

## Analysis of Soil Water Erosion Risk Using Machine Learning Technics – A Case Study of Ourika Watershed in Morocco

Ilham L'Arfouni<sup>1\*</sup>, Ahmed Algouti<sup>1</sup>, Abdellah Algouti<sup>1</sup>, Rachid Es-Sadiq<sup>2</sup>

<sup>1</sup> Geosciences Geotourism Natural Hazards and Remote Sensing Laboratory, Department of Geology, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco

<sup>2</sup> The Chamber of Agriculture of Marrakech, Safi, Morocco

\* Corresponding author's e-mail: [ilham.larfouni@gmail.com](mailto:ilham.larfouni@gmail.com)

### ABSTRACT

Soil erosion is a major environmental problem with detrimental consequences. In this article, we present a detailed study on the analysis of soil water erosion using machine learning (ML) techniques in the Oued Ourika watershed in Morocco. We collected data on various factors that may influence the mechanisms of soil water erosion events. Subsequently, we developed machine learning models to predict the potential for soil water erosion based on these factors. Finally, field studies were conducted compared to the obtained results. A historical inventory of water erosion has been created through fieldwork, satellite imagery, and historical water erosion events. Models were constructed using the training data, and their performance and accuracy in predicting susceptibility to water erosion were evaluated using the validation data. This data division allowed for a fair assessment of the models' ability to make accurate predictions. Using a Geographic Information System (GIS) and programming in the R language, four supervised machine learning algorithms were applied, including k-nearest neighbor (KNN), extreme gradient boosting (XGB), random forest (RF), and naive bayes (NB). The results show that the NB model exhibited the highest accuracy in predicting and evaluating the effectiveness of these algorithms in forecasting susceptibility to water erosion in the study area. Accuracy was assessed using the area under the curve (AUC) metric, with an AUC of 98%. The XGB algorithm had an AUC of 96%, followed by RF with an AUC of 87%, and KNN with an AUC of 84%. Thus, the Naive Bayes model proved to be the best for determining susceptibility to water erosion in the study area. The analysis of water erosion reveals that 43% of the total area of the Oued Ourika watershed is exposed to a high to very high risk of erosion in the Ourika region. These findings can assist regional and local authorities in reducing the risk of water erosion and implementing effective measures to prevent potential damages. The goal is to protect the communities and infrastructure located along the course of the Ourika. Overall analysis of natural disasters, the accuracy of the results heavily depends on the availability and quality of data, which must encompass an adequate number of parameters.

**Keywords:** water erosion, natural disasters, machine learning, Ourika, high atlas of Marrakech, Morocco.

### INTRODUCTION

Soil erosion is a natural phenomenon that occurs due to the wear and tear of the Earth's surface under the influence of forces such as water, wind, soil characteristics, precipitation, anthropogenic factors, vegetation, slopes, lithology, and gravity.

In Morocco, a country characterized by arid and semi-arid regions, this process holds particular significance. However, human activities such

as intensive agriculture, deforestation, and urbanization have significantly accelerated this natural process. As a result, we are facing massive losses of arable soil, water quality issues, and disruptions in the delicate ecosystems of these regions.

Over the past few decades, several empirical models have been developed to assess soil erosion and sediment transport based on physical and geomorphological parameters (Wischmeier, Smith, 1978; Renard et al., 1997). Empirical

models can be used in the evaluation of soil water erosion. However, these models may have limitations in certain regions. For example, they are constrained by their uncertainty and a significant gap between predicted and measured values (Alewell et al., 2019). ML techniques play a crucial role in the management of soil water erosion for several reasons. They enable the identification and analysis of complex datasets that encompass factors contributing to erosion. These methods provide advanced and powerful tools for a more effective understanding and modeling of soil erosion susceptibility.

Currently, multiple studies have been conducted to investigate soil erosion management using machine learning investigations (Pradhan 2013; Padarian et al., 2020; Sahour et al., 2021; Folharini et al., 2023). These studies have demonstrated the effectiveness of machine learning techniques in assessing soil water erosion and in overall natural risk management.

In order to address these critical environmental issues, it is imperative to conduct an in-depth analysis of the factors contributing to soil erosion and develop highly accurate prediction tools. These prediction tools are of paramount importance in guiding the implementation of effective and sustainable conservation measures. In

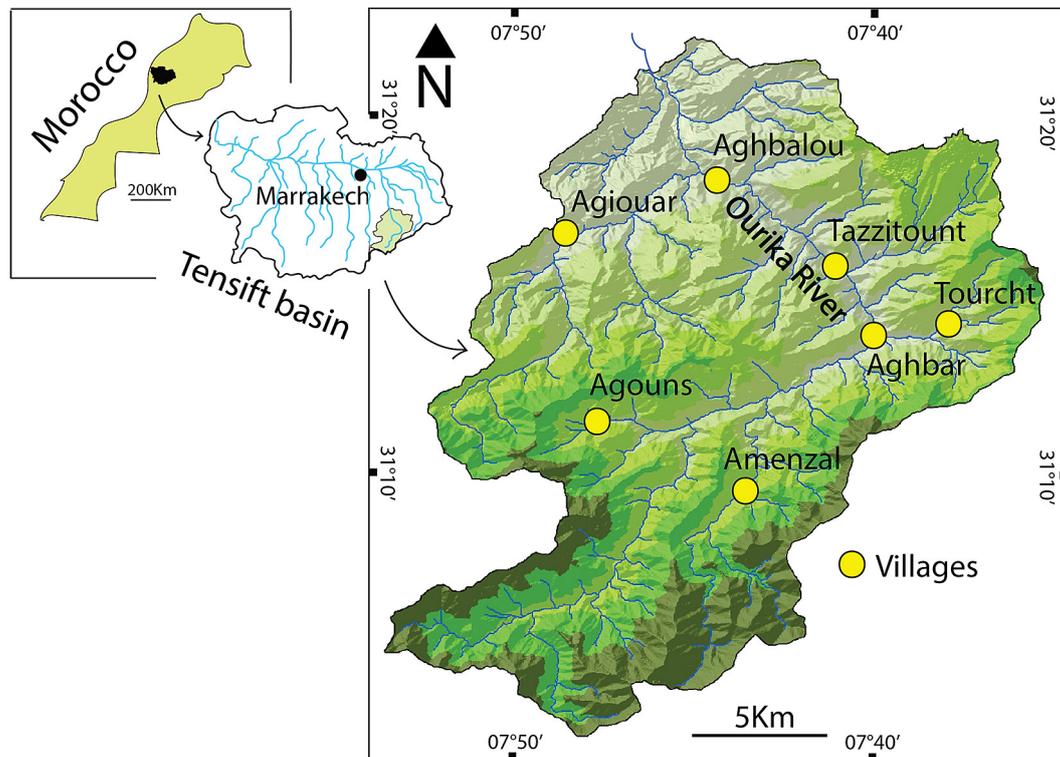
the context of this study, the Oued Ourika watershed was chosen as a case study due to its unique geographical and climatic characteristics, with a particular focus on semi-arid regions where vulnerability to erosion is often heightened.

The ultimate goal of this study is to utilize machine learning techniques to analyze and predict soil erosion based on various environmental factors within the Ourika watershed. The overarching aim is to provide valuable information to farmers and relevant stakeholders, assisting them in better managing and preventing water erosion in these arid and semi-arid regions of the High Atlas. By comprehending the underlying factors contributing to erosion, we can implement targeted and sustainable conservation measures, thus contributing to soil preservation, safeguarding water resources, and preserving the fragile ecosystems of these critical regions essential for food and environmental security.

## MATERIALS AND METHODS

### Study area

The Ourika watershed (Fig. 1) is located between latitude  $31^{\circ}\text{N}$  and  $31^{\circ}21'\text{N}$  and longitude



**Figure 1.** Geographical location of the study area

7°30'W and 7°60'W. It is bounded to the south by the Souss watershed, to the north by the Haouz plain, to the east by the Zat watershed, and to the west by the Rheraya watershed. The slopes in the watershed are generally steep, which enhances runoff and erosion. It is a continental watershed with an area of approximately 582 km<sup>2</sup>. It is situated on the northwestern front of the high atlas in Marrakech, Morocco. This watershed is known for its devastating floods and land movements that primarily affect the less competent Triassic terrains.

## Data collection

We gathered detailed data on topography, vegetation, precipitation, and agricultural practices. Subsequently, we developed Machine Learning models to predict areas susceptible to water erosion in the Ourika watershed, located on the southern slopes of the high atlas in Marrakech.

## Geology

A geological map (Fig. 3a) is digitized from Proust's (1961), geological map, and the various lithological facies are grouped into four classes: clays and unconsolidated materials, shales, conglomerates, and hard Precambrian materials (Biron et al., 1982; Nefly 1998). Subsequently, each lithology type is assigned an index that depends on its contribution to erosion.

## Rainfall

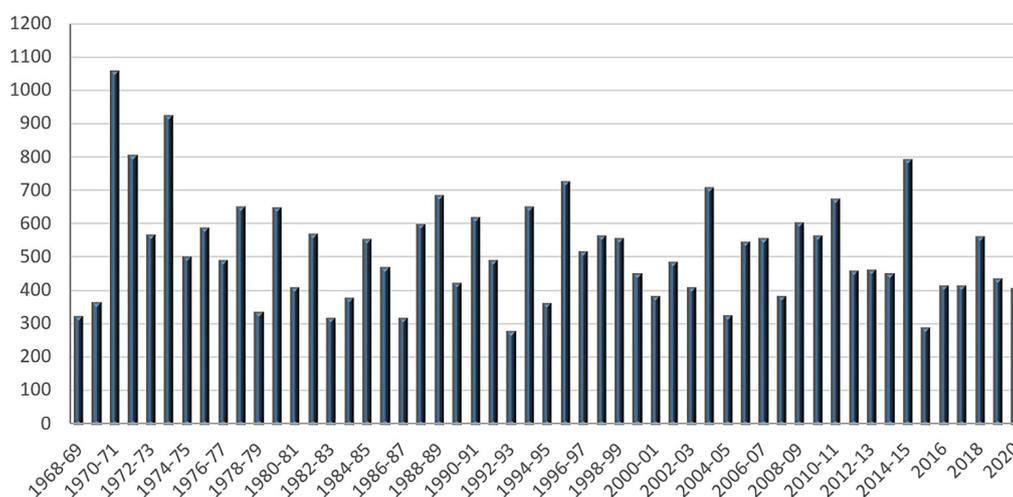
The rainfall map was established using historical precipitation data for the Ourika region. The region is characterized by spatiotemporal

variability in precipitation and relative irregularity in surface streamflow of rivers (Saidi, 1994; 2010). Altitudes disturbances, particularly concentrated on high peaks, result in predominantly summer thunderstorms (Delannoy, 1981).

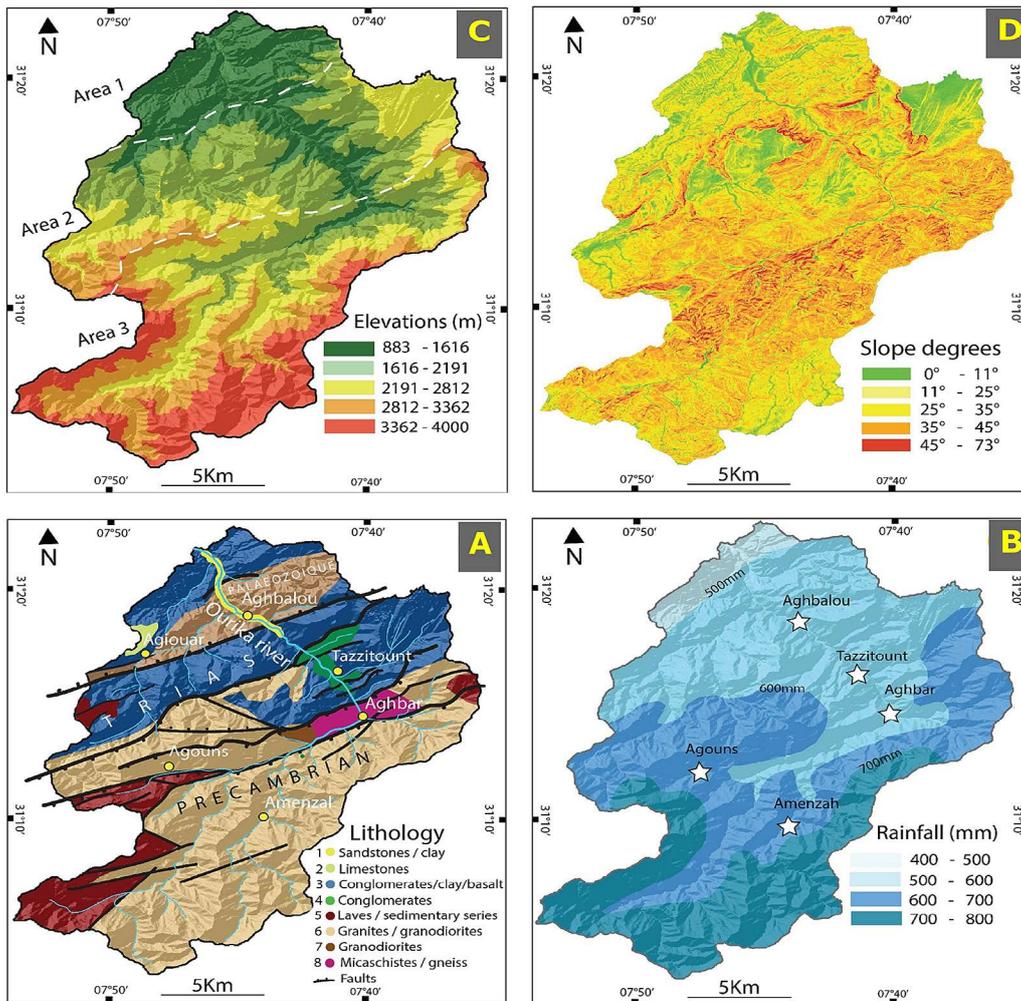
The average annual rainfall (Fig. 2 and Fig. 3b) in the Ourika watershed varies from 400 to over 650 mm (semiarid to subhumid bioclimate) in the foothills and sub-Atlas zone, and from 800 to 1000 mm (upper subhumid bioclimate) on the high peaks exposed to oceanic winds (Ouhammou, 1986). The rainiest months are February, March, and April, while July and August are the driest. The analysis of average monthly rainfall shows that April is the wettest month with an average of 88.4 mm, and July is the driest month with an average of 3.5 mm. The monthly distribution of rainfall indicates that the dry season is in summer, and the rainy season is in winter, accounting for nearly 40% of the annual precipitation, with autumn and spring having almost equal precipitation rates.

## Hypsometry

In the Ourika watershed, the main watercourse originates at an altitude of approximately 3600 m. The hypsometric map (Fig. 3c) illustrates that the watershed encompasses high altitudes ranging from 800 m downstream to 4000 m at its upstream extremity. The Aghbalou to Imin Taddert region is characterized by a morphology dominated by extensive, deeply incised tabular surfaces. Over 50% of the watershed's area features very high altitudes, with rugged terrain and peaks reaching as high as 3800 to 4000 m in some places.



**Figure 2.** Annual precipitation from 1969 to 2020 at the Aghbalou station



**Figure 3.** (a) Geological map (based on Proust., 1961), (b) rainfall map, (c) map of elevations in, and (d) distribution of slope angles in the Ourika watershed

*The slopes*

This is the most important parameter; the slope map is generated based on the digital elevation model (DEM) and reclassified into four slope categories. The weight assigned to the slope is the highest because it is the most critical factor in slope instability (Sarkar et al., 2008, Wachal, Hudack 2000).

In general, the slopes in the Ourika watershed are characterized by steep gradients. Indeed, over 60% of its surface exhibits slopes exceeding 25 degrees. The slope distribution map of the watershed highlights three distinct zones (Fig. 3d):

- Zone 1 – this area is characterized by moderate slopes in the valley bottoms and on the slopes of Amassine and Isk-n-Tanoumri. Gentle slopes are observed in the Tamzendirt basin.
- Zone 2 – it features gentle slopes in the valley bottoms and on the plateaus of Timmenkar and Yagour. However, along the edges of

these plateaus, the slopes are steep, exceeding 45 degrees.

- Zone 3 – the slope distribution reflects an immature relief. Except for the Oukaïmdene region, where slopes are gentle, the entire zone has steep slopes ranging from 35 to 74 degrees. These steep slopes are more pronounced on the right bank of the Ourika river than on the left bank.

In the Ourika watershed, several erosion surfaces (or plateaus) are observed. They are located at various levels along the valley, but they are primarily concentrated on the Triassic terrains in the Tougalkhir-Aït Lmchkour section. Most of these plateaus are discontinuous and are found at varying altitudes. The mapping shows the presence of several plateaus, which can be categorized into two main families: numerous plateaus located at approximately 1200 m elevation and plateaus at around 1700 m altitude.

The vegetation index

Vegetation has always been a significant factor in slope stability. The vegetation cover map is categorized into four classes based on the vegetation index, which depends on its density (forests, cultivated lands, etc.). The Normalized Difference Vegetation Index (NDVI) has been used to assess the density of the vegetation area and its correlation with occurrences of water erosion. NDVI provides a measure of vegetation distribution and facilitates the assessment of changes progressively (Liu 2005; 2019). In this case study, the NDVI map was obtained from Landsat 8 OLI images, and the components used to calculate the index are detailed in Equation 1.

$$NDVI = (BNIR - BR) / (BNIR + BR) \quad (1)$$

Subsequently, the NDVI was reclassified into five different classes (Fig. 4a). The infrared band of the spectrum is represented by BNIR, and the red band by BR. The NDVI classes range from -1 to +1. Areas most susceptible to floods and water erosion are arid lands (sand and friable rock), which have values below 0.1. Low NDVI values correspond to grasslands (0.2–0.3), while high values indicate tropical forests (0.6–0.8) (Campolo et al., 2003).

Distance to faults

In a mountainous, actively uplifting context such as the study area, faults represent a significant factor that destabilizes the slopes they intersect. In the

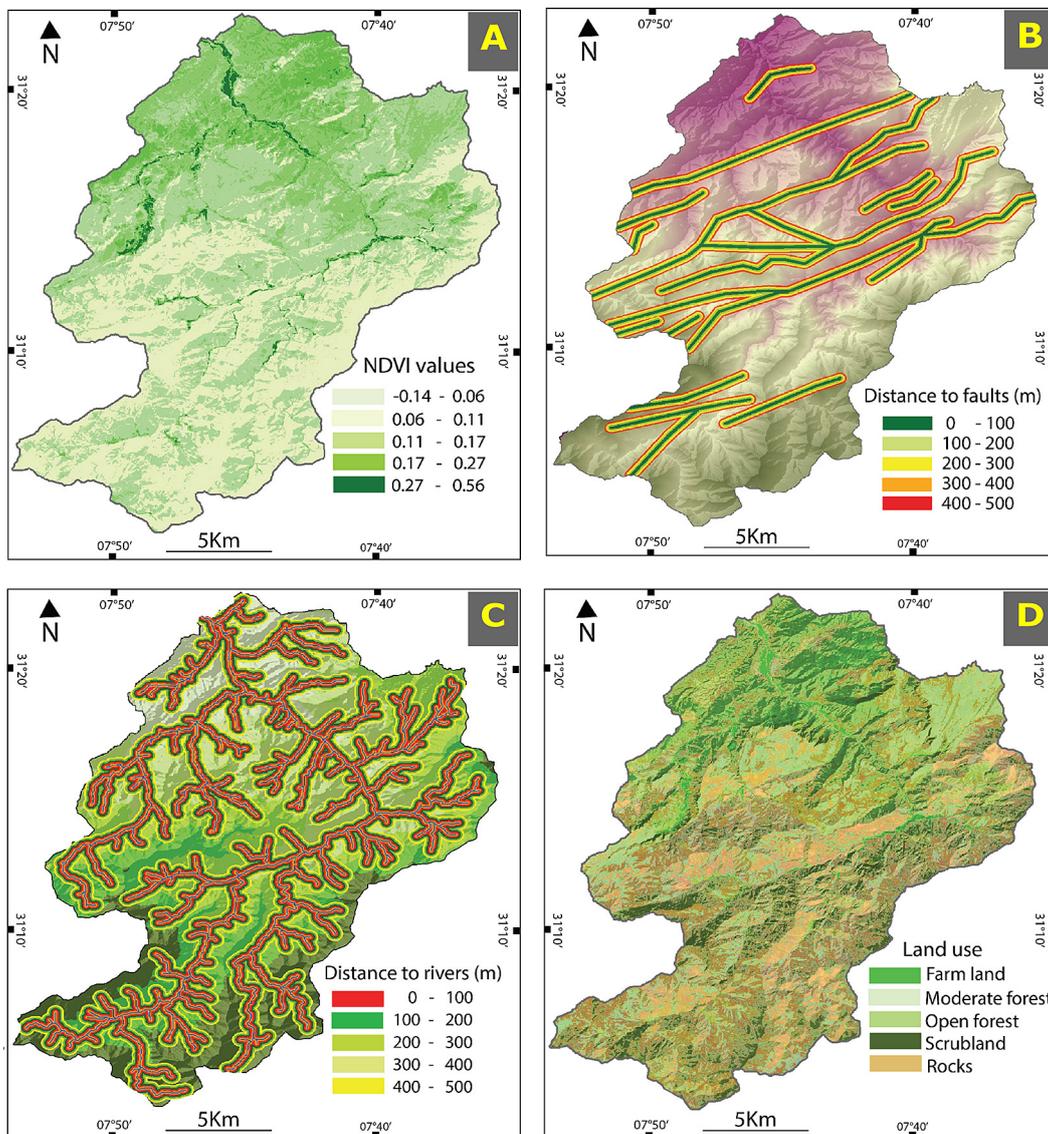


Figure 4. (a) NDVI map, (b) map of distances to faults, (c) map of distances to rivers, and (d) land use map of the Ourika watershed

Ourika region, they play a crucial role in triggering landslides by increasing the slope along tectonic escarpments (as in the case of the Amlouggui landslide). Indices are assigned based on the distance from major faults in the watershed (Fig. 4b).

#### Distance to rivers

The distance from rivers and their tributaries (Fig. 4C) is a crucial element in assessing the risk of water erosion. In our study, the proximity to rivers was divided into five categories, ranging from 0–100 m, 100–200 m, 200–300 m, 300–400 m, to 400–500 m.

#### Land use

The relationship between land use, land cover, and water erosion is significant. Areas with lower vegetation density experience more frequent erosion than those with dense vegetation cover. The land use map was used to generate a thematic layer (Figure 4D), which was then divided into five categories: agricultural land, moderate forest, open forest, shrubland, and rocky areas.

#### Data preparation

Mapping water erosion involves creating maps of various factors (Fig. 3 and 4) that control this natural risk. Subsequently, a general hazard map is generated by combining the different maps and inventory using scripts in R-studio. The initiation of water erosion depends on several factors, including topography, vegetation, precipitation, and agricultural practices, among others. Other factors may come into play, but with a relatively lesser impact, such as human activity and erosion at the base of slopes. These parameters do not govern the hazard in the same way. Among them, there are static parameters (non-variable over time), changing parameters (variables over time), and triggering parameters. The combination of the primary parameters for each hazard determines its degree of susceptibility. The machine learning models used in this study are based on eight (8) essential factors (Fig. 3 and 4): slope, lithology, precipitation, distance to faults, and distance to rivers, altitudes, NDVI, and land use.

#### Inventory map

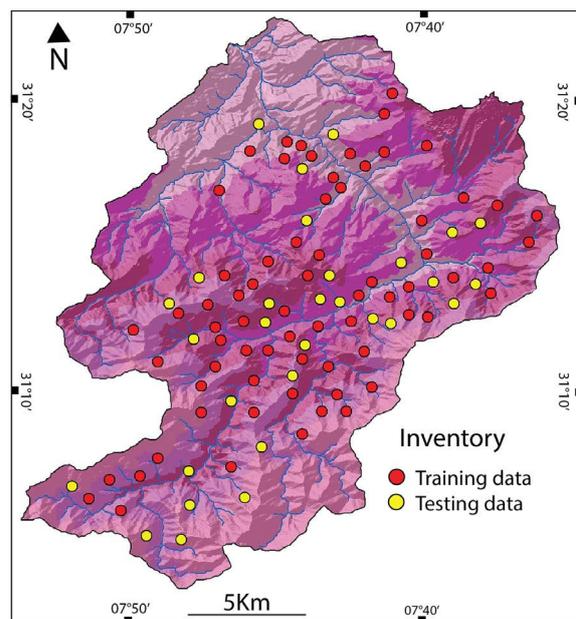
Locations of water erosion have been identified based on areas previously susceptible to erosion, and historical flood events have been

collected from satellite imagery and field data. Information layers representing the conditioning factors have an impact on the reliability of susceptibility mapping.

It is widely recognized and accepted that future erosion events and landslides tend to occur under similar conditions to previous erosions. The inventory map (Fig. 5) is crucial for susceptibility modeling as it establishes the relationship between high-risk areas and contributing factors to the hazard.

The distribution of previous erosion areas (inventory map) was created using various data sources, including historical record, satellite image analysis, and field surveys, showing where events occurred in the watershed region. The table below lists the data sources used to generate the water erosion inventory map (Table 1). These occurrences were selected as inventory and used with unique conditioning elements for susceptibility modeling. The inventory was conducted by identifying approximately 100 water erosion areas.

In general, according to (Tehrany et al., 2014), there are no strict rules for choosing the percentages of training and test data sets. The resulting list was randomly divided into 70 random locations to run the model, and the last 30 points were used to validate the model's sensitivity. Historical data on water erosion were assessed and interpreted using medium-resolution images (Landsat images acquired on the dates: 2010, 2015, and



**Figure 5.** The inventory data of water erosion used to test and validate the models

**Table 1.** Overview of data collected and used in the present study

Data type	Data source	Spatial data or map	Resolution
Monthly rainfall	Tensift agency data	Rainfull distribution of different return periods	Daily data
Historical spatial flood	Google earth images/ field survey	Flood inventory	15 m
Digital elevation model	SRTM-DEM	Slope, elevations, and river network	30 m
Geological data	Marrakech geological	Geological units	1/200.000
	Map		
Satellite images	Landsat 8 Oli Tirs	Mapping of geomorphological units, NDVI	30 m and 15 m
		Landuse	
		Verify the erosion locations after the occurred events	
Field investigations	Water erosion inventory and damaged areas. Satellite images	Eroded and damaged areas of previous events	-

2020). Risk areas were identified by comparing satellite data and field surveys. Using R scripts, the location points were randomly divided into two groups. Naimi and Araújo (2016) explain that the random partitioning method involves randomly dividing inventory points into training and validation data sets.

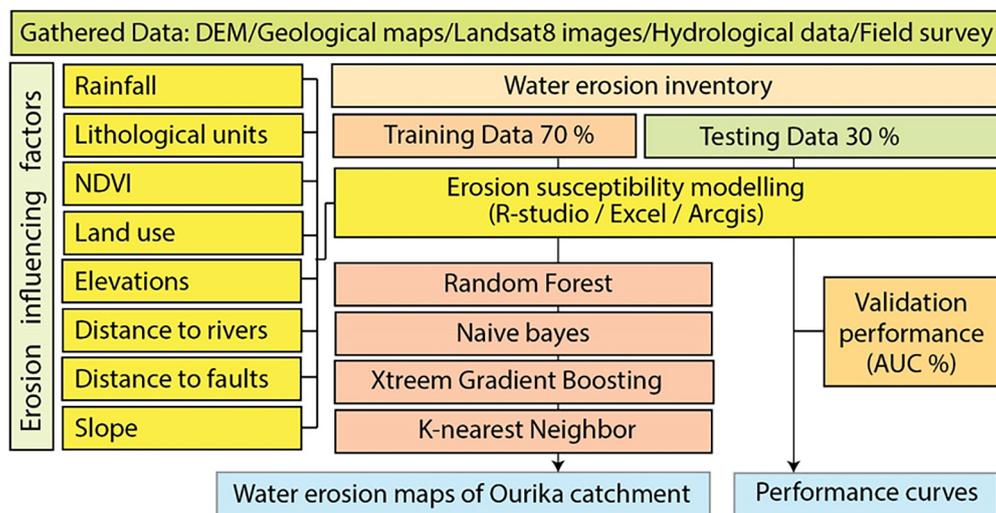
The literature suggests that a commonly used ratio for data set partitioning is 70% for training and 30% for validation.

**Data processing**

Studies focused on modeling susceptibility to erosion involve a combination of data and methodologies. This includes the use of remote

sensing data, GIS tools, and hydrological analyses. The goal is to improve our understanding of water erosion and its effects on the environment in the Ourika watershed. The integration of various techniques and tools allows for a more comprehensive assessment of event behavior and enables the development of effective erosion management strategies.

The methodology, as illustrated in Figure 6, adopted in this research comprises four steps: (1) Identification of historical erosion locations in the study area. (2) Selection and classification of relevant factors with an associated map for each variable based entirely on available data sources. (3) Production of sensitivity maps by applying four models, including RF, naive-bayes (NB), extreme



**Figure 6.** Framework of the data and modeling steps leading to the mapping of susceptibility to water erosion in the Ourika watershed

gradient boosting (XGB), and K-nearest neighbor (KNN). (4) Validation of results using the receiver operating characteristic (ROC) curve method to compare different models and field surveys.

Software such as ArcGIS, QGIS, and RStudio is used to process the data, generate thematic maps of sub-factors, and produce the final sensitivity maps. Statistical measures are performed using Excel.

### Predictive capacities

In this study, four models were used to determine the importance of variables impacting water erosion in the Oued Ourika watershed. The results, presented in Figure 7, which shows 8 causative factors and their average values in order of importance, are as follows: slopes with an average value of 0.78, vegetation index with an average value of 0.31, altitudes with 0.24, precipitation with 0.20, lithological units with 0.19, distance from rivers with a value of 0.17, land use with 0.16, distance from faults with 0.13. The higher the value, the more significant the factor is in impacting erosion in the Oued Ourika watershed.

The most significant factors affecting the occurrence of risk are slopes, NDVI, altitudes, and precipitation, in that order. The remaining factors, in order of importance, are lithology, distance from rivers, land use, and distance from faults. Therefore, the analysis of average values suggests that all of these factors positively contribute and can be integrated into the modeling of water erosion in the High Atlas of Marrakech.

### The models used in this study

1. Random forest (RF) – the random forest algorithm, initially introduced by (Avand et al., 2019), is a machine learning technique that uses an ensemble of decision trees to classify input datasets. Random Forest has gained significant attention due to its ability to achieve excellent classification results while maintaining fast processing speeds. In the Random Forest, a random subset of features is selected at each step of the output prediction, and the outputs are weighted based on the votes obtained. The final classification is determined by a majority vote among the outputs of individual decision trees (Ghorbanzadeh et al., 2019; Liu et al., 2019). This ensemble approach improves prediction accuracy and overcomes the uncertainty issues associated with using a single decision tree (Valdez et al., 2017). In the context

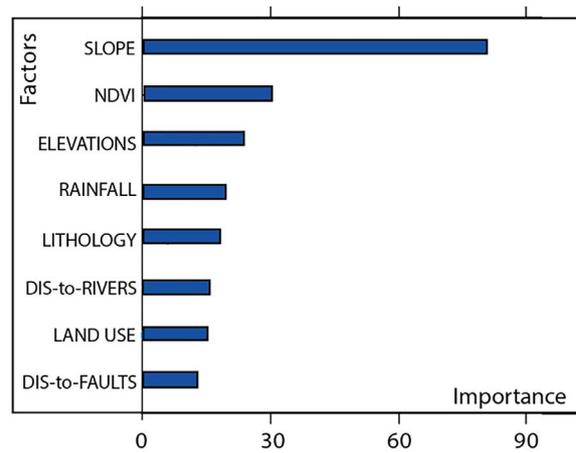


Figure 7. Predictive capabilities of the 8 conditioning factors

of susceptibility mapping, the Random Forest is considered a leading non-parametric ensemble learning method. Training the Random Forest model involves determining the maximum number of trees, the number of variables used in the splitting search, and the variant of the sampling process (Chen et al., 2016). In this analysis, the maximum number of trees was set to 500. The Random Forest model incorporates both the first and second training options for the splitting search. The use of a large number of trees and a diverse set of variables contributes to obtaining high variability and improves the accuracy of susceptibility mapping using Random Forest.

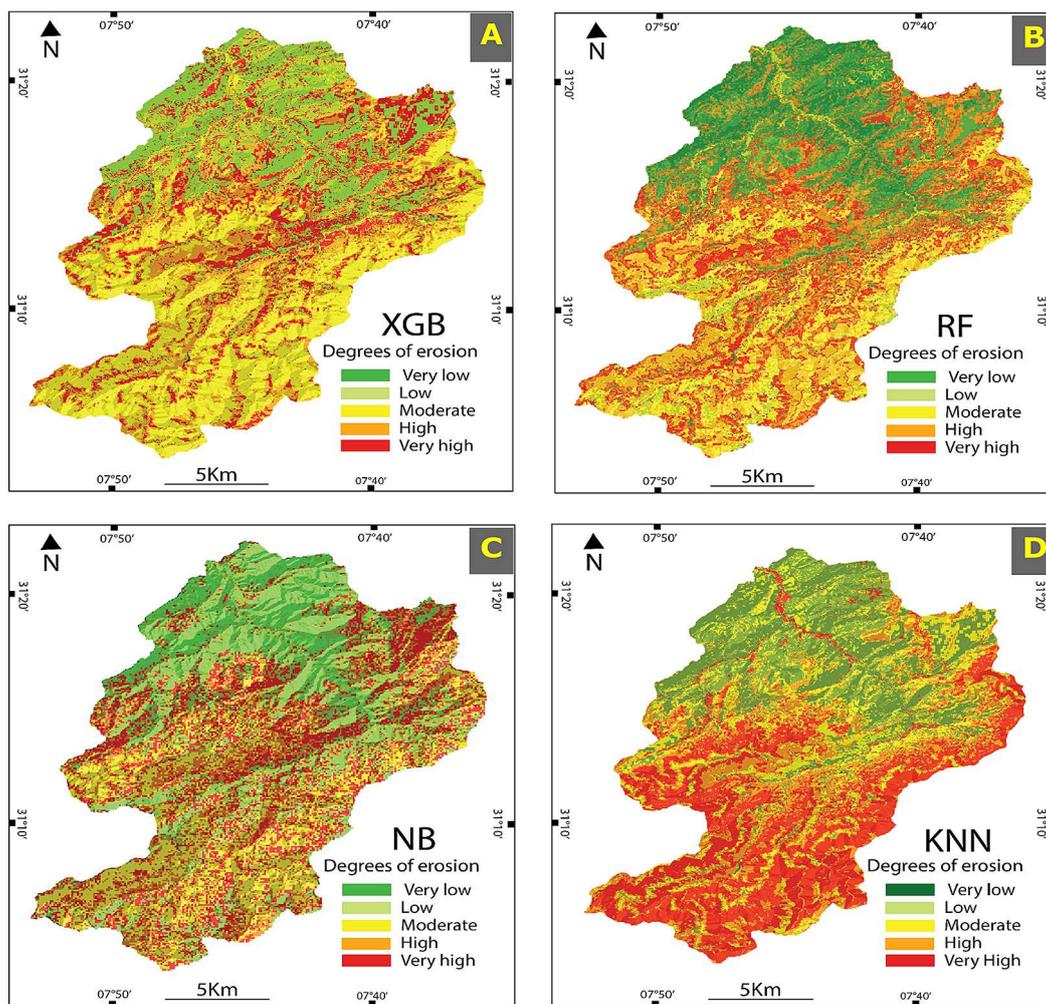
2. Naive bayes – a naive bayes classifier is a classification system based on Bayes' theorem, which assumes that all attributes are completely independent given the output class, a concept known as the conditional independence assumption (Talei et al., 2010). The main advantage of the NB classifier is that it is straightforward to construct without requiring complicated iterative parameter estimation procedures (Froude et al., 2018). Additionally, the NB classifier is robust to noise and irrelevant attributes. This method has been successfully applied in many fields (Beven et al., 1979). In a given context, where an observation is composed of  $k$  attributes  $x_i$ ,  $i = 1, 2, \dots, k$ , where  $x_i$  represents the factors conditioning landslides, and  $y_i$ ,  $j =$  landslide or no landslide is the output class, the Naive Bayes classifier estimates the probability  $P_{y_j/x_i}$  for all possible output classes.

3. Extreme gradient boosting (XGBoost) – XGBoost, a popular gradient boosting framework, is a machine learning model that incorporates tree pruning techniques and effectively handles missing values. It aims to minimize a regularized objective function that includes a penalty term for model complexity. In XGBoost, trees are fit to residual probabilities, as mentioned in the work of (Wu et al., 2019). To do this, examples are grouped based on the similarity of their residuals, and the branching process proceeds accordingly. The equation defines a similarity score for each threshold, providing a measure of similarity between examples based on the chosen threshold.
4. K-Nearest neighbor (KNN) – the K-Nearest neighbor is a well-established and commonly used classifier in the field of artificial intelligence (AI), as mentioned by (Sun et al., 2022). It is a simple and non-parametric method of

classification. KNN operates by comparing a given test dataset with a similar training dataset, allowing categorization based on similarity (Avand et al., 2019). One of the main advantages of the KNN classification system is its simplicity. The user has the option to choose the number of neighbors, denoted as  $k$ , and the distance measure to use. More detailed information about the KNN algorithm can be found in the works of (Abraham et al., 2021).

## RESULTS AND DISCUSSION

The results showed that the susceptibility maps generated by the models used exhibited different overall patterns (Fig. 8), with the highest susceptibility levels observed in the southern part of the watershed. However, the KNN model displayed a highly detailed pattern, indicating high



**Figure 8.** Susceptibility maps to water erosion derived from XGB (a); RF (b); NB (c); and KNN (d) models of the Ourika watershed

susceptibility in the high-altitude regions of the watershed. These variations in the mapping results can be attributed to differences in modeling structures and parameters.

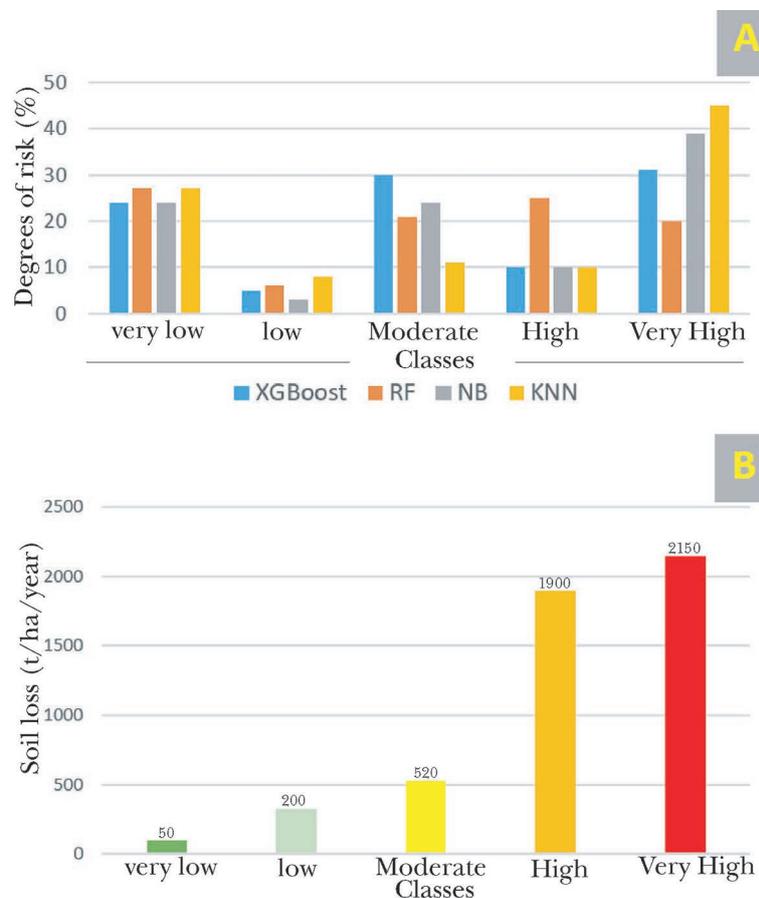
This study provides valuable insights for disaster management, enabling resource prioritization and the development of effective disaster response plans. Furthermore, the susceptibility maps for water erosion serve as essential tools for policymakers in land-use planning, including urban development, infrastructure construction, and land conservation. Additionally, these maps can serve as educational resources for communities, further raising awareness about water erosion risks, promoting preparedness, and reducing vulnerability. The results emphasize the importance of using and comparing multiple models and approaches when creating such maps.

According to the results from the susceptibility maps, the XGB, RF, and NB models reveal that erosion is particularly pronounced in the southern tributaries of the Oued Ourika basin, with a focus on the central part of the region. This

area is characterized by a combination of factors contributing to erosion, including the presence of various geological formations such as Triassic conglomerates, sandstones, and clays, as well as Precambrian mica schists and granites.

Furthermore, this area is characterized by moderate vegetation, sinuous river courses, slopes exceeding 35 degrees, altitudes exceeding 2191 meters, precipitation exceeding 600 mm, and a concentration of major faults. The KNN model, on the other hand, reveals a concentration of erosion at the southern end of the study area. This disparity in results can be explained by the differences in the parameters used in each algorithm. Specifically, the emphasis on the concentration of erosion in this area is attributed to several factors, including very high altitudes, frequent snowfall, heavy precipitation, and steep slopes.

Overall, the analysis of water erosion reveals that approximately 43% of the total area of the Oued Ourika watershed is exposed to a high to very high risk of erosion, as illustrated in Figures 8 and 9. This high to very high threat is mainly



**Figure 9.** Percentage distribution of water erosion susceptibility in the Ourika watershed: (a) distribution of degrees of risk for each model; (b) Soil loss histogram

concentrated in the central part of the watershed, in the high plateau area, as well as in the south-eastern end of the axial zone. However, moderate to low-level risks cover approximately 57% of the total area of the watershed, observed in the sub-altas zone, on the tops of the high plateaus, as well as in the northern and northwestern parts of the watershed.

In summary, these results highlight a concentration of water erosion in the Oued Ourika watershed, with a predominance of high-risk areas at the core of the region, while moderate to low-risk areas are found in other parts of the watershed. These observations have significant implications for the management and preservation of this geographical area, emphasizing the areas that require particular attention in terms of soil conservation and erosion prevention.

The evaluation of the effectiveness of different models in predicting erosion-prone locations was carried out using success rate curves and prediction rate curves. The validation of these techniques is considered crucial to ensure the quality of the applied approaches, as it provides the model with scientific credibility (Chung., 2003).

The validation process involves using a receiver operating characteristic (ROC) curve, which measures the model's fit based on the area

under the curve (Arabameri et al., 2019; Traoré et al., 2019; Basu et Pal., 2018). There are five categories of AUC values, ranging from excellent (0.90–1.00) to poor (0.50–0.60), to determine the accuracy of the models. The ROC curves of the four models used to produce erosion maps of the Oued Ourika are presented in Figure 10.

The AUC values for the prediction models, including KNN, XGB, NB, and RF, were found to be 0.81, 0.91, 0.87, and 0.82, respectively for prediction rates, and 0.84, 0.96, 0.98, and 0.87 for success rates. These results reveal that machine learning algorithms, especially NB and XG-Boost, demonstrate a higher degree of accuracy compared to other models. Therefore, these machine learning models can be considered reliable for further research and studies. Additionally, the achieved accuracy levels are deemed acceptable, and either of the models can be used for disaster preparedness. The erosion susceptibility maps can provide valuable information to engineers and planners for making informed decisions regarding land use and disaster planning.

Different models can also be applied in different regions, depending on the specific needs and circumstances. (Stefanidis, Stathis, 2013) and (Tehrany et al., 2015) conducted separate studies using different models, with Pradhan (2013)

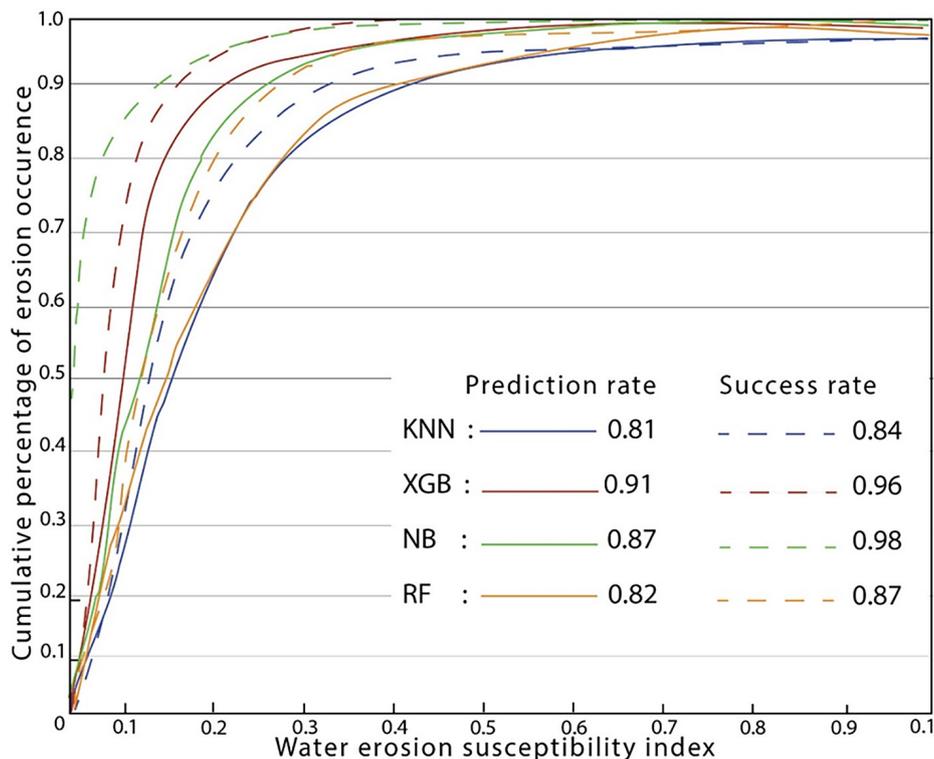


Figure 10. Validation of water erosion susceptibility maps using the area under the curve

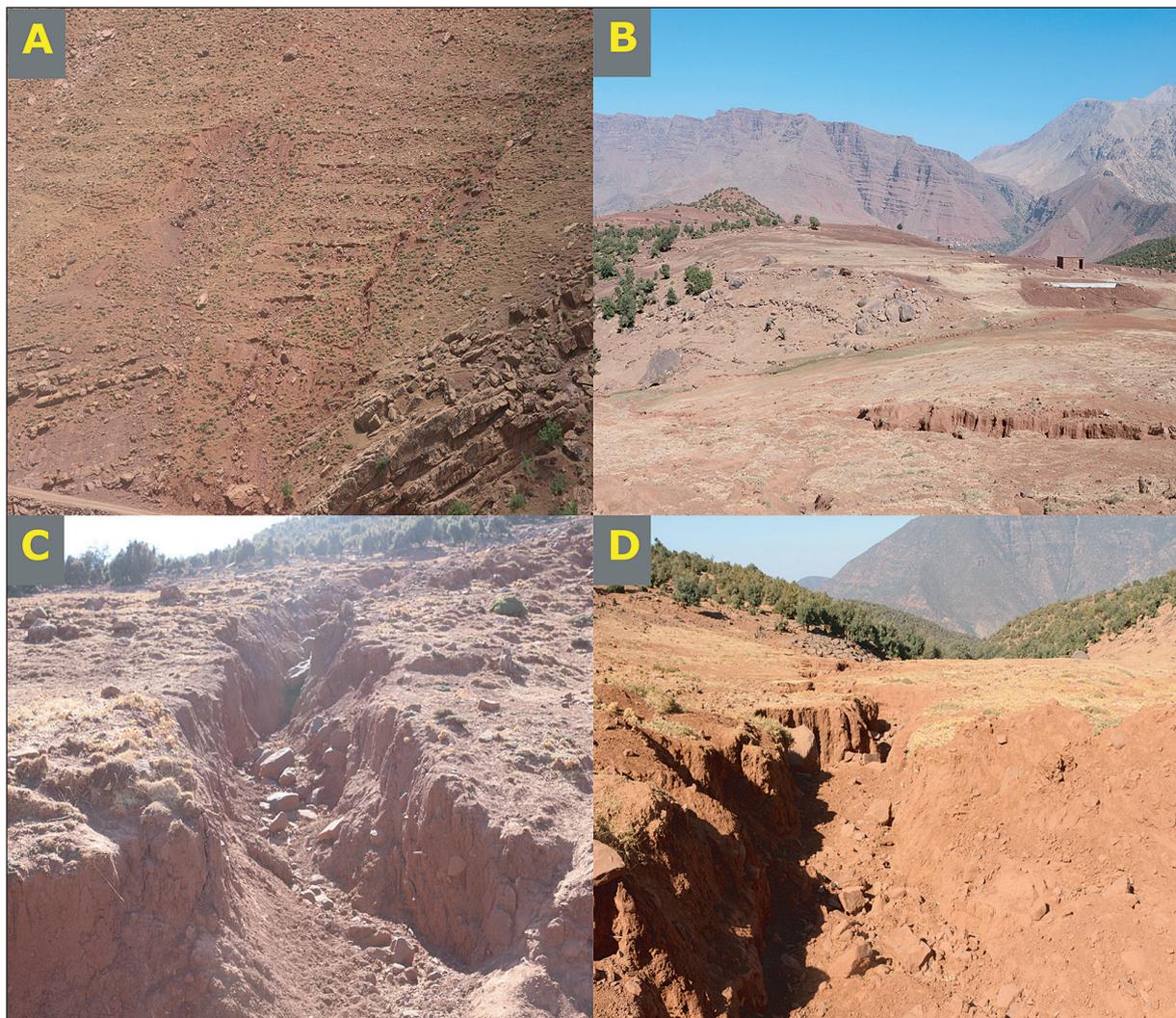
using support vector machine (SVM) for flood mapping in the Terengganu basin in Malaysia and Stefanidis and Stathis using analytic hierarchy process (AHP) in Kassandra, Greece.

Therefore, it is necessary to conduct comparative research to validate the performance of these models under similar conditions and compare their effectiveness (Goetz et al., 2015). Previous research on susceptibility mapping based on machine learning has mainly focused on determining the most suitable model to predict this natural risk (Pradhan, 2013; Chen, 2018).

This study emphasizes that the choice of susceptibility method can influence the outcome of erosion susceptibility mapping, even when using similar inventory data. This highlights the inherent uncertainty in predictions of natural risks by many techniques, despite their good accuracy in terms of AUC values.

## Field study

The field investigations have revealed a positive correlation with the results obtained through Machine Learning models. Within the Tarzaza, Tikhferte, and Oussane micro-watersheds (Fig 11), we identified signs of water erosion, manifesting as deeply incised gullies. These gullies are characterized by moderate dimensions, with widths and depths around two meters, extending up to a hundred meters in length at times. By analyzing these features, one can grasp the considerable extent of eroded soil in this region, potentially reaching several tons annually. The formation of these gullies is primarily influenced by soil lithological properties (especially the prevalence of clays), vegetation deficiency, steep topography, and the variable of precipitation.



**Figure 11.** Photos showing the phenomenon of water erosion in the study area: (a) Tarzaza micro-basin, (b) Tikhferte micro-basin, and (c) and (d) Oussane micro-basin

## CONCLUSIONS

In this article, we have presented an in-depth analysis of soil erosion utilizing Machine Learning techniques. The Ourika watershed stands out as a geographical area prone to erosion phenomena due to its hydro-geomorphological features.

The findings of this study reveal that all four models have effectively identified erosion-prone areas in the Oued Ourika watershed. The susceptibility maps created by these models have accurately depicted the spatial diversity of potentially vulnerable areas. The analysis of water erosion reveals that more than 43% of the total area of the Oued Ourika watershed is exposed to a high to very high risk of erosion with a predominance of high-risk areas at the core of the region, while moderate to low-risk areas are found in other parts of the watershed.

The results suggest that machine learning algorithms offer excellent performance in mapping soil erosion potential and natural risks. This advantage is due to these algorithms' ability to capture the relationships between relevant factors and manage the inherent uncertainties in statistical models (such as frequency ratio and logistic regression) that rely on historical inventory maps of the phenomenon. While all models employed in this study exhibited good to very good accuracy, it's essential to recognize that variations exist in the data quality and type needed for each model. The Naive bayes model consistently performed well, demonstrating a high level of accuracy that effectively addressed the study's limitations. Despite the intricacy and data transformation demands associated with machine learning models like XGBoost and KNN, they come highly recommended for assessing susceptibility to floods, erosion, and landslides due to their superior accuracy compared to bivariate statistical models. Furthermore, it's crucial to underscore that the quality of the results is primarily contingent on data availability and quality.

However, while complete elimination of natural or human-induced erosion factors is rare, if not impossible, measures and recommendations can be implemented to minimize soil erosion risks to the greatest extent possible. In the context of the Ourika watershed, it would be wise to prioritize erosion prevention measures, acting upstream to reduce the consequences of soil loss. It is also essential to promote sustainable agricultural practices aimed at protecting soil resources,

infrastructure, and the Ourika watershed community. Finally, a comprehensive and integrated assessment of erosion risk, considering environmental vulnerability and the pressures on it, is highly recommended.

## Acknowledgments

The authors would like to extend their deepest gratitude to the anonymous reviewers whose constructive comments, critiques, and suggestions have greatly contributed to enhancing the quality of this article.

## REFERENCES

1. Abraham M.T, Satyam R, Pradhan B., 2021. Factors affecting landslide susceptibility mapping: Assessing the influence of different machine learning approaches, sampling strategies and data splitting. *Land*, 10. <https://doi.org/10.3390/land10090989>
2. Alewell C., Borrelli P., Meusburger K., Panagos P., 2019. Using the USLE: Chances, challenges and limitations of soil erosion modelling. *Int Soil Water Conserv Res*; 7, 203–25. <https://doi.org/10.1016/j.iswcr.2019.05.004>
3. Arabameri A., Rezaei K., Cerda A., Conoscenti C., Kalantari Z.A., 2019. Comparison of statistical methods and multi-criteria decision-making to map the susceptibility to flood risks in northern Iran. *Sci. Total Environment*. 660, 443–458. doi:10.1016/j.scitotenv. 01.02.
4. Avand M., Janizadeh S., Naghibi S.A., Pourghasemi H.R., Khosrobeigi Bozchaloei S., Blaschke T.A., 2019. Comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. *Water*, 11, 2076.
5. Basu T., Pal S., 2018. Identification of landslide susceptibility zones in Gish River basin, West Bengal, India. *Georisk*, 12, 14–28. <https://doi.org/10.1080/17499518.2017.1343482>
6. Beven K.J., Kirkby M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. J.*, 24, 43–69.
7. Biron P., 1982. The permo-triassic of the Ourika region (high atlas of Marrakech, Morocco). Lithostratigraphy, sedimentology, tectonics and mineralization. 3rd cycle thesis, University of Grenoble, 170.
8. Campolo M., Soldati A., Andreussi P., 2003. Artificial neural network approach to flood forecasting in the River Arno. *Hydrol. Sci. J.* 48, 381–398.
9. Chen W., Peng J., Hong H., Shahabi H., Pradhan B., Liu J., Zhu A., Pei X., Duan Z., 2018. Landslide

- susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi-Province, China. *Sci. Total Environ.* 626, 1121–1135. <https://doi.org/10.1016/j.scitotenv>
10. Chen W., Chai H., Sun X., Wang Q., Ding X., Hong H.A., 2016. GIS-based comparative study of frequency ratio, statistical index and weights-of-evidence models in landslide susceptibility mapping. *Arab. J. Geosci.* 9, 204. <https://doi.org/10.1007/s12517-015-2150-7>
  11. Chung C.J.F., Fabbri A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. *Nat. Hazards*, 30, 451–472.
  12. Delannoy H., 1981. Some thoughts about the factor analysis of the correspondences of the monthly rainfall of Moroccan coastal stations. *Waters and Climates*. Grenoble. 165–178.
  13. Folharini S., Vieira A., Bento-Gonçalves A., Silva S., Marques T., Novais, J., 2023. Soil erosion quantification using machine learning in sub-watersheds of Northern Portugal. *Hydrology*, 10, 7. <https://doi.org/10.3390/hydrology10010007>
  14. Froude M.J., Petley D.N., 2018. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* 18, 2161–2181. <https://doi.org/10.5194/nhess-18-2161>
  15. Ghorbanzadeh O., Blaschke T., Gholamnia K., Meena S., Tiede D., Aryal J., 2019. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*, 11. <https://doi.org/10.3390/rs11020196>
  16. Goetz J.N., Brenning A., Petschko H., Leopold P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* 81, 111. <https://doi.org/10.1016/j.cageo.2015.04.007>
  17. Liu J., Xu Z., Chen F., Chen F., Zhang L., 2019. Flood hazard mapping and assessment on the Angkor world heritage site, Cambodia. *Remote Sens.*, 11, 98. <https://doi.org/10.3390/rs11010098>
  18. Liu Y., De Smedt F., 2005. Flood modeling for complex terrain using GIS and remote sensed information. *Water ResourManag.* 19, 605–624.
  19. Naimi B., Araújo M.B., 2016. sdm. A reproducible and extensible R platform for species distribution modelling. *Ecography*, 39, 368–375. <https://doi.org/10.1111/ecog.01881>
  20. Nefly M., 1998. The Precambrian crystallophylian massif of the Ourika (high atlas of Marrakech, Morocco): example of a gneissic dome of diapiric origin. State thesis, Casablanca (Morocco), 165.
  21. Ouhammou A., 1986. Research on the staging of vegetation in the Ourika basin (central high atlas, Morocco). Thesis III° cycle, Cadi Ayyad University, Fac. Sciences Marrakech, 181.
  22. Padarian J., Minasny B., and McBratney A.B., 2020. Machine learning and soil sciences: a review aided by machine learning tools, *SOIL*, 6, 35–52, <https://doi.org/10.5194/soil-6-35-2020>
  23. Pradhan B.A., 2013. Comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* 51, 350–365. <https://doi.org/10.1016/j.cageo.2012.08.023>
  24. Proust F., 1961. Stratigraphic, petrographic and structural study of the Eastern Block of the Ancient Massif of the High Atlas (Morocco). Thesis, Montpellier, France, 272.
  25. Proust F., 1973. Stratigraphic, petrographic and structural study of the Eastern Block of the Ancient High Atlas Massif. Morocco. Memo Note. Serv. Geol. Morocco, 254, 15–53.
  26. Renard K.G., Foster G.R., Weesies G.A., McCool D.K., Yoder D.C., 1997. Predicting soil erosion by water: a guide to conservation planning with the Revised Universal Soil Loss equation (RUSLE). Washington, D.C.
  27. Sahour, H., Gholami, Vazifedan, V., Saeedi, M., Saeedi, S., 2021. Machine learning applications for water-induced soil erosion modeling and mapping. *Soil and Tillage Research.* 211. 105032. <https://doi.org/10.1016/j.still.2021.105032>
  28. Saidi M.E., Daoudi L., Aresmouk M.E., Fnguire F., Boukrim S., 2010. The floods of the Wadi Ourika (High Atlas, Morocco): Extreme events in a semi-arid mountain context. *Comunicações Geológicas*, T. 97, 113–128.
  29. Saidi M.E., 1994. Genesis and propagation of floods in a sub-arid environment: Example of the wadi-Souss (Morocco). *Bulletin of the Association of French Geographers*, volume 1, 94–111, Paris.
  30. Sarkar S., Kanungo D.P., Patra A.K., Kumar P., 2008. GIS based spatial data analysis for landslide susceptibility mapping, *Journal of Mountain Science*, 5, 52–62.
  31. Sarkar S., Kanungo D.P., Mehrotra G.S., 2008. Landslide hazard zonation: a case study in Garhwal Himalaya, India. *Mountain Research and Development* 15, 301–309.
  32. Stefanidis S., Stathis D., 2013. Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). *Nat. Hazards*, 68, 569–585. <https://doi.org/10.1007/s11069-013-0639-5>.
  33. Sun Z., Sandoval L., Crystal-Ornelas R., Mousavi S.M., Wang J., Lin C., Cristea N., Tong, D., Carande W.H., Ma X., Rao Y.A., 2022. Review of earth artificial intelligence. *Comput. Geosci.* 105034.
  34. Talei A., Chua L.H.C., Quek C.A., 2010. Novel

- application of a neuro-fuzzy computational technique in event-based rainfall–runoff modeling. *Expert Syst. Appl.* 37, 7456–7468. <https://doi.org/10.1016/j.eswa.2010.04.015>
35. Tehrany M.S., Pradhan B., Mansor S., Ahmad N., 2014. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena* 2015, 125, 91–101. <https://doi.org/10.1016/j.catena.2014.10.017>
36. Tourani A.I., 1988. Etude stratigraphique, sédimentologique et ichnologique du Carbonifère de l’Atlas de Marrakech (région Ourika-Zat, Maroc). Thèse Université. Université de Marrakech, Maroc, 150.
37. Traoré M.K., Zacharewicz G., Duboz R., Zeigler B., 2019. Modeling and simulation framework for value-based healthcare systems. *Simulation*, 95, 481–497. <https://doi.org/10.1177/0037549718776765>
38. Valdez M.C., Chang K.T., Chen C.F., Chiang S.H., Santos J.L., 2017. Modelling the spatial variability of wildfire susceptibility in Honduras using remote sensing and geographical information systems. *Geomatics, Natural Hazards and Risk*, 8, 876–892.
39. Wachal D.J., Hudak P.F., 2000. Mapping landslide susceptibility in Travis Country, Texas, USA. *Geo-Journal*, 51, 245–253.
40. Wischmeier W.H., Smith D.D., 1978. Predicting rainfall erosion losses: a guide to conservation planning with Universal Soil Loss Equation (USLE). Washington, D.C.: Agriculture Handbook 537, Department of Agriculture.
41. Wu S., Chen J., Zhou W., Iqbal J., Yao L.A., 2019. Modified Logit model for assessment and validation of debris-flow susceptibility. *Bull. Eng. Geol. Environ.* 78, 4421–4438.