

# Spatial prediction of flood-susceptible zones in the Ourika watershed of Morocco using machine learning algorithms

Floods  
prediction in  
the Ourika  
watershed

Modeste Meliho

*Laboratoire Ingénierie des Systèmes Avancés (ISA), Université Ibn Tofail,  
Kenitra, Morocco*

Abdellatif Khattabi

*Ecole Nationale Forestière d'Ingénieurs Salé, Salé, Morocco*

Zejli Driss

*Laboratoire Ingénierie des Systèmes Avancés (ISA), Université Ibn Tofail,  
Kenitra, Morocco, and*

Collins Ashianga Orlando

*Independent Researcher, Salé, Morocco*

Received 27 September 2021

Revised 7 December 2021

14 January 2022

Accepted 19 January 2022

## Abstract

**Purpose** – The purpose of the paper is to predict mapping of areas vulnerable to flooding in the Ourika watershed in the High Atlas of Morocco with the aim of providing a useful tool capable of helping in the mitigation and management of floods in the associated region, as well as Morocco as a whole.

**Design/methodology/approach** – Four machine learning (ML) algorithms including k-nearest neighbors (KNN), artificial neural network, random forest (RF) and x-gradient boost (XGB) are adopted for modeling. Additionally, 16 predictors divided into categorical and numerical variables are used as inputs for modeling.

**Findings** – The results showed that RF and XGB were the best performing algorithms, with AUC scores of 99.1 and 99.2%, respectively. Conversely, KNN had the lowest predictive power, scoring 94.4%. Overall, the algorithms predicted that over 60% of the watershed was in the very low flood risk class, while the high flood risk class accounted for less than 15% of the area.

**Originality/value** – There are limited, if not non-existent studies on modeling using AI tools including ML in the region in predictive modeling of flooding, making this study intriguing.

**Keywords** Natural hazards, Flood susceptibility, Predictive mapping, Machine learning algorithms, Ourika watershed

**Paper type** Research paper

## 1. Introduction

Natural disasters, including floods, pose a real threat to human life and can result in significant human losses and devastating economic consequences [1]. The potential magnitude of damage caused by floods has led some authors to consider them the most widespread and damaging of natural hazards [2]. Between 2000 and 2008, floods affected over 99 million people worldwide [3]. Due to heavy rainfall causing water to overflow into

© Modeste Meliho, Abdellatif Khattabi, Zejli Driss and Collins Ashianga Orlando. Published in *Applied Computing and Informatics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>



Applied Computing and  
Informatics  
Emerald Publishing Limited  
e-ISSN: 2210-8327  
p-ISSN: 2634-1964  
DOI 10.1108/ACI-09-2021-0264

---

riverbeds, floods damage both natural ecosystems and infrastructure, including the transportation system [4].

Floods are the most widespread natural disaster in Morocco. They rank second only to earthquakes in terms of the number of victims and injured individuals [5]. The watersheds of the Moroccan High Atlas are highly vulnerable to flood risks [6]. Indeed, the Ourika watershed has experienced flooding due to scouring in the valley, killing more than 200 people in 1995 [5]. In 2002, torrential rains resulted not only in the loss of more than 60 lives but also in significant property damage in the Ourika Valley [5]. Similarly, flooding in the Tougha Gorge resulted in two deaths and the partial and total destruction of 26 and 114 houses, respectively.

In the Ourika watershed, flood hazards are caused by rising waters and overflowing rivers that originate in the High Atlas Mountains. After the widespread devastation caused by the 1995 floods, the Moroccan government took steps to reduce the impact of flooding. These measures include improved weather observations and forecasts, the establishment of monitoring systems, measures to protect land uses, measures to combat water erosion and the construction of anti-erosion structures. Nevertheless, flood susceptibility mapping is critical to addressing flood risks in the Ourika watershed.

To develop flood susceptibility maps, geospatial analysis techniques based on geographic information system (GIS), remote sensing (RS) and statistics have been extensively used [7, 8]. Additionally, hydrological models such as Soil & Water Assessment Tool [9] and HYDROTEL [10], combined with GIS and RS have been employed for analyzing the various flood predisposing factors. A review of the literature shows the popularity of bivariate statistical methods, such as frequency ratio, value of information, weight of evidence and multivariate statistical methods, including logistic regression and multivariate adaptive spline regression, as approaches used in predictive mapping of natural disasters [11–14]. For flood prediction, Franci *et al.* [15] used high-resolution satellite imagery, GIS, and then multi-criteria analysis to produce flood risk maps with satisfactory results in Cyprus. In an alternative approach, Ziarh *et al.* [16] explored a data-driven multi-criteria decision analysis incorporating catastrophe and entropy theories in an effort to provide an unbiased assessment of flood risk distribution in Peninsular Malaysia.

While physical and statistical approaches may be sufficient in some cases, they remain limited for assessing the complex processes and interactions that influence natural phenomena such as floods. The recent success of machine learning (ML) models lies in their ability to not only account for nonlinearity issues related to physical processes, but also make it easier to model them at reduced costs [17]. Recent advances in ML techniques have made a considerable contribution to the enhancement of predictive flood hazard mapping [18]. Using ML algorithms, the limitations of traditional approaches can be addressed and the accuracy of predictions greatly improved [19]. Traditional techniques struggle to translate physical processes into mathematical terms [20]. Several ML algorithms have been successfully used to predict flood risks. These include artificial neural network (ANN), which is one of the most widely used ML algorithms for flood risk prediction [21–27]. Other ML algorithms have been used to predict flood risk such as support vector machine [28–30], random forest (RF) [17, 31, 32], logistic regression [7], adaptive neuro-fuzzy inference system [33] and Long Short-term Memory [34, 35].

Due to the high vulnerability of the Ourika watershed to flood hazards [5] and the lack of predictive flood hazard mapping in the region, this study is aimed at using ML algorithms such as K-nearest neighbors (KNN), ANN, RF and extreme gradient boost (XGB) to predict flash floods and subsequently generate a flood susceptibility map of the Ourika watershed. Additionally, this study aims to compare the performance and results of the models. This will be done by (1) identifying the flood predisposing factors, (2) comparing the accuracy of the models, (3) producing the flood susceptibility map using ML algorithms and (4) comparing

the model results. The susceptibility maps developed will serve as useful tools in flood prevention and mitigation in the Ourika watershed.

## 2. Materials and method

### 2.1 Study area

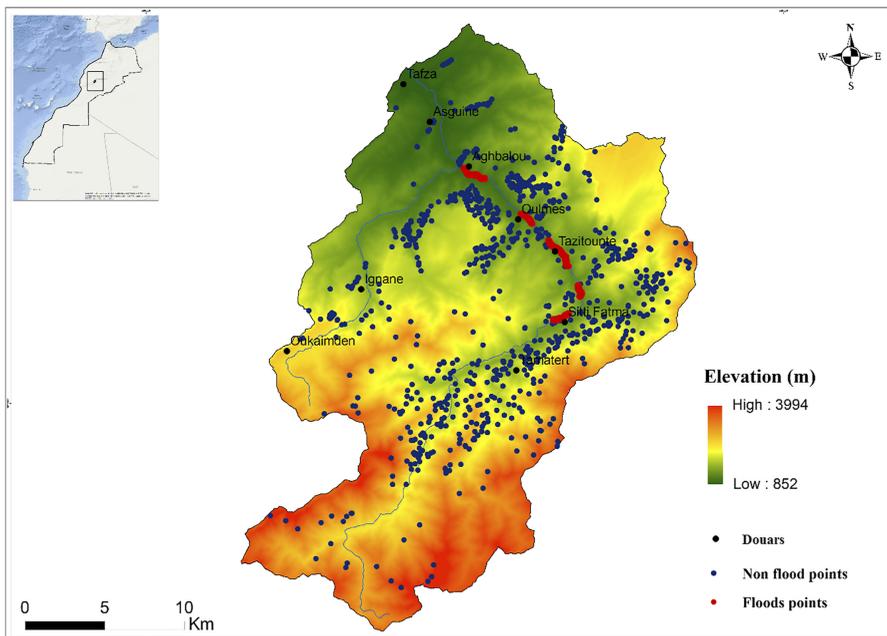
The Ourika watershed is located about 40 km south of Marrakech and covers an area of 576 km<sup>2</sup> (Figure 1). The mean annual rainfall is about 541 mm, with a coefficient of variability of 34%. This variability depends largely on the month and the season, and can reach monthly and seasonal coefficients of about 55% and 50%, respectively. Geologically, the area is mainly made up of magmatic rocks in its upstream section and sedimentary rocks downstream. The land use of the watershed is characterized by a highly diversified vegetation cover.

### 2.2 Data acquisition

Analysis of past flood events is crucial to predicting future floods. In this study, data on historical flooding in the Ourika watershed were provided by the Agence du Bassin Hydraulique du Tensift. The flood points represented as polygons of areas that experienced flooding in past covered 1076 pixels each measuring 30m by 30m. Additionally, non-flood points were randomly selected from the watershed map in areas with a slope greater than 50%, covering a similar number of pixels (1076), resulting in a total of 2152 pixels for both flooded and non-flooded polygons. The 30 m resolution digital elevation model (DEM) used in this study was obtained from the ASTER GDEM website (<https://asterweb.jpl.nasa.gov/gdem.asp>).

### 2.3 Floods parameters and conditioning factors

Selecting flood conditioning factors to be used in predicting floods is often challenging, making it imperative to select the most suitable ones for flood susceptibility mapping [29, 36].



**Figure 1.**  
Location of the study  
area and floods  
inventory

---

The factors identified as numerical variables include: curvature, elevation, distance to rivers, drainage density, flow accumulation, rainfall, slope, topographic wetness index (TWI), normalized difference vegetation index (NDVI), stream power index (SPI) and wetness index (WI). Conversely, the categorical variables selected for this study include: aspect, flow direction, geology, land use and substrate erodibility. Detailed descriptions of both numerical and categorical variables as well as their spatial representation in the study area are included as [supplementary material](#).

---

#### *2.4 Machine learning algorithms used*

Four algorithms were used for the modeling of flood susceptibility in the Ourika watershed. They include RF, extreme gradient boost (XGB), ANN and KNN. A detailed description of the algorithms has been included in the [supplementary material](#). Four models were developed corresponding to each of the four selected algorithms. Thus, 16 models were used for prediction in this study. For a given algorithm (e.g. KNN), the created models were as follows:

- (1) KNN: neither one-hot encoding nor variable selection was performed
- (2) KNN.TR: only one-hot encoding was performed
- (3) KNN.SE: only variable selection was performed
- (4) KNN.TR.SE: both one-hot encoding and variable selection were performed

#### *2.5 Cross-validation and feature selection*

Feature selection involves identifying and selecting a subset of variables from the original data set, to use as inputs of ML models. It helps tackle the issue of overfitting and makes the models simpler to interpret while shortening training times, which reduces the computational cost. Feature selection is accomplished using wrapper methods, which help examine possible feature combinations to identify the optimal feature set. Using these methods, features are removed one at a time, or are added one at a time, an ML model is built and the performance is determined. The selection procedure ends when the best performing model is found. In this study, forward feature selection together with the leave-one-out cross-validation (LOOCV) method was adopted to filter out variables that cause overfitting. The CAST package for R was used as a wrapper for forward feature selection.

#### *2.6 Handling of categorical variables*

Real-world data often involve discrete variables including categorical variables. The non-numerical nature of these variables presents several challenges when applying an ML algorithm. Thus, it is necessary to find a way to transform the data into numerical values.

One-hot encoding is the most popular method for transforming a categorical variable into a numerical variable. Its popularity lies mainly in the ease of application. Moreover, for many problems, it yields good results. Consider a categorical variable  $X$  which has  $K$  modalities  $m_1, m_2, \dots, m_K$ . One-hot encoding involves creating  $K$  indicator variables such that a vector of size  $K$  has 0 values everywhere and 1 at position  $i$  corresponding to modality  $m_i$ . Thus, the categorical variable is replaced by  $K$  numerical variables.

#### *2.7 Models performance assessment*

It is essential to assess the accuracy and overall performance of ML models. In this study, data splitting was used to separate the data into two distinct data sets: training and validation sets. From 2,152 data points, a subset of 70% of the data corresponding to 1,506 points was selected as the training data. The remaining 30% corresponding to 646 points were used to

evaluate performance. The model was trained through input of factor-flooding relationships and the resulting model applied to the entire watershed. The ROC-AUC curve was used to measure model performance. Receiver operating characteristic (ROC) is a graph illustrating the performance of a classification model at all classification thresholds while area under the curve (AUC) measures the entire two-dimensional area underneath the ROC curve and represents the degree of separability. Five AUC classes were highlighted in this study: poor (0.5–0.6), medium (0.6–0.7), good (0.7–0.8), very good (0.8–0.9) and excellent (0.9–1).

### 2.8 Flood susceptibility mapping

Based on the evaluation and correlations between each conditioning factor and the occurrence of floods in the Ourika watershed, and after validation of the models, the estimation of flood susceptibility values was carried out. Subsequently, the latter were reclassified into five susceptibility classes: very low, low, moderate, high and very high. This resulted in 16 flood susceptibility maps being produced – one for each of the four different models of the four ML algorithms employed.

### 2.9 Model similarity assessment

An important concept in modeling is the assessment for similarities and dissimilarities between results. Indeed, these are expected when working with models that differ in accuracy and overall performance. Identifying these points of divergence is critical because it underscores the confidence in the model predictions of flood-prone areas. To determine areas of similarity and dissimilarity in the model results, spatial comparisons of the model results were performed by overlaying the 16 susceptibility maps developed for each of the four models of the four ML algorithms used.

## 3. Results

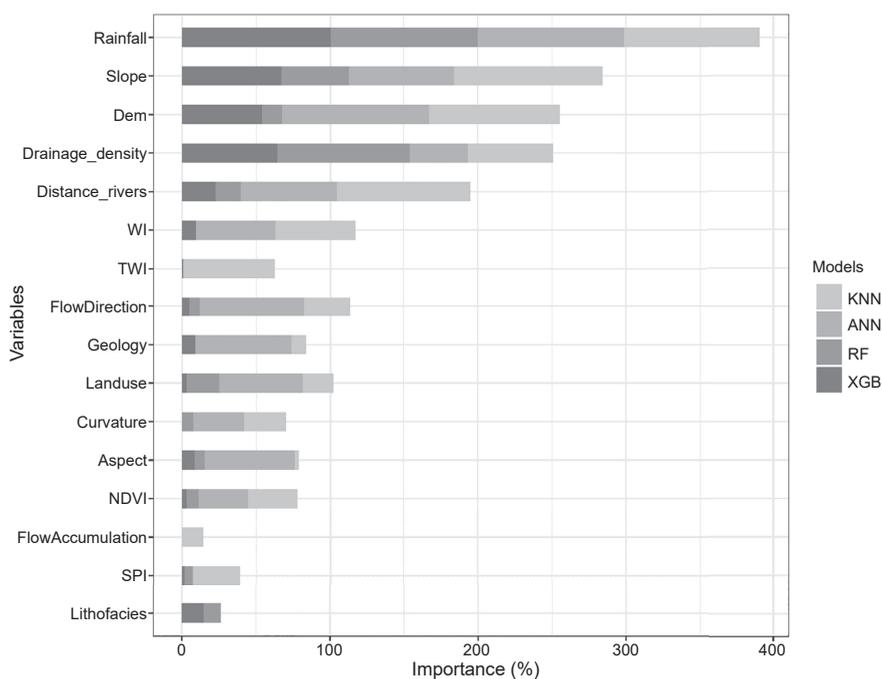
### 3.1 Variable selection and variable importance

**3.1.1 Selection of variables.** The variables selected for each model following forward feature selection are presented in [Table 1](#). Drainage density, rainfall and distance to river were the predominant predisposing factors in the Ourika watershed. Drainage density was selected as input for all models while rainfall was selected for all models except XGB.TR.SE. Distance to rivers was selected for all models but ANN.SE and XGB.TR.SE.

**3.1.2 Variables importance.** [Figure 2](#) shows the importance of variables for each model. Rainfall and slope were among the most important variables in all models, presenting scores above 90% and 65%, respectively, in all models. Drainage density was a significant predictor for the RF and KNN models with scores of 98.5% and 90.9%, respectively, while DEM was an

Models	Selected variables
KNN	KNN.SE Drainage density, rainfall, distance to rivers and geology
	KNN.TR.SE Drainage density, rainfall, distance to rivers, slope, DEM and TWI
ANN	ANN.SE Drainage density, rainfall and geology
	ANN.TR.SE Drainage density, rainfall, distance to rivers, west, granite, limestone and open juniperus
RF	RF.SE Drainage density, rainfall, distance to rivers and geology
	RF.TR.SE Drainage density, rainfall, distance to rivers, limestone and sandstone/marl
XGB	XGB.SE Drainage density, rainfall, distance to rivers, slope and DEM
	XGB.TR.SE Drainage density, DEM and open <i>Tetraclinis articulata</i> stands

**Table 1.**  
Selected variables for  
the models



**Figure 2.**  
Variable importance of  
the ML models

important variable for predicting flooding using the ANN and KNN models with scores of 100% and 88.3%, respectively.

### 3.2 Model performance analysis

The evaluation of a model's performance is an integral step in modeling. In this study, the ROC-AUC curve was used to assess model accuracy. The resulting AUC scores for each model are shown in [Table 2](#). A 10-factor cross-validation with three replications was used for training control. The hyperparameters of the different prediction models and the optimal parameters revealed during parameter tuning are included as part of the [supplementary material](#).

The most accurate models were under the RF (RF and RF.SE) and ANN (ANN and ANN.SE) algorithms, each recording AUC scores of 99.9%. Nevertheless, on average, RF and XGB were the best performing algorithms, with average AUC scores of 99.1% and 99.2% for their respective models. Conversely, KNN was on average the worst performing algorithm, with an AUC score of 94.4%. Indeed, the worst performing model across all algorithms was KNN, with KNN.TR recording an AUC score of 86.3%. Overall, the highest scores were observed for models that did not undergo one-hot encoding and for which variable selection was performed (ANN, RF, KNN and XGB), and for models associated with variable selection alone (ANN.SE, RF.SE, KNN.SE and XGB.SE). The models were sufficiently accurate, as evidenced by their AUC scores, which were above the prediction rate of 80% for the test data.

The overall classification accuracy of the models is included as [supplementary material](#). Consistent with the AUC scores, XGB and RF exhibited the highest average accuracy values at 96.4% (Kappa = 92.7%) and 91.8% (Kappa = 83.6%), respectively. Conversely, ANN had the lowest average accuracy at 90.3% (Kappa = 80.5%). Additionally, the sensitivity and specificity results of the models are included as [supplementary material](#).

Models		AUC (%)	Floods prediction in the Ourika watershed
KNN	KNN	97.9	
	KNN.SE	97.8	
	KNN.TR	86.3	
	KNN.TR.SE	95.5	
ANN	ANN	99.9	
	ANN.SE	99.9	
	ANN.TR	89.2	
	ANN.TR.SE	92.6	
RF	RF	99.9	
	RF.SE	99.9	
	RF.TR	98.5	
	RF.TR.SE	98.2	
XGB	XGB	99.7	
	XGB.SE	99.8	
	XGB.TR	99.6	
	XGB.TR.SE	97.7	

### 3.3 Floods susceptibility mapping

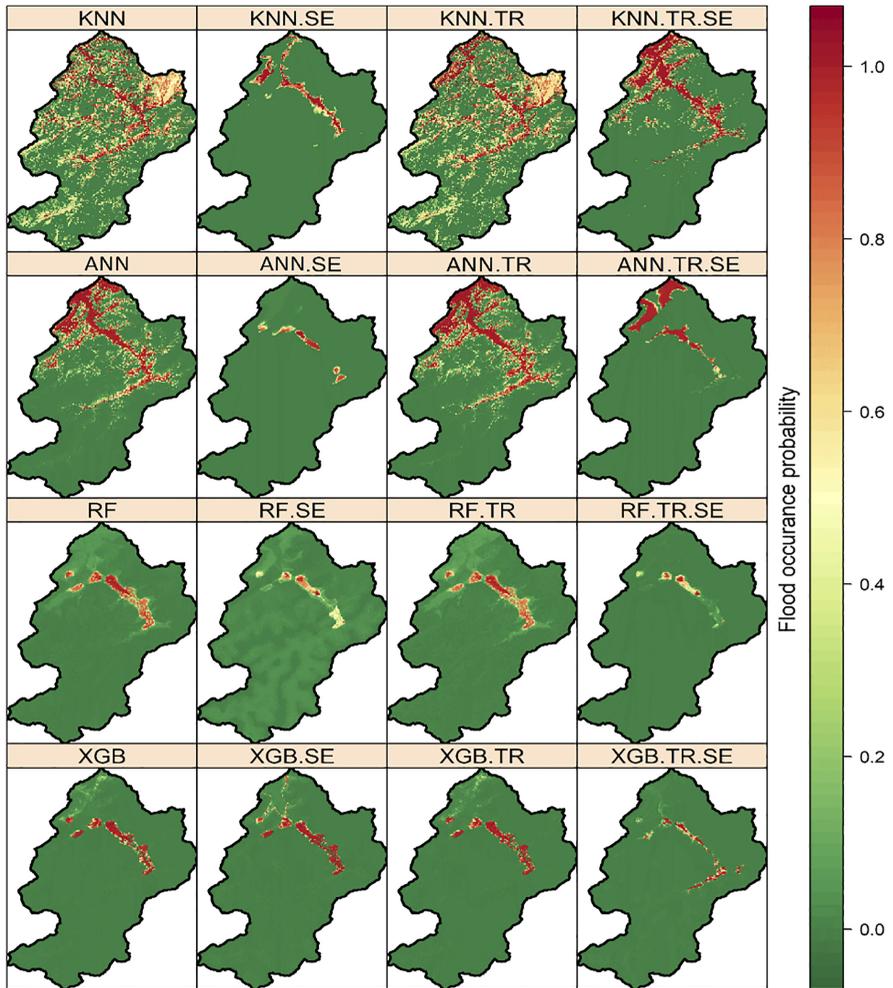
The resulting 16 flood susceptibility maps for the Ourika watershed for each model of the four ML algorithms employed are presented in [Figure 3](#), while the susceptibility classes are shown in [Table 3](#).

As expected, most high-risk areas for flooding are located near rivers. In addition, they are more concentrated near the watershed outlet in the NW section. These areas represent the low-lying areas of the watershed. The majority of the watershed area is classified in the very low susceptibility class ([Table 3](#)). Indeed, the very low class accounted for 95.87%, 81.73%, 61.99% and 96.72% of the watershed area for RF, ANN, KNN and XGB, respectively. KNN showed the most flood-prone areas, with most of the high (9.11%) to very high (13.77%) classes predicted by the model. Overall, the southern region of the watershed was predicted to be the least flood-prone, with the ANN, KNN and RF models showing virtually no areas susceptible to flooding ([Figure 3](#)). For the most part, the built-up areas belong to the very low sensitivity class (81.07%). However, a significant part of the built-up areas (14.91%) is located in areas of very high flood vulnerability.

### 3.4 Models results comparison

[Figure 4](#) shows the areas of the Ourika watershed with similar observations for the very low and very high flood risk classes, as well as the areas where model predictions were inconsistent. Additional information corresponding to these areas across the watershed is included as [supplementary material](#).

For the lowest performing algorithm, KNN, the corresponding models (KNN, KNN.TR, KNN.SE and KNN.TR.SE) were consistent in predicting the same areas of the watershed as being classified as very low (55.16% of the total area) and very high (1.20% of the total area) risk of flooding. By contrast, varying predictions across the four models were observed over the remaining 43.64% of the watershed area. For RF, which was the best performing algorithm, the corresponding models (RF, RF.TR, RF.SE and RF.TR.SE) were consistent in predicting 95.79% and 0.24% of the watershed as belonging to the very low and very high susceptibility classes, respectively. Conversely, inconsistencies in prediction were observed for the remaining 3.97% of the watershed. Overall, the 16 models were consistent in predicting 53.79% and 0.11% of the watershed as belonging to the very low and very high flood risk classes, respectively, while inconsistencies were observed for the remaining 46.10%.



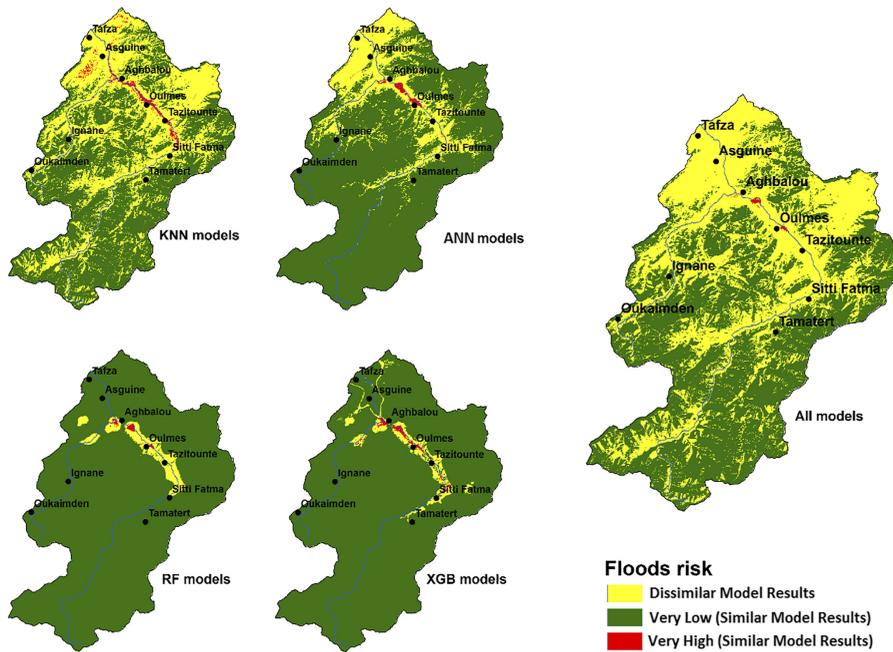
**Figure 3.** Floods occurrence probability maps of all the models

Susceptibility classes	Area (%)				
	RF	ANN	KNN	XGB	
Very low	95.87	81.73	61.99	96.72	
Low	0.72	3.07	8.72	0.24	
Moderate	0.24	2.24	6.41	0.22	
High	0.77	2.40	9.11	0.38	
Very high	2.40	10.56	13.77	2.43	

#### 4. Discussion

An important step in developing mitigation plans and allocating appropriate resources in response to future floods is the identification and delineation of areas prone to flooding. Generating dependable flood susceptibility maps remains a challenge notwithstanding the

## Floods prediction in the Ourika watershed



**Figure 4.** Models results comparison for each kind of model and for all models

popular and widespread adoption of ML techniques for flood prediction. In this study, we adopted four ML algorithms, including RF, ANN, KNN and XGB, and compared their prediction performance in the Ourika watershed in Morocco.

A total of 11 flood-conditioning factors, including curvature, elevation, distance to river, drainage density, slope, flow accumulation, precipitation, TWI, NDVI, SPI and WI, were selected as numerical variables, while five factors, including aspect, flow direction, geology, substrate resistance to erosion, and land use, were identified as categorical variables in mapping flood susceptibility based on the literature [30, 36–45].

The final susceptibility maps did not show considerable variation and were spatially consistent between models, with the exception of KNN, which predicted the greatest areas of the Ourika watershed as being susceptible to flooding. Indeed, RF, ANN and XGB predicted that the central and northwestern part of the watershed would be moderately to highly susceptible to flooding, while almost the entire southern half was predicted to be the least susceptible. Moderately to highly prone areas are dominated by substrates that have low resistance to erosion while being dominated by human activities such as agriculture and construction. These are likely to create conditions that favor hydrologic processes such as runoff, thereby increasing the likelihood of flood events. Overall, most of the study area was classified as being at very low risk of flooding. However, a significant portion was predicted by ANN and KNN to be very highly prone to flooding, with the distribution of built-up areas revealing that nearly 15% of these areas fall into this class. This highlights the potential impact of urbanization in influencing flooding in the watershed. Indeed, urbanization has been confirmed as a driving factor in its role in increasing in flood risk [46–48]. It leads to a significant increase in impervious surface, which generally reduces the hydrologic response time and thus increases the risk of flooding [49]. Moreover, studies conducted by Al-Ghamdi *et al.* [50] and Zhao *et al.* [51] have noted its role in significantly increasing peak flood events.

The models used in this study presented satisfactory results and were considered appropriate for the establishment of flood susceptibility maps for the Ourika watershed. Indeed, the lowest AUC score was recorded for the KNN algorithm at 86.3%. XGB and RF were found to be the most reliable for predicting flooding, with AUC scores above 99%. The superior performance of RF in particular over other algorithms has been highlighted in other studies [52–56]. Indeed, it has been shown to be robust to noise and outliers, which are some of the common problems in flood susceptibility modeling. RF is not only capable of predicting the role of input factors in the modeling process, but also of handling huge data composed of varying inputs without factor suppression [54, 55]. This was observed in the study on multi hazard mapping conducted Salzburg (Austria) by [57] where it performed in flood prediction compared to support vector machines. As for the equally strong performance by XGB, Rampali *et al.* [58] obtained comparable results, finding it to have the highest predictive power in their work on flood risk assessment in India. Correspondingly, Abedi *et al.* [56] in their study in Romania also noted its high performance, although, much like in our study, better than RF. These results are in line with our observations, showing that the selected algorithms can provide sufficiently accurate models for flood prediction in the region.

In order to remedy the devastating situation caused by floods, the improvement of flood forecasting and prevention remains an essential step. Better information of the exposed populations and the reduction of the vulnerability of the goods located in the floodable zones are to be privileged. Although the application of ML methods for flood susceptibility modeling often comes with inherent challenges such as the selection of appropriate model inputs, the algorithms used, their power, ease of application and relatively low cost compared to traditional methods can be leveraged effectively, making them a useful tool in flood risk management.

## 5. Conclusions

Floods are one of the most devastating and damaging events. In the Mediterranean region, the extent of past floods and the forecast of an increase in their frequency require that the risk of flooding be taken into consideration by local planners and decision makers. In this context, the determination of areas likely to be affected by floods and the subsequent elaboration of flood susceptibility maps is essential for a better management of this risk.

Our approach offers a tool for flood risk assessment at the watershed scale. It is based on the identification and analysis of factors influencing flooding followed by use of ML techniques to predict floods. Thus, models based on the four ML algorithms were used to develop flood susceptibility maps, with XGB and RF exhibiting the highest predictive power. Overall, the models were highly accurate, thus showing they can be applied with confidence in the region.

Although flood susceptibility modeling remains a challenging endeavor with many inherent complexities, our results can help regional planners and decision makers to implement mitigation and development strategies that would be useful for optimal flood management, not only in the region, but also in the Moroccan and ultimately Mediterranean contexts.

## References

1. Samanta RK, Bhunia GS, Shit PK, Pourghasemi HR. Flood susceptibility mapping using geospatial frequency ratio technique: a case study of Subarnarekha River Basin, India. *Model Earth Syst Environ.* 2018; 4(1): 395-408.
2. Doocy S, Daniels A, Packer C, Dick A, Kirsch TD. The human impact of earthquakes: a historical review of events 1980-2009 and systematic literature review. *PLOS Currents Disasters.* 2013; 5(1).
3. Opolot E. Application of remote sensing and geographical information systems in flood management: a review. *Res J Appl Sci Eng Technol.* 2013; 6(10): 1884-94.
4. Kron W. Flood risk = hazard  $\times$  exposure  $\times$  vulnerability. In: Wu M, *et al.* (Eds). *Flood defence.* New York: Science Press; 2002. p. 82-97.

5. Rapport final de l'Etude préparatoire pour le Projet de Système de Prévision et d'Alerte aux Crues dans la région du Haut Atlas Royaume du Maroc, 2011.
6. Zkhiri W, Trambly Y, Hanich L, Berjamy B. Regional flood frequency analysis in the High Atlas mountainous catchments of Morocco. *Nat. Hazards*. 2016; 86(2): 953-67.
7. Pradhan B. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *J Spat Hydrol*. 2009; 9: 1-18.
8. Pradhan B, Shafiee M, Pirasteh S. Maximum flood prone area mapping using RADARSAT images and GIS: kelantan river basin. *Int J Geoinformatics*. 2010; 5: 11.
9. Jayakrishnan R, Srinivasan R, Santhi C, Arnold J. Advances in the application of the SWAT model for water resources management. *Hydrol. Process*. 2005; 19: 749-62.
10. Jutras S, Rousseau A, Clerc C. Implementation of a peatland-specific water budget algorithm in HYDROTEL. *Can Water Resour J*. 2009; 34(4): 349-64.
11. Ayalew L, Yamagishi H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains. *Geomorphology*. 2005; 65: 15-31.
12. Tien Bui D, Khosravi K, Shahabi H, Daggupati P, Adamowski JF, Melesse MA, Lee S. 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. *Rem Sens*. 2019; 11(13): 1589.
13. Tehrany HM, Kumar L, Shabani F. A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia. *Peer J*. vol. 7; 2019. p. 1-32.
14. Al-Abadi AM, Al-Najar NA. Comparative assessment of bivariate, multivariate and machine learning models for mapping flood proneness. *Nat Hazards*. 2020; 100: 461-91.
15. Franci F, Bitelli G, Mandanici E, *et al*. Satellite remote sensing and GIS-based multi-criteria analysis for flood hazard mapping. *Nat Hazards*. 2016; 83: 31-51. doi: [10.1007/s11069-016-2504-9](https://doi.org/10.1007/s11069-016-2504-9).
16. Ziarh G, Asaduzzaman M, Dewan A, Nashwan M, Shahid S. Integration of catastrophe and entropy theories for flood risk mapping in peninsular Malaysia. *J Flood Risk Manag*. 2020; 14. doi: [10.1111/jfr3.12686](https://doi.org/10.1111/jfr3.12686).
17. Mosavi A, Ozturk P, Chau K. Flood prediction using machine learning models: literature review. *Water*. 2018; 10(11): 1536.
18. Wang Z, Lai C, Chen X, Yang B, Zhao S, Bai X. Flood hazard risk assessment model based on random forest. *J Hydrol*. 2015; 527: 1130-41.
19. Xu ZX, Li JY. Short-term inflow forecasting using an artificial neural network model. *Hydrol Process*. 2002; 16: 2423-39.
20. Hosseiny H, Nazari F, Smith V, Nataraj C. A framework for modeling flood depth using a hybrid of hydraulics and machine learning. *Sci Rep*. 2020; 10: 8222.
21. Shu C, Burn DH. Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resour Res*. 2004; 40: W09301.
22. Seckin N, Cobaner M, Yurtal R, Haktanir T. 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 Resour Manag*. 2013; 27(7): 2103-24.
23. Campolo M, Soldati A, Andreussi P. Artificial neural network approach to flood. *Hydrol Sci J*. 2003; 48: 381-98.
24. Kim S, Matsumi Y, Pan S, Mase H. A real-time forecast model using artificial neural network for after- runner storm surges on the Tottori coast, Japan. *Ocean Eng*. 2016; 122: 44-53.
25. Liu R, Chen Y, Wu J. Assessing spatial likelihood of flooding hazard using naïve Bayes and GIS: a case study in Bowen Basin, Australia. *Stoch Environ Res Risk Assess*. 2016; 3: 1575-90.

26. Jahangir MH, Mousavi Reineh SM, Abolghasemi M. Spatial predication of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm. *Weather Clim Extremes*. 2019; 25: 100215.
27. Rahman M, Chen N, Islam MM, Mahmud G, Pourghasemi H, Alam M, Rahim M, Baig M, Bhattacharjee A, Dewan. Development of flood hazard map and emergency relief operation system using hydrodynamic modeling and machine learning algorithm. *J Clean Prod*. 2021; 311: 127594. doi: [10.1016/j.jclepro.2021.127594](https://doi.org/10.1016/j.jclepro.2021.127594).
28. Tehrany MS, Pradhan B, Jebur MN. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J Hydrol*. 2014; 512: 332-43.
29. Tehrany MS, Pradhan B, Mansor S, Ahmad N. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*. 2015; 125: 91-101.
30. Dazzi S, Vacondio R, Mignosa P. Flood stage forecasting using machine-learning methods: a case study on the parma river (Italy). *Water*. 2021; 13: 1612. doi: [10.3390/w13121612](https://doi.org/10.3390/w13121612).
31. Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ Model Softw*. 2017; 95: 229-245.
32. Farzaneh SH, Choubin B, Mosavi A, Nabipour N, Band S, Darabi H, Haghghi AT. Flash-flood hazard assessment using Ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. *Sci Total Environ*. 2019; 711: 135161. doi: [10.1016/j.scitotenv.2019.135161](https://doi.org/10.1016/j.scitotenv.2019.135161).
33. Ahmadlou M, Karimi M, Alizadeh S, Shirzadi A, Parvinnejhad D, Shahabi H, Panahi M. Susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and bat algorithms (BA). *Geocarto Int*. 2018; 34: 1-21.
34. Dazzi S, Vacondio R, Mignosa P. Flood stage forecasting using machine-learning methods: a case study on the parma river (Italy). *Water*. 2021; 13: 1612. doi: [10.3390/w13121612](https://doi.org/10.3390/w13121612).
35. Apaydin H, Feizi H, Sattari M, Çolak MS, Band S, Chau KW. Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. *Water*. 2020; 12(5, Supp. 1500). doi: [10.3390/w12051500](https://doi.org/10.3390/w12051500).
36. Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WNA, Moradi A. An artificial neural network model for flood simulation using GIS: johor River Basin, Malaysia. *Environ Earth Sci*. 2012; 67(1): 251-264.
37. Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR. A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. *Nat Hazards*. 2016; 83: 947-987.
38. Mojaddadi H, Pradhan B, Nampak H, Ahmad N, Ghazali AHB. Ensemble machine-learning-based geospatial approach for flood risk assessment using multisensory remote-sensing data and GIS. *Geomat. Nat. Haz. Risk*. 2017; 8(2): 1080-1102. doi: [10.1080/19475705.2017.1294113](https://doi.org/10.1080/19475705.2017.1294113).
39. Lee S, Kim JC, Jung HS, Lee MJ, Lee S. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Nat Hazards Risk*. 2017; 8(2): 1185-203.
40. Yariyan P, Avand M, Abbaspour RA, Haghghi AT, Costache R, Ghorbanzadeh O, Janizadeh S, Blaschke T. Flood susceptibility mapping using an improved analytic network process with statistical models. *Geomat. Nat. Hazards Risk*. 2020; 11: 2282-2314.
41. Islam ARMT, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, Kuriqi A, Linh NT. Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci Front*. 2020; 12(3).
42. Robinson M, Dupeyrat A. Effects of commercial forest felling on streamflow regimes at Plynlimon, mid-Wales. *Hydrol. Process.*. 2003; 19: 1213-26.
43. O'Connell PE, Beven KJ, Carney JN, Clements RO, Ewen J, Fowler H, Harris GL, Hollis J, Morris J, O'Donnell GM, Packman JC, Parkin A, Quinn PF, Rose SC, Shepherd M, Tellier S. Project FD2114:

- review of impacts of rural land use and management on flood generation. Defra R&D Technical Report FD2114. Defra, London; 2004.
44. FAO/IIASA/ISRIC/ISSCAS/JRC. Harmonized world Soil database (version 1.2). Rome, Italy and IIASA. Laxenburg, Austria: FAO; 2012.
  45. Calder IR, Aylward B. Forests and Floods: moving to an evidence-based approach to watershed and integrated flood management. *Water Int.* 2006; 31(1): 87-99.
  46. Liu YB, De Smedt F, Hoffmann L, Pfister L. Assessing land use impacts on flood processes in complex terrain by using GIS and modeling approach. *Environ Model Assess.* 2005; 9(4): 227-35.
  47. Saghafian B, Farazjoo H, Bozorgy B, Yazdandoost F. Flood intensification due to changes in land use. *Water Resour Manag.* 2008; 22: 1051-67.
  48. Huang Q, Wang J, Li M, Fei M, Dong J. Modeling the influence of urbanization on urban pluvial flooding: a scenario-based case study in Shanghai, China. *Nat Hazards.* 2017; 87: 1035-55.
  49. Feng B, Zhang Y, Bourke R. Urbanization impacts on flood risks based on urban growth data and coupled flood models. *Nat Hazards.* 2021; 106: 613-27. doi: [10.1007/s11069-020-04480-0](https://doi.org/10.1007/s11069-020-04480-0).
  50. Al-Ghamdi KA, Elzahrany RA, Mirza MN, Dawod GM. Impacts of urban growth on flood hazards in Makkah City, Saudi Arabia. *Int J Water Resour Environ Eng.* 2012; 4: 23-34.
  51. Zhao G, Gao H, Cuo L. Effects of urbanization and climate change on peak flows over the San Antonio River Basin, Texas. *J Hydrometeorol.* 2016; 17: 2371-89.
  52. Sameen MI, Pradhan B, Lee S. Self-learning random forests model for mapping groundwater yield in data-scarce areas. *Nat Resour Res.* 2019; 28: 757-75.
  53. Vafakhah M, Mohammad Hasani Loor S, Pourghasemi H, *et al.* Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping. *Arab J Geosci.* 2020; 13: 417. doi: [10.1007/s12517-020-05363-1](https://doi.org/10.1007/s12517-020-05363-1).
  54. Naghibi SA, Pourghasemi HR, Dixon B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* 2016; 188: 44. doi: [10.1007/s10661-015-5049-6](https://doi.org/10.1007/s10661-015-5049-6).
  55. Naghibi SA, Pourghasemi HR. A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. *Water Resour. Manag.* 2015; 29: 5217-36.
  56. Abedi R, Costache R, Shafizadeh-Moghadam H, Bao Pham Q. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto Int.* 2021. doi: [10.1080/10106049.2021.1920636](https://doi.org/10.1080/10106049.2021.1920636).
  57. Rampalli M, Sistla S, Raju K. Application of machine learning algorithms for flood susceptibility assessment and risk management. *Journal of Water and Climate Change.* 2021; 12. doi: [10.2166/wcc.2021.051](https://doi.org/10.2166/wcc.2021.051).
  58. Nachappa TG, Ghorbanzadeh O, Gholamnia K, Blaschke T. Multi-hazard exposure mapping using machine learning for the state of Salzburg, Austria. *Remote Sens.* 2020; 12: 2757. doi: [10.3390/rs12172757](https://doi.org/10.3390/rs12172757).

## Appendix

Supporting data including tables and figures relative to the study has been made available at: <https://github.com/melmos44/flooding-ourika>

## Corresponding author

Modeste Meliho can be contacted at: [modestemeliho@yahoo.fr](mailto:modestemeliho@yahoo.fr)

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)