

Article

Integration of High-Accuracy Geospatial Data and Machine Learning Approaches for Soil Erosion Susceptibility Mapping in the Mediterranean Region: A Case Study of the Macta Basin, Algeria

Hamza Bouguerra ^{1,*}, Salah Eddine Tachi ², Hamza Bouchehed ², Gordon Gilja ³, Nadir Aloui ²,
Yacine Hasnaoui ², Abdelmalek Aliche ², Saâdia Benmamar ² and Jose Navarro-Pedreño ^{4,*}

- ¹ Water Resources and Sustainable Development Laboratory, Department of Geology, Faculty of Earth Sciences, Badji Mokhtar—Annaba University, P.O. Box 12, Annaba 23000, Algeria
 - ² Department of Hydraulics, Laboratoire de Recherche des Sciences de L'eau, National Polytechnic School, 10 Rue des Frères OUDEK, El Harrach, Algiers 16200, Algeria; salah_eddine.tachi@g.enp.edu.dz (S.E.T.); h.bouchahed@ensh.dz (H.B.); nadir.aloui@g.enp.edu.dz (N.A.); yacine.hasnaoui@g.enp.edu.dz (Y.H.); abdelmalek.aliche@g.enp.edu.dz (A.A.); saïda.benmamar@g.enp.edu.dz (S.B.)
 - ³ Department of Hydrosience and Engineering, Faculty of Civil Engineering, University of Zagreb, Kaciceva 26, HR-10000 Zagreb, Croatia
 - ⁴ Department of Agrochemistry and Environment, University Miguel Hernández of Elche, 03202 Elche, Spain
- * Correspondence: hamza.bouguerra@univ-annaba.dz (H.B.); jonavar@umh.es (J.N.-P.)

Abstract: Erosion can have a negative impact on the agricultural sustainability and grazing lands in the Mediterranean area, especially in northern Algeria. It is useful to map the spatial occurrence of erosion and identify susceptible erodible areas on large scale. The main objective of this research was to compare the performance of four machine learning techniques: Categorical boosting, Adaptive boosting, Convolutional Neural Network, and stacking ensemble models to predict the occurrence of erosion in the Macta basin, northwestern Algeria. Several climatologic, morphologic, hydrological, and geological factors based on multi-sources data were elaborated in GIS environment to determine the erosion factors in the studied area. The conditioning factors encompassing rainfall erosivity, slope, aspect, elevation, LULC, topographic wetness index, distance from river, distance from roads, clay mineral ratio, lithology, and geology were derived via the integration of topographic attributes and remote sensing data including Landsat 8 and Sentinel 2 within a GIS framework. The inventory map of soil erosion was created by integrating data from the global positioning system to locate erosion sites, conducting extensive field surveys, and analyzing satellite images obtained from Google Earth through visual interpretation. The dataset was divided randomly into two sets with 60% for training and calibrating and 40% for testing the models. Statistical metrics including sensitivity, specificity, accuracy, and the area under the receiver operating characteristic curve (ROC) were used to assess the validity of the proposed models. The results revealed that machine learning and deep learning, as well stacking ensemble techniques, showed outstanding performance with accuracy over 98% with sensitivity 0.98 and specificity 0.98. Policy makers and local authorities can utilize the predicted erosion susceptibility maps to promote sustainable use of water and soil conservation and safeguard agricultural activities against potential damage.

Keywords: erosion susceptibility; GIS-machine learning; land use-land cover; land degradation; Macta basin (Algeria); satellite images



Citation: Bouguerra, H.; Tachi, S.E.; Bouchehed, H.; Gilja, G.; Aloui, N.; Hasnaoui, Y.; Aliche, A.; Benmamar, S.; Navarro-Pedreño, J. Integration of High-Accuracy Geospatial Data and Machine Learning Approaches for Soil Erosion Susceptibility Mapping in the Mediterranean Region: A Case Study of the Macta Basin, Algeria. *Sustainability* **2023**, *15*, 10388. <https://doi.org/10.3390/su151310388>

Academic Editor: Matteo Gentilucci

Received: 5 May 2023

Revised: 21 June 2023

Accepted: 27 June 2023

Published: 30 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Water erosion is defined as the process in which surface runoff forms channels, establishing a dominant flow zone, and removes the soil from these restricted areas to great depths over short periods of time [1–3]. Aggressive soil erosion can lead to formation of

deep gullies [4,5]; gully erosion contributes to soil loss rates ranging from 10 to 94 percent of the total sediment output resulting from water-induced erosion. Soil erosion causes important land degradation and significant damage to agricultural lands on a large scale as well as to construction sites such as bridges, roads, and villages [6–10]. Gully erosion is generated by several factors, mainly rainfall caused by extreme climatic events like heavy rains in short times or thunderstorms [6,11]. Besides the runoff, other factors contribute to soil erosion, such as vegetation, soil properties, subsurface flow [11] overland water flow [12], and wind [13,14]. Anthropogenic activities such as unsustainable agriculture practices [15–17], deforestation [18,19], and road construction [20,21] can also contribute to increased soil erosion rates.

Recently, gully erosion has attracted growing interest because of its negative impact on the environment. To study this phenomenon, various factors should be considered, including topographical factors (e.g., elevation, slope, and aspect), hydrological factors (e.g., rainfall, distance to river, and stream density), soil characteristics (e.g., soil type and structure), geological factors (e.g., lithology), and environmental factors (e.g., distance to road) [4,22].

According to the scholarly literature cited in [23], the development of gully erosion inventory maps has been investigated using a range of mathematical models, including both bivariate and multivariate approaches. In addition, the utilization of GIS and remote sensing has been found to be effective in this context. Among the used statistical techniques, there are empirical models such as USLE, RUSLE, and MUSLE that were designed to estimate long-term average annual soil loss caused by water erosion from specific field slopes in range of land-use applications and management systems (i.e., crops, rangeland, recreational areas, etc.) [24–27].

Despite the established efficacy of the methods, estimating soil loss and sediment discharge is a time-consuming process, which is complicated by the influence of multiple factors [28,29]. Thus, in the last decade some researchers have adopted a variety of mathematical, machine learning, and data-mining approaches that have since been built to map and analyze gully erosion and associated processes [4,29–33].

Many scientists describe a number of GIS-based models that have been utilized for erosion susceptibility mapping, including the frequency ratio model [34,35], weights of evidence [29,34], linear and logistic regression [31,36,37], and the analytical hierarchy process [38,39].

Avand et al. compared a random forest and a K-nearest neighbor classifier for gully erosion susceptibility mapping in the Hobaturck watershed in Iran [40]. The study was carried out on 242 gully erosion locations and 12 conditioning factors. The ROC-AUC results indicated that the random forest algorithm performed better than the K-nearest neighbor. Rainfall, altitude, and distance from the river were identified as the most influential parameters in mapping gully erosion susceptibility in the study. These findings highlight the significant role of these factors in shaping the spatial distribution and intensity of gully erosion. Saha et al. studied the vulnerability of fully erosion using MLP, MLF bagging, and ML bagging methods in the Hinzolo river basin in India [41]. The study indicated that the use of a hybrid method improved the accuracy of MLP models.

These findings indicate that elevation has the strongest influence on gully erosion susceptibility, followed by rainfall and NDVI. On the other hand, geology, soil type, and sediment transportation index (STI) were found to have relatively less influence on gully erosion susceptibility. Arabameri et al. [42] evaluated the accuracy of using a hybrid artificial intelligence model in mapping gully erosion susceptibility based on the use of 18 conditioning factors in the Kohpayeh-Sagazi river basin in Iran. The results indicated that the hybrid GE-XG boost model has better performance than the other benchmark solution. According to their results, and using gain ratio information, the highest information gain ratios are the soil depth, soil type, TWI, lithology, and NDVI. Additionally, the moderate values of information gain ratio were elevation, plan curvature, slope, TPI, and drainage density.

Another work was carried out by Yang et al. [43] to investigate gully erosion mapping in complex terrain in the Mizhigou watershed, China, by using the random forest (RF), gradient boosted decision tree (GBDT), and extreme gradient boosting (XGBoost) algorithms separately or combined with the statistical weight of evidence model (WoE). During this work, 14 conditioning factors were considered for mapping erosion susceptible area, and the results showed that slope gradient, land use, and altitude were the main factors. The results indicated that the area under curve (AUC) values of different models used were higher than 0.925 indicating high prediction of the models; it was also shown that the AUC values for marching learning regression methods were higher without the WoE model. The XGBoost algorithm performed better than the RF and GBDT, and the main factors for gully mapping were slope gradient, land use, and altitude. Goetz et al. [44] compared statistical and machine learning models for regional-scale landslide susceptibility modeling in Lower Austria. They used spatial K-fold cross-validation and variable importance assessment to evaluate the models. Random forest and bundling classification techniques exhibited the best predictive performances, with overall median estimated AU-ROC differences ranging from 2.9 to 8.9 percentage points. Slope angle, surface roughness, and plan curvature consistently emerged as highly influential variables. This evaluation framework offers valuable guidance for selecting appropriate modeling techniques for landslide susceptibility mapping.

These results provide valuable insights into the key factors driving gully erosion and can guide future research and management strategies. However, it is important to note that different machine learning models can yield varying results in terms of feature importance. To address this gap and improve the accuracy of mapping areas susceptible to soil erosion, the application of a stacking method can be explored. By integrating multiple machine learning models and considering their respective feature importance, the stacking method has the potential to enhance the accuracy of mapping and prediction for susceptible areas prone to soil erosion. Further research in this direction can contribute to the development of more robust and reliable erosion susceptibility models.

Different studies worldwide showed the high erosion rate in semi-arid regions, which shows the importance of this phenomenon and its impact on water resources and land development (Table 1).

Table 1. Statistics of erosion rate in semi-arid regions around the world.

Area	Method	Erosion Rate	Reference
Tunisian Dorsal, Tunisia	Reservoir siltation measurement	Average rate 14.5 t ha ⁻¹ year ⁻¹ Maximum rate 36.4 t ha ⁻¹ year ⁻¹	[45]
Andipatti Taluk, India	RUSLE	Average rate 5.26 t ha ⁻¹ year ⁻¹ Maximum rate 95.54 t ha ⁻¹ year ⁻¹	[46]
Madhya Pradesh, India	RUSLE	Average rate 6.42 t ha ⁻¹ year ⁻¹ Maximum rate 179.9 t ha ⁻¹ year ⁻¹	[47]
Machados County, Brazil	USLE	Average rate 8.11 t ha ⁻¹ year ⁻¹ Maximum rate above 20 t ha ⁻¹ year ⁻¹	[48]
Seybouse basin, Algeria	RUSLE	Average rate (20 y): 13 t ha ⁻¹ year ⁻¹ Maximum rate: over 50 t ha ⁻¹ year ⁻¹	[49]

In Algeria, water erosion poses a significant threat to the country's agricultural productivity, leading to soil loss, depletion of fertilizers, and nutrient degradation. Moreover, the decline in water reserves in Algerian dams is a major concern due to sedimentation caused by erosion in watersheds and accumulation in reservoirs. Bathymetric surveys conducted by the National Agency for Dams and Transfers (ANBT 2005) on 31 dams revealed an average loss of 980 million cubic meters in storage capacity, equivalent to approximately 13% of the initial capacity. This research holds implications for land management, environmental planning, and decision-making processes in Mediterranean regions, while offering the potential to identify high-risk erosion areas and implement targeted control measures.

It can also support policymakers in formulating sustainable land-use policies to mitigate soil erosion and promote effective land management practices.

The objectives of this study were: (i) To map erosion prone areas in the Macta basin Northwestern Algeria (ii) Investigate and compare the feature importance of several machine learning and deep learning techniques, while highlighting the significance of using stacking ensemble techniques to improve soil erosion mapping. (iii) To evaluate maximum conditioning factors that significantly control the erosion phenomenon and elaborate on the minimum factors needed to avoid over-fitting problems that could occur in the modeling of soil erosion.

2. Materials and Methods

2.1. Studied Area

The study area comprises the basin of Macta, which includes the wilayas of Mascara, SidiBel Abbes, Mostaganem, Tlemcen, Oran, and Saida. It is located between latitudes 34°34' and 35°79' N and between longitude 1°06' W and 0°56' E. The Macta basin is bounded by the Mediterranean Sea to the north, by the Tighenif plain and the Saida mountains to the east, by the highlands of Ras El Ma and the lowlands of Maalif to the south, by the plain of Telagh, the mountains of Tessala, and the mountains of Tlemcen to the west (Figure 1). The area covers 14,458 km², and the perimeter is 717 km. It has a semi-arid climate [50]. It is composed of two tributaries of Mediterranean rivers, the Mekerra wadi to the west and the El Hammam wadi to the east. The topography and altitude vary in such a way that soil erosion occurs in most of the area. For this study, a total of 400 points (200 points of erosion location and 200 points of non-erosion locations) were randomly selected in the studied area.

The analysis of annual average precipitation recorded at rainfall stations in the Macta watershed, from 1980 to 2015, reveals variations in precipitation distribution across the entire basin. The precipitation values range from 289 mm to 486 mm, with an average of approximately 378 mm. The central area of the basin, characterized by higher altitudes, receives higher rainfall compared to the lower areas in the south and north, as altitude increases relative to sea level.

The spatial distribution of land use in the Macta basin exhibits several categories of land occupation. Grasslands dominate the basin, covering 8749 km², which represents 60.51% of the total surface area. Croplands also occupy a significant portion, covering 4376 km², accounting for 30.27% of the basin. Forest formations cover approximately 4% of the total area and are primarily found in regions with moderately rugged terrain. The remaining portion of the basin comprises unproductive lands such as rocky areas, bare lands, water bodies, and urbanized zones (Table 2).

Table 2. Distribution of land use/land cover area in the Macta basin.

LULC Class	Area (km ²)	Area (%)
Grasslands	8749	60.51
Croplands	4376	30.27
Forest	607	4.20
Urbanization	432	2.99
Bare lands	280	1.94
Water Bodies	13	0.09
Total	14,458	100

The main soil types in the Macta basin are Calcisols, Luvisols, Vertisols, and Leptosols according to the latest version of the World Reference Base for soil resource (WRB). These soil types occupy a substantial area, accounting for 42.31%, 27.99%, 14.26%, and 11.23% of the total surface area, respectively. Cambisols and Kastanozems, on the other hand, constitute only 3.28% of the basin's area. Therefore, it can be concluded that the majority of the soil in this watershed exhibits moderate resistance to water erosion (Table 3).

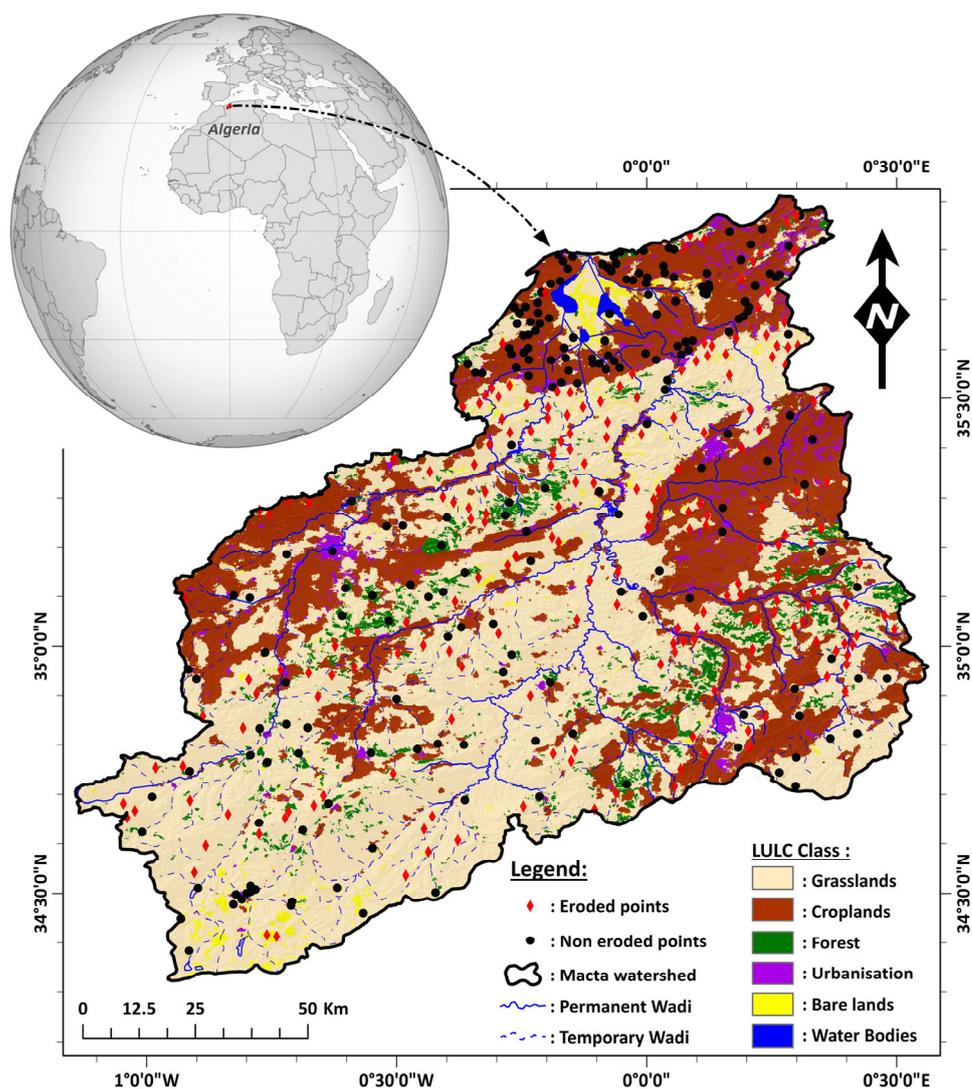


Figure 1. Location map of the study area.

Table 3. Distribution of soil types in the Macta basin.

Soil Types	Area (km ²)	Area (%)
Calcisols	6116.75	42.307
Luvisols	4047.23	27.993
Vertisols	2062.43	14.265
Leptosols	1623.92	11.232
Cambisols	325.16	2.249
Kastanozems	149.21	1.032
Phaeozems	51.76	0.358
Regosols	50.89	0.352
Fluvisols	15.61	0.108
Acrisols	14.46	0.100
Solonchaks	0.58	0.004
Total	14,458	100

2.2. Machine Learning Methods

The methods used in this study are based on artificial intelligence (AI). The techniques used for modeling gully erosion include Adaptive boosting (AdaBoost), Categorical boosting (CatBoost), Convolutional Neural Network (CNN), the stacking method, and geospatial data processing.

- AdaBoost is machine learning technique initiated by Freund and Schapire [51]; many algorithms are derived from AdaBoost either for classification or applied to regression [52,53]. The AdaBoost algorithm is an iterative approach that seeks to construct a robust classifier through the combination of weak learners generated in prior iterations. The algorithm modifies the learning pattern in accordance with the error returned by the weak learners, with the ultimate goal of achieving a final hypothesis that exhibits low error relative to a given distribution [51,54].
- CatBoost is new gradient boosting based on decision tree [55], and its characteristic is that it requires small data training comparing to other models and deals with different data formats [56]. The CatBoost model employs the generation of random permutations of the dataset and gradients to inform the selection of an optimal tree structure, thereby enhancing the robustness of the algorithm and mitigating overfitting [57].
- CNN is a type of deep learning architecture that imitates the natural visual perception of living beings [58]. CNN comprises several layers, including the convolutional layer, non-linearity layer, pooling layer, and fullyconnected layer. While the convolutional and fullyconnected layers are parameterized, the pooling and non-linearity layers are not. Among the various forms of artificial neural networks, CNN is particularly remarkable [59]. As reported in the literature, the name “Convolutional Neural Network” (CNN) is derived from the mathematical operation of convolution, which involves the multiplication of matrices [60].
- The stacking method was implemented in this study to improve the performance of developed predictive model. By leveraging ensemble learning methods, such as the stacking method, a meta-model is used to combine predictions generated by several base models [61]. Stacking, which is also referred to as stacked generalization, is a widely used ensemble learning technique that combines multiple base models to improve prediction accuracy. Here, three different algorithms were used as base models: CNN as a powerful deep learning architecture that has the ability to capture spatial features from input data and CatBoost and AdaBoost to combine weak learners to create a strong learner. Categorical boosting is specifically designed for categorical data, while Adaptive boosting is a general-purpose method that can be used for both categorical and numerical data.

2.3. GIS and Geospatial Data Processing

In this section, we provide a detailed description of the GIS environment and geospatial data preprocessing methods employed in our research (Figure 2). The utilization of GIS allowed us to effectively elaborate and process the collected data, which encompassed various climatological, morphological, hydrological, and geological factors. The integration of these factors within the GIS framework facilitated the generation of thematic maps that visually represented their spatial distribution across the study area.

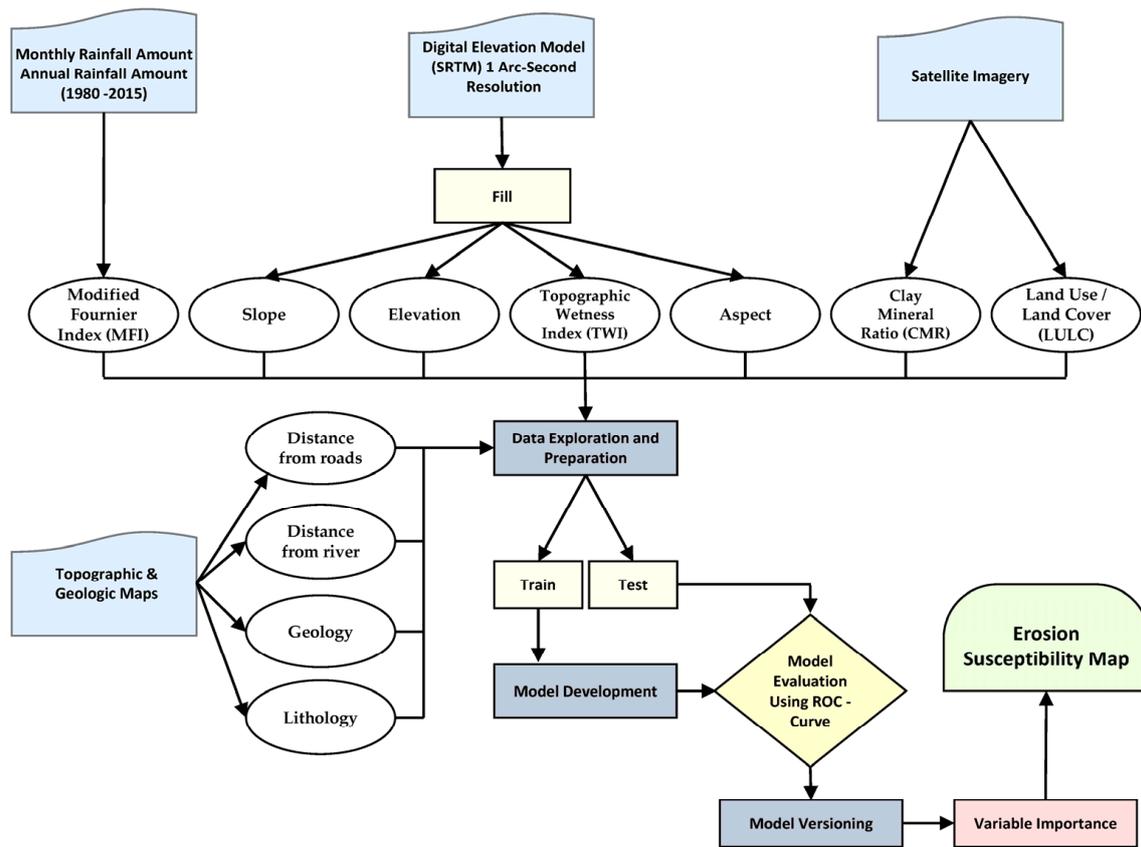


Figure 2. Flowchart methodology for erosion susceptibility mapping.

Slope and aspect were calculated in the GIS Environment using a Digital Elevation Model (DEM). The DEM was imported into GIS to initiate the process. The “Slope” and “Aspect” were utilized to derive information about the steepness and orientation of the terrain. The topographic wetness index was calculated using rasters of flow accumulation and slope. TWI provided information on landscape wetness based on topographic characteristics and supported hydrological assessments.

The modified Fournier index values were calculated using the equation of (71) as explained in the Section titled ‘selection of variables’.

The distance from a river and the distance from roads were calculated by uploading the rivers and roads in the studied area and then calculating the distance of each pixel point to these targets using Euclidean distance measurement. The results can be visualized and used for various geospatial analyses, providing valuable information about proximity to rivers and roads separately.

The calculation of the clay mineral ratio was carried out on the selected spectral bands of Landsat 8 (Table 4). The Clay Index formula, defined as $(\text{Band 7} - \text{Band 5}) / (\text{Band 7} + \text{Band 5})$, is applied to quantify clay mineral abundance in the selected area.

The Raster Calculator in GIS software was a powerful tool used for performing mathematical operations on raster layers. It provided the capability to create new raster layers by applying various mathematical expressions or formulas to existing raster layers.

Table 4. Geospatial data sources.

Parameter	Source	Link	Spatial Resolution	Temporal Periods
MFI	National Agency for Hydraulic Resources (ANRH)	-	-	1980–2015
Soil Class	Soil Grids	https://soilgrids.org/ (accessed on 4 May 2022)	190 m	2016
LULC	Esri Sentinel-2	https://livingatlas.arcgis.com/landcover/ (accessed on 4 May 2022)	10 m	2022
DEM	USGS Earth Explorer	https://earthexplorer.usgs.gov/ (accessed on 4 May 2022)	1 Arc-Second	2014
Satellite Imagery	Landsat 8 OLI/TIRS	https://earthexplorer.usgs.gov/ (accessed on 4 May 2022)	30 m	05/2022
Topographic and Geologic Maps	National Institute of Cartography	-	1/50,000	-

To ensure the reliability and accuracy of our analysis, we adopted a comprehensive approach to geospatial data preprocessing. The initial step involved the extraction of relevant factors using an inventory map that classified sites as either land degradation or non-degradation. This inventory map provided the foundation for gathering essential information for subsequent analysis.

Following factor extraction, we conducted a pre-treatment analysis of the statistical data. This involved applying appropriate statistical techniques to evaluate and preprocess the extracted values. The goal was to ensure data quality, consistency, and suitability for further modeling.

Subsequently, we employed classification modeling techniques to develop models capable of predicting erosion occurrence based on the identified factors. Machine learning algorithms were integrated within the GIS environment to facilitate this modeling process. By leveraging the power of machine learning, we aimed to capture the complex relationships and patterns between the factors and erosion susceptibility (Figure 2).

To assess the performance of the developed models, we utilized performance criteria as outlined in our methodology. These criteria allowed us to evaluate the accuracy and reliability of the models in predicting erosion susceptibility within the study area. The performance evaluation provided valuable insights into the strengths and limitations of the models and their applicability for practical use.

Furthermore, we employed feature importance analysis to determine the relative significance of each factor in contributing to the erosion susceptibility models. This analysis allowed us to prioritize and weigh the importance of different factors in the final susceptibility maps generated by the models. By identifying the most influential factors, we aimed to enhance the accuracy and effectiveness of the models' predictions.

The integration of GIS and machine learning techniques in our research enabled us to leverage the spatial data and develop models that accurately predicted erosion susceptibility. This combination facilitated a comprehensive analysis of the study area, providing valuable insights into the factors influencing erosion occurrence. The incorporation of GIS and machine learning techniques showcased the potential for their synergy in addressing complex environmental issues and supporting informed decision-making processes.

2.4. Model Evaluation

The assessment of model accuracy in this study involved the evaluation of both goodness-of-fit, which reflects how well the model fits the calibration subset, and predictive performance, which measures the model's ability to accurately predict the validation subset. To quantify model performance, we employed the area under the curve and

receiver operating characteristic (AUC-ROC) metrics. As reported by Williams [62], a confusion matrix was generated to compare the final model's predictions with the actual outcomes of the observations (Table 4). The actual observations were represented in the rows of the matrix, while the columns corresponded to the model's predictions, and the cell counts indicated the numbers of observations for each variable.

As stated in the literature [63], a confusion matrix is typically a square matrix of size $n \times n$ that is used to evaluate the performance of a classifier by comparing its predicted and actual classifications. Here, n represents the number of different classes. For instance, a confusion matrix for binary classification with $n = 2$ typically has four entries, each with a specific meaning as shown in Table 5:

Table 5. Example of a confusion matrix for $n = 2$.

	Predicted Negative	Predicted Positive
Actual negative	a	b
Actual positive	c	d

where a = number of correct negative predictions, b = number of incorrect positive predictions, c = number of incorrect negative predictions, d = number of correct positive predictions.

Accuracy is a metric that measures the overall performance of a classifier and indicates the fraction of total samples that are correctly classified [56]. The formula to calculate accuracy (τ) and the error (ϵ) are as follows:

$$\tau = (a + d)/(a + b + c + d), \quad (1)$$

$$\epsilon = (b + c)/(a + b + c + d), \quad (2)$$

This study used ROC curve analysis as another statistical technique to evaluate the goodness-of-fit and prediction performance of each model [64]. The ROC curve shape provides an indication of the accuracy of a model, where a curve closer to the upper left corner ($AUC = 1$) represents higher accuracy, while a curve closer to 0.5 indicates model inaccuracy [65]. According to AUC values, the predictive performance was classified as acceptable for $AUC \geq 0.7$, excellent for $AUC \geq 0.8$, and outstanding for $AUC \geq 0.9$ [66]. To assess the robustness of the models, the positive and negative calibration and validation datasets with optimal pixel size were changed three times [33].

2.5. Selection of Variables

In order to select the most relevant variables for the classification model, a feature selection process was carried out considering 19 variables (elevation, aspect, slope, curvature, hill-shade, stream density, distance from rivers, distance from roads, modified Fournier index, NDVI, topographic wetness index, topographic roughness index, sediment transport index, stream power index, LULC, soil type, geology, lithology, and clay mineral ratio). To avoid the double effect of the same factors on the modeling, the NDVI and sediment transport index (STI) variables were eliminated based on their high correlation with clay mineral ratio and stream power index (Table 6). Then, the top 11 variables were selected using feature importance scores obtained from a CNN (see Table 6). The CNN was trained on the input data and was able to learn the relevant features through a series of convolutional and pooling layers. The feature importance scores were then calculated by evaluating the impact of each feature on the model's accuracy. The 11 variables with the highest feature importance scores were retained for the final classification model. This approach ensures that only the most informative variables are included in the model for predicting soil erosion in the Macta basin.

Table 6. Feature selection using Convolutional Neural Network (CNN).

Ranking	Features	Correlation	Importance (%)
1	Slope		59.65
2	LULC		5.31
3	Lithology		4.36
4	TWI		3.68
5	MFI		3.55
6	Geology		3.00
7	D_F_Roads		2.49
8	CMR		2.25
9	D_F_Rivers		1.94
10	Elevation		1.89
11	Aspect		1.87
12	Stream_Den		1.64
13	HillShade		1.63
14	Soil_Type		1.59
15	NDVI	CMR = 76%	1.44
16	TRI		1.40
17	Curvature		1.07
18	SPI	STI = 82%	0.81
19	STI		0.43

To assess the susceptibility of the given area to erosion, a series of parameters and their relationship to the studied phenomena must be considered. It is worth noting that there is no conventional method for the selection of erosion conditioning factors. These vary from place to place depending on the study area and data availability.

In this work, a total of eleven conditioning factors are selected and mapped by using the Geographic Information System (GIS). The selected parameters are defined as follows:

Slope is a parameter that represents the degree of topographic change. The slope and water flow velocity are strongly related, where a higher slope increases the surface runoff velocity. Therefore, the risk of land erosion becomes more important. The slope map was determined using GIS, it was found that mild slopes ($<7^\circ$) are located in the extreme south, north, and east of the region, meanwhile the steepest slopes ($>12^\circ$) occur in the center of the basin. The south-west region is characterized by moderated slope values ($7\text{--}12^\circ$) (Figure 3).

Aspect is defined as the slope orientation. In our case the slope directions occur irregularly in the basin, which means that there is no privileged direction in any part of the basin. The Telagh high lands, Tessala Mountains, and Tlemcen mountains in the west of the basin all have slopes that are directed towards the north-east and towards the south-east. The Ghriss plain region and the central massif of the basin are characterized mainly by slopes oriented towards the north-west.

The elevation map shows that the studied region topography is divided into three sections. The first one (elevation < 260 m) is located in low coastal plains in the north. The second section (260 m $<$ elevation < 700 m) is in the center (Figure 3). The last section (elevation > 700 m) is located in the south, where we find the Tlemcen mountains (1412 m in Djebel Ouargla) and Dhaya Mountains (1455 m in Djebel Mezioud).

Land use/land cover (LULC) have been identified as factors that can impact runoff and soil loss [67,68]. Changes in land use and land cover can have a significant impact on erosion-prone areas, as they affect various hydrological processes such as infiltration, evaporation, evapotranspiration, and runoff. This can either accelerate or decelerate the erosion process in watersheds. In this study, the LULC map consists of several thematic maps, including water bodies, trees, flooded vegetation, crops, built-up areas, bare ground, snow/ice, and rangeland (Figure 3).

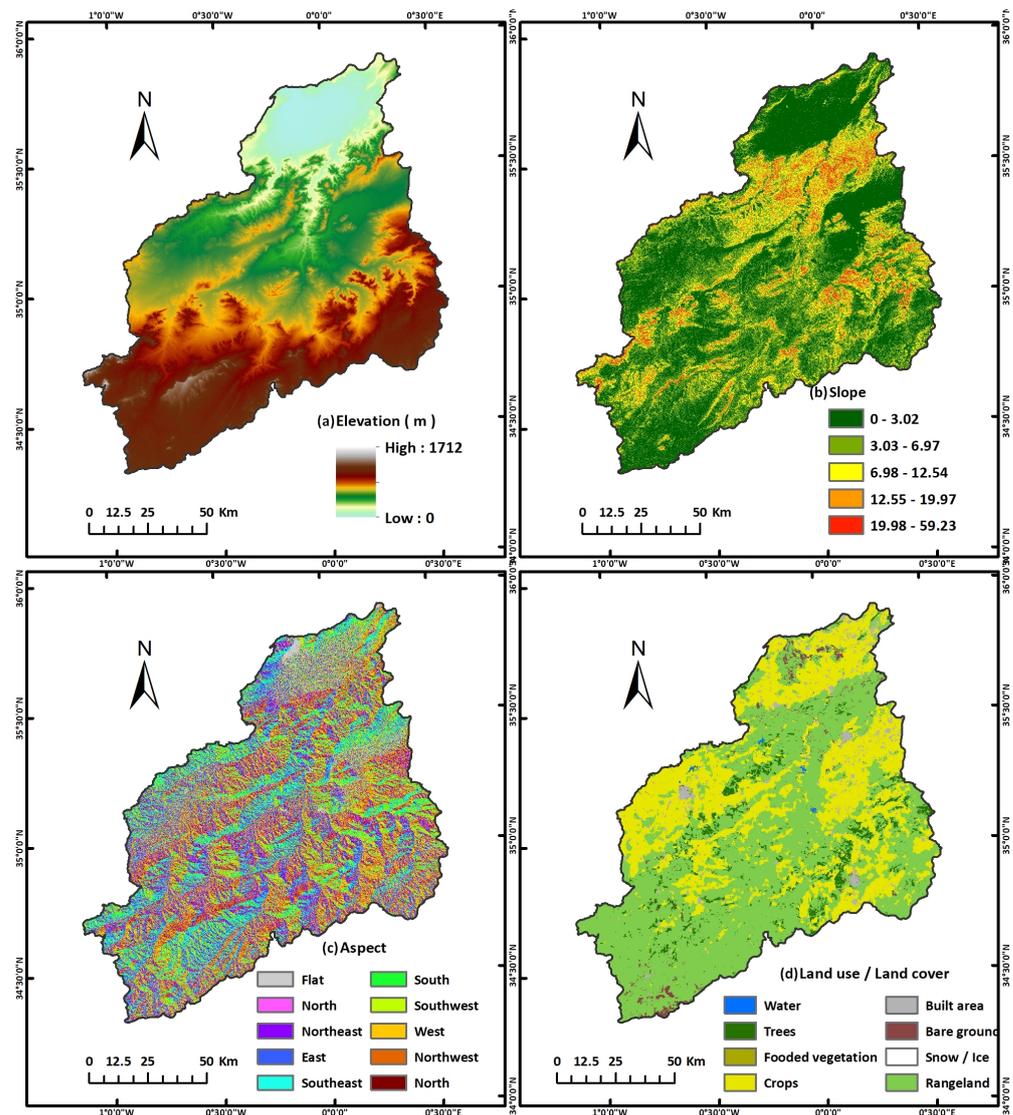


Figure 3. Conditioning factors: (a) elevation, (b) slope, (c) aspect, and (d) land use/land cover.

Topographic wetness index (TWI) is used to evaluate the impact of the topography on the hydrological process. The TWI map shows that low TWI values (3.018–8.4) are present in the central and south-east regions of the basin; therefore, a weak humidity is present in the two regions. Moderate TWI values ranging from 8.4 to 10.84 are present in north, south, and south-west of the basin, indicating the presence of an average humidity (Figure 4). A high humidity is detected in the outlet of and along the course of waterways due to the presence of runoff. The TWI values at this region varied from 10.84 to 25.20.

Modified Fournier index (MFI) has been demonstrated as a crucial factor in accurately estimating the R factor in areas that experience high-intensity rainfall events, which is essential for assessing the risk of soil erosion in the context of future changes in land use and climate, particularly under the Revised Universal Soil Loss Equation (RUSLE) framework [69]. The erosivity index in the current study was determined by utilizing average precipitation data collected from the national agency for hydraulic resources for one hundred and six (106) meteorological stations located in the Macta basin (Table A1). Analytical equations were employed to evaluate R factors based on the amount of rainfall. R values were calculated using the equations described by [70]. The calculated modified Fournier index ranges from 28 to 61, with the highest values located in the middle part of the basin from east to west (Figure 4). The lowest values of the index are found in the southern part of the Macta basin, where a semi-arid climate prevails.

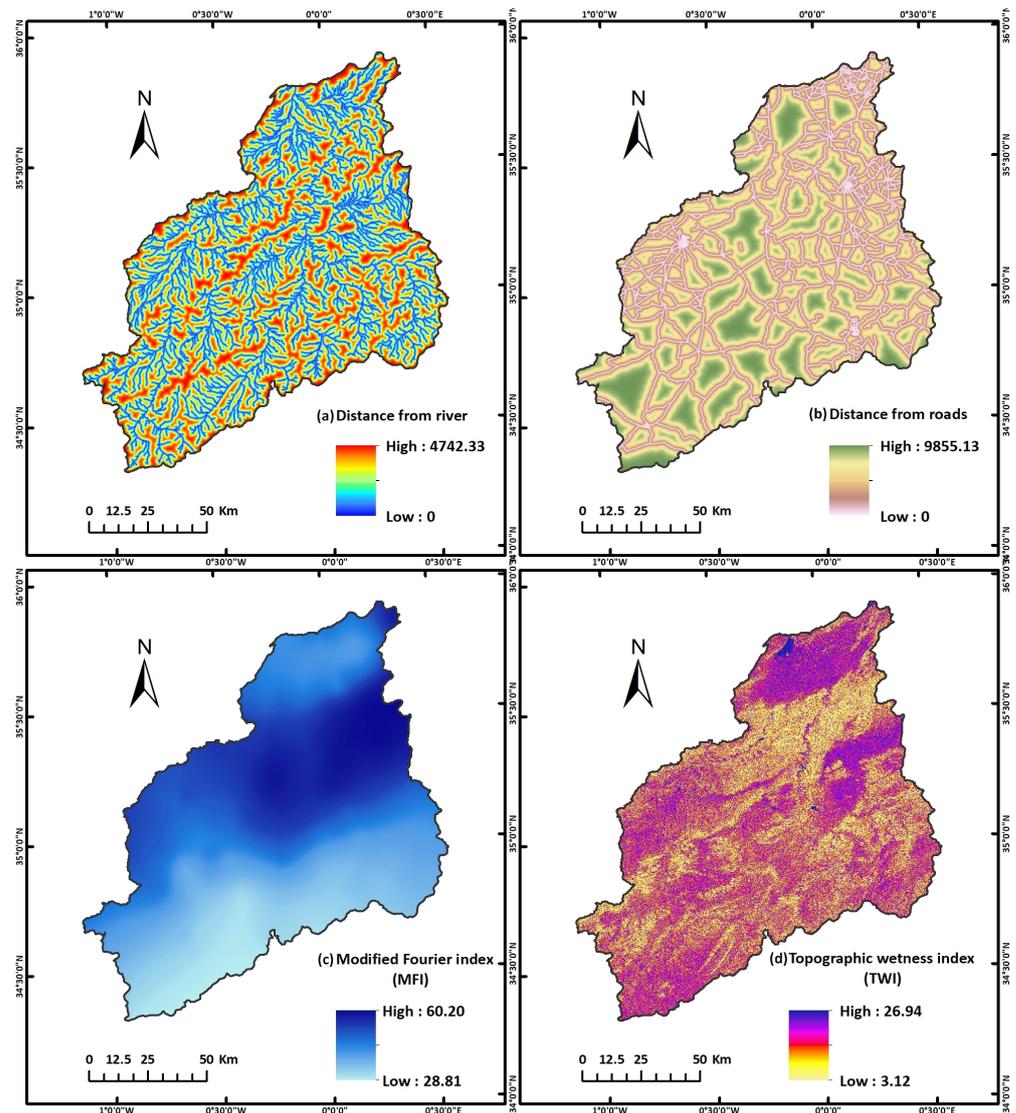


Figure 4. Conditioning factors: (a) distance from river, (b) distance from roads, (c) modified Fournier index, and (d) topographic wetness index.

Distance from roads, the proximity of a location to roads plays a crucial role in determining its susceptibility to erosion. This is because roads can hinder the absorption of water into the ground, leading to an increase in surface runoff and erosion-prone areas. As a result, areas situated closer to roads are more vulnerable to erosion due to reduced infiltration rates and faster runoff.

Distance from river can significantly impact the severity and extent of water-induced soil erosion. The distance from the river is considered by many researchers as a key factor in assessing erosion risk. When a location is situated close to a river, it becomes more vulnerable to soil erosion due to the increased water flow volume and velocity, which accelerate the process of erosion by flash floods. As a result, areas located in close proximity to a river are at a higher risk of experiencing soil erosion.

Lithology represents the geological composition of the region, characterized by a variety of quaternary formations with different lithological properties (Figure 5). The degrees of rock compaction and alteration, as well as the occurrence of fractures and joints in the subsurface or exposed rock, have a significant influence on the recharge of fractured aquifers [71–73]. The Wadi El Hammam basin is characterized by a diverse geology, with quaternary formations dominating the region. These formations have varying lithological characteristics, with the compactness and alteration state of the rock, as well as the presence

of diacalse and joint, playing a crucial role in recharging fractured aquifers. The Ghriss–Mascara plain, which is drained by Wadi Ain Fekane, is mainly composed of detrital formations such as marl-clay and sandy-clay with gravel passages. The region’s massifs are made up of carbonate rocks, including Cretaceous limestone in the Tessala and Beni-Chougrane mountains and limestone and/or Jurassic dolomites in the Tlemcen and Dhaya mountains. In the western zone of the region, horsts and grabens oriented ENE-WSW are present, extending from the Tlemcen Mountains to the Traras massif, with large normal faults bounding the compartments [74].

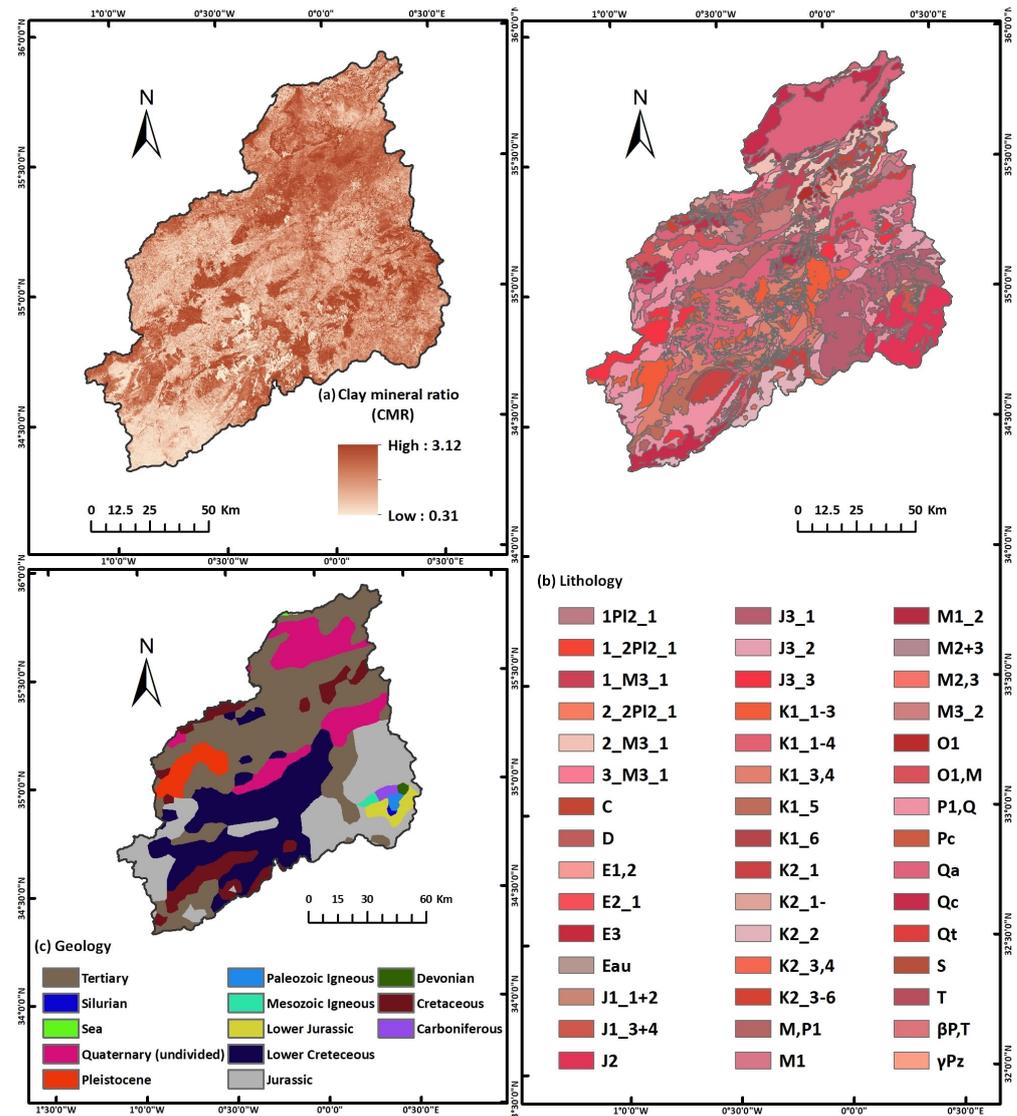


Figure 5. Conditioning factors: (a) clay mineral ratio, (b) lithology, and (c) geology.

Geology of the basin of Macta occupies the western part of the Tellian Atlas, encompassing a multitude of geological formations ranging from the primary age to the Quaternary, with marl and limestone facies predominating (Figure 5). Quaternary and Plio-Quaternary formations occupy depressions in the north and northeast as well as the hollows of valleys. The majority of Pliocene formations crop out to the west, while Upper Jurassic and Lower Cretaceous geological formations are found in the center, south, and southeast. The geological formations of the primary age appear in the south-eastern part, consisting mainly of schists and quartzites.

Clay Mineral Ratio (CMR) in soil has a significant impact on soil erosion. The presence of clay minerals in the soil increases its ability to retain water and reduce soil erosion

(Figure 5). A higher clay mineral ratio in soil can enhance its ability to resist erosion by slowing down the rate of water infiltration, thus minimizing the velocity of surface runoff. On the other hand, soils with a lower clay mineral ratio are more prone to erosion as they tend to have a higher rate of water infiltration and faster surface runoff velocity [75,76].

3. Results

The CatBoost, AdaBoost, CNN, and stacking ensemble methods were trained on 60% of the data in the training sample and evaluated on 40% of the data in the validation sample. The ROC-AUC curves for each method are presented in (Figure 6). Following this, the susceptibility values were classified into four levels of risk (i.e., very low, low, medium, and high) using the quantile classification method, as previously described in the literature [44]. The resulting risk map is shown in (Figure 7).

Observations indicate that the hybrid method based on the stacking ensemble technique, using the combination of CNN, AdaBoost, and CatBoost, exhibits the highest performance in terms of area under the curve (AUC), achieving a score of 98% (Figure 6). The individual models, CNN, AdaBoost, and CatBoost, exhibit AUC values of 97%, 96%, and 94%, respectively (Table 7). These results demonstrate that machine learning and deep learning techniques can be effectively used for predicting erosion-prone areas, yielding robust and high-performing results.

The AdaBoost model was employed to model erosion occurrence in selected pixels based on the influencing factors. The obtained results (Table 6) demonstrate that the topographic wetness index (TWI) has the highest impact on the erosion modeling in the Macta basin, with an influence percentage of 16%. This is followed by aspect, distance from river, clay mineral ratio, and modified Fournier index, with percentage influences of 14%, 13%, 12%, and 11%, respectively. In contrast, elevation, geology, distance from roads, slope, lithology, and land use/land cover had a lesser influence on the erosion modeling process in the studied region.

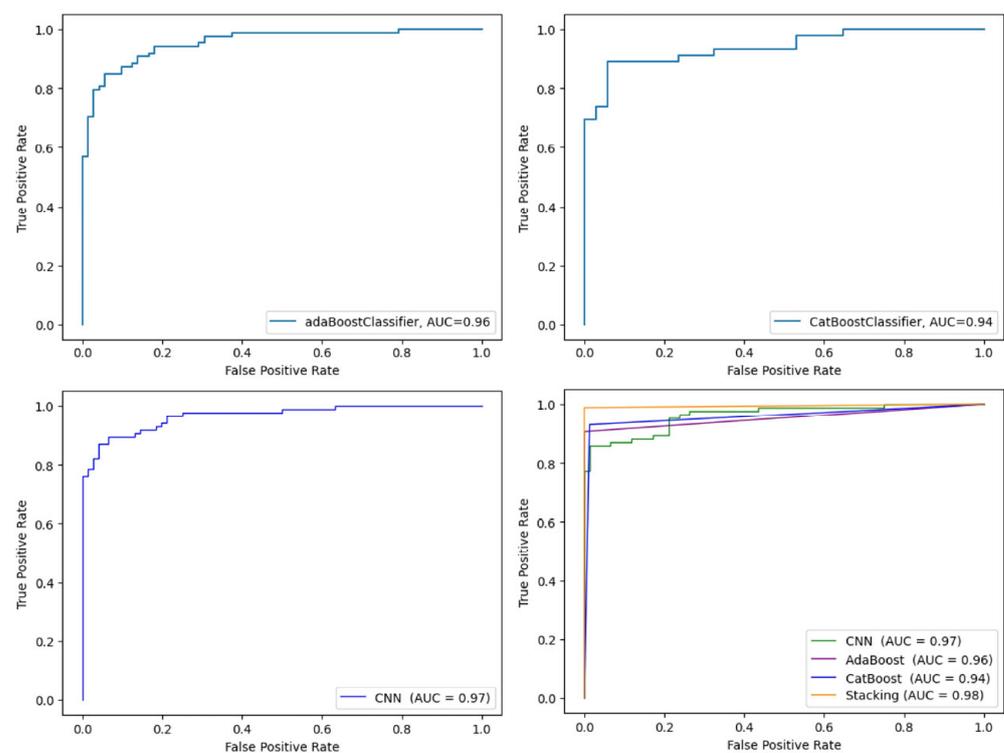


Figure 6. ROC-AUC curves.

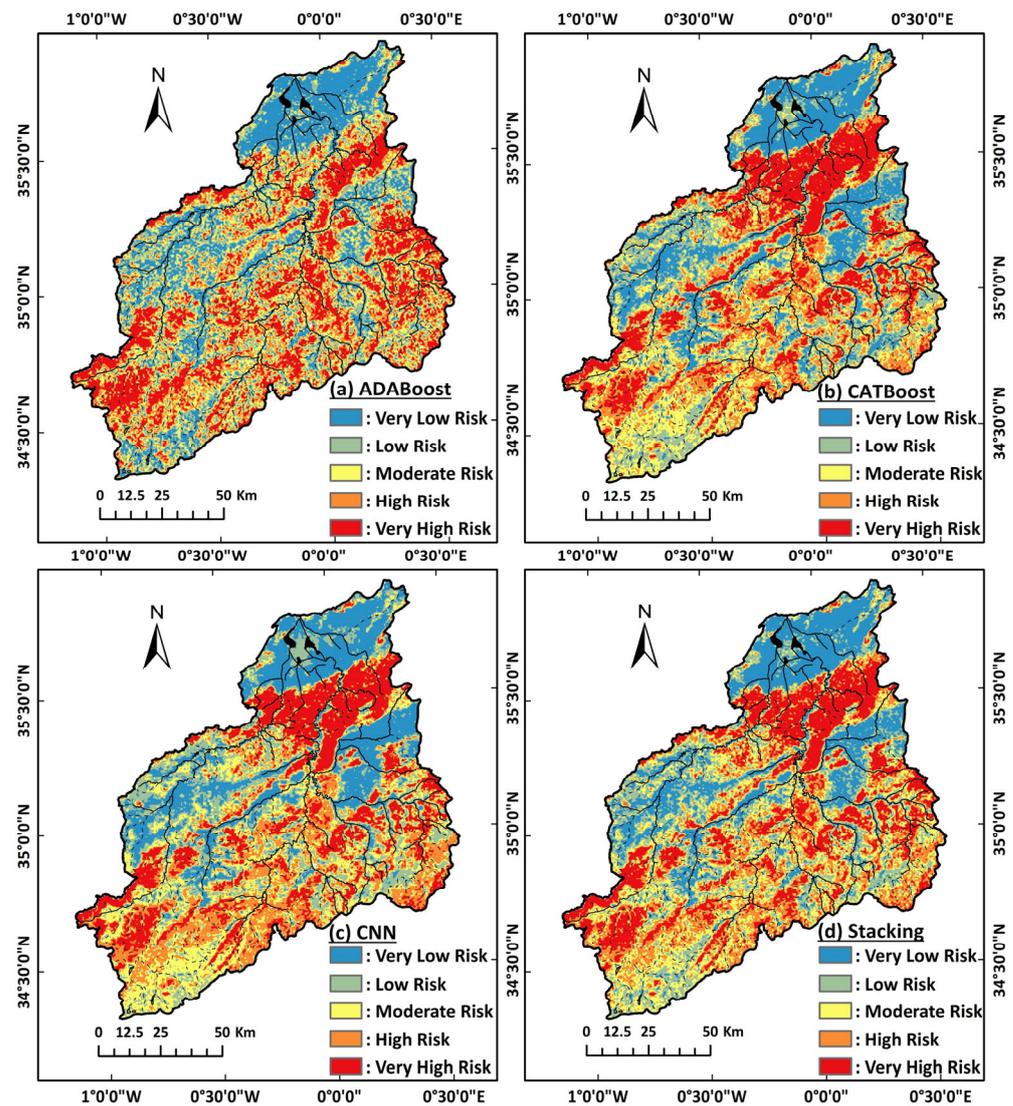


Figure 7. Erosion susceptibility maps using (a) AdaBoost, (b) CatBoost, (c) CNN, and (d) stacking ensemble.

Table 7. Evaluation metrics of the machine learning models.

Statistics	CatBoost	AdaBoost	CNN	Stacking
TP	82	79	69	78
TN	64	62	72	78
FP	4	9	4	2
FN	10	10	15	2
Sensitivity	0.89	0.89	0.82	0.98
Specificity	0.94	0.87	0.95	0.98
F1 score	0.92	0.89	0.88	0.98
Recall	0.89	0.89	0.82	0.98
Precision	0.95	0.90	0.95	0.98

The CatBoost model exhibited the lowest performance. The results were still considered outstanding in comparison to the other models. The obtained results (Table 6) indicate that the slope is the most influential parameter associated with erosion phenomena in this

model, with an importance value of 60%. This is followed by land use/land cover, modified Fournier index, distance from roads, and distance from river, with importance values of 9.17%, 5.36%, 4.39%, and 4.18%, respectively.

CNN exhibited the best model performance compared to other individual ensemble methods due to its complexity and deep understanding of the phenomenon. The obtained results demonstrate that slope is the most influential parameter associated with erosion, with an importance value of 49.6%. This is followed by land use/land cover (LULC), lithology, and topographic wetness index (TWI), with importance values of 13.82%, 12.91%, and 11.17%, respectively. The remaining factors demonstrated insignificant importance, with percentages varying between 0% and 3.5%.

Stacking ensemble method is of great interest in combining multiple algorithms to model complex phenomena. The obtained results indicate that the use of stacking improves performance and demonstrates superiority over other models. The importance of the stacking ensemble method was calculated based on the aggregation of feature importance scores across all models. One approach to achieve this is to calculate the average importance score for each feature while considering the performance of each base model. The importance of each factor was ranked as follows: slope, topographic wetness index (TWI), LULC, lithology, aspect, distance from river, modified Fournier index, elevation, clay mineral ratio, distance from road, and geology.

4. Discussion

The Macta basin located in north-western Algeria has been experiencing a significant issue with erosion, resulting in substantial soil loss and impeding the implementation of sustainable land management practices. As a result, it is crucial to identify areas of high vulnerability using the most effective modeling techniques to enable the implementation of appropriate soil and water conservation measures.

To achieve our objective of accurately estimating erosion in this area and identify the most appropriate model, we applied four artificial intelligence models (CNN, CatBoost, AdaBoost, and stacking). Our research aimed to determine the primary factors contributing to gully erosion, as this phenomenon is influenced by various factors. Our analysis is based on the hybridization of the used models, combining the ensemble methods and deep learning techniques. This identified that slope is a crucial factor—as expected—additionally, TWI, LULC, and lithology are the most dominant factors in mapping area susceptible to erosion. Although variations between the importance of variables were observed between models, the accuracy of the performance results is still outstanding in the elaborated models. These results are consistent with the results of previous studies [22,39,40].

According to the results obtained from the stacking ensemble model shown in Figure 7, it can be inferred that the very high susceptibility level covers 19.97% of the total area of the basin, while 20.04% of the area is categorized as having a high susceptibility, which is mainly located in grasslands. Moreover, 20.74% of the total area faces moderate susceptibility, 19.86% faces low susceptibility, and 19.38% faces very low susceptibility.

Regarding the results obtained by the CNN and CatBoost models (Figure 7), they showed that 19.60% and 19.55% of the area are classified as very high susceptibility, 20.23% and 20.44% high susceptibility, 20.42% and 19.71% moderate susceptibility, 20.09% and 20.99% low susceptibility, and 19.66% and 19.31% having very low susceptibility, respectively. The CNN and CatBoost exhibited a highly similar distribution of erosion prone area across different LULC classes (Table 8).

In contrast, the results obtained by the AdaBoost model (Figure 7) showed that 19.23% of the area is classified as having very high susceptibility, 20.01% high susceptibility, 20.41% moderate susceptibility, 20.66% low susceptibility, and 19.69% having very low susceptibility.

Table 8. Statistics of Erosion prone area in different Land Use/Land Cover classes.

Model	LULC Class	Grasslands	Croplands	Forest	Urbanization	Bare Lands	Water Bodies	Total
AdaBoost	Very Low Risk	5.93	11.48	0.39	1.24	0.60	0.05	19.69
	Low Risk	10.65	7.90	0.74	0.90	0.47	0.01	20.66
	Moderate Risk	13.33	5.26	0.95	0.48	0.37	0.01	20.41
	High Risk	14.83	3.60	1.05	0.23	0.30	0.00	20.01
	Very High Risk	15.87	2.00	1.05	0.12	0.19	0.00	19.23
CatBoost	Very Low Risk	2.86	14.57	0.33	1.24	0.29	0.03	19.31
	Low Risk	9.45	9.14	0.78	1.01	0.60	0.02	20.99
	Moderate Risk	14.42	3.53	0.91	0.38	0.46	0.01	19.71
	High Risk	16.73	2.13	1.10	0.21	0.27	0.01	20.44
	Very High Risk	17.15	0.88	1.07	0.12	0.33	0.00	19.55
CNN	Very Low Risk	2.62	15.19	0.28	1.29	0.24	0.03	19.66
	Low Risk	8.94	8.57	0.99	0.95	0.61	0.02	20.09
	Moderate Risk	14.53	3.89	1.08	0.42	0.48	0.01	20.42
	High Risk	17.00	1.80	0.93	0.20	0.29	0.01	20.23
	Very High Risk	17.50	0.79	0.89	0.10	0.32	0.01	19.60
Stacking	Very Low Risk	3.17	14.17	0.28	1.32	0.40	0.04	19.38
	Low Risk	8.96	8.67	0.77	0.95	0.50	0.02	19.86
	Moderate Risk	14.37	4.41	1.10	0.41	0.45	0.01	20.74
	High Risk	16.40	2.11	1.05	0.19	0.29	0.01	20.04
	Very High Risk	17.70	0.90	0.98	0.10	0.29	0.00	19.97

Based on these results, it can be concluded that more than 60% of the basin is classified as having low to moderate erosion according to the stacking ensemble method, CNN, AdaBoost and CatBoost models, whereas around 40% of the area corresponds to the classes of high to very high erosion. The results of this study are consistent with the findings of Taye et al. [77], who reported significantly higher seasonal runoff coefficient values for grasslands compared to croplands. This aligns with our observation of high to very high susceptibility areas exhibiting similar patterns, indicating the potential influence of land cover on soil erosion dynamics (Table 8).

The understanding of the influential factors by each model differs, which implies that less precise factors can lead to high accuracy models. For a better understanding of this phenomenon, it is better to use complex algorithms for feature selection, which search for the associations between the studied phenomena and the influencing factors.

Creating an erosion susceptibility map is a valuable tool for mitigating the risks of water-induced soil erosion. Areas identified as having high or very high susceptibility, yet to experience erosion, indicate conditions favorable for erosion development. Therefore, these areas are particularly vulnerable to soil erosion, underscoring the significance of including them in the erosion susceptibility map and promoting sustainable practices, both in agriculture and forestry, to preserve soil quality.

Moreover, it is important to consider that the mapping of erosion-susceptible areas using AI techniques reveals variations in the machine's understanding of the phenomena. The results indicate that the Adaptive boosting model exhibits different influencing factors compared to the CNN and CatBoost models, yet still achieves an overall accuracy higher than 94%.

To further improve the accuracy, a stacking method was employed, which combines the previous machine learning techniques in a hybrid model. This stacking approach yielded exceptionally high results, with an accuracy reaching 99% (Table 7). By leveraging the strengths of multiple models, the stacking method enhances the predictive power and reliability of the erosion susceptibility mapping process.

Overall, the integration of AI techniques and the application of the stacking method have proven to be effective in accurately identifying erosion-prone areas and providing valuable insights for soil conservation and management strategies.

5. Conclusions

This research showcases the significant potential of machine learning and deep learning tools, as well as meta-models, in improving the identification, visualization, and interpretation of areas susceptible to erosion. By focusing on the Macta basin in Algeria, this study successfully developed soil erosion susceptibility maps using four distinct machine learning algorithms: CNN, Adaptive boosting, Categorical boosting (CatBoost), and stacking ensemble methods.

The findings shed light on the factors influencing erosion-prone areas in the Macta basin. AdaBoost identified TWI, aspect, distance from river, and clay mineral ratio as key parameters in mapping erosion-prone areas. CatBoost highlighted slope, LULC, modified Fournier index, distance from road, and distance from river as the most influential factors. The CNN model emphasized slope, LULC, lithology, and TWI as critical factors in mapping erosion-susceptible areas.

Moreover, the utilization of stacking ensemble methods demonstrated exceptional accuracy and significantly improved the prediction and mapping of erosion-susceptible areas. By combining predictions from multiple base models, the stacking ensemble approach provided a more robust and reliable estimation of erosion susceptibility. This hybrid model leveraged the strengths of individual machine learning algorithms while effectively mitigating their weaknesses, resulting in a highly accurate and comprehensive assessment of erosion risks.

The reliable erosion susceptibility maps generated in this study serve as invaluable tools for decision-makers and government officials involved in erosion risk management. The integration of machine learning and deep learning techniques, along with the stacking ensemble method, offers a promising approach to better delineate, visualize, and interpret erosion-prone areas.

Moving forward, future research should focus on refining machine learning algorithms and ensemble methods in erosion modeling to mitigate water-induced soil erosion and enhance sustainable land use planning. An important direction for future studies is the analysis of erosion susceptibility over longer time periods using climatic models. This approach will provide valuable insights into the long-term impact of climate change and enable proactive measures to ensure the sustainability of soil resources. By integrating climate projections, erosion modeling can support decision-making and protect soil resources in watershed management and conservation, promoting sustainable practices. The combination of machine learning, ensemble methods, and climate projections has the potential to enhance our understanding of erosion dynamics and guide effective prevention and mitigation strategies, thereby ensuring the long-term sustainability of land use and ecosystem preservation.

Author Contributions: Conceptualization: S.E.T. and H.B. (Hamza Bouguerra); Methodology: S.E.T., H.B. (Hamza Bouguerra) and S.B.; Software: S.E.T., H.B. (Hamza Bouguerra), H.B. (Hamza Bouchehed) and Y.H.; Validation: S.E.T., A.A. and H.B. (Hamza Bouchehed); Formal analysis: H.B. (Hamza Bouguerra), H.B. (Hamza Bouchehed), Y.H. and A.A.; Investigation: S.E.T. and H.B. (Hamza Bouguerra); Resources: S.E.T., H.B. (Hamza Bouguerra) and N.A.; Writing—original draft preparation: S.E.T., G.G. and J.N.-P.; Writing—review and editing: J.N.-P., G.G. and S.E.T.; Supervision: S.E.T., S.B., H.B. (Hamza Bouguerra) and J.N.-P.; Project administration: H.B. (Hamza Bouguerra) and S.E.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Ministry of Higher Education and Scientific Research (MESRS—Algeria).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to express our sincere gratitude to Hamza Atoui, from the department of electronics, Badji Mokhtar Annaba University, Algeria.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Modified Fournier index results for the 106 meteorological stations located in the Macta basin.

No.	Station		Lambert Coordinate		MFI
	Code	Name	X (m)	Y (m)	(mm)
1	110102	RAS EL MA	177,450	139,500	29.68
2	110201	SIDI ALI BEN YOUB	186,550	192,200	44.18
3	110202	MOULAY SLISSENE MF	181,200	171,550	41.79
4	110203	EL HACAIBA	183,500	161,650	35.78
5	110208	SLISSENE CENTRE	183,650	174,650	42.56
6	110209	TAMFOUSSET	192,900	183,350	35.93
7	110305	SIDI BEL ABBES	194,250	214,150	46.33
8	110306	SIDI BRAHIM	203,230	222,480	46.87
9	110307	BEN BADIS	170,850	190,800	46.72
10	110308	SIDI ALI BOUSSIDI	178,250	206,150	46.49
11	110309	HASSI DAHO	204,800	204,100	44.76
12	110310	LAMTAR	181,400	203,000	45.96
13	110311	SIDI KHALED	188,500	207,500	45.24
14	110312	MOSTEFA BEN BRAHIM	221,700	214,740	49.03
15	110313	TESSALA	184,500	222,050	46.25
16	110314	AIN TRID	193,000	226,000	46.28
17	110315	AIN EL BERD	208,400	234,300	48.62
18	110317	HASSI ZEHANA	172,700	198,200	46.66
19	110318	SIDI LAHCENE	191,200	212,900	45.19
20	110319	CAID BELARBI	212,900	210,700	46.07
21	110322	TABIA	186,800	196,700	44.52
22	110328	SULLY	201,500	206,400	44.61
23	110329	LES TREMBLES	204,800	227,260	47.28
24	110334	CHETOUANE	175,300	191,250	45.96
25	110401	BOUDJEBAA (Dar Esba)	226,200	233,000	49.16
26	110402	CHEURFAS Bge	232,100	238,300	49.47
27	110501	MERINE	216,300	170,500	28.31
28	110502	TELAGH	200,650	170,150	32.76
29	110503	TEGHALINET	203,450	181,600	39.19
30	110504	TENIRA	205,500	196,250	42.29
31	110505	EL HADJIRA	199,400	195,600	41.72
32	110507	FERME CHABRIER	194,800	190,450	41.13
33	110509	SIDI AHMED	204,050	190,050	42.08
34	110510	DOMAINE ZERROUKI	204,650	185,000	40.34
35	110514	AIN CHAFIA	210,700	185,250	38.36
36	110602	OUED SEFFIOUN	221,150	201,100	46.84
37	110603	AIN FRASS	237,750	215,000	51.73
38	110605	HASSI EL ABD	226,750	189,200	43.40
39	110701	TOUAZIZINE M.F. (Dhaya)	191,150	155,200	32.55
40	110702	DOUAHILA	228,700	155,350	29.51
41	110703	TOUAZIZINE (Dhaya)	196,300	157,450	30.36
42	110802	DAOUD YOUB	234,500	185,000	43.25
43	110902	HASSI AYOUN MF	241,750	161,250	29.18
44	110903	DOUI THABET	252,100	181,700	34.21
45	110904	BOU EL FERID	245,730	169,150	31.44

Table A1. Cont.

No.	Station		Lambert Coordinate		MFI
	Code	Name	X (m)	Y (m)	(mm)
46	111002	FERME EL HARIG	245,590	192,450	44.10
47	111102	MEFTAH SIDI BOUBEKEUR	259,500	195,750	42.94
48	111103	AIN EL HADJAR	266,500	165,200	31.70
49	111105	SID AMAR	263,850	195,100	41.28
50	111106	KILOMETRE 50	268,450	192,000	38.31
51	111112	HAMMAM RABI	270,400	184,500	36.44
52	111113	DJEBEL KAROUS	264,700	181,200	33.87
53	111114	REBAHIA FERME 917	272,600	180,500	34.55
54	111120	FERME DU SYNDICAT	263,700	165,500	31.14
55	111128	AIN ZERGA FERME	273,900	176,400	33.58
56	111130	SAIDA ANRH	266,750	174,400	33.80
57	111201	OUED TARIA	262,350	204,850	45.95
58	111202	OUM EL DJIRANE	283,000	173,400	34.24
59	111203	AIN BALLOUL	296,850	190,550	38.90
60	111204	AIN TIFRIT	290,050	182,450	36.80
61	111205	AIN SOLTANE	281,400	188,400	37.44
62	111208	SIDI MIMOUN	289,100	196,100	39.76
63	111209	BLED EL BEIDA	283,300	183,100	35.98
64	111210	TAMESNA	295,600	174,500	35.58
65	111211	SIDI BEN KADOUR MF	291,500	164,100	33.64
66	111213	EL HAZEM	272,200	168,600	32.20
67	111215	BOUCHERID MOHAMED	276,750	172,600	32.81
68	111217	BENIANE	275,000	203,150	44.52
69	111219	HASNA Dne BOUCHIKHI	277,350	194,550	39.84
70	111401	MAOUSSA	277,300	233,920	58.17
71	111402	FROHA	266,100	226,000	54.25
72	111404	AOUF M.F.	287,150	211,800	45.55
73	111405	MATEMORE	273,970	228,350	53.66
74	111407	TIGHENNIF	285,100	237,900	55.75
75	111408	KHAOUILA	282,150	243,100	60.20
76	111409	AIN FARES	277,500	245,100	60.20
77	111413	TIZI	261,500	227,800	54.71
78	111414	SIDI KADA	285,900	228,300	51.87
79	111415	AIN FEKAN MN	255,600	217,200	52.53
80	111416	SIDI ALI KERROUCHA	290,100	214,600	45.66
81	111418	NESMOTH M.F.	289,250	219,700	49.02
82	111422	MASCARA Pedo.	271,400	232,600	55.37
83	111424	GHRISS	269,200	219,800	51.34
84	111502	SAHOUEZ OUIZERT	247,620	215,800	50.51
85	111503	BOU HANIFIA Bge	249,000	223,600	50.38
86	111508	SFISSEF	233,750	218,800	53.65
87	111509	HACINE	255,550	243,500	50.20
88	111512	FERGOUG	259,100	250,150	49.13
89	111513	BOUHNIFIA MN	250,200	227,700	50.34
90	111517	MOHAMMADIA SAEF	261,750	257,370	41.49
91	111601	MACTA	245,450	279,700	41.09
92	111603	SIG	237,720	252,000	45.20
93	111604	OGGAZ	232,200	255,800	43.18
94	111605	BOU HENNI	247,500	255,400	42.44
95	111606	FORNAKA	250,850	278,500	41.85
96	111607	SAMOURIA	265,950	261,200	43.27
97	111608	EL GHOMRI	274,000	268,000	40.85
98	111609	BOUGHIRAT	278,000	275,000	41.30
99	111610	MOCTA DOUZ	251,250	260,200	41.41
100	111611	FERME BLANCHE	256,800	265,350	40.73

Table A1. Cont.

No.	Station		Lambert Coordinate		MFI
	Code	Name	X (m)	Y (m)	(mm)
101	111612	BLED TAOURIA	277,000	284,600	44.64
102	111614	AIN MOUISSY	260,300	281,500	42.87
103	111615	FORNAKA	254,950	275,500	40.59
104	111616	MARAIS DE SIRAT	269,300	275,600	39.50
105	111617	FERME ASSORAIN	281,250	291,850	48.03
106	111618	SOUAFFLIOS	285,200	285,650	50.88

References

- Kettler, T.A.; Doran, J.W.; Gilbert, T.L. Simplified method for soil particle-size determination to accompany soil-quality analyses. *Soil Sci. Soc. Am. J.* **2001**, *65*, 849–852. [\[CrossRef\]](#)
- Bouguerra, H.; Bouanani, A.; Khanchoul, K.; Derdous, O.; Tachi, S.E. Mapping erosion prone areas in the Bouhamdane watershed (Algeria) using the Revised Universal Soil Loss Equation through GIS. *J. Water Land Dev.* **2017**, *32*, 13–23. [\[CrossRef\]](#)
- Tachi, S.E.; Bouguerra, H.; Derdous, O.; Djabri, L.; Benmamar, S. Estimating suspended sediment concentration at different time scales in Northeastern Algeria. *Appl. Water Sci.* **2020**, *10*, 118. [\[CrossRef\]](#)
- Arabameri, A.; Pradhan, B.; Rezaei, K.; Sohrabi, M.; Kalantari, Z. GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. *J. Mt. Sci.* **2019**, *16*, 595–618. [\[CrossRef\]](#)
- Poesen, J.W.; Vandaele, K.; Van Wesemael, B. Contribution of gully erosion to sediment production on cultivated lands and rangelands. *IAHS Publ.-Ser. Proc. Rep.-Int. Assoc. Hydrol. Sci.* **1996**, *236*, 251–266.
- Jahantigh, M.; Pessarakli, M. Causes and effects of gully erosion on agricultural lands and the environment. *Commun. Soil Sci. Plant Anal.* **2011**, *42*, 2250–2255. [\[CrossRef\]](#)
- Bouguerra, H.; Tachi, S.E.; Derdous, O.; Bouanani, A.; Khanchoul, K. Suspended sediment discharge modeling during flood events using two different artificial neural network algorithms. *Acta Geophys.* **2019**, *67*, 1649–1660. [\[CrossRef\]](#)
- Bouhaddeb, C.E.; Menani, M.R.; Bouguerra, H.; Derdous, O. Assessing soil loss using GIS based RUSLE methodology. Case of the BouNamoussa watershed–North-East of Algeria. *J. Water Land Dev.* **2018**, *36*, 27–35. [\[CrossRef\]](#)
- Liu, H.; Zhang, T.; Liu, B.; Liu, G.; Wilson, G.V. Effects of gully erosion and gully filling on soil depth and crop production in the black soil region, northeast China. *Environ. Earth Sci.* **2013**, *68*, 1723–1732. [\[CrossRef\]](#)
- Poesen, J.; Nachtergaele, J.; Verstraeten, G.; Valentin, C. Gully erosion and environmental change: Importance and research needs. *Catena* **2003**, *50*, 91–133. [\[CrossRef\]](#)
- Lal, R. Restoring land degraded by gully erosion in the tropics. *Soil Restor.* **1992**, *17*, 123–152.
- Xiong, X.; Zhang, K.; Chen, X.; Shi, H.; Luo, Z.; Wu, C. Sources and distribution of microplastics in China’s largest inland lake–Qinghai Lake. *Environ. Pollut.* **2018**, *235*, 899–906. [\[CrossRef\]](#)
- Skidmore, E.L. Soil loss tolerance. *Determ. Soil Loss Toler.* **1982**, *45*, 87–93.
- Duniway, M.C.; Pfennigwerth, A.A.; Fick, S.E.; Nauman, T.W.; Belnap, J.; Barger, N.N. Wind erosion and dust from US drylands: A review of causes, consequences, and solutions in a changing world. *Ecosphere* **2019**, *10*, e02650. [\[CrossRef\]](#)
- Tarolli, P.; Preti, F.; Romano, N. Terraced landscapes: From an old best practice to a potential hazard for soil degradation due to land abandonment. *Anthropocene* **2014**, *6*, 10–25. [\[CrossRef\]](#)
- Hou, G.; Bi, H.; Huo, Y.; Wei, X.; Zhu, Y.; Wang, X.; Liao, W. Determining the optimal vegetation coverage for controlling soil erosion in Cynodondactylon grassland in North China. *J. Clean. Prod.* **2020**, *244*, 118771. [\[CrossRef\]](#)
- Mostazo, P.; Asensio-Amador, C.; Asensio, C. Soil Erosion Modeling and Monitoring. *Agriculture* **2023**, *13*, 447. [\[CrossRef\]](#)
- Gholami, V. The influence of deforestation on runoff generation and soil erosion (Case study: Kasilian Watershed). *J. For. Sci.* **2013**, *59*, 272–278. [\[CrossRef\]](#)
- Kouassi, J.L.; Gyau, A.; Diby, L.; Bene, Y.; Kouamé, C. Assessing land use and land cover change and farmers’ perceptions of deforestation and land degradation in South-West Côte d’Ivoire, West Africa. *Land* **2021**, *10*, 429. [\[CrossRef\]](#)
- Nyssen, J.; Poesen, J.; Moeyersons, J.; Luyten, E.; Veyret-Picot, M.; Deckers, J.; Haile, M.; Govers, G. Impact of road building on gully erosion risk: A case study from the northern Ethiopian highlands. *Earth Surf. Process. Landf. J. Br. Geomorphol. Res. Group* **2002**, *27*, 1267–1283. [\[CrossRef\]](#)
- Frick, W.F.; Kingston, T.; Flanders, J. A review of the major threats and challenges to global bat conservation. *Ann. N. Y. Acad. Sci.* **2020**, *1469*, 5–25. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mosavi, A.; Sajedi-Hosseini, F.; Choubin, B.; Taromideh, F.; Rahi, G.; Dineva, A.A. Susceptibility mapping of soil water erosion using machine learning models. *Water* **2020**, *12*, 1995. [\[CrossRef\]](#)
- Saha, S.; Sarkar, R.; Thapa, G.; Roy, J. Modeling gully erosion susceptibility in Phuentsholing, Bhutan using deep learning and basic machine learning algorithms. *Environ. Earth Sci.* **2021**, *80*, 295. [\[CrossRef\]](#)

24. Renard, K.G.; Freimund, J.R. Using monthly precipitation data to estimate the R-factor in the revised USLE. *J. Hydrol.* **1994**, *157*, 287–306. [[CrossRef](#)]
25. Wischmeier, W.H.; Smith, D.D. Predicting rainfall erosion losses from cropland east of the rocky mountains: Guide for selection of practices for soil and water conservation. In *Agriculture Handbook*; United States Department of Agriculture: Washington, DC, USA, 1965; Volume 282, p. 58.
26. Wischmeier, W.H.; Smith, D.D. Predicting rainfall erosion losses: A guide to conservation planning. In *Agriculture Handbook*; United States Department of Agriculture: Washington, DC, USA, 1978; Volume 537.
27. Almagro, A.; Thomé, T.C.; Colman, C.B.; Pereira, R.B.; Junior, J.M.; Rodrigues, D.B.B.; Oliveira, P.T.S. Improving cover and management factor (C-factor) estimation using remote sensing approaches for tropical regions. *Int. Soil Water Conserv. Res.* **2019**, *7*, 325–334. [[CrossRef](#)]
28. Hudson, N.W. Instrumentation for studies of the erosive power of rainfall. In *Erosion and Sediment Transport Measurement (Proceedings of the Florence Symposium)*; IAHS Publication: Florence, Italy, 1981; pp. 383–390.
29. Gayen, A.; Saha, S. Application of weights-of-evidence (WoE) and evidential belief function (EBF) models for the delineation of soil erosion vulnerable zones: A study on Pathro river basin, Jharkhand, India. *Model. Earth Syst. Environ.* **2017**, *3*, 1123–1139. [[CrossRef](#)]
30. Gayen, A.; Pourghasemi, H.R.; Saha, S.; Keesstra, S.; Bai, S. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Sci. Total Environ.* **2019**, *668*, 124–138. [[CrossRef](#)]
31. Cama, M.; Schillaci, C.; Kropáček, J.; Hochschild, V.; Bosino, A.; Märker, M. A probabilistic assessment of soil erosion susceptibility in a head catchment of the Jemma Basin, Ethiopian Highlands. *Geosciences* **2020**, *10*, 248. [[CrossRef](#)]
32. Zakerinejad, R.; Märker, M. Prediction of Gully erosion susceptibilities using detailed terrain analysis and maximum entropy modeling: A case study in the Mazayegan Plain, Southwest Iran. *Geogr. Fis. Din. Quat.* **2014**, *37*, 67–76.
33. Conoscenti, C.; Agnesi, V.; Cama, M.; Caraballo-Arias, N.A.; Rotigliano, E. Assessment of gully erosion susceptibility using multivariate adaptive regression splines and accounting for terrain connectivity. *Land Degrad. Dev.* **2018**, *29*, 724–736. [[CrossRef](#)]
34. Rahmati, O.; Pourghasemi, H.R.; Melesse, A.M. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *Catena* **2016**, *137*, 360–372. [[CrossRef](#)]
35. Zabihi, M.; Mirchooli, F.; Motevalli, A.; Darvishan, A.K.; Pourghasemi, H.R.; Zakeri, M.A.; Sadighi, F. Spatial modelling of gully erosion in Mazandaran Province, northern Iran. *Catena* **2018**, *161*, 1–13. [[CrossRef](#)]
36. Chaplot, V.; Le Brozec, E.C.; Silvera, N.; Valentin, C. Spatial and temporal assessment of linear erosion in catchments under sloping lands of northern Laos. *Catena* **2005**, *63*, 167–184. [[CrossRef](#)]
37. Conoscenti, C.; Angileri, S.; Cappadonia, C.; Rotigliano, E.; Agnesi, V.; Märker, M. Gully erosion susceptibility assessment by means of GIS-based logistic regression: A case of Sicily (Italy). *Geomorphology* **2014**, *204*, 399–411. [[CrossRef](#)]
38. Hembram, T.K.; Saha, S. Prioritization of sub-watersheds for soil erosion based on morphometric attributes using fuzzy AHP and compound factor in Jainti River basin, Jharkhand, Eastern India. *Environ. Dev. Sustain.* **2020**, *22*, 1241–1268. [[CrossRef](#)]
39. Bouamrane, A.; Bouamrane, A.; Abida, H. Water erosion hazard distribution under a Semi-arid climate Condition: Case of Mellah Watershed, North-eastern Algeria. *Geoderma* **2021**, *403*, 115381. [[CrossRef](#)]
40. Avand, M.; Janizadeh, S.; Naghibi, S.A.; Pourghasemi, H.R.; KhosrobeigiBozchaloei, S.; Blaschke, T. A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. *Water* **2019**, *11*, 2076. [[CrossRef](#)]
41. Saha, S.; Roy, J.; Arabameri, A.; Blaschke, T.; TienBui, D. Machine learning-based gully erosion susceptibility mapping: A case study of Eastern India. *Sensors* **2020**, *20*, 1313. [[CrossRef](#)]
42. Arabameri, A.; ChandraPal, S.; Costache, R.; Saha, A.; Rezaie, F.; Seyed Danesh, A.; Pradhan, B.; Lee, S.; Hoang, N.D. Prediction of gully erosion susceptibility mapping using novel ensemble machine learning algorithms. *Geomat. Nat. Hazards Risk* **2021**, *12*, 469–498. [[CrossRef](#)]
43. Yang, A.; Wang, C.; Pang, G.; Long, Y.; Wang, L.; Cruse, R.M.; Yang, Q. Gully erosion susceptibility mapping in highly complex terrain using machine learning models. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 680. [[CrossRef](#)]
44. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [[CrossRef](#)]
45. Jebari, S.; Berndtsson, R.; Bahri, A.; Boufaroua, M. Spatial soil loss risk and reservoir siltation in semi-arid Tunisia. *Hydrol. Sci. J.-J. Sci. Hydrol.* **2010**, *55*, 121–137. [[CrossRef](#)]
46. Balasubramani, K.; Veena, M.; Kumaraswamy, K.; Saravanabavan, V. Estimation of soil erosion in a semi-arid watershed of Tamil Nadu (India) using revised universal soil loss equation (rusle) model through GIS. *Model. Earth Syst. Environ.* **2015**, *1*, 10. [[CrossRef](#)]
47. Mohapatra, R. Application of revised universal soil loss equation model for assessment of soil erosion and prioritization of ravine infested sub basins of a semi-arid river system in India. *Model. Earth Syst. Environ.* **2022**, *8*, 4883–4896. [[CrossRef](#)]
48. Falcão, C.J.L.M.; Duarte, S.M.D.A.; Da Silva Veloso, A. Estimating potential soil sheet Erosion in a Brazilian semiarid county using USLE, GIS, and remote sensing data. *Environ. Monit. Assess.* **2020**, *192*, 47. [[CrossRef](#)]
49. Bouzeria, H.; Tachi, S.E.; Bouguerra, H.; Derdous, O.; Benmamar, S. Evaluating the Effect of Land Use Land Cover Changes on Soil Loss Distribution in the Seybouse Basin, Northeastern Algeria. *Dokl. Earth Sci.* **2023**, *510*, 335–348. [[CrossRef](#)]
50. Derdous, O.; Tachi, S.E.; Bouguerra, H. Spatial distribution and evaluation of aridity indices in Northern Algeria. *Arid Land Res. Manag.* **2021**, *35*, 1–14. [[CrossRef](#)]

51. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
52. Mathanker, S.K.; Weckler, P.R.; Bowser, T.J.; Wang, N.; Maness, N.O. AdaBoost classifiers for pecan defect classification. *Comput. Electron. Agric.* **2011**, *77*, 60–68. [[CrossRef](#)]
53. Solomatine, D.P.; Shrestha, D.L. AdaBoost. RT: A boosting algorithm for regression problems. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 1163–1168.
54. Margineantu, D.D.; Dietterich, T.G. Pruning adaptive boosting. *ICML* **1997**, *97*, 211–218.
55. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 1–11. Available online: https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf (accessed on 4 May 2023).
56. Jabeur, S.B.; Gharib, C.; Mefteh-Wali, S.; Arfi, W.B. CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technol. Forecast. Soc. Change* **2021**, *166*, 120658. [[CrossRef](#)]
57. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363.
58. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [[CrossRef](#)]
59. O’Shea, K.; Nash, R. An introduction to convolutional neural networks. *arXiv* **2015**, arXiv:1511.08458.
60. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology, Antalya, Turkey, 21–23 August 2017; pp. 1–6.
61. Nguyen, K.A.; Chen, W.; Lin, B.S.; Seeboonruang, U. Comparison of ensemble machine learning methods for soil erosion pin measurements. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 42. [[CrossRef](#)]
62. Williams, B.K. Adaptive management of natural resources—Framework and issues. *J. Environ. Manag.* **2011**, *92*, 1346–1353. [[CrossRef](#)]
63. Visa, S.; Ramsay, B.; Ralescu, A.L.; Van Der Knaap, E. Confusion matrix-based feature selection. *MAICS* **2011**, *710*, 120–127.
64. Lasko, T.A.; Bhagwat, J.G.; Zou, K.H.; Ohno-Machado, L. The use of receiver operating characteristic curves in biomedical informatics. *J. Biomed. Inform.* **2005**, *38*, 404–415. [[CrossRef](#)]
65. Akgün, A.; Türk, N. Mapping erosion susceptibility by a multivariate statistical method: A case study from the Ayvalık region, NW Turkey. *Comput. Geosci.* **2011**, *37*, 1515–1524. [[CrossRef](#)]
66. Hosmer, D.W.; Lemeshow, S. Goodness of fit tests for the multiple logistic regression model. *Commun. Stat.-Theory Methods* **1980**, *9*, 1043–1069. [[CrossRef](#)]
67. Mitchell, J.F.B.; Manabe, S.; Meleshko, V.; Tokioka, T. Equilibrium climate change and its implications for the future. *Clim. Chang. IPCC Sci. Assess.* **1990**, *131*, 172.
68. Nunes, A.N.; De Almeida, A.C.; Coelho, C.O. Impacts of land use and cover type on runoff and soil erosion in a marginal area of Portugal. *Appl. Geogr.* **2011**, *31*, 687–699. [[CrossRef](#)]
69. Abdelsamie, E.A.; Abdellatif, M.A.; Hassan, F.O.; El Baroudy, A.A.; Mohamed, E.S.; Kucher, D.E.; Shokr, M.S. Integration of RUSLE Model, Remote Sensing and GIS Techniques for Assessing Soil Erosion Hazards in Arid Zones. *Agriculture* **2022**, *13*, 35. [[CrossRef](#)]
70. Arnoldus, J.M.J. Methodology used to determine the maximum potential average annual soil loss due to sheet and rill erosion in Morocco. *Food Agric. Organ. Soils Bull.* **1977**, *34*, 39–51.
71. Krishnamurthy, J.; Venkatesa Kumar, N.; Jayaraman, V.; Manivel, M. An approach to demarcate ground water potential zones through remote sensing and a geographical information system. *Int. J. Remote Sens.* **1996**, *17*, 1867–1884. [[CrossRef](#)]
72. Krishnamurthy, J.; Mani, A.; Jayaraman, V.; Manivel, M. Groundwater resources development in hard rock terrain—an approach using remote sensing and GIS techniques. *Int. J. Appl. Earth Obs. Geoinf.* **2000**, *2*, 204–215. [[CrossRef](#)]
73. Bhuiyan, C.; Singh, R.P.; Flügel, W.A. Modelling of ground water recharge-potential in the hard-rock Aravalli terrain, India: A GIS approach. *Environ. Earth Sci.* **2009**, *59*, 929–938. [[CrossRef](#)]
74. Maizi, D.; Boufekane, A.; AitOuali, K.; Aoudia, M. Identification of potential area of recharge using geospatial and multi-criteria decision analysis in the Macta watershed (Western Algeria). *Arab. J. Geosci.* **2020**, *13*, 127. [[CrossRef](#)]
75. Sabins, F.F. Remote sensing for mineral exploration. *Ore Geol. Rev.* **1999**, *14*, 157–183. [[CrossRef](#)]
76. Van der Meer, F.D.; Van der Werff, H.M.; Van Ruitenbeek, F.J.; Hecker, C.A.; Bakker, W.H.; Noomen, M.F.; Van Der Meijde, M.; Carranza, E.J.M.; De Smeth, J.B.; Woldai, T. Multi- and hyperspectral geologic remote sensing: A review. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *14*, 112–128. [[CrossRef](#)]
77. Taye, G.; Poesen, J.; Wesemael, B.V.; Vanmaercke, M.; Tekle, D.; Deckers, J.; Goosse, T.; Maetens, W.; Nyssen, J.; Hallet, V.; et al. Effects of land use, slope gradient, and soil and water conservation structures on runoff and soil loss in semi-arid Northern Ethiopia. *Phys. Geogr.* **2013**, *34*, 236–259. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.