Check for updates

Journal of
Water & Climate Change

Integrated approaches for flash flood susceptibility mapping: spatial modeling and comparative analysis of statistical and machine learning models. A case study of the Rheraya watershed, Morocco

Akram Elghouat [ID][a,b,*], Ahmed Algouti[a], Abdellah Algouti[a], Soukaina Baid[a], Salma Ezzahzi[a], Salma Kabili[a] and Saloua Agli[a]

[a] Faculty of Sciences Semlalia, Department of Geology, 2GRNT Laboratory (Geosciences, Geotourism, Natural Hazards and Remote Sensing), Cadi Ayyad University, Marrakech, Morocco
[b] Interuniversity Institute for Earth System Research (IISTA), University of Granada, Granada 18006, Spain
*Corresponding author. E-mail: elghouatakram@go.ugr.es

[ID] AE, 0000-0003-0744-4715

ABSTRACT

Flash floods are highly destructive disasters, posing severe threats to lives and infrastructure. In this study, we conducted a comparative analysis of bivariate and multivariate statistical models and machine learning to predict flash flood susceptibility in the flood-prone Rheraya watershed. Six models were utilized, including frequency ratio (FR), logistic regression (LR), random forest (RF), extreme gradient boosting (XGBoost), K-nearest neighbors (KNN), and naïve Bayes (NB). We considered 12 flash flood conditioning variables, such as slope, elevation, distance to the river, and others, as independent variables and 246 flash flood inventory points recorded over the past 40 years as dependent variables in the modeling process. The area under the curve (AUC) of the receiver operating characteristic was used to validate and compare the performance of the models. The results indicated that distance to the river was the most contributing factor to flash floods in the study area. Moreover, the RF outperformed all the other models, achieving an AUC of 0.86, followed by XGBoost (AUC = 0.85), LR (AUC = 0.83), NB (AUC = 0.76), KNN (AUC = 0.75), and FR (AUC = 0.72). The RF model effectively pinpoints highly susceptible zones, which is critical for establishing precise flash flood mitigation strategies within the region.

Key words: flash floods, frequency ratio, GIS, logistic regression, machine learning, remote sensing

HIGHLIGHTS

- Comparative assessment of bivariate and multivariate statistical models and machine learning algorithms to classify locations as flash-flooded or not.
- Distance to the river and drainage density have the strongest influence on flash floods in the Rheraya watershed.
- The RF model outperformed all the other models.
- Highly susceptible zones within the ungauged and flood-prone Rheraya watershed were effectively identified.

GRAPHICAL ABSTRACT
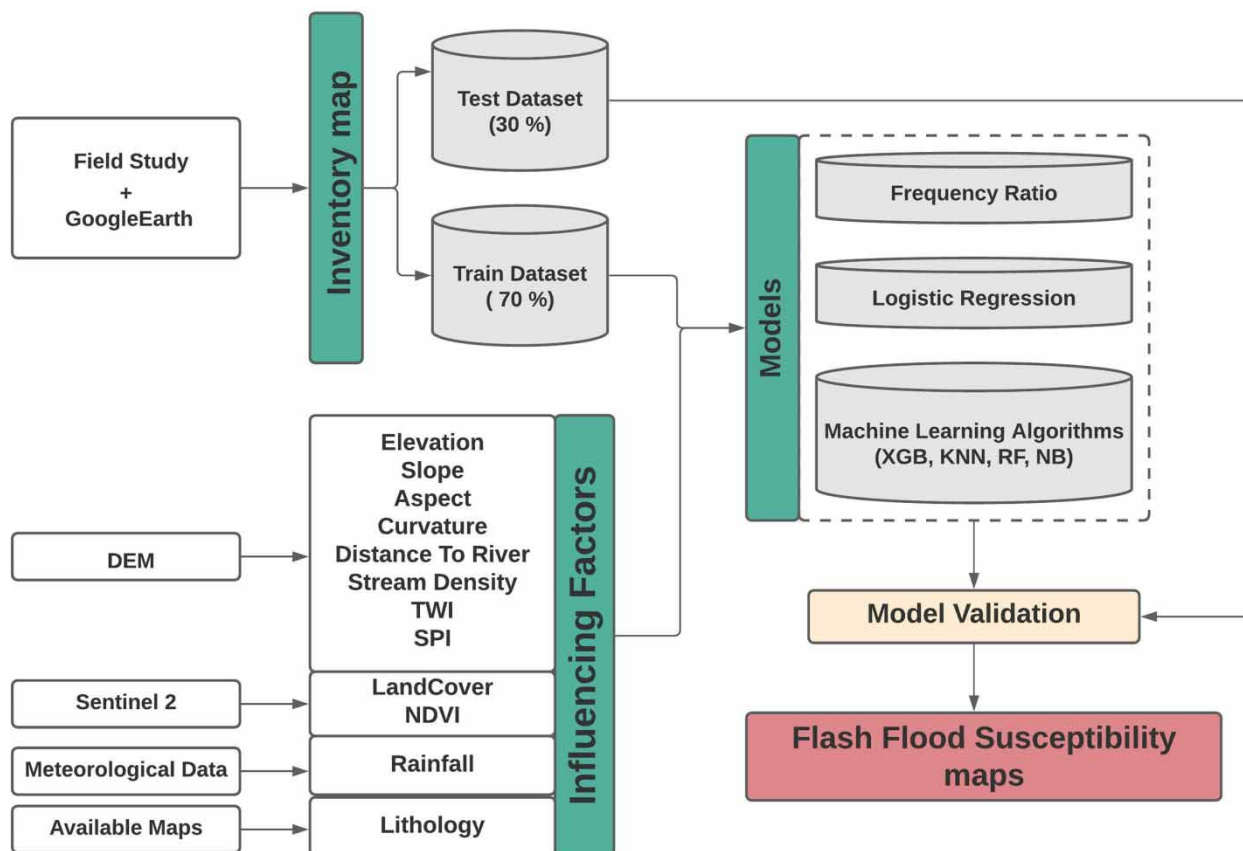
1. INTRODUCTION

A flash flood is a rapid rise of water along a stream or low-lying urban area. Flash flooding occurs within 6 h of a significant rain event and is usually caused by intense storms that produce heavy rainfall in a short amount of time (Doswell 2015). They are one of the most devastating, costly, and frequent disasters, resulting in extensive damage to infrastructure and properties and even the loss of human lives (Jeyaseelan 2003; Tingsanchali 2012; Kuenzer *et al.* 2013; Dahri & Abida 2017; Nogueira *et al.* 2018; Mohanty *et al.* 2020). More specifically, they are responsible for approximately 84% of global deaths (Jamali *et al.* 2020). In recent years, the increasing frequency of extreme weather events related to global warming, urbanization, loss of natural land, deforestation, and changes in land use patterns such as canalization of water streams has made flash floods a growing concern for scientific communities worldwide, as they expect the severity of this phenomenon to increase (Kundzewicz *et al.* 2014; Guha-Sapir *et al.* 2016; Mekonnen & Hoekstra 2016). Effective flood management strategies and accurate identification of high-risk areas are therefore necessary to reduce the impact of these events on human lives and properties (Grothmann & Reusswig 2006; European Union 2007). However, predicting flooding remains a challenging task due to the complex nature of the phenomenon (Kalantari *et al.* 2014).

In the context of climate change, the severity and intensity of floods are increasing in Morocco (Loudyi *et al.* 2022). Over recent decades, the country, like many others worldwide, has experienced numerous destructive hydrological events that have impacted many areas of the country. In fact, flash floods are the most common and dangerous disasters in the country due to how often they happen, their magnitude, and their sudden onset (Mallouk *et al.* 2016; Karmaoui & Balica 2019). Moreover, there is a 95% chance that the country will face either earthquakes or floods over the next 30 years (The World Bank 2013). This is mainly due to the geographical location of the country, which frequently experiences intense rainfall events. The country has experienced over 35 major flash floods between 1951 and 2015 (TARGA-AIDE & Zurich Insurance 2015). More specifically, the Rheraya watershed witnessed one of the most destructive flooding events in 1995 due to

heavy rainfall in a short period of time, killing more than 150 people, including 60 tourists (Digby 2000). Moreover, in 2019, the basin encountered another severe flash flood, which resulted in significant damage to properties. Similarly, flooding in the Ourika basin, which limits the Rheraya basin to the east, resulted in the deaths of more than 200 people in 1995 and more than 60 people in 2002 (Kingdom of Morocco 2011). The high risk and significant loss of life in these specific areas can also be attributed to the fact that residents often inhabit these exposed regions for tourism purposes (El Alaoui El Fels et al. 2018). Therefore, pinpointing highly susceptible zones to flash floods in the Rheraya watershed is of paramount relevance for developing precise mitigation strategies in this touristy zone.

Utilizing flood susceptibility mapping (FSM) is essential for identifying flood-prone areas by considering various environmental factors influencing floods (Wang et al. 2019). Rainfall-runoff models are valuable for flood forecasting but come with limitations, including the need for calibration and extensive gauging data, which can be resource-intensive (Ludwig et al. 2003; Peel & McMahon 2020). Researchers have utilized diverse methods for creating flood susceptibility maps, including geospatial analysis through geographic information system tools and remote sensing. Recent advancements include the adoption of statistical methods and machine learning (ML) algorithms. Statistical methods are commonly applied for studying spatial phenomena and mitigating flood risks. Tehrany et al. (2017), for instance, used bivariate and multivariate statistical models to map flood susceptibility in Busan City, South Korea. Similarly, Bui et al. (2019) conducted flood susceptibility modeling in the Haraz catchment, Iran, using a multivariate logistic regression (LR) model. Both studies showed high performance of their models in mapping the susceptibility to floods. ML algorithms, on the other side such as artificial neural networks (ANNs) (Shu & Burn 2004; Seckin et al. 2013; Liu et al. 2016; Jahangir et al. 2019; Rahman et al. 2021), support vector machines (Tehrany et al. 2014a, 2015; Dazzi et al. 2021), random forests (RFs) (Chapi et al. 2017; Lee et al. 2017; Farzaneh et al. 2019; Vafakhah et al. 2020; Abedi et al. 2022; Ghanim et al. 2023), adaptive neuro-fuzzy inference systems (Ahmadlou et al. 2018), and long-term memory (Apaydin et al. 2020; Dazzi et al. 2021), have enhanced flood risk prediction accuracy by addressing non-linearity. For example, Seydi et al. (2022) and Sellami et al. (2022) conducted comparative analyses of multiple ML models to assess flash flood susceptibility (FFS) in Iran and Tetouan (Morocco), respectively. Their findings demonstrated that while there were minor differences in algorithm performance, all models effectively identified areas at very high risk of flooding. Nonetheless, the literature lacks a comprehensive comparative study that incorporates both statistical models and ML techniques. Specifically, there is limited research assessing the performance of bivariate and multivariate statistical models alongside various ML algorithms in the same geographical area, with only one study focusing on just two ML algorithms based on regression trees (Al-Abadi & Al-Najar 2020).

This study aims to compare the effectiveness of statistical and ML models for predicting FFS and developing an FFS map for the Rheraya watershed, Morocco. Our specific objectives were (1) to identify the most contributing factors to flash flood occurrence in the region; (2) to create flood susceptibility maps of the region using various models; (3) to assess and compare the performance of the models; and (4) to pinpoint high-risk zones within the study area. The novelty of this study lies in being the first study to conduct a comparative analysis of bivariate and multivariate statistical models and various ML algorithms to assess FFS within a single study and for a specific area. Second, it addresses a crucial gap by developing the first FFS map of the ungauged Rheraya watershed, an area renowned for its historically destructive flash floods. Through this study, we seek to provide valuable insights that will aid in the effective management and mitigation of flash flood risk in the Rheraya watershed.

The subsequent sections of the paper are organized as follows: Section 2 describes the study area, methodological approach, data sources for modeling, and background information on the statistical and ML models used in this work. Section 3 outlines the obtained results, while Section 4 presents a discussion of the outcomes, including a comparison between the various models and the final FFS maps. Finally, Section 5 summarizes the research findings with closing remarks.

2. MATERIALS AND METHODS

2.1. Study area

The Rheraya watershed is located in the south of Marrakech city in the Western High Atlas region. Its outlet is defined by the hydrometric station of Tahanaout at 1,030.0 m (31.3°N 7.9°E) (Figure 1). Major population centers in the area include Asni, Imlil, and Moulay Brahim, each hosting various tourism amenities such as restaurants, cafes, and accommodations. The Rheraya is considered to be one of the main catchment areas in the High Atlas range, where the socio-economic status of the region is characterized by a significant reliance on tourism, recreational activities, and agriculture. It covers an area of
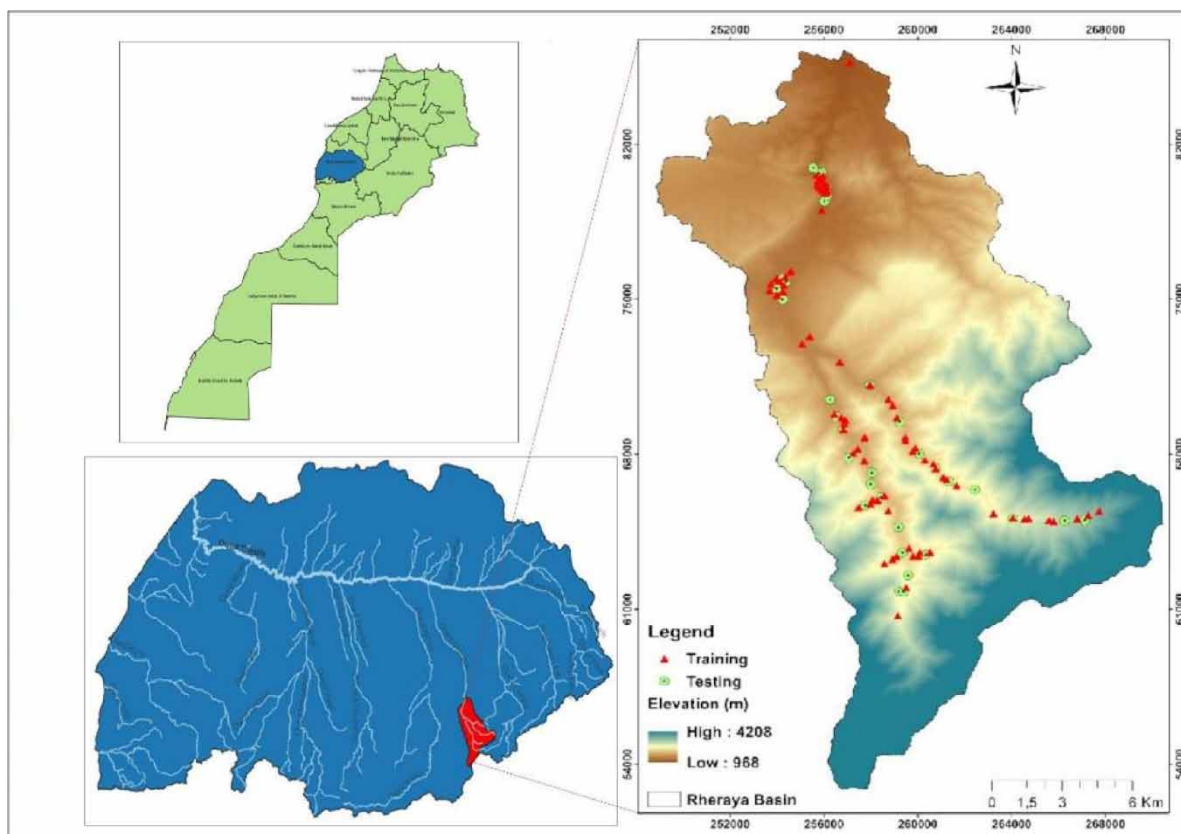
Figure 1 | The location of the study area with the training and testing flash flood inventory points.

224 km^2, with elevations ranging from 1,084.0 to 4,167.0 m and slopes of up to 80°. Rheraya exhibits significant spatio-temporal variability in precipitation, ranging from 600 mm· year^{-1} in high mountain areas to 356 mm year^{-1} in low mountain regions (Chaponnière *et al.* 2008). Approximately 30–50% of the precipitation occurs as snow (Boudhar *et al.* 2009). At the Tahanaout gauge station, the average runoff is recorded at 1.15 m^3 s^{-1}. The hydrological pattern of the Rheraya River is characterized as pluvio-nival, featuring a unimodal distribution peaking in April, which correlates with snowmelt. The temperatures typically range from 18 to 38 °C. Approximately 50% of the land is covered by forests, 10% by urban areas, 4% comprises orchards, and 28% is dedicated to agriculture. In detail, the hillslopes in the study area are characterized by degraded rangelands with minimal flora and a stone cover, while the valley has a limited strip of irrigated crops on either side of the river (Boudhar *et al.* 2007).

2.2. Data and methods

A flowchart was developed to outline and summarize the adopted methodology of our research (Figure 2). Two statistical methods (i.e., frequency ratio (FR) and LR) and four advanced ML algorithms (i.e., extreme gradient boosting (XGBoost), RF, naïve Bayes (NB), K-nearest neighbors (KNN)) were used for the present study. The process consisted of the following steps: (a) generate a flash flood inventory map (dependent variable); (b) identify flash flood conditioning factors (independent variables); (c) analyze the spatial relationship between each influencing factor and flash flood events; (d) determine the optimal parameters and create a map of FFS; and (e) evaluate the performance of the models.

2.2.1. Inventory map

Identifying and creating a flash flood inventory map of the study area is a critical step in investigating the relationship between flash flood events and various influencing factors. Current approaches generally rely on historical records, field surveys, and satellite imagery. We first used the historical flood records of the Tensift Hydraulic Basin Agency (ABHT), followed by a field survey in which local residents contributed to the development of a historical flash flood inventory map, and finally, Google
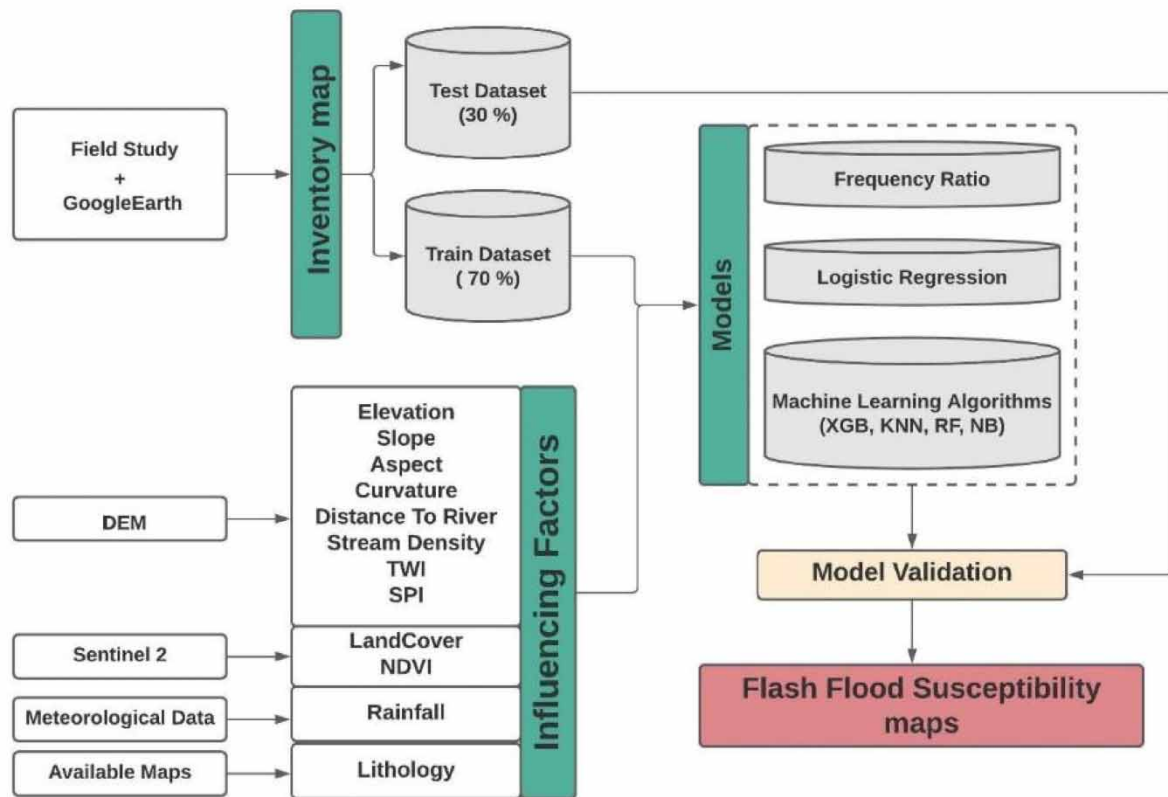
Figure 2 | Methodological approach applied for flash flood susceptibility modeling.

Earth images of both pre- and post-flash floods were exploited for further verification. Through these steps, we were able to identify 123 flash flood locations. In addition, 123 points were selected as 'non-flash flood' from the areas where there was no evidence of flash floods occurring. Flash FSM is a binary classification; therefore, to generate the training data, we assigned binary values of 1 to flash flood points and 0 to non-flash flood points for the modeling process. The resulting dataset (i.e., 246 points) was then randomly split into a 70% training set and a 30% testing set (Figure 1).

2.2.2. Flash flood influencing factors

The identification of flash flood conditioning factors can greatly affect the accuracy of the mathematical models (Kia *et al.* 2012). We thoughtfully selected the conditioning factors based on a comprehensive literature review of previous studies (e.g., Bentivoglio *et al.* 2022), data availability, and the characteristics of flash floods in the Rheraya watershed. We made sure to consider different aspects of the study area, including topographic, hydrological, geological, and land cover features.

First, a 12.5 m digital elevation model (DEM) of the study area was downloaded from the ALOS PALSAR sensor and processed using ArcGIS 10.2. Using this DEM, topographic and hydrological variables such as elevation (Figure 3(a)), slope (Figure 3(b)), aspect (Figure 3(c)), curvature (Figure 3(d)), stream power index (SPI) (Figure 3(f)), drainage density (Figure 4(b)), and topographic wetness index (TWI) (Figure 3(e)) were generated. The Euclidean distance tool in ArcGIS was used to create a map of the distance to the river (Figure 4(a)). In addition, 40 years of precipitation data from two stations in the study area were used to create an annual rainfall map using the inverse distance weighted (IDW) interpolation approach (Figure 4(c)). To generate land cover and NDVI maps, Sentinel-2B satellite data with a resolution of 10 m were used. Finally, lithology, which characterizes the soil and underlying rocks that can influence catchment infiltration and runoff (Janizadeh *et al.* 2019; Zhao *et al.* 2019), was incorporated into the modeling process using a lithology map of the study area, created by digitizing a 1:100,000 geological map of Tahanaout (Figure 4(f)). The role of each factor in flash flood occurrences is discussed in further detail in the following sections.
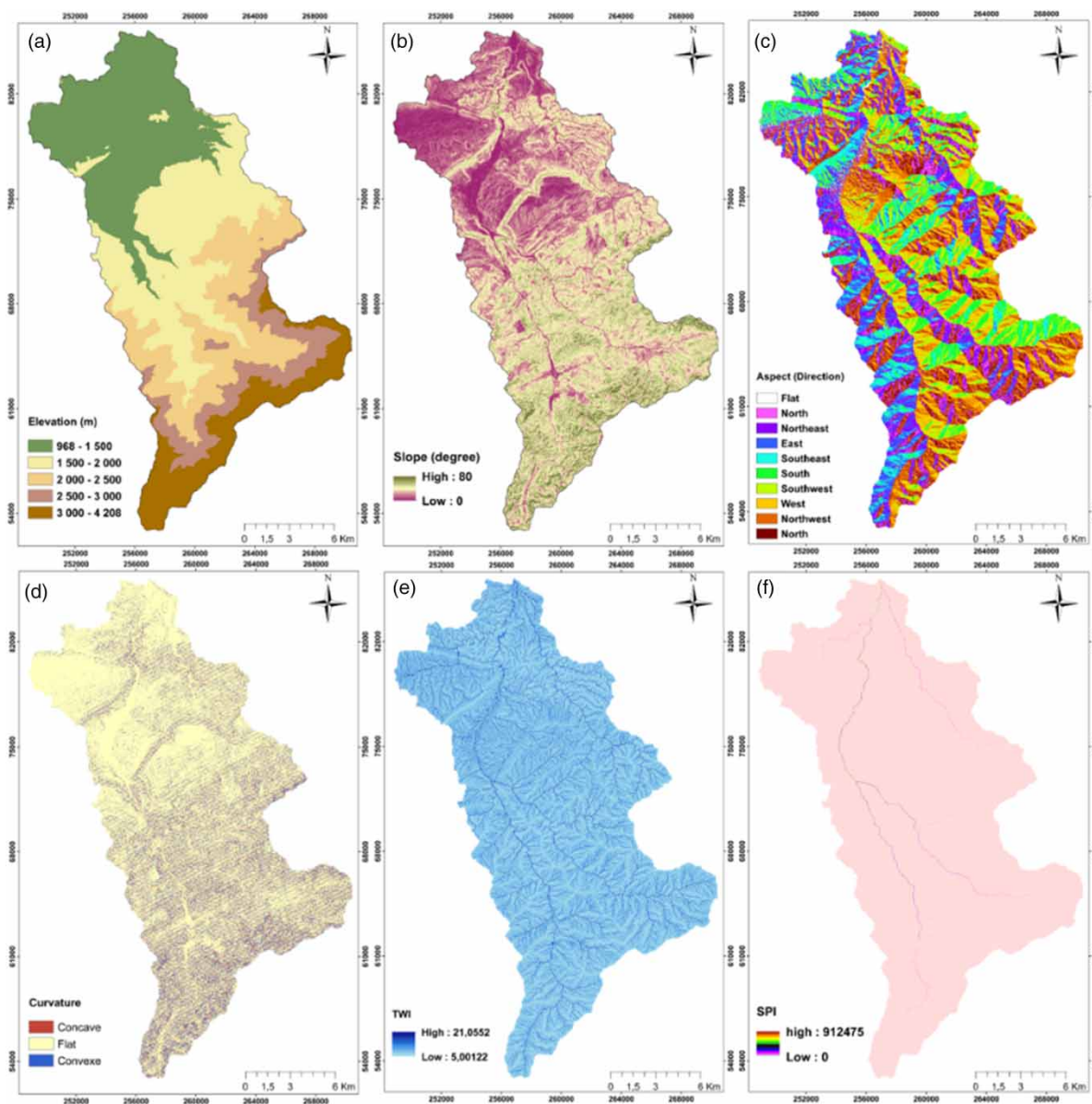
Figure 3 | Flash flood condition factors used in this study are: (a) elevation, (b) slope, (c) aspect, (d) curvature, (e) TWI, and (f) SPI.

2.2.2.1. Topographical factors. Topographical features included elevation, slope, aspect, curvature, and distance to the river. Elevation is a key factor in flash flood modeling (Bui *et al.* 2020; Dodangeh *et al.* 2020). It is inversely related to flash floods (Fernández & Lutz 2010), meaning that as elevation decreases, the terrain becomes flatter and the amount of water carried by rivers increases (Cao *et al.* 2016). Slope is an important factor in flash floods because it affects the speed of flowing water (Stevaux *et al.* 2020). In general, a steeper slope angle leads to higher flow velocity, which can decrease the rate of infiltration and increase water stagnation. Aspect influences floodwater flow directions, which helps to maintain the humidity of the soil (Chu *et al.* 2020). It indirectly affects the flooding. Slope curvature separates diverging and converging runoff regions, which influences water flow (Torcivia & López 2020). Depending on the slope, runoff accelerates or decelerates. Convex slopes tend to increase overland flow, potentially affecting infiltration and soil saturation (Cao *et al.* 2016), while concave slopes can slow down overland flow and potentially improve infiltration (Young & Mutchler 1969). Distance to the river is a critical factor in determining an area's vulnerability to flooding in a basin (Tehrany *et al.* 2015). Areas closer to rivers are more prone to flooding than those farther away (Butler *et al.* 2006; Chapi *et al.* 2017).
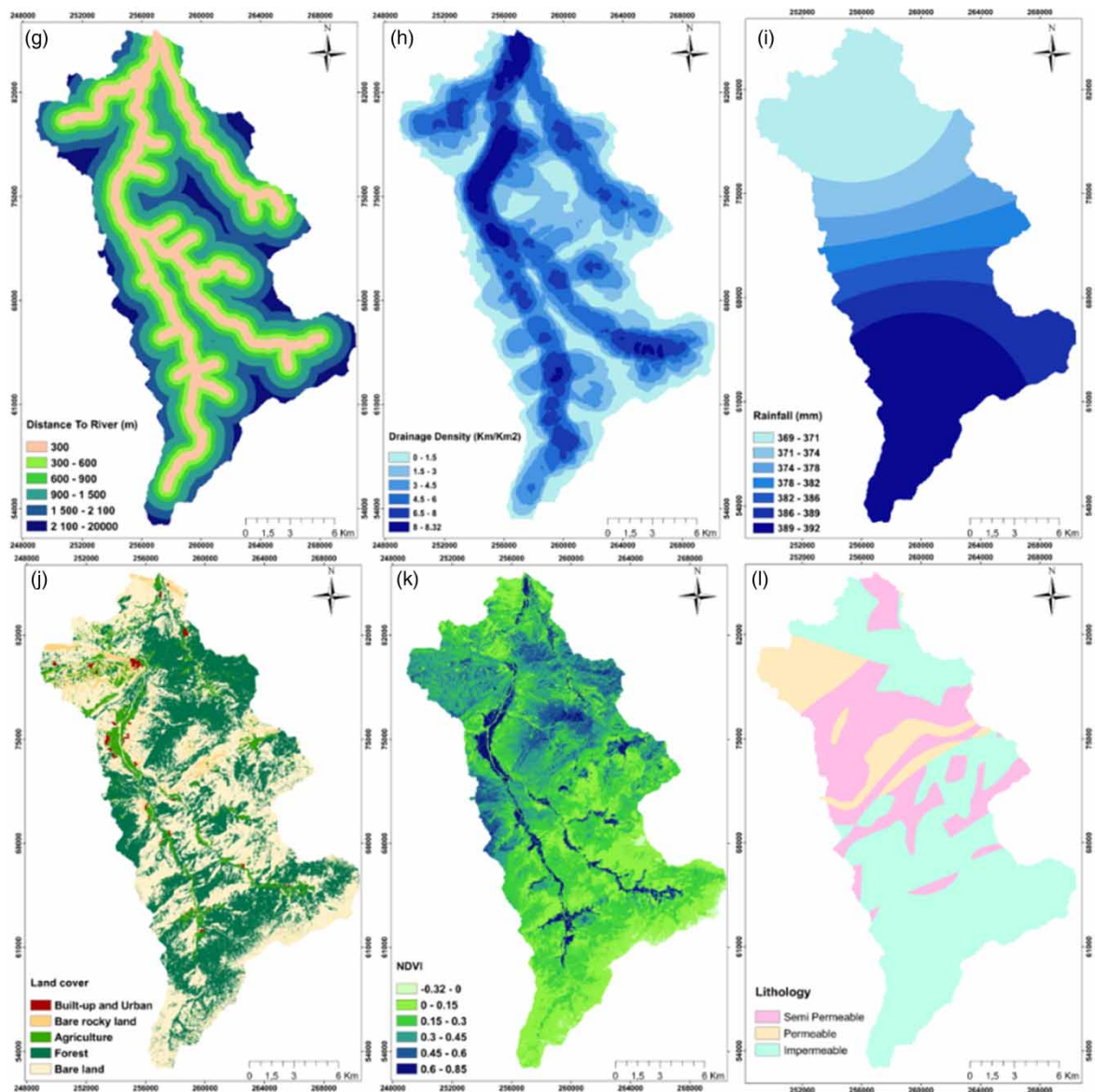
Figure 4 | Flash flood condition factors used in this study are: (a) distance to the river, (b) drainage density, (c) rainfall, (d) land cover, (e) NDVI, and (f) lithology.

2.2.2.2. Hydrological and meteorological factors. Hydrological and meteorological variables directly influence the occurrence and severity of flash floods. Various factors were taken into consideration in this study, such as SPI, TWI, drainage density, and rainfall. SPI is a measure that assesses the potential flow erosion at a specific topographic surface point. It has a significant impact on the fluvial system (Knighton 1999). TWI is a hydrological measure corresponding to the ratio of the area of a specific basin to the angle of the slope (Wilson & Gallant 2000; Nhu *et al.* 2020). It reflects the amount of water present in each pixel of the area (Zhang *et al.* 2020). Drainage density is the sum of the stream length per unit watershed area (Elmore *et al.* 2013; Nguyen *et al.* 2020). High stream densities have a higher risk of flooding than low stream densities, assuming all other conditions are equal (Chapi *et al.* 2017). Rainfall is one of the most significant factors that can cause floods (Pourghasemi *et al.* 2020). When the intensity of the rain exceeds the ground's infiltration capacity, flash floods occur. To generate the precipitation map, data acquired over a 40-year period (1983–2023) from two meteorological stations, Tahanaout and Armed, were used, with Tahanaout located in the lower catchment and Armed positioned in the upper catchment. The average annual rainfall over 40 years for each station was computed, and a rainfall map was generated using the IDW interpolation method in ArcGIS 10.2.

2.2.2.3. Geological and land cover factors. Geological and land cover variables included lithology, land cover, and NDVI. The variety of lithologic structures in a study area can significantly increase or decrease the level of flood risk since the permeability and porosity of these different structures directly influence infiltration and runoff. Land cover affects surface runoff and sediment transport, which directly influence flood frequency (Benito *et al.* 2010). Flooding is more common in urban areas, whereas vegetation, particularly forests, intercepts precipitation and slows runoff velocity. We created a land cover map using Sentinel-2B images acquired in May 2023. These images were chosen because they exhibited lower cloud coverage and reduced snow cover. The images were classified into five classes, including agriculture, bare land, bare rocky soil, forest, and built-up areas, using the maximum likelihood classification in SNAP software (Figure 4(d)). Field surveys and Google Earth images were used to validate the obtained map. The NDVI is defined as a dimensionless index that describes the difference between near-infrared and red light, which has values ranging from -1 (e.g., low vegetation density) to $+1$ (e.g., high vegetation density). NDVI values can indicate changes in vegetation and surface water cover over time (Ahmed & Akter 2017) and reveal the relationship between flooding and vegetation in a basin (Tehrany *et al.* 2013). To create the NDVI map of the Rheraya watershed, Sentinel-2B images from May 2023 were used, and the map was classified into six classes (Figure 4(e)).

2.3. Flash flood susceptibility modeling

2.3.1. Statistical approaches

2.3.1.1. Frequency ratio. The FR is defined as a bivariate statistical analysis approach for analyzing the impact of different flood conditioning factors on the likelihood of future floods (Lee *et al.* 2012). It examines the correlation between flash flood locations and their related factors in a specific area. The equation used to calculate the FR is:

$$FR = [N_{pix}(SX_i) / \sum_{i=1}^{m} SX_i] / [N_{pix}(X_j) / \sum_{j=1}^{n} N_{pix}(X_j)]$$

where $N_{pix}(SX_i)$ is the number of flash flood points in class i of variable X, $N_{pix}(X_j)$ is the number of pixels in variable X_j, m is the total number of classes in variable X_i, and n is the total number of flood conditioning factors (Regmi *et al.* 2013; Jaafari *et al.* 2014). The FR values for each class of each conditioning factor are then used to create a susceptibility map using the following equation:

$$\text{Flash flood susceptibility index (FSI)} = \sum_{j=1}^{n} FR$$

The FR is one of the most commonly used and trustworthy methods to evaluate the susceptibility to floods worldwide (Rahmati *et al.* 2016a, 2016b; Samanta *et al.* 2018). An FR value higher than 1 indicates that the factors have a substantial influence on flash flooding, while an FR value less than 1 means that there is a negative correlation between the flash flood frequency and the conditioning factors (Lee & Talib 2005).

2.3.1.2. Logistic regression. LR is a multivariate statistical analysis approach applied to estimate and predict the likelihood of an event occurring by analyzing multiple variables. The method uses dichotomous data such as 1 or 0 (presence or absence) to identify the relationship between flash floods and the factors that influence them (Helsel & Hirsch 2002). The main objective of this approach is to find the most accurate model that explains the connection between the variables. To calculate the LR model, we used the equation:

$$P = 1/(1 + e^{-z})$$

where z is the linear regression model:

$$z = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n$$

The intercept of the model is b_0, the number of independent variables is n, the coefficients are $b_1, b_2,\dots b_n$, and the flash flood conditioning factors are $x_1, x_2,\dots x_n$. The P-value indicates the probability of vulnerability and ranges from 0 to 1. A P-value close to 1 suggests a high vulnerability, while a P-value close to 0 represents a low vulnerability.

2.3.2. ML algorithms

2.3.2.1. Extreme gradient boosting. XGBoost is an advanced ML algorithm that minimizes a regularized objective function. It handles missing values and uses tree pruning. Examples are initially grouped so that similar residues are in the same cluster, and then they are branched off. A similarity score for each threshold is given by the following equation:

$$\text{Similarity score} = \frac{(\sum \text{Residuals})^2}{\sum [pi' \times (1 - pi')] + \lambda}$$

where pi' is the previous probability calculated for the i training example, and λ is the regularization parameter.

The minimum number of residuals in each leaf, as well as the minimum child weight (cover), must also be considered (Madhuri *et al.* 2021). Generally, the minimum cover value is 1. The equation for calculating cover is:

$$\text{Cover} = \sum [pi' \times (1 - pi')]$$

The leaf is removed if the cover is less than the minimum. The tree complexity parameter facilitates tree pruning. For all leaves, the output values (W) are given by the following equation:

$$W = \frac{\sum \text{Residuals}}{\sum [pi' \times (1 - pi')] + \lambda}$$

The output value is multiplied by the rate of learning, subtracted from the prior probability, and expressed in terms of the log of the odds. After one iteration, the probability can be calculated by changing the obtained value back to probability. In the end, XGBoost aims to reduce the objective function in the equation given below:

$$O(y_i, p_i, w) = \sum_{i=1}^{n} L\,(y_i,\ p_i)\ +\ \frac{1}{2}\,\lambda w2$$

The loss function, $L\,(y_i, p_i)$, was used as the negative log-likelihood function in the equation that follows to derive all the equations and expressions above:

$$L(y_i,\ p_i) = [-y_i \log (p_i) + (1-y) \log (1-p_i)]$$

The optimization procedure in XGBoost starts with the creation of the first learner for the variable dataset, followed by the creation of the model according to the residuals. When it reaches the stopping criteria, the procedure ends. The algorithm becomes stronger when there is missing data in the dataset compared with other models. The Caret package in the R statistical software (R.3.6.2., R Core Team) was used to apply the XGBoost algorithm.

2.3.2.2. Random forest. The RF algorithm (Breiman 2001) is a binary decision tree-based ensemble classifier system. It is primarily a statistically based approach and can handle many variables. The model creates numerous trees based on random bootstrapping using the training dataset (Goetz *et al.* 2015) and is commonly employed to investigate the complex relationships between explanatory and response variables. Therefore, the RF model has been used to identify hierarchical and non-linear interactions in large datasets as well as to better predict novel data cases (Olden *et al.* 2008). Three parameters must be tuned in the RF algorithm: the number of variables, the number of trees, and the maximum number of terminal nodes (Yang *et al.* 2016). The model is built on tree-structured classifiers and is represented as follows:

$$h(x, i_k), \quad k = 1, 2, \ldots n$$

where 1, 2… n are the input vectors, and x and i_k are the conditioning factors. The general error of the algorithm is defined as follows (Masetic & Subasi 2016):

$$GE = P_{x,y}(mg(x, y) < 0)$$

where x and y indicate the different conditioning factors of the flash flood, and mg represents the margin function. The RF model is generated using the Caret package in R statistical software (R.3.6.2, R Core Team).

2.3.2.3. K-nearest neighbors.   KNN is an easy-to-use supervised ML algorithm that can be used for both regression and classification tasks. It is a non-parametric and lazy algorithm, meaning it does not make assumptions about the dataset and only calculates the KNN based on distance for prediction. This is especially useful when modeling hydrological phenomena, such as floods, where there is little prior knowledge of data distribution (Wettschereck *et al.* 1997). The optimal number of neighbors typically depends on the regression and classification metrics used. For continuous variables, the Euclidean distance is the most commonly used distance metric, while for discrete variables, the Hamming distance is the most typical. The value of K is usually set to the square root of the number of samples, and it can vary depending on the dataset (Duda *et al.* 2012; Guo *et al.* 2023).

2.3.2.4. Naïve Bayes.   NB is a simple and easy-to-implement algorithm that is based on Bayes probability theory. It is used for classifying phenomena based on the probability of the phenomenon occurring or not occurring. It does not require extensive tuning of hyperparameters, performs well on small-scale data, and is capable of handling various classification issues (Zhu *et al.* 2020). In addition, the algorithm has a strong mathematical base and consistent classification accuracy. The NB algorithm assumes that the impact of a given category (c) of the various predictor values is neutrally affected by the predictor cost (x). This presumption is called 'conditional independence':

$$P(c|x) = \frac{P(x|c) \times P(c)}{P(x)}$$

$$P(x|c) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c)$$

where $P(c)$ is the prior probability of the class, $P(x)$ is the prior probability of the predictor, and $P(c|x)$ is the posterior probability of the target (Zhang 2004).

2.4. Model validation

Assessing the performance of a model is a critical step in probabilistic modeling, as it ensures the reliability of the output. There are various metrics that have been used to evaluate the performance of flash flood prediction models. The area under the receiver operating characteristic curve (AUC ROC) is used to assess the performance of our models. AUC computes the entire two-dimensional area below the ROC curve and provides an aggregate measure of classification performance across all potential thresholds. It quantifies the probability that the model will correctly rank a randomly chosen positive instance higher than a randomly chosen negative instance (Hanley 1989). A higher AUC value indicates a better performance of the model (Shirzadi *et al.* 2019).

3. RESULTS AND ANALYSIS

3.1. FR model

The susceptibility of the Rheraya watershed to flash floods was assessed using the FR bivariate statistical method with geospatial techniques. The FR for various classes of each factor was used to understand and determine the significance or probability of a subclass under flash flood occurrences (Table 1).

   Analysis shows that flash floods frequently occur at an elevation of 981.0–1,610.0 m with an FR value of 1.46, which indicates that they are unlikely to occur at higher altitudes in the Rheraya basin. The slope angle was also found to be a significant factor, with the highest probability of flash flood occurrence observed in the 0–7-degree and 7–18-degree classes, with FR values of 1.50 and 1.52, respectively. In addition, it was observed that as the slope angle increases, the FR values decrease, indicating a reduced probability of flash floods at higher slope angles. The relationship between flash floods and the slope

Table 1 | Spatial relationship between the flooded area and its related factors using the FR method

Factors	Factor class	Number of flash flood pixels	Percentage of flash flood	Number of pixels in class	Percentage of domain	FR
Elevation	400–981	3,593.75	27.06	705,148	28.66	0.94
	981–1,610	4,218.75	31.76	536,973	21.83	1.46
	1,610–2,188	5,156.25	38.82	737,981	30.00	1.29
	2,188–2,928	312.5	2.35	317,140	12.89	0.18
	2,928–4,211		0.00	162,723	6.61	0.00
Slope	0–7	4,062.5	30.59	495,495	20.14	1.52
	7–18	4,062.5	30.59	502,247	20.42	1.50
	18–28	2,031.25	15.29	371,190	15.09	1.01
	28–39	2,812.5	21.18	738,683	30.03	0.71
	39–80	312.5	2.35	352,350	14.32	0.16
SPI	0–26,678	12,500	94.12	2,453,021	99.72	0.94
	26,678–92,923	0.00	0.00	2,868	0.12	0.00
	92,923–199,745	312.5	2.35	1,373	0.06	42.16
	199,745–352,565	468.75	3.53	1,136	0.05	76.43
	352,565–913,374	0.00	0.00	1,567	0.06	0.00
TWI	5_6	3,437.5	25.88	1,094,759	44.50	0.58
	6_8	3,906.25	29.41	793,326	32.25	0.91
	8_10	2,656.25	20.00	399,365	16.23	1.23
	10_13	1,093.75	8.24	136,934	5.57	1.48
	13_21	2,187.5	16.47	35,581	1.45	11.39
Aspect	Flat	2,968.75	22.35	462,063	18.78	1.19
	North	3,437.5	25.88	351,613	14.29	1.81
	Northeast	1,093.75	8.24	262,371	10.67	0.77
	East	625	4.71	285,843	11.62	0.40
	South	2,968.75	22.35	368,264	14.97	1.49
	Southwest	937.5	7.06	389,575	15.84	0.45
	West	1,250	9.41	340,236	13.83	0.68
Distance To river	0–300	10,781.25	81.18	492,783	20.03	4.05
	300–600	1,406.25	10.59	813,071	33.05	0.32
	600–800	0.00	0.00	565,760	23.00	0.00
	800–1,000	156.25	1.18	337,585	13.72	0.09
	1,000–1,200	937.5	7.06	250,472	10.18	0.69
Drainage density	0–1.5	625	4.71	484,182	19.68	0.24
	1.5–3	312.5	2.35	550,492	22.38	0.11
	3–4.5	625	4.71	543,162	22.08	0.21
	4.5–6	3,906.25	29.41	488,064	19.84	1.48
	6–8.32	7,812.5	58.82	387,534	15.75	3.73
NDVI	−0.32	156.25	1.18	24,787	1.01	1.17
	0–0.15	3,906.25	29.41	774,171	31.47	0.93
	0.15–0.3	5,468.75	41.18	819,295	33.31	1.24
	0.3–0.45	1,093.75	8.24	454,964	18.49	0.45
	0.45–0.85	2,656.25	20.00	305,485	12.42	1.61
Rainfall	369–373	4,062.5	30.59	713,544	29.01	1.05
	373–378	1,562.5	11.76	658,802	26.78	0.44
	378–384	781.25	5.88	226,069	9.19	0.64
	385–389	2,812.5	21.18	278,829	11.33	1.87
	389–392	4,062.5	30.59	582,427	23.68	1.29
Land cover	Agriculture	4,687.5	35.29	208,855	8.49	4.16
	Built-up and urban	312.5	2.35	24,114	0.98	2.40
	Bare soil	5,000	37.65	1,333,323	54.20	0.69
	Forest	2,812.5	21.18	781,870	31.78	0.67

(*Continued.*)

Table 1 | Continued

Factors	Factor class	Number of flash flood pixels	Percentage of flash flood	Number of pixels in class	Percentage of domain	FR
	Rocky bare soil	468.75	3.53	111,679	4.54	0.78
Lithology	Impermeable	6,718.75	50.59	1,106,856	44.99	1.12
	Permeable	3,906.25	29.41	770,101	31.31	0.94
	Semipermeable	2,656.25	20.00	501,845	20.40	0.98
Curvature	Concave	468.75	3.53	223,604	9.08	0.39
	Flat	1,093.75	8.24	2,058,021	83.66	0.10
	Convex	11,718.75	88.24	178,340	7.24	12.17

aspect reveals that flash floods are frequent on flat (1.19), north (1.81), and south (1.49) terrain surfaces. Convex slope curvatures had the highest probability of flood occurrence, with an FR value of 12.17. The proximity to a river was found to be a significant factor, with the highest FR values observed within a distance of 0–300 m from a river, as flash floods are more likely to occur close to the riverbank. For land cover, the highest FR value observed is in the agriculture and built-up classes. This might be due to the impermeable surfaces and sparse vegetation in these areas. Drainage density was also found to have an impact on flash floods, with the highest probability of occurrence at densities of 6–8.32, with an FR value of 3.7. The study found that impermeable rocks were the most sensitive to flash floods, with an FR value of 1.12. The TWI was found to be proportional to the FR; as TWI values increased, so did the FR values. The river power index (SPI) indicated that areas with high flow power (199,745–352,565) had a high likelihood of flash flood occurrence, with an FR value of 76.43. The relationship between the vegetation index (NDVI) and flash floods showed that areas with high vegetation values had the highest FR values. Lastly, the study found that rainfall values between 389 and 392 mm had a high FR value of 1.29, indicating a higher occurrence of flash floods.

The FR method was used to generate a flash flood susceptibility index (FSI), from which a susceptibility map was created (Figure 5(a)). The map was created using 12 conditioning factors and classified into 5 susceptibility classes using the natural break classification technique: very high (1.62%), high (12.66%), moderate (13.64%), low (16.23%), and very low (55.84%) (Figure 6). The main strength of the FR approach is its ability to conceptualize the impact of subclasses of each conditioning factor on flash flood occurrence (Tehrany et al. 2014a). Thus, the results show that bivariate statistical analysis approaches are useful and efficient for flood susceptibility assessment and flood mitigation.

3.2. LR model

LR is a commonly used technique for determining the magnitude of the relationship between flash flood conditioning factors and locations. The relative importance of each conditioning factor is expressed by the LR coefficients, which were computed using R software (Table 2).

The results revealed that curvature, distance to the river, drainage density, rainfall, SPI, and NDVI show positive correlations with flash flooding in the study area, while the remaining factors show negative correlations. A positive value means that the influence of the variable increases the likelihood of flash flooding, while a negative value means that the presence of the variable reduces the likelihood of flash flooding. It is clear that flash floods in the study area are not controlled by a single factor but rather by a combination of factors. Based on the absolute values of the LR coefficients, the factor with the greatest impact on flash flood occurrence in the study area was found to be drainage density.

Although slope, elevation, land cover, and lithology seem like factors that would influence FFS (Al-Juaidi et al. 2018; Anucharn 2019), their significance in the LR model can vary depending on the study area and the scale of analysis. In some cases, other variables such as rainfall intensity, distance to the stream, and drainage density patterns may overshadow the effects of these variables. For instance, in areas with high rainfall intensity, even relatively flat terrain can experience flash floods due to poor drainage or urbanization (Shao & Shao 2019). The lack of significance of land cover in the analysis could be due to the complexity of land cover types and their interaction with other variables. In some cases, land cover may indirectly influence FFS through factors such as soil infiltration rates. Similar to land cover, lithology's lack of contribution
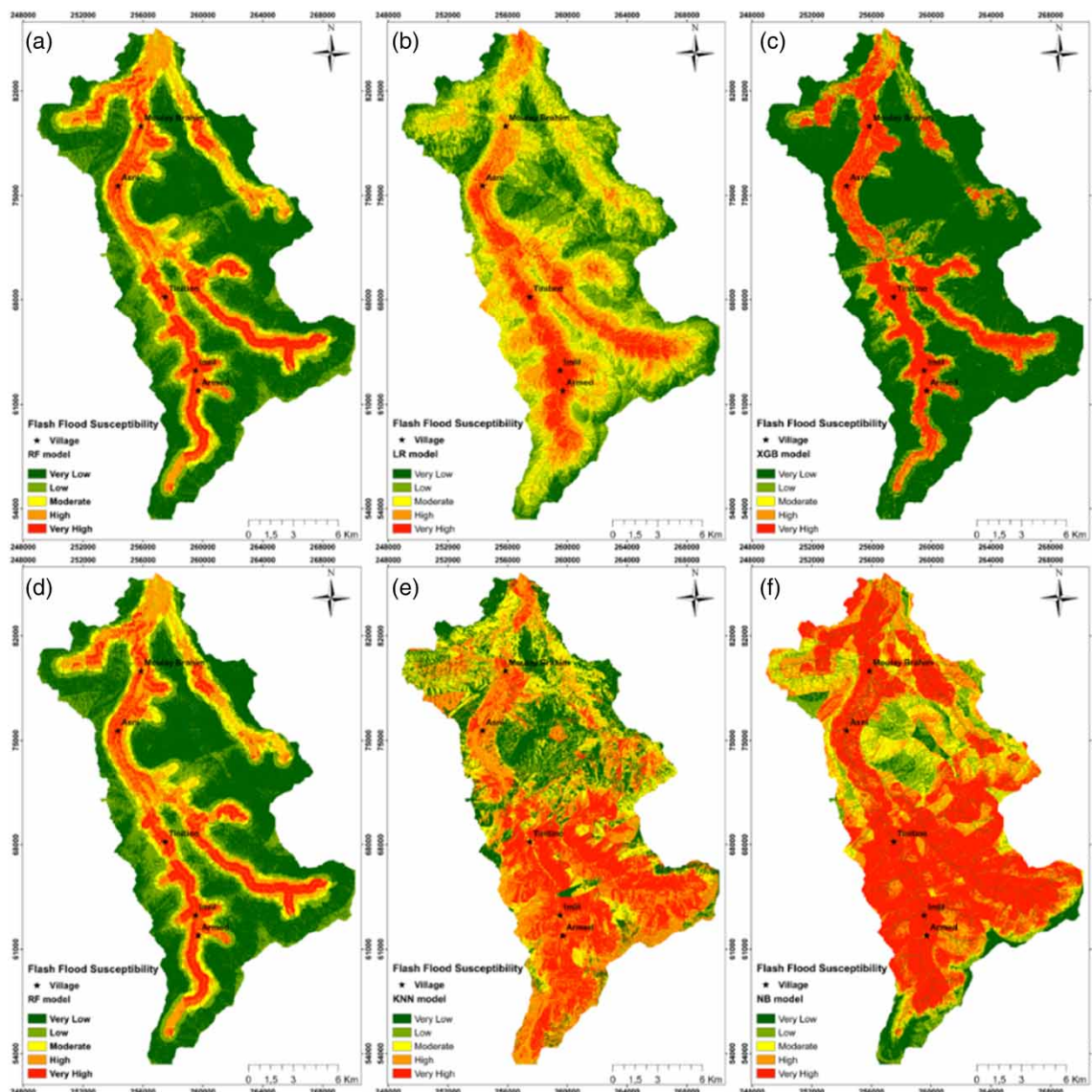
Figure 5 | Flash flood susceptibility models using (a) FR, (b) LR, (c) XGBoost, (d) RF, (e) KNN, and (f) NB.

might be attributed to its interaction with other variables or the specific geological characteristics of the study area. Certain lithological units may promote rapid runoff and contribute to flash floods, but their significance may not always be captured in the model due to collinearity with other predictors.

Finding the same example in other studies is difficult, but in some studies, certain factors that have a high influence on flash floods have either values close to 0 or negative values. Tehrany *et al*. (2014), for instance, showed that soil drainage was almost the least influencing factor, and the slope and soil effect were not statistically significant in the model. Nandi *et al*. (2016) also revealed that the distance to the stream was the least contributing factor to floods in their study area. Therefore, in summary, the lack of significance of slope, elevation, land cover, and lithology in our LR analysis may be due to the interplay of various factors and the specific characteristics of the study area.

An FFS map was created by multiplying the LR coefficients with their corresponding conditioning factors (Figure 5(b)). These values were then classified into five classes using the natural break classification technique, i.e., very high (10.06%), high (19.81%), moderate (26.95%), low (27.6%), and very low (15.58%) (Figure 6).
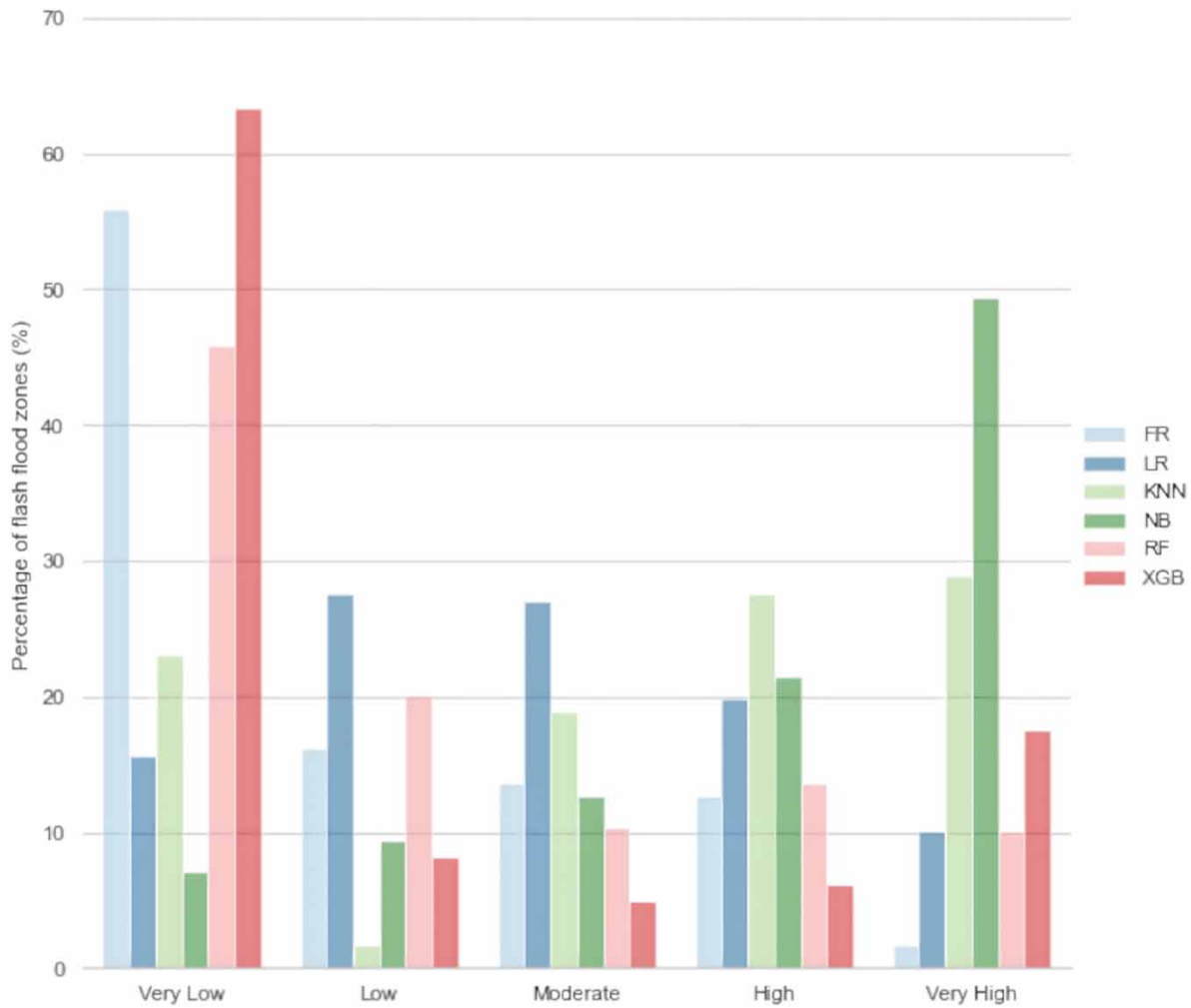
Figure 6 | The percentage of flash flood susceptible areas under different classes of the six models.

Table 2 | LR coefficients of different factors

Factors	LR coefficient
Constant	-8.594×10^{-1}
Elevation	-1.437×10^{-3}
Slope	-8.159×10^{-4}
Aspect	-4.935×10^{-3}
Curvature	2.881×10^{-2}
Distance to the river	7.35×10^{-5}
Drainage density	1.016
Rainfall	2.353×10^{-2}
TWI	-4.34×10^{-2}
SPI	9.947×10^{-6}
Land cover	-6.329×10^{-7}
NDVI	2.585×10^{-1}
Lithology	-8.371×10^{-7}

3.3. ML models

In the present section, four ML algorithms were used to examine the impact of each flood conditioning factor on the likelihood of flash floods occurring in the Rheraya basin. The analysis results indicate that the distance to the river, drainage density, and TWI are the most influential factors, followed by SPI, rainfall, elevation, lithology, aspect, NDVI, land cover, slope, and curvature (Figure 7). These values demonstrate that all conditioning factors play a significant role in flash flood occurrence; however, the distance to the river and drainage density have a greater impact on the study area.

XGBoost, RF, NB, and KNN algorithms were utilized to evaluate the susceptibility of flash floods for each pixel of the basin. The RF model proved to be the most successful in terms of prediction performance. A number of classification methods were applied in several studies to classify the FFS maps, such as the equal interval, regular interval, standard deviation, quantile, natural break, and manual approach. Among these methods, the natural break and quantile approaches are the most popular in the literature (Tehrany *et al*. 2019a, 2019b; Tien Bui *et al*. 2019a, 2019b), and thus, this study used the natural break classification method to classify the FFS maps into five classes: very low, low, moderate, high, and very high (Figure 5(c)–5(f)).

Based on the results, the KNN model showed that the lowest percentage of area (1.62%) belonged to the low-risk class, followed by the moderate (18.83%), very low (23.05%), high (27.6%), and very high (28.9%) classes. The NB model revealed that the percentages of areas for the very high, high, moderate, low, and very low classes are 49.35, 21.43, 12.66, 9.42, and 7.14%, respectively. The RF model shows that the percentage of areas in the very high flash flood susceptible class is 10.06%, followed by moderate (10.39%), high-risk (13.64%), low-risk (20.13%), and the very low flash flood susceptible classes (45.78%). Lastly, for the XGBoost model, the area percentages were 17.53% for the very high FFS class, 6.17% for the high, 4.87% for the moderate, 8.12% for the low, and 63.31% for the very low (Figure 6).

The findings indicate that the KNN and NB models tend to overestimate FFS in regions with high risk while underestimating it in areas with low risk, in contrast to the RF and XGBoost models. Nevertheless, all four models unanimously indicate a very high flash flood risk in locations near the main river within the study area.

3.4. Validation and comparison

The accuracy and prediction ability of the six FFS models were evaluated using the area under the ROC curve metric using both training and testing data. The higher the AUC value, the better the model's prediction performance, and vice versa.
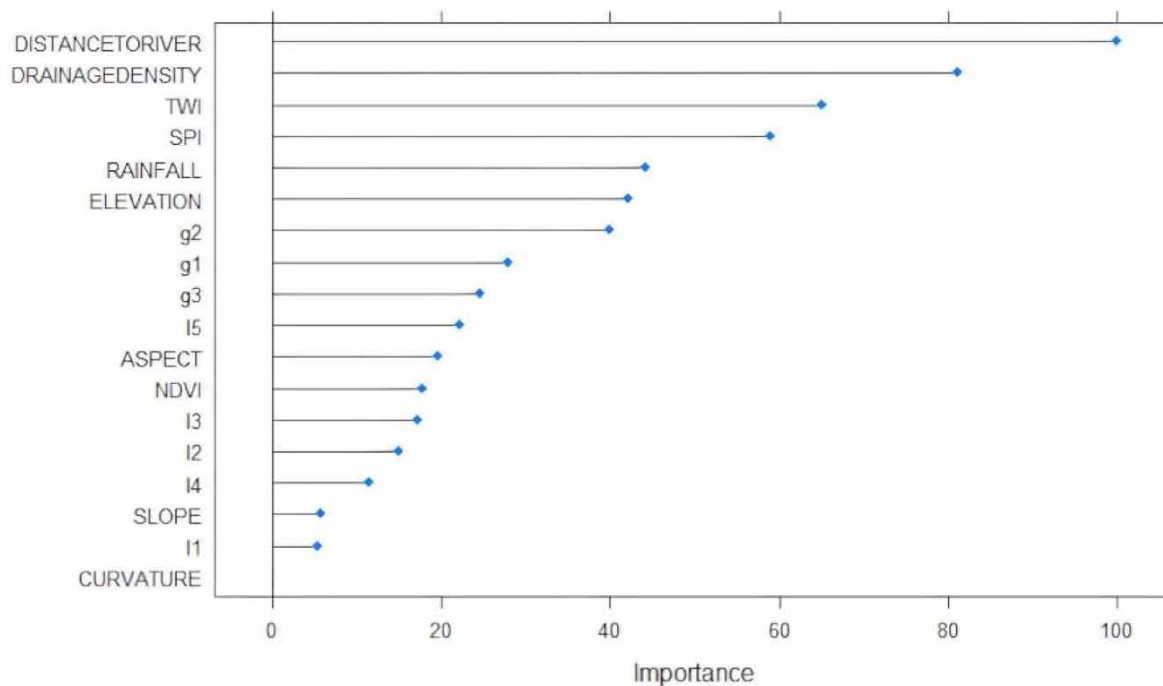

Figure 7 | The importance of flash flood conditioning factors. g1: Semipermeable, g2: impermeable, g3: permeable, l1: built-up and urban, l2: bare rocky land, l3: agriculture, l4: forest, and l5: bare land.

According to the success rate curve (e.g., training dataset), the RF model (AUC = 0.90) surpassed the XGBoost (AUC = 0.86), LR (AUC = 0.85), NB (AUC = 0.81), KNN (AUC = 0.79), and FR (AUC = 0.77) models (Figure 8(a)). The results of the prediction rate curve (e.g., testing dataset) also showed similar results, with the RF model having a higher AUC value of 0.86 in comparison to the XGBoost (AUC = 0.85), LR (AUC = 0.83), NB (AUC = 0.76), KNN (AUC = 0.75), and FR (AUC = 0.72) (Figure 8(b)).

Although all six FFS models exhibited high to moderate prediction accuracy with AUC values greater than 0.70, the RF model was determined to be the most effective for predicting FFS in the study area.

4. DISCUSSION

This study highlights the multifaceted nature of flash flood occurrences, the interplay of conditioning factors, and the potential of bivariate and multivariate statistical models and ML techniques for flash FSM in the Rheraya watershed, a flood-prone region. These susceptibility maps are invaluable resources for a wide range of stakeholders, including hazard managers, urban planners, and policymakers, to prevent flash flood-related injuries and property losses (Figure 5; Shokouhifar et al. 2022). Moreover, the increasing risk of flash floods, driven by various factors such as rapid urbanization, deforestation, canalization, changes in land use, and the effects of climate change (i.e., changes in the intensity and frequency of heavy precipitation), highlights the critical need for improved mapping of FFS (Hapuarachchi et al. 2011; Badraq Nejad et al. 2019; Prasad et al. 2021). Therefore, our study holds significant implications for understanding and managing FFS in the Rheraya basin and similar ungauged regions known for their past destructive flash flood occurrences.

The results indicate that all the models unveiled several significant factors influencing FFS in the Rheraya watershed, such as elevation, slope, distance to the river, NDVI, and rainfall. This highlights the complex nature of flash floods, reflecting the combined impact of terrain features, land use patterns, and hydrological conditions. Identifying the most contributing factor to flash floods in our area is so significant as it will allow for targeted and effective mitigation strategies, resource allocation, and risk assessment. Distance to the river emerged as the most influential factor for flash floods in the region. Many studies have also confirmed the important contribution of this factor to flood occurrence (Rahmati et al. 2016a, 2016b; Pham et al. 2020; Bansal et al. 2022; Chaulagain et al. 2023). This is evident as areas near riverbanks are more susceptible due to their increased exposure to rapid water flow as well as their role as natural drainage paths. Therefore, urban development in the Rheraya watershed should consider safe locations, particularly those away from riverbanks.

Our FFS modeling involved six approaches, including two statistical methods (i.e., FR and LR) and four ML algorithms (i.e., RF, XGBoost, NB, and KNN). The comparative analysis revealed the RF model as the best-performing model, with a consistently high AUC value of 0.86, followed by XGBoost (AUC = 0.85), LR (AUC = 0.83), NB (AUC = 0.76), KNN (AUC = 0.75), and FR (AUC = 0.72). Multiple studies have aligned with our findings, emphasizing the robustness of RF in modeling FFS (Lee et al. 2017). For instance, Islam & Chowdhury (2024) conducted a local-scale FFS assessment in northeastern Bangladesh using RF and support vector machine algorithms, demonstrating RF's superior performance. Similarly,
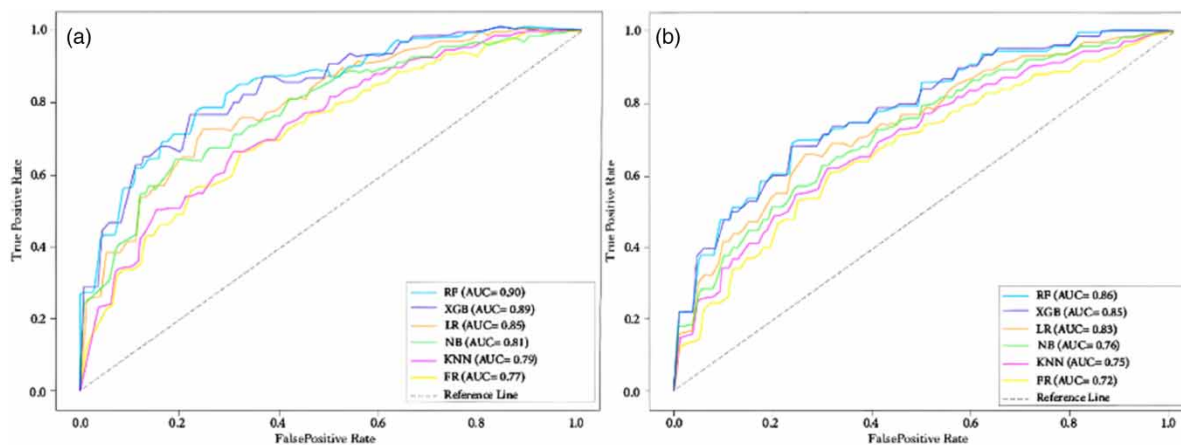

Figure 8 | The validation of the six-flash flood susceptible models using the ROC curve based on the training point (a) and the validation point (b).

Ghanim *et al.* (2023), Abedi *et al.* (2022), and Vafakhah *et al.* (2020) assessed various ML algorithms and statistical techniques to delineate FFS at regional scales in Jeddah City (Saudi Arabia), the Bâsca Chiojdului river basin (Romania), and Gilan province (Iran), respectively, also finding RF to be the most effective. In addition, Zhao *et al.* (2018) mapped flood susceptibility in mountainous areas on a national scale in China using RF, artificial neural network, and support vector machine methods, with RF emerging as the optimal model. Hence, RF emerges as a consistently effective model for FFS modeling. However, it is noteworthy to mention that all models in our study exhibited AUC values exceeding 0.70, indicating their efficacy in assessing FFS.

Our FFS maps of the Rheraya watershed, generated by the different models, offer valuable insights into high-risk areas. Consistently, regions such as Moulay Brahim (Figure 9(a)), Asni (Figure 9(b)), Tinitine (Figure 9(c)), Imlil (Figure 9(d)), and Armed (Figure 9(e)) have been identified as highly vulnerable to flash floods. These areas, known for their tourist attractions and urbanization, need immediate attention in terms of disaster preparedness and mitigation measures. It is imperative to strengthen the infrastructure for disaster management in these regions, including the establishment of additional meteorological stations for enhanced rainfall measurement and flood forecasting. In addition, policies should be implemented to regulate urban development, ensuring that construction activities adhere to flood-resistant standards and are located away from high-risk zones. Moreover, public awareness campaigns and community engagement initiatives should be undertaken to educate residents about the risks associated with flash floods and the necessary safety measures to mitigate their impact. By implementing these policies and measures, we can effectively safeguard lives and properties in these vulnerable areas in the Rheraya watershed.

Although this study showed the effectiveness and robustness of statistical methods and ML techniques, particularly RF, in identifying flash flood high-risk areas in the ungauged and flood-prone Rheraya watershed, in addition to identifying the most contributing factors to flash floods in our region, it is important to mention that our approach is not without limitations. The models applied rely on various assumptions, and the results may be sensitive to the choice of conditioning factors and their classification (Tehrany *et al.* 2019a, 2019b). Future research could explore the use of additional data sources and the validation of the models for different conditions to enhance the accuracy of the FFS maps. Climate models, for instance, provide data for different future projections, which could lead to better FFS mapping. In addition, collaborative efforts in data collection and sharing can enhance data accessibility, spanning from local to global flood inventories. This expanded dataset's availability can help in the evaluation of models performance across diverse spatial scales, assessing their adaptability and generalizability. The flashiness of floods is significantly influenced by the characteristics of rainfall events. Saharia *et al.* (2021) showed that spatial variability of rainfall impacts flash flood severity as much as basin geomorphology
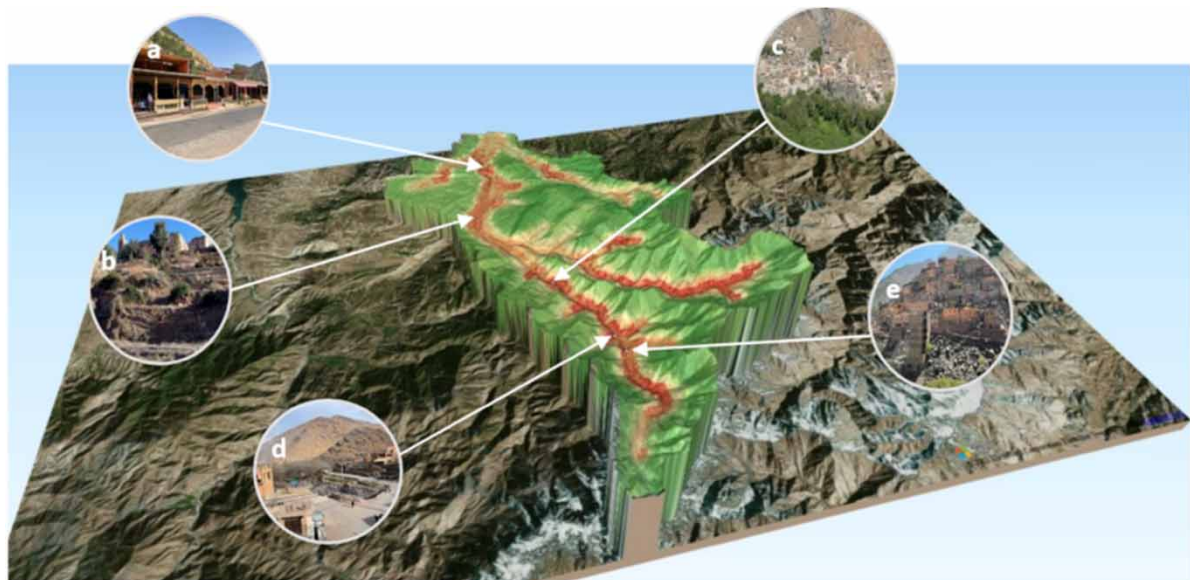

Figure 9 | 3D representation of the RF model with the area's most vulnerable to flash floods: (a) Moulay Brahim, (b) Asni, (c) Tinitine, (d) Imlil, and (e) Armed.

and climatology. Therefore, future work should also investigate adding rainfall event-related factors such as intensity, duration, spatial distribution, and temporal patterns to the modeling and consequently developing more robust FFS maps. It is crucial to note that our study utilized data from just one study site, which limits the generalizability of our findings to other ungauged basins. Nonetheless, to ensure the scalability of our approach, it is essential to assess the accuracy of our methods across basins with diverse topographical, hydrological, geological, geographical, and anthropogenic characteristics. Future research should aim to confirm the reliability of ML compared with statistical models across various watersheds, considering the variability in factors that could impact model outcomes.

5. CONCLUSION

Flooding risks are constantly increasing due to climate change and land use changes. Flash floods, in particular, are a very difficult phenomenon to predict, and the impacts and consequences can be very extreme. Creating an FFS map through the use of mathematical and statistical methods based on analyzing geospatial data is a significant advancement in managing the hazards of flash floods.

The Rheraya basin is an area susceptible to flash flooding due to its steep slopes, high altitude, low permeability, high density of drainage, etc. To mitigate the severity of flash floods in this region, effective methods for identifying the most vulnerable locations are necessary. Recent advancements in statistical approaches and ML algorithms have proven to be valuable tools in this endeavor. In this study, six models were evaluated and applied, including LR, FR, RF, NB, XGBoost, and KNN. Input data for the models consisted of two major variables: 12 independent factors such as elevation, land use, slope, NDVI, rainfall, lithology, distance to the river, and drainage density, and one dependent factor, flash flood inventory, which represents 246 locations recorded over the past 40 years, obtained from historical archives, field surveys, and validated with satellite data. Flash flood inventory was divided into 70% for training the models and 30% for validating them. Results indicated that the RF model performed the best based on the ROC curve evaluation metric (AUC = 0.86), followed by XGBoost (AUC = 0.85), LR (AUC = 0.83), NB (AUC = 0.76), KNN (AUC = 0.75), and FR (AUC = 0.72). The analysis also revealed that distance to the river and drainage density were the most influential factors in the occurrence of flash floods in the Rheraya basin.

The current study has demonstrated the robustness and effectiveness of the methodology adopted for the analysis of susceptibility to flash floods. Therefore, it is strongly recommended to extend the application of the RF model to predict the susceptibility to flash floods in other basins, especially in the case of ungauged watersheds that lack hydrological data. This will enable a more accurate assessment of vulnerability in areas with high human traffic, allowing for the precise identification of high-risk zones. Furthermore, these models can also be utilized to guide the installation of meteorological equipment and evaluate the suitability of the location of flood warning stations. Future work should broaden the factors considered in susceptibility modeling to better understand flash flood vulnerability. Collaborative data collection and improved data accessibility will help incorporate additional information, making flood management and disaster risk reduction more effective on a larger scale.

ETHICAL APPROVAL

The authors confirm that this article is original research and has not been published or presented previously in any journal or conference in any language (in whole or in part).

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

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 International* **37** (19), 5479–5496. doi:10.1080/10106049.2021.1920636.

Ahmadlou, M., Karim, M., Alizadeh, S., Shirzadi, A., Parvinnejhad, D., Shahabi, H. & Panahi, M. 2018 Susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and bat algorithms (BA). *Geocarto Inernational* **34**, 1–21.

Ahmed, K. R. & Akter, S. 2017 Analysis of land cover change in Southwest Bengal delta due to floods by NDVI, NDWI, and K-means cluster with Landsat multi-spectral surface reflectance satellite data. *Remote Sensing Applications: Society and Environment* **8**, 168–181.

Al-Abadi, A. M. & Al-Najar, N. A. 2020 Comparative assessment of bivariate, multivariate and machine learning models for mapping flood proneness. *Natural Hazards* **100** (2), 461–491.

Al-Juaidi, A. E., Nassar, A. M. & Al-Juaidi, O. E. 2018 Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. *Arabian Journal of Geosciences* **11** (24), 765.

Anucharn, T. 2019 A study of physical factors affecting flood susceptibility in Khlong Nathawi Subbasin using multiple logistic regression model. *Burapha Science Journal* 284–298.

Apaydin, H., Feizi, H., Sattari, M., Çolak, M. S., Band, S. & Chau, K. W. 2020 Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. *Water* **12** (5 Supp.), 1500. doi:10.3390/w12051500.

Badraq Nejad, A., Sarli, R., Babaii, M. & Basiri, M. 2019 Evaluating and analyzing the spatial distribution of rural inhabitants with emphasis on biological and activity risk taking using GIS and SPSS (the area under study: Aq Su rural area). *Journal of the Studies of Human Settlements Planning* **14**, 735–756.

Bansal, N., Mukherjee, M. & Gairola, A. 2022 Evaluating urban flood hazard index (UFHI) of Dehradun city using GIS and multi-criteria decision analysis. *Modeling Earth Systems and Environment* 1–14.

Benito, G., Rico, M., Sánchez-Moya, Y., Sopeña, A., Thorndycraft, V. R. & Barriendos, M. 2010 The impact of late Holocene climatic variability and land use change on the flood hydrology of the Guadalentín River, southeast Spain. *Global and Planetary Change* **70** (1–4), 53–63.

Bentivoglio, R., Isufi, E., Jonkman, S. N. & Taormina, R. 2022 Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrology and Earth System Sciences* **26** (16), 4345–4378.

Boudhar, A., Duchemin, B., Hanich, L., Chaponnière, A., Maisongrande, P., Boulet, G., Stitou, J. & Chehbouni, A. 2007 Snow covers dynamics analysis in the Moroccan High-Atlas using SPOT-VEGETATION data. *Sécheresse* **18**, 278–288.

Boudhar, A., Hanich, L., Boulet, G., Duchemin, B., Berjamy, B. & Chehbouni, A. 2009 Evaluation of the snowmelt runoff model in the Moroccan High Atlas Mountains using two snow-cover estimates. *Hydrological Sciences Journal* **54** (6), 1094–1113.

Breiman, L. 2001 *Random Forests Machine Learning*, Vol. 45. Springer, Berlin/Heidelberg, Germany, pp. 5–32.

Bui, Q. T., Nguyen, Q. H., Nguyen, X. L., Pham, V. D., Nguyen, H. D. & Pham, V. M. 2020b Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *Journal of Hydrology* **581**, 124379.

Butler, D., Kokkalidou, A. & Makropoulos, C. K. 2006 Supporting the siting of new urban developments for integrated urban water resource management. In: *Integrated Urban Water Resources Management*. Springer, pp. 19–34.

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.

Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T. & Khosravi, K. 2017 A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software* **95**, 229–245.

Chaponnière, A., Boulet, G., Chehbouni, A. & Aresmouk, M. 2008 Understanding hydrological processes with scarce data in a mountain environment. *Hydrological Processes: An International Journal* **22** (12), 1908–1921.

Chaulagain, D. *et al.* 2023 Flood susceptibility mapping of Kathmandu metropolitan city using GIS-based multi-criteria decision analysis. *Ecological Indicators* **154**, 110653.

Chu, H., Wu, W., Wang, Q. J., Nathan, R. & Wei, J. 2020 An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environmental Modelling & Software* **124**, 104587.

Costache, R. & Bui, D. T. 2020 Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Science of the Total Environment* **712**, 136492.

Dahri, N. & Abida, H. 2017 Monte Carlo simulation-aided analytical hierarchy process (AHP) for flood susceptibility mapping in Gabes Basin (southeastern Tunisia). *Environmental Earth Sciences* **76**, 1–1.

Dazzi, S., Vacondio, R. & Mignosa, P. 2021 Flood stage forecasting using machine-learning methods: A case study on the Parma river (Italy). *Water* **13**, 1612. doi:10.3390/w13121612.

Digby, B. 2000 *Changing Environments*. Heinemann.

Dodangeh, E., Choubin, B., Eigdir, A. N., Nabipour, N., Panahi, M., Shamshirband, S. & Mosavi, A. 2020 Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Science of the Total Environment* **705**, 135983.

Doswell, C. A., 2015 Flooding. In: *Encyclopedia of Atmospheric Sciences*, 2nd edn In: (North, G. R., Pyle, J. & Zhang, F., eds.). Academic Press, Oxford, pp. 201–208. https://doi.org/10.1016/B978-0-12-382225-3.00151-1.

Duda, R. O., Hart, P. E. & Stork, D. G. 2012 *Pattern Classification*. John Wiley & Sons, Hoboken, NJ, USA.

El Alaoui El Fels, A., Alaa, N., Bachnou, A. & Rachidi, S. 2018 Flood frequency analysis and generation of flood hazard indicator maps in a semi-arid environment, case of Ourika watershed (western High Atlas, Morocco). *Journal of African Earth Sciences* **141**, 94–106.

Elmore, A. J., Julian, J. P., Guinn, S. M. & Fitzpatrick, M. C. 2013 Potential stream density in mid-Atlantic US watersheds. *PLoS ONE* **8** (8), e74819.

European Union 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. Available from: https://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:288:0027:0034:EN:PDF (accessed 18 February 2023).

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X. & Xiang, Y. 2018 Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management* **164**, 102–111.

Farzaneh, S. H., Choubin, B., Mosavi, A., Nabipour, N., Band, S., Darabi, H. & Haghighi, A. T. 2019 Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Science of the Total Environment* **711**, 135161. doi:10. 1016/j.scitotenv.2019.135161.

Fernández, D. & Lutz, M. 2010 Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. *Engineering Geology* **111** (1–4), 90–98.

Ghanim, A. A. *et al.* 2023 An improved flood susceptibility assessment in Jeddah, Saudi Arabia, using advanced machine learning techniques. *Water* **15** (14), 2511.

Goetz, J. N., Brenning, A., Petschko, H. & Leopold, P. 2015 Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & Geosciences* **81**, 1–11.

Grothmann, T. & Reusswig, F. 2006 People at risk of flooding: Why some residents take precautionary action while others do not. *Natural Hazards* **38** (1–2), 101–120.

Guha-Sapir, D., Below, R. & Hoyois, P. 2016 *EM-DAT: The CRED/OFDA International Disaster Database*.

Guo, G., Wang, H., Bell, D., Bi, Y. & Greer, K. 2023 KNN model-based approach in classification. In *Proceedings of the OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, 3–7 November 2003, Catania, Italy. Springer, Berlin/Heidelberg, Germany, pp. 986–996.

Hanley, J. A. 1989 Receiver operating characteristic (ROC) methodology: The state of the art. *Critical Reviews in Diagnostic Imaging* **29** (3), 307–335.

Hapuarachchi, H., Wang, Q. & Pagano, T. 2011 A review of advances in flash flood forecasting. *Hydrological Processes* **25**, 2771–2784. https://doi.org/10.1002/hyp.8040.

Helsel, D. R. & Hirsch, R. M. 2002 *Statistical Methods in Water Resources – Hydrologic Analysis and Interpretation*. U.S. Geological Survey Techniques of Water-Resources Investigations. Book 4, Chapter A3, p. 510.

Islam, R. & Chowdhury, P. 2024 Local-scale flash flood susceptibility assessment in northeastern Bangladesh using machine learning algorithms. *Environmental Challenges* **14**, 100833.

Jaafari, A., Najaf, A., Pourghasemi, H. R., Rezaeian, J. & Sattarian, A. 2014 GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *International Journal of Environmental Science and Technology* **11** (4), 909–926.

Jahangir, M. H., Mousavi Reineh, S. M. & Abolghasemi, M. 2019 Spatial predication of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm. *Weather and Climate Extremes* **25**, 100215.

Jamali, B., Bach, P. M. & Deletic, A. 2020 Rainwater harvesting for urban flood management – An integrated modelling framework. *Water Research* **171**, 115372.

Janizadeh, S., Avand, M., Jaafari, A., Phong, T. V., Bayat, M., Ahmadisharaf, E., Prakash, I., Pham, B. T. & Lee, S. 2019 Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran. *Sustainability* **11** (19), 5426.

Jayakrishnan, R., Srinivasan, R., Santhi, C. & Arnold, J. 2005 Advances in the application of the SWAT model for water resources management. *Hydrological Processes* **19**, 749–762.

Jeyaseelan, A. T. 2003 Droughts & floods assessment and monitoring using remote sensing and GIS. In: *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology*, Vol. 291. World Meteorological Organization.

Jutras, S., Rousseau, A. & Clerc, C. 2009 Implementation of a peatland-specific water budget algorithm in HYDROTEL. *Canadian Water Resources Journal* **34** (4), 349–364.

Kalantari, Z., Nickman, A., Lyon, S. W., Olofsson, B. & Folkeson, L. 2014 A method for mapping flood hazard along roads. *Journal of Environmental Management* **133**, 69–77.

Karmaoui, A. & Balica, S. 2019 A new flood vulnerability index adapted for the pre-Saharan region. *International Journal of River Basin Management*. https://doi.org/10.1080/15715124.2019.1583668.

Kia, M. B. *et al.* 2012 An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences* **67**, 251–264.

Knighton, A. D. 1999 Downstream variation in stream power. *Geomorphology* **29** (3–4), 293–306.

Kuenzer, C., Guo, H., Huth, J., Leinenkugel, P., Li, X. & Dech, S. 2013 Flood mapping and flood dynamics of the Mekong delta: ENVISAT-ASAR-WSM based time series analyses. *Remote Sensing* **5** (2), 687–715.

Kundzewicz, Z. W. *et al.* 2014 Flood risk and climate change: Global and regional perspectives. *Hydrological Sciences Journal* **59** (1), 1–28.

Lee, S. & Talib, J. A. 2005 Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology* **47** (7), 982–990.

Lee, M. J., Kang, J. E. & Jeon, S. 2012 Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, pp. 895–898.

Lee, S., Kim, J.-C., Jung, H.-S., Lee, M. J. & Lee, S. 2017 Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk* **8** (2), 1185–1203. doi:10.1080/19475705.2017.1308971.

Liu, R., Chen, Y. & Wu, J. 2016 Assessing spatial likelihood of flooding hazard using naïve Bayes and GIS: A case study in Bowen Basin, Australia. *Stochastic Environmental Research and Risk Assessment* **3**, 1575–1590.

Loudyi, D., Hasnaoui, M. D. & Fekri, A. 2022 *Flood Risk Management Practices in Morocco: Facts and Challenges*. Wadi Flash Floods, p. 35.

Madhuri, R., Sistla, S. & Srinivasa Raju, K. 2021 Application of machine learning algorithms for flood susceptibility assessment and risk management. *Journal of Water and Climate Change*.

Mallouk, A. *et al.* 2016 A multicriteria approach with GIS for assessing vulnerability to flood risk in urban area (Case of Casablanca City, Morocco). In: *Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015: MedCT 2015 Volume 1*. Springer International Publishing.

Masetic, Z. & Subasi, A. 2016 Congestive heart failure detection using random forest classifier. *Computer Methods and Programs in Biomedicine* **130**, 54–64.

Mekonnen, M. M. & Hoekstra, A. Y. 2016 Four billion people facing severe water scarcity. *Science Advances* **2** (2), e1500323.

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R. & Abderrahmane, B. 2020 Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews* **207**, 103225.

Mohanty, M. P., Mudgil, S. & Karmakar, S. 2020 Flood management in India: A focussed review on the current status and future challenges. *International Journal of Disaster Risk Reduction* 101660.

Mosavi, A., Ozturk, P. & Chau, K. 2018 Flood prediction using machine learning models: Literature review. *Water* **10** (11), 1536.

Naghibi, S. A., Hashemi, H., Berndtsson, R. & Lee, S. 2020 Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. *Journal of Hydrology* 125197.

Nguyen, V.-N., Yariyan, P., Amiri, M., Tran, A. D., Pham, T. D., Do, M. P., Ngo, P.-T. T., Nhu, V.-H., Long, N. Q. & Tien Bui, D. 2020 A New modeling approach for spatial prediction of flash flood with biogeography optimized CHAID tree ensemble and remote sensing data. *Remote Sensing* **12** (9). https://doi.org/10.3390/rs12091373.

Nhu, V.-H., Ngo, P.-T. T., Pham, T. D., Dou, J., Song, X., Hoang, N.-D., Tran, D. A., Cao, D. P., Aydilek, İ. B., Amiri, M., Costache, R., Hoa, P. V. & Tien Bui, D. 2020 A new hybrid firefly–PSO optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible mapping. *Remote Sensing* **12** (7). https://doi.org/10.3390/rs12172688.

Nogueira, K., Fadel, S. G., Dourado, I. C., Werneck, R. d. O., Munoz, J. A., Penatti, O. A., Calumby, R. T., Li, L. T., dos Santos, J. A. & Torres, R. d. S. 2018 Exploiting convNet diversity for flooding identification. *IEEE Geoscience and Remote Sensing Letters* **15** (9), 1446–1450.

Olden, J. D., Kennard, M. J. & Pusey, B. J. 2008 Species invasions and the changing biogeography of Australian freshwater fishes. *Global Ecology and Biogeography* **17**, 25–37.

Pham, B. T. *et al.* 2020 GIS based hybrid computational approaches for flash flood susceptibility assessment. *Water* **12** (3), 683.

Pourghasemi, H. R., Termeh, S. V. R., Kariminejad, N., Hong, H. & Chen, W. 2020 An assessment of metaheuristic approaches for flood assessment. *Journal of Hydrology* 124536.

Pradhan, B. 2009 Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *Journal of Spatial Hydrology* **9**, 1–18.

Pradhan, B., Shafiee, M. & Pirasteh, S. 2010 Maximum flood prone area mapping using RADARSAT images and GIS: Kelantan river basin. *International Journal of Geoinformatics* **5**, 11.

Prasad, R., Charan, D., Joseph, L., Nguyen-Huy, T., Deo, R. C. & Singh, S. 2021 Daily flood forecasts with intelligent data analytic models: Multivariate empirical mode decomposition-based modeling methods. In: *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation*. Springer, Cham, pp. 359–381.

Rahman, M., Chen, N., Islam, M. M., Mahmud, G., Pourghasemi, H., Alam, M., Rahim, M., Baig, M., Bhattacharjee, A. & Dewan. 2021 Development of flood hazard map and emergency relief operation system using hydrodynamic modeling and machine learning algorithm. *Journal of Cleaner Production* **311**, 127594. doi:10.1016/j.jclepro.2021.127594.

Rahmati, O., Haghizadeh, A., Pourghasemi, H. R. & Noormohamadi, F. 2016a Gully erosion susceptibility mapping: The role of GIS based bivariate statistical models and their comparison. *Natural Hazards* **82** (2), 1231–1258.

Rahmati, O., Pourghasemi, H. R. & Zeinivand, H. 2016b Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International* **31** (1), 42–70.

Rapport final de l'Etude preparatoire pour le Projet de Systeme de Prevision et d'Alerte aux Crues dans la region du Haut Atlas Royaume du Maroc 2011.

R Core Team 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Regmi, A. D., Devkota, K. C., Yoshida, K., Pradhan, B., Pourghasemi, H. R., Kumamoto, T. & Akgun, A. 2013 Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arabian Journal of Geosciences* **7** (2), 725–742.

Rish, I. 2001 An empirical study of the naive Bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, pp. 41–46.

Saharia, M., Kirstetter, P. E., Vergara, H., Gourley, J. J., Emmanuel, I. & Andrieu, H. 2021 On the impact of rainfall spatial variability, geomorphology, and climatology on flash floods. *Water Resources Research* **57** (9), e2020WR029124.

Samanta, R. K., Bhunla, 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. *Modeling Earth Systems and Environment*. https://doi.org/10.1007/s40808-018-0427-z.

Seckin, N., Cobaner, M., Yurtal, R. & Haktanir, T. 2013 Comparison of artificial neural network methods with L-moments for estimating flood flow at ungauged sites: The case of east Mediterranean river basin, Turkey. *Water Resources Management* **27** (7), 2103–2124.

Sellami, E. M., Maanan, M. & Rhinane, H. 2022 Performance of machine learning algorithms for mapping and forecasting of flash flood susceptibility in Tetouan, Morocco. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 46, pp. 305–313.

Seydi, S. T., Kanani-Sadat, Y., Hasanlou, M., Sahraei, R., Chanussot, J. & Amani, M. 2022 Comparison of machine learning algorithms for flood susceptibility mapping. *Remote Sensing* **15** (1), 192.

Shao, J. & Shao, L. 2019 The causes of floods and control measures during the rainstorm period in cities with flat terrain. *Journal of Architectural Research and Development* **3** (3).

Shirzadi, A., Solaimani, K., Roshan, M. H., Kavian, A., Chapi, K., Shahabi, H., Keesstra, S., Ahmad, B. B. & Bui, D. T. 2019 Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. *CATENA* **178**, 172–188.

Shokouhifar, Y., Lotfirad, M., Esmaeili-Gisavandani, H. & Adib, A. 2022 Evaluation of climate change effects on flood frequency in arid and semi-arid basins. *Water Supply* **22** (8), 6740–6755. https://doi.org/10.2166/ws.2022.271.

Shu, C. & Burn, D. H. 2004 Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resources Research* **40**, W09301.

Stevaux, J. C., de Azevedo Macedo, H., Assine, M. L. & Silva, A. 2020 Changing fluvial styles and backwater flooding along the Upper Paraguay River plains in the Brazilian Pantanal wetland. *Geomorphology* **350**, 106906.

Tehrany, M. S., Pradhan, B. & Jebur, M. N. 2013 Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology* **504**, 69–79.

Tehrany, M. S., Pradhan, B. & Jebur, M. N. 2014a Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology* **512**, 332–343.

Tehrany, M. S., Lee, M.-J., Pradhan, B., Jebur, M. N. & Lee, S. 2014b Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environmental Earth Sciences* **72** (10), 4001–4015. doi:10.1007/s12665-014-3289-3.

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.

Tehrany, M. S., Kumar, L. & Shabani, F. 2019a A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia. *PeerJ* **7**, e7653.

Tehrany, M. S., Jones, S. & Shabani, F. 2019b Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *CATENA* **175**, 174–192.

Termeh, S. V. R., Kornejady, A., Pourghasemi, H. R. & Keesstra, S. 2018 Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of the Total Environment* **615**, 438–451.

Theilen-Willige, B., Charif, A., El Ouahidi, A., Chaibi, M., Ayt Ougougdal, M. & AitMalek, H. 2015 Flash floods in the Guelmim area/ southwest Morocco—Use of remote sensing and GIS-tools for the detection of flooding-prone areas. *Geosciences* **5** (2), 203–221. doi:10.3390/geosciences5020203.

The World Bank 2013 *Building Morocco's Resilience: Inputs for an Integrated Risk Management Strategy*.

Tien Bui, D., Khosravi, K., Shahabi, H., Daggupati, P., Adamowski, J. F., Melesse, M. A. & Lee, S. 2019a Flood spatial modeling in northern Iran using remote sensing and GIS: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote Sensing* **11** (13), 1589.

Tien Bui, D., Shirzadi, A., Shahabi, H., Chapi, K., Omidavr, E., Pham, B. T., Talebpour Asl, D., Khaledian, H., Pradhan, B., Panahi, M. & Bin Ahmad, B. 2019b A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (Iran). *Sensors* **19** (11), 2444.

Tingsanchali, T. 2012 Urban flood disaster management. *Procedia Engineering* **32**, 25–37.

Torcivia, C. E. G. & López, N. N. R. 2020 Preliminary morphometric analysis: Río Talacasto Basin, Central Precordillera of San Juan, Argentina. In: *Advances in Geomorphology and Quaternary Studies in Argentina*. Springer, Cham, pp. 158–168.

Tucker, C. & Sellers, P. 1986 Satellite remote sensing of primary production. *International Journal of Remote Sensing* **7** (11), 1395–1416.

UNISDR 2015 *International Strategy for Disaster Reduction (ISDR): Making Development Sustainable: The Future of Disaster Risk Management*. Global Assessment Report on Disaster Risk Reduction, Geneva, Switzerland (accessed 05 November 2019).

Vafakhah, M., Mohammad Hasani Loor, S., Pourghasemi, H. & Katebikord, A. 2020 Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping. *Arabian Journal of Geosciences* **13** (11), 1–16.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S. & Bai, X. 2015 Flood hazard risk assessment model based on random forest. *Journal of Hydrology* **527**, 1130–1141.

Wang, Y., Hong, H., Chen, W., Li, S., Panahi, M., Khosravi, K., Shirzadi, A., Shahabi, H., Panahi, S. & Costache, R. 2019 Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. *Journal of Environmental Management* **247**, 712–729.

Werren, G., Reynard, E. & El Khalki, Y. 2012 Cartographie des phénomènes en vue de la realisation de la carte indicative des dangers hydrologiques dans la ville de Beni Mellal, Maroc. In *Actes du Colloque 'La Montagne Marocaine : Géomorphologie, Environnement et Développement'*, 5–6 mai 2010.

Wettschereck, D., Aha, D. W. & Mohri, T. 1997 A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review* **11**, 273–314.

Wilson, J. P. & Gallant, J. C. 2000 *Terrain Analysis: Principles and Applications*. JohnWiley & Sons.

Wu, L., Peng, Y., Fan, J. & Wang, Y. 2019 Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on cross-station and synthetic data.

Yang, R.-M., Zhang, G.-L., Liu, F., Lu, Y.-Y., Yang, F., Yang, F., Yang, M., Zhao, Y.-G. & Li, D.-C. 2016 Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators* **60**, 870–878.

Young, R. A. & Mutchler, C. K. 1969 Soil movement on irregular slopes. *Water Resources Research* **5** (5), 1084–1089.

Zhang, H. 2004 The optimality of naive Bayes. *American Association for Artificial Intelligence*. (www aaai org).

Zhang, G., Chen, W., Li, G., Yang, W., Yi, S. & Luo, W. 2020 Lake water and glacier mass gains in the northwestern Tibetan Plateau observed from multi-sensor remote sensing data: Implication of an enhanced hydrological cycle. *Remote Sensing of Environment* **237**, 111554.

Zhao, G., Pang, B., Xu, Z., Yue, J. & Tu, T. 2018 Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of the Total Environment* **615**, 1133–1142.

Zhao, Y., Han, Q., Zhao, Y. & Liu, J. 2019 Soil pore identification with the adaptive fuzzy C-means method based on computed tomography images. *Journal of Forestry Research* **30** (3), 1043–1052.

Zhu, R., Hu, X., Hou, J. & Li, X. 2020 Application of machine learning techniques for predicting the consequences of construction accidents in China. *Process Safety and Environmental Protection* **145**, 293–302.

Zkhiri, W., Tramblay, Y., Hanich, L. & Berjamy, B. 2016 Regional flood frequency analysis in the High Atlas mountainous catchments of Morocco. *Natural Hazards* **86** (2), 953–967.

Zurich Insurance Company Ltd, Targa-AIDE 2015 *Morocco Floods of 2014: What We Can Learn From Guelmim and Sidi Ifni*. Zurich.