphase flowing bottom-hole pressure of vertical oil wells using three machine learning techniques. *Pet Res* 6(4):417–422. <https://doi.org/10.1016/j.ptlrs.2021.05.004>
- Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2(3):160. <https://doi.org/10.1007/s42979-021-00592-x>
- Sestras P, Mircea S, Roșca S, Bilașco Ș, Sălăgean T, Dragomir LO et al (2023) GIS based soil erosion assessment using the USLE model for efficient land management: a case study in an area with diverse pedo-geomorphological and bioclimatic characteristics. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 51(3):13263–13263. <https://doi.org/10.15835/nbha51313263>
- Seydi ST, Kanani-Sadat Y, Hasanlou M, Sahraei R, Chanussot J, Amani M (2022) Comparison of machine learning algorithms for flood susceptibility mapping. *Remote Sens* 15(1):192. <https://doi.org/10.3390/rs15010192>
- Shah K, Patel H, Sanghvi D, Shah M (2020) A comparative analysis of logistic regression, random forest and KNN models for the text classification. *Augment Hum Res* 5:1–16. <https://doi.org/10.1007/s41133-020-00032-0>
- Sivakumar MV (2005) Impacts of natural disasters in agriculture, rangeland and forestry: an overview. *Nat Disasters Extreme Events Agric Impacts Mitig*. https://doi.org/10.1007/3-540-28307-2_1
- Soulaïmani A, Bouabdelli M (2005) Le Plateau de Lakhssas (Anti-Atlas occidental, Maroc): un graben finiprécambrien réactivé à l'hercynien. *Ann Soc Géol Nord* 2(2):177–184
- Tan ML, Gassman PW, Yang X, Haywood J (2020) A review of SWAT applications, performance and future needs for simulation of hydro-climatic extremes. *Adv Water Resour* 143:103662. <https://doi.org/10.1016/j.advwatres.2020.103662>
- Tariq A, Yan J, Ghaffar B, Qin S, Mousa B, Sharifi A et al (2022) Flash flood susceptibility assessment and zonation by integrating analytic hierarchy process and frequency ratio model with diverse spatial data. *Water* 14(19):3069. <https://doi.org/10.3390/w14193069>
- Tehrany MS, Pradhan B, Jebur MN (2014) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J Hydrol* 512:332–343. <https://doi.org/10.1016/j.jhydrol.2014.03.008>
- Thi Thuy Linh N, Pandey M, Janizadeh S, Sankar Bhunia G, Norouzi A, Ali S et al (2022) Flood susceptibility modeling based on new hybrid intelligence model: optimization of XGboost model using GA metaheuristic algorithm. *Adv Space Res* 69(9):3301–3318. <https://doi.org/10.1016/j.asr.2022.02.027>
- Towfiqul Islam ARM, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB et al (2021) Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci Front* 12(3):101075. <https://doi.org/10.1016/j.gsf.2020.09.006>
- Verma S, Bhatla R, Shahi N, Mall R (2022) Regional modulating behavior of Indian summer monsoon rainfall in context of spatio-temporal variation of drought and flood events. *Atmos Res* 274:106201. <https://doi.org/10.1016/j.atmosres.2022.106201>
- 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>

- Wang Y, Hong H, Chen W, Li S, Pamučar D, Gigović L et al (2018) A hybrid GIS multi-criteria decision-making method for flood susceptibility mapping at Shangyou, China. *Remote Sens* 11(1):62. <https://doi.org/10.3390/rs11010062>
- Youssef AM, Mahdi AM, Pourghasemi HR (2023a) Optimal flood susceptibility model based on performance comparisons of LR, EGB, and RF algorithms. *Nat Hazards* 115(2):1071–1096. <https://doi.org/10.1007/s11069-022-05584-5>
- Youssef B, Bouskri I, Brahim B, Kader S, Brahim I, Abdelkrim B, Spalević V (2023b) The contribution of the frequency ratio model and the prediction rate for the analysis of landslide risk in the Tizi N'tichka area on the national road (RN9) linking Marrakech and Ouarzazate. *CATENA* 232:107464. <https://doi.org/10.1016/j.catena.2023.107464>
- 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>

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

Authors and Affiliations

Youssef Bammou¹ · Brahim Benzougagh² · Brahim Igmoullan¹ · Abdessalam Ouallali³ · Shuraik Kader^{4,5}  · Velibor Spalevic⁶ · Paul Sestras^{7,8} · Paolo Billi^{9,10} · Slobodan B. Marković¹¹

✉ Shuraik Kader
shuraik10@gmail.com; shuraik.mohamedabdulkader@griffithuni.edu.au; s.mohamedabdulkader@griffith.edu.au

Youssef Bammou
youssef.bammou@ced.uca.ma

Brahim Benzougagh
brahim.benzougagh@is.um5.ac.ma; benzougaghbrahim@gmail.com

Brahim Igmoullan
igmoullan@gmail.com

Abdessalam Ouallali
abdessalam.ouallali@univh2c.ma

Velibor Spalevic
velibor.spalevic@gmail.com

Paul Sestras
psestras@yahoo.com; psestras@mail.utcluj.ro

Paolo Billi
bli@unife.it

¹ Laboratory of Geo-Resources, Department of Geology, Faculty of Science and Technology, Geo-Environment and Civil Engineering (L3G), Cadi Ayyad University, Marrakech, Morocco

² Department of Geomorphology and Geomatics, Scientific Institute, Mohammed V, University in Rabat, Avenue Ibn Battouta, Agdal, P.B. 703, 10106 Rabat-City, Morocco

³ Process Engineering and Environment Laboratory, Faculty of Sciences and Techniques of Mohammedia, Hassan II University of Casablanca, BP 146, 28806 Mohammedia, Morocco

⁴ School of Engineering and Built Environment, Griffith University, Nathan, QLD 4111, Australia

⁵ Green Infrastructure Research Labs (GIRLS), Cities Research Institute, Griffith University,

Nathan, QLD 4111, Australia

- ⁶ Biotechnical Faculty, University of Montenegro, 81000 Podgorica, Montenegro
- ⁷ Faculty of Civil Engineering, Technical University of Cluj-Napoca, 400020 Cluj-Napoca-Napoca, Romania
- ⁸ Academy of Romanian Scientists, Ilfov 3, 50044 Bucharest, Romania
- ⁹ Department of Earth Sciences, University of Ferrara, 44121 Ferrara, Italy
- ¹⁰ International Platform for Dryland Research and Education, Tottori University, 680-0945 Tottori, Italy
- ¹¹ Department of Geography, Tourism and Hotel Management, Faculty of Sciences, University of Novi Sad, Trg Dositeja Obradovića 3, 21000 Novi Sad, Serbia