

Article

Susceptibility Mapping of Soil Water Erosion Using Machine Learning Models

Amirhosein Mosavi ^{1,2,*} , Farzaneh Sajedi-Hosseini ^{3,*}, Bahram Choubin ⁴ ,
Fereshteh Taromideh ⁵, Gholamreza Rahi ⁶ and Adrienn A. Dineva ⁷ 

¹ Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

² Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam

³ Department of Reclamation of Arid and Mountainous Regions, Faculty of Natural Resources, University of Tehran, Karaj 31585-77871, Iran

⁴ Soil Conservation and Watershed Management Research Department, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia 57169-63963, Iran; b.choubin@areeo.ac.ir

⁵ Department of Irrigation, Sari Agricultural Sciences and Natural Resources University, Sari 48181-68984, Iran; ftaromide@stu.sanru.ac.ir

⁶ Natural Resources and Watershed Management Research Department, Bushehr Agricultural and Natural Resources Research and Education Center, Bushehr 75156-43373, Iran; gh.raghi@areeo.ac.ir

⁷ Kalman Kando Faculty of Electrical Engineering, Obuda University, 1034 Budapest, Hungary; dineva.adrienn@kvk.uni-obuda.hu

* Correspondence: amirhosein.mosavi@tdtu.edu.vn (A.M.); farzsajedi@ut.ac.ir (F.S.-H.)

Received: 3 May 2020; Accepted: 10 July 2020; Published: 14 July 2020



Abstract: Soil erosion is a serious threat to sustainable agriculture, food production, and environmental security. The advancement of accurate models for soil erosion susceptibility and hazard assessment is of utmost importance for enhancing mitigation policies and laws. This paper proposes novel machine learning (ML) models for the susceptibility mapping of the water erosion of soil. The weighted subspace random forest (WSRF), Gaussian process with a radial basis function kernel (Gausspradial), and naive Bayes (NB) ML methods were used in the prediction of the soil erosion susceptibility. Data included 227 samples of erosion and non-erosion locations through field surveys to advance models of the spatial distribution using predictive factors. In this study, 19 effective factors of soil erosion were considered. The critical factors were selected using simulated annealing feature selection (SAFS). The critical factors included aspect, curvature, slope length, flow accumulation, rainfall erosivity factor, distance from the stream, drainage density, fault density, normalized difference vegetation index (NDVI), hydrologic soil group, soil texture, and lithology. The dataset cells of samples (70% for training and 30% for testing) were randomly prepared to assess the robustness of the different models. The functional relevance between soil erosion and effective factors was computed using the ML models. The ML models were evaluated using different metrics, including accuracy, the kappa coefficient, and the probability of detection (POD). The accuracies of the WSRF, Gausspradial, and NB methods were 0.91, 0.88, and 0.85, respectively, for the testing data; 0.82, 0.76, and 0.71, respectively, for the kappa coefficient; and 0.94, 0.94, and 0.94, respectively, for POD. However, the ML models, especially the WSRF, had an acceptable performance regarding producing soil erosion susceptibility maps. Maps produced with the most robust models can be a useful tool for sustainable management, watershed conservation, and the reduction of soil and water loss.

Keywords: water erosion; susceptibility; Gaussian process; climate change; radial basis function kernel; weighted subspace random forest; extreme events; extreme weather; naive Bayes; feature selection; machine learning; hydrologic model; simulated annealing; earth system science

1. Introduction

Soil conservation is of utmost importance for sustainable development, food security, and environmental protection [1]. Understanding soil erosion is considered to be an essential practice for soil conservation programs around the world [2]. Currently, soil erosion has increasingly become known as a severe concern for sustainable agriculture, water resource management, and modern civilization [3]. Soil erosion is a significant menace for soil, ecology, and for humanity since the long-term production of soil productive capacity is profoundly affected by the destruction and leaching of soil's organic and topsoil matters [4]. Soil erosion is an intricate process that depends on the plant cover and land use, watershed topography, soil properties, climate, and land management practices. In the last century, soil erosion has intensified due to human activity and is an environmental problem [5]. Primary soil segregates when the rainfall or water flow power is greater than the soil's resistance to corrosion [6]. Generally, there are different types of water erosion, such as sheet, gully, landslide, debris flow, streambank, etc. [7].

In semiarid regions, such as Iran, soil erosion is a significant crisis [8] and can be considered to be one of the critical problems concerning agricultural development, natural resources, and the environment [9]. In such regions, water is limited, and there are many sources of sediment [10]. The high input of sediment in upstream rivers increases the water turbidity, reduces the lifespan of dams owing to reservoir siltation, and negatively affects water quality and biological activity [8]. According to scholars, the mean annual rate of soil erosion in Iran is about 25 tons/ha/year, which is four times more than the mean yearly rate around the world [11,12]. Therefore, the susceptibility mapping of soil erosion is necessary for controlling this critical problem.

Rather than using traditional and experimental models, such as the universal soil loss equation (USLE) [13] and multi-criteria decision-making methods [14], that have been used in water erosion assessments, machine learning (ML) models are known to be successful methods [15,16]. Different ML methods, such as support vector machine (SVM), boosted regression trees (BRT), random forest (RF), naive Bayes (NB), and artificial neural network (ANN), have been used for landslides [17–23], debris flows [24–26], and gully erosion [27–30]. For instance, Angileri et al. [15] used the stochastic gradient tree boost (SGT) for water erosion susceptibility mapping in central-northern Sicily, Italy. The results indicated that the applied model had excellent reliability (accuracy from 0.87 to 0.92). Recently, Garosi et al. [31] applied the RF, SVM, and NB models, along with the generalized additive model (GAM), to predict the gully erosion susceptibility in the Ekbatan Dam drainage basin, Iran. The results indicated that the RF model had the highest performance (accuracy = 92.4%) among the models tested. Svoray et al. [16] used different ML models, namely, SVM, ANN, and decision trees (DT), for predicting the gully erosion in a watershed scale in Israel and compared them with the results from topographic threshold (TT) and analytic hierarchy process (AHP) methods. The results indicated that the ML models produced better performances than the AHP and TT methods. Mao et al. [32] evaluated the soil erosion in the Shiqiaopu catchment, Hubei province, China, using SVM and ANN models. They optimized the parameters of the SVM using the particle swarm optimization (PSO) algorithm. The results indicated that the SVM had higher accuracy in comparison with the ANN model. Rahmati et al. [28] compared the ML models of SVM, ANN, RF, and BRT when predicting the gully erosion susceptibility in the Kashkan watershed, Iran. The results indicated that the performance of the RF and SVM models for predicting the gully occurrences in the watershed were better than the other models.

Due to the advancement of ML models, applying and evaluating novel methods in water erosion studies can help to accurately predict hazardous areas, especially in developing countries where soil erosion data are incomplete. The current study tried to predict water erosion susceptibility using two novel ML models, namely, a weighted subspace random forest (WSRF) and a Gaussian process with a radial basis function kernel (Gaussprradial), for the first time and compared their results with the NB model. Therefore, the primary purposes of this study were: (i) to identify the more significant factors regarding soil erosion through feature selection, (ii) to compare the performance of the novel

predictive models (i.e., WSRF and Gausspradial) with a model previously used for this application (i.e., NB), and (iii) the prediction of the spatial susceptibility of soil erosion induced by water.

2. Materials and Methods

2.1. Study Area

The Nur-Rood watershed is located in the southwest of the Haraz watershed, in the north of Iran. The watershed lies within $51^{\circ}26'–52^{\circ}19'$ E and $36^{\circ}01'–36^{\circ}16'$ N (Figure 1). The elevation of the watershed ranges from 732 to 4333 m. There are six rain gauge stations in the region provided the long-term mean annual data from 1976 to 2016 that were used in this study. The study area is about 1297 km² and is located upstream of the Haraz dam. The main application of this dam is to provide drinking and agriculture water for five cities (i.e., Amol, Babol, Babolsar, Nur, and Mahmoodabad) in the Mazandaran province. According to the literature, the watershed generates water with a high sediment load such that it causes a reduction in the dam's capacity [8,33]. Therefore, identifying the hazardous areas can help to control the upstream erosion and aid with providing sustainable watershed management in the Nur-Rood watershed.

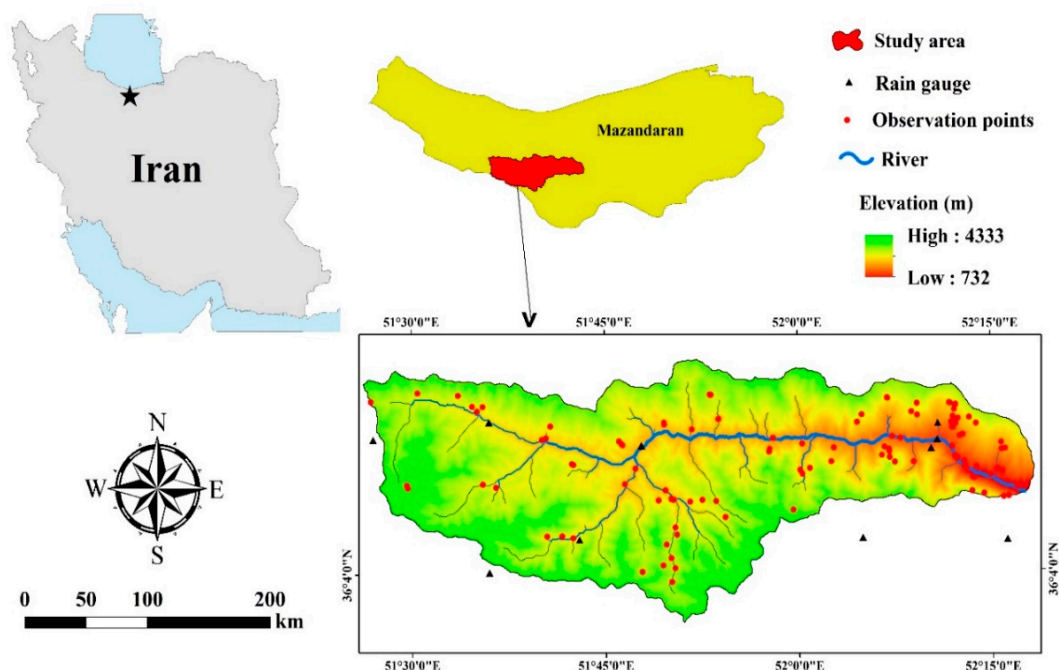


Figure 1. Location of the Nur-Rood watershed, Mazandaran province, Iran.

2.2. Methodology

The methodology consisted of several fundamental building blocks to ensure the accuracy of the susceptibility prediction. Figure 2 presents the schematic of the methodology workflow from data sampling to the susceptibility prediction. The method consisted of five sections: (i) preparation and collection of the relevant factors for soil erosion modeling; (ii) extraction of the erosion and non-erosion locations by the field observations; (iii) selection of the essential factors using the simulated annealing feature selection (SAFS) algorithm; (iv) water erosion modeling using the Gausspradial, NB, and WSRF models in the Nur-Rood watershed; and (v) evaluating the models' performance.

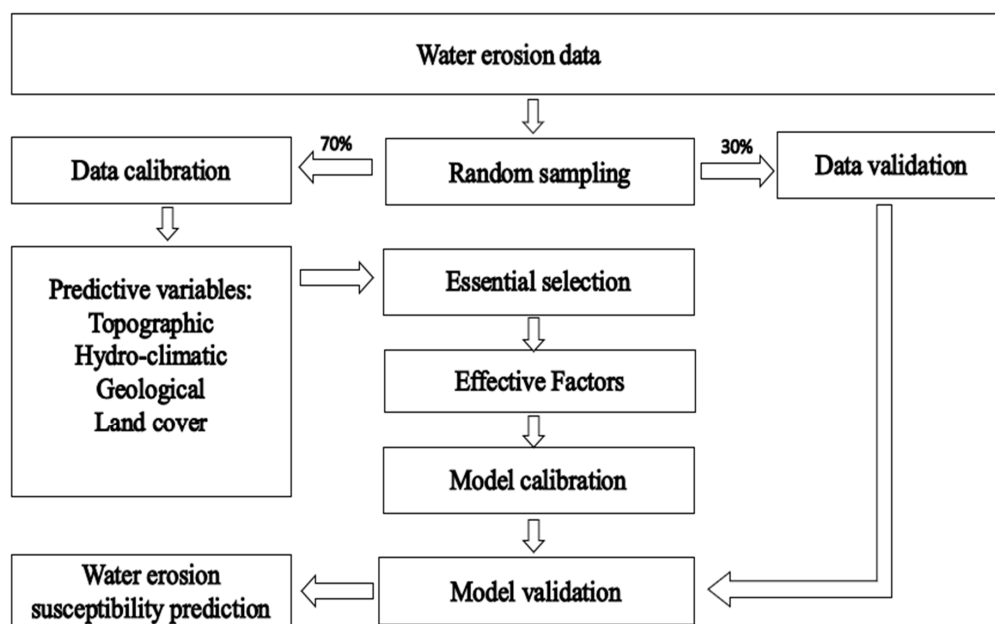


Figure 2. Schematic representation of the proposed method.

2.2.1. Field Data

It was necessary to know the locations of eroded and non-eroded areas for susceptibility mapping of the Nur-Rood watershed. Therefore, the locations (i.e., x and y coordinates) of 227 area (116 erosion locations and 111 non-erosion locations) were sampled through field surveys to model the water erosion susceptibility based on a binary scale (occurrence/non-occurrence). According to Sajedi-Hosseini et al. [8], the recorded soil erosion areas include different kinds of water erosions (such as sheet, rill, gully, and mass movements).

2.2.2. Predictive Variables

In this study, according to the literature review, 19 relevant factors regarding soil erosion were collected and prepared, including the topographic, hydro-climatic, geological, and land cover factors. The attributes of the factors are presented in Table 1. A brief description of each of the predictive factors is presented afterward.

Table 1. Characteristics of the considered factors for susceptibility mapping of water erosion.

Factors	Range/Class
Topographic factors:	
Elevation	732 to 4333 (m)
Slope	0 to 473.8 (%)
Aspect	Flat, north, northeast, east, southeast, south, southwest, west, northwest
Slope length (SL)	0 to 5613 (m)
Curvature	−35 to 25
Hydro-climate factors:	
Drainage density (DD)	0 to 3 (km/km ²)
Distance from stream (DFS)	0 to 5135 (m)
Topographic wetness index (TWI)	6 to 21.3
Stream power index (SPI)	0 to 150,768
Flow accumulation (FA)	0 to 1,630,717 (pixel)
Precipitation (PCP)	0 to 768 (mm)
Rainfall erosivity factor (R)	272 to 2078
Hydrologic soil group (HSG)	B ¹ , C ² , D ³

Table 1. Cont.

Factors	Range/Class
Geological factors ⁴ :	
Fault density (FD)	0 to 2.4 (km/km ²)
Lithology	TRJs, Pr, Mm.s.l, Pd, Odi, Tre, PZ2bvt, Tre1, Qs.D, Ebv, Tra.bv, Jl, Ek, K1bvt, Ktzt, Pldv, Jk, K2l2, Eksh
Soil texture	Sandy loam, loamy sand, loam, clay loam, sandy clay loam, clay
Land-cover factors:	
Normalized difference vegetation index (NDVI)	−0.07 to 0.63
Land use	Rangeland, residential, forest, agriculture, rock
Distance from road (DFR)	0 to 18,978 (m)

¹ Silt loam types of soils with a moderate infiltration rate. ² Sandy clay loam types of soils with low infiltration rates. ³ Clay loam, silty clay loam, sandy clay, silty clay, or clay with the highest runoff potential. ⁴ Definition of the geological factors include; TRJs: Dark grey shale and sandstone; Pr: Dark grey medium - bedded to massive limestone; Mm.s.l: Marl, calcareous sandstone, sandy limestone and minor conglomerate; Pd: Red sandstone and shale with subordinate sandy limestone; Odi: Diorite; Tre: Thick bedded grey o'olitic limestone; PZ2bvt: Basaltic volcanic tuff; Tre1: Thin bedded, yellow to pinkish argillaceous limestone with worm tracks; Qs.D: Unconsolidated wind-blown sand deposit including sand dunes; Ebv: Basaltic volcanic rocks; Tra.bv: Triassic, andesitic and basaltic volcanics; Jl: Light grey, thin-bedded to massive limestone; Ek: Well bedded green tuff and tuffaceous shale; K1bvt: Basaltic volcanic tuff; Ktzt: Thick bedded to massive, white to pinkish orbitolina bearing limestone; Pldv: Rhyolitic to rhyodacitic volcanics; Jk: Conglomerate, sandstone and shale with plant remains and coal seams; K2l2: Thick-bedded to massive limestone; Eksh: Greenish-black shale, partly tuffaceous with intercalations of tuff.

Topographic Parameters

The topographic parameters included the elevation, slope, aspect, slope length (SL), and curvature (Figure 3). These factors are influential regarding soil erosion velocity [34]. The different elevations (Figure 3a), aspects (Figure 3c), and curvature (Figure 3e) cause different conditions of evaporation, soil temperature, soil moisture, and solar radiation, which have different effects on the soil erosion. Furthermore, slope (Figure 3b) and SL (Figure 3d) affect the runoff velocity and volume, where a steeper slope or a longer SL can increase the soil erosion by water [8].

The topographic factors were produced using a digital elevation model (DEM) with a cell size of 30 m in the ArcGIS 10 software (Environmental Systems Research Institute, Redlands, CA, USA).

Hydro-Climate Factors

The hydro-climate factors included the drainage density (DD), distance from the stream (DFS), topographic wetness index (TWI), stream power index (SPI), flow accumulation (FA), precipitation (PCP), rainfall erosivity factor (R), and hydrologic soil group (HSG) (Figure 4). The DD (Figure 4a) is calculated from the sum of the length of all streams in the watershed area. The DD values depend on the permeability and resistance of the surface and deeper soil layers that affect water erosion [8]. Regarding the DFS (Figure 4b), the regions near streams are more susceptible to soil erosion [35]. The DD and DFS layers were created using line density and Euclidian distance tools, respectively, in geographic information system (GIS). TWI (Figure 4c) shows the soil moisture and water-saturated area of the watershed. SPI (Figure 4d) indicates the potential for erosion due to the water flow, in which higher values indicate a higher potential. TWI and SPI were produced using the SAGA GIS 2.0.7 software (SAGA User Group Association, Hamburg, Germany). The flow accumulation (FA) function (Figure 4e) computes the sum of the weight of all accumulated pixels upstream [36], which is most important for showing the water-accumulated pixels that affect the water erosion. The PCP (Figure 4f) and R (Figure 4g) were the climate factors considered to affect soil erosion. Their effects depend on soil attributes such as the soil texture, soil organic matter, and soil structure. The PCP map is produced by the mean annual precipitation of the gauge stations in the study area. The R factor is directly related to the soil erodibility. The best method for calculating it is a direct measurement of soil erosion in plots [37]. However, in this study, according to Takal et al. [38], an empirical equation was used to calculate this factor, as follows.

$$R = 0.0483P^{1.61}, \quad (1)$$

where R is the precipitation erosivity index ($\text{MJ}\cdot\text{mm}\cdot\text{ha}^{-1}\cdot\text{hr}^{-1}$) and P is the mean annual precipitation (mm).

The HSG (Figure 4h) indicates the infiltration and runoff generation rates that affect soil erosion. This layer is extracted from the digital soil map of the world [39] and it includes three groups: B, C, and D. Group B has moderately low runoff potential when completely humid. Soils in this group have 50 to 90% sand, 10 to 20% clay, and have sandy loam or loamy sand textures. Water transition across the soil is unrestricted. Group C soils have moderately high runoff potential when completely humid. They have less than 50% sand, 20 to 40% clay, and include sandy clay loam, silty clay loam, loam, silt loam, and clay loam textures. Group D soils have high runoff potential and the infiltration across the soil is very limited [40].

Geological Factors

Geological factors include the fault density (FD), lithology, and soil texture (Figure 5). The FD affects infiltration and runoff, which can affect soil erosion. Furthermore, the existence of a fault can accelerate the mass movements [41]. The layer of FD (Figure 5a) was produced in the ArcGIS environment by using the line density tool on the fault layer. The lithology has the greatest effect on erosion control. Erosion depends on the exposed material weathering attributes or the lithology [42,43]. The lithology map (Figure 5b) was taken from a geological survey done by the Iranian department of environment and had a scale of 1:100,000. The other important factor is soil texture (Figure 5c). Porosity and soil texture, along with the soil profile and surface, are the dominant soil attributes that influence soil erosion. An increase in the clay value of the soil causes a decrease in soil erosion [44]. The soil textures of the study area were clay, clay loam, loam, loamy sand, sandy clay loam, and sandy loam (Figure 5c).

Land Cover Factors

The land cover factors considered were the normalized difference vegetation index (NDVI), land-use, and distance from road (DFR) (Figure 6). The NDVI (Figure 6a) was extracted from Landsat satellite images for June 2018. The NDVI values range from -1 to 1 [45]. A watershed with a higher NDVI provides higher resistance against soil erosion [9,46]. The land uses of the study area included rangeland, residential, forest, agriculture, and rock (Figure 6b). The land-use map was received from the Iranian Water Resources Management Company (IWRMC). Roads are one of the man-made features that increase the availability of materials for transformation and increase the sediment yield in the watershed. Moreover, roads increase the runoff speed through collecting and concentrating the surface runoff in the given areas (such as near bridges); therefore, faster flows increase the erosion. The DFR layer (Figure 6c) was calculated using the line density tool within the ArcGIS environment.

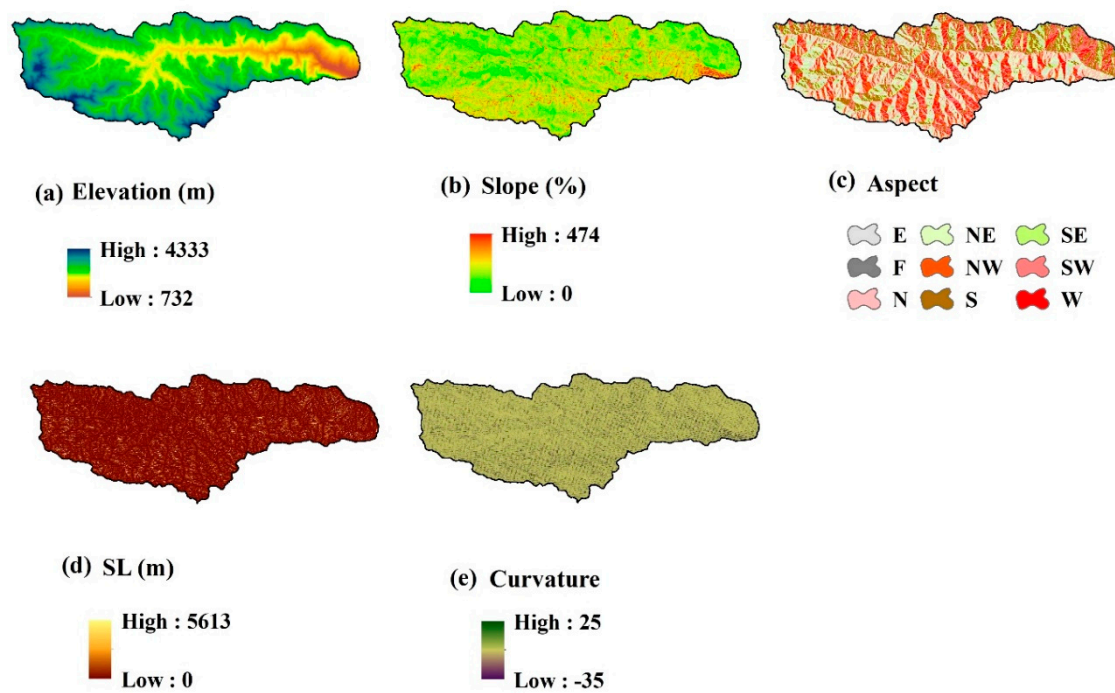


Figure 3. Topographic factors: (a) elevation, (b) slope, (c) aspect, (d) slope length (SL), and (e) curvature.

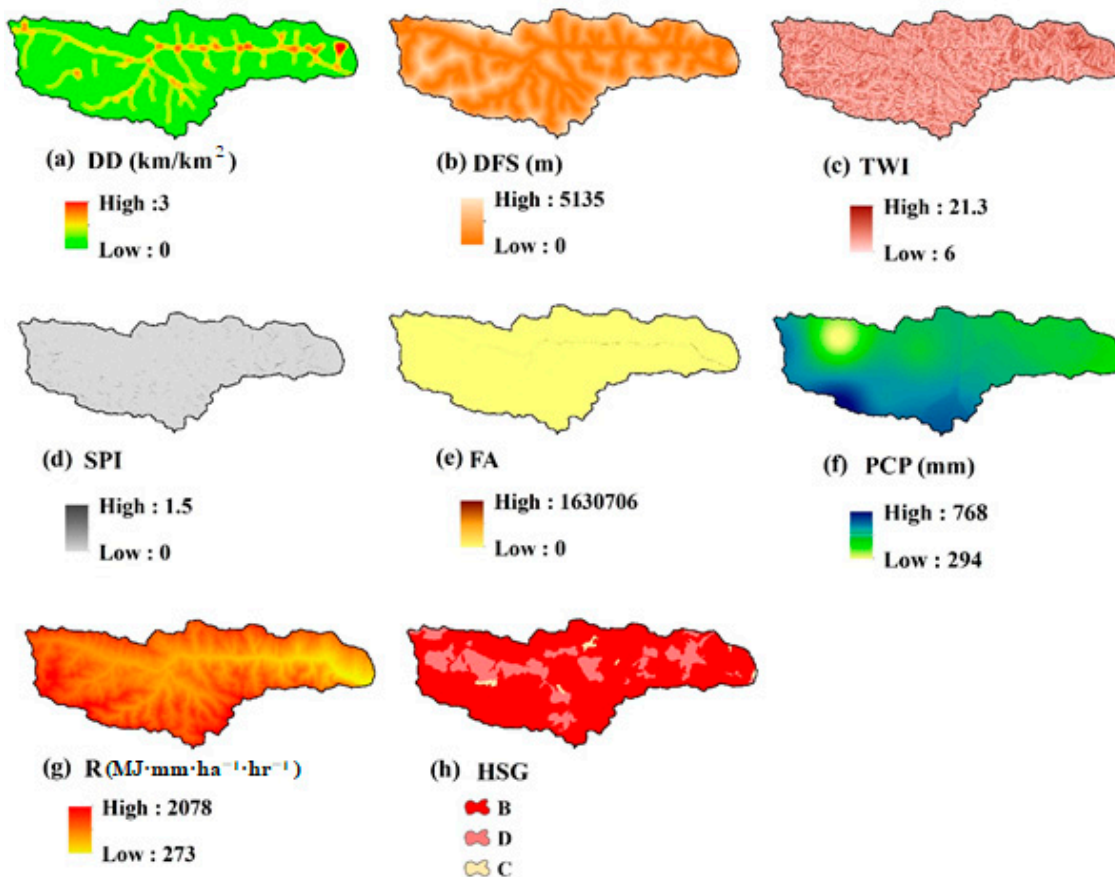


Figure 4. Hydro-climate factors: (a) drainage density (DD), (b) distance from the stream (DFS), (c) topographic wetness index (TWI), (d) stream power index (SPI), (e) flow accumulation (FA), (f) precipitation (PCP), (g) rainfall erosivity factor (R), and (h) hydrologic soil group (HSG).

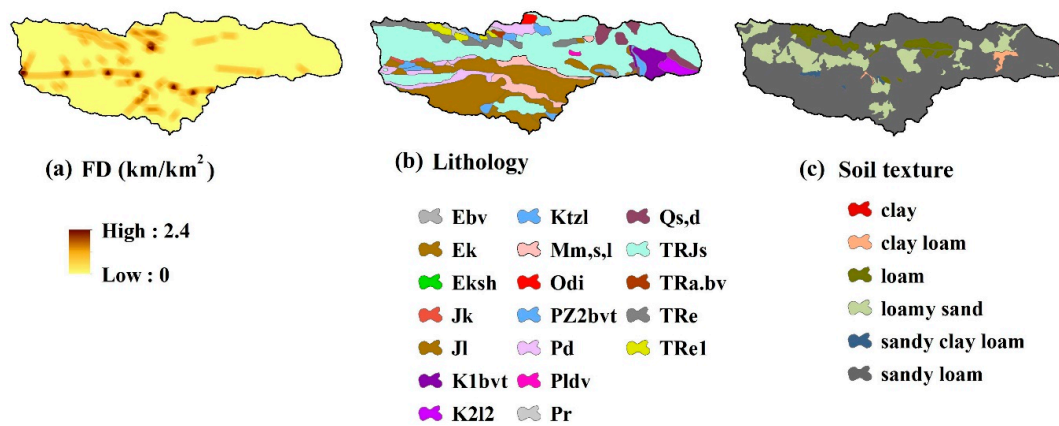


Figure 5. Geological factors: (a) fault density (FD), (b) lithology, and (c) soil texture.

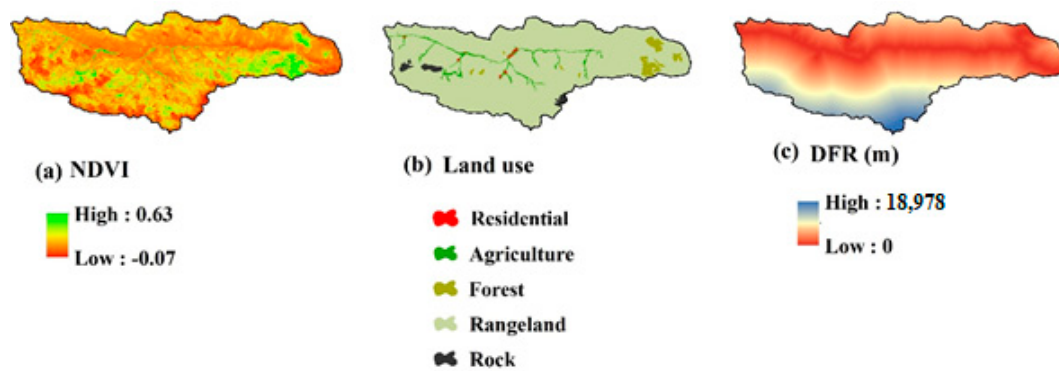


Figure 6. Land cover factors: (a) normalized difference vegetation index (NDVI), (b) land use, and (c) distance from road (DFR).

2.2.3. Feature Selection

To select the most important factors in the water erosion of soil based on parsimonious objectives from the large number of factors considered, the simulated annealing feature selection (SAFS) model was used. The SAFS method is based on the minimum energy configuration theory, whereby a solid is gradually cooled such that its structure is frozen [47]. Many studies have used this method for feature selection in environmental fields, such as flash-flood hazard assessment [48], dust and air quality evaluation [49], and earth fissure hazard prediction [50]; see Bertsimas and Tsitsiklis [47] for more details of the SAFS method.

In the current research, the SAFS was conducted using the k-fold ($k = 10$) cross-validation methodology and it was implemented in the Caret package [51] of the R software (4.0.2, R Core Team, Vienna, Austria).

2.2.4. Weighted Subspace Random Forest (WSRF)

Xu et al. [52] suggested a new random forest, namely, the WSRF model, which involves weighting the input variables and afterward opting for the variables that ensure each subspace always includes informative attributes. The WSRF model is implemented as multi-thread processes. This algorithm categorizes very high-dimensional data and sparse data with random forests made using small subspaces. A new variable weighting manner is applied for the variable subspace choice rather than the traditional random variable sampling in the random forest model [53]. More details of the WSRF model are presented in Xu et al. [52] and Zhao et al. [53]. The WSRF model was implemented using the “wsrf” package [53] in the R software using the k-fold ($k = 10$) cross-validation procedure.

2.2.5. Naive Bayes (NB)

The NB classifiers are a set of assortment algorithms that use Bayes' Theorem. This is a family of algorithms, where every pair of features being categorized is independent of each other; on the other hand, all of them share a common principle. The dataset is categorized into two sections:

- A response vector, which includes the value of the class variable.
- The feature matrix, which includes all the rows of the dataset and each row contains all the dependent features.

According to the primary naive Bayes hypothesis, each element must be independent and equal [54,55]; see Webb et al. [56] for more details of the NB model. The NB model was done using the k-fold ($k = 10$) cross-validation method in the "klar" package [57,58] within the R software.

2.2.6. The Gaussian Process with a Radial Basis Function Kernel (Gaussprradial)

Gaussian process regression is a vigorous, non-parametric Bayesian method used for solving regression problems and modeling unknown functions [59,60]. It can capture the different relationships between inputs and output variables by applying a hypothetically infinite number of parameters and allowing the dataset to determine the level of complexity via Bayesian inference [61]. The Gaussian process is parametrized using a kernel. One of the benefits of Gaussian process regression is the flexibility in choosing the kernel; furthermore, the different kernels can be combined to perform the regression [59]. In this study, the radial basis function network (RBF) was used to perform the Gaussian process. The Gaussprradial was performed in the R software using the "kernlab" package [62] using the k-fold ($k = 10$) cross-validation approach.

2.2.7. Model Calibration and Validation

The database, including the predictand and predictors, was randomly divided into the training (70%) and testing (30%) sets. A k-fold ($k = 10$) cross-validation methodology was used to calibrate the models. The models were assessed using testing datasets after the calibration using the features selected by the SAFS. Here, for the assessment of the models' performances, three classification evaluation metrics were used: accuracy, kappa, and the probability of detection (POD). The models' performances were represented as accuracy percentages. Kappa indicates the probability of agreement by chance using the likelihood of the model classification [63]. The metrics are computed as follows:

$$\text{Accuracy} = \frac{H + CN}{H + FA + M + CN} \quad (2)$$

where H (the number of hits), FA (the number of false alarms), M (the number of misses), and CN (the number of correct negatives) were computed from a contingency table.

$$\text{Kappa} = \frac{\text{Accuracy} - Pe}{1 - Pe}, \quad (3)$$

where Pe is the expected probability of chance agreement [64] that is computed using Equation (4):

$$Pe = \frac{(H + FA)(H + M) + (M + CN)(FA + CN)}{(H + FA + M + CN)^2} \quad (4)$$

The POD is a metric used to quantify the possibility of finding a specific detect. The POD is significantly linked to the subject of risk evaluation and probabilistic analyses of the components' integrity. The POD is the ratio of the correct predicted data to the total number observed occurrences. It ranges from 0 to 1, where 1 indicates a perfect score [49,50]. The metric is calculated using Equation (5):

$$\text{POD} = \frac{H}{H + M}. \quad (5)$$

3. Results and Discussion

3.1. Feature Selection Results

A relatively large number of factors, such as elevation, slope, aspect, SL, curvature, DD, DFS, TWI, SPI, FA, PCP, R, HSG, FD, lithology, soil texture, NDVI, land use, and DFR, were used in the current study to predict water erosion. The results of the feature selection using the SAFS algorithm are shown in Table 2. As can be seen, the minimum and maximum selected features were 8 and 14 variables, respectively, in the folds number of 8 (accuracy = 0.84, Kappa = 0.67) and 3 (accuracy = 0.92, Kappa = 0.83). The fold number 6 provided the worst performance (accuracy = 0.74, Kappa = 0.48), whereas the fold number 10 provided the best performance (accuracy = 0.93, Kappa = 0.87).

Table 2. Selected factors in each fold using the simulated annealing feature selection (SAFS) method.

Fold	Number of Selected Features	Selected Features	Accuracy	Kappa
1	10	Aspect, elevation, DFR, FA, lithology, HSG, NDVI, R, SL, soil texture	0.85	0.69
2	9	Aspect, DF, DFS, FA, lithology, NDVI, PCP, slope, TWI	0.75	0.49
3	14	DD, DF, DFR, DFS, FA, lithology, HSG, NDVI, R, PCP, slope, TWI, soil texture, SL	0.92	0.83
4	10	aspect, curvature, DD, DF, DFS, lithology, NDVI, SL, SPI, soil texture	0.91	0.82
5	9	Curvature, elevation, DF, lithology, HSG, NDVI, R, SL, SPI	0.86	0.72
6	9	Curvature, aspect, DD, DFR, FA, HSG, land use, NDVI, R, SL, slope, soil texture	0.74	0.48
7	9	DFR, DFS, FA, lithology, HSG, land use, NDVI, R, soil texture	0.90	0.81
8	8	DF, DFS, FA, NDVI, R, PCP, SL, TWI	0.84	0.67
9	13	aspect, curvature, DD, DF, DFS, lithology, land use, NDVI, R, SL, slope, soil texture, TWI	0.89	0.78
10	11	Aspect, DD, DF, FA, lithology, land use, NDVI, R, SL, SPI, soil texture	0.93	0.87
Average	10.2	-	0.86	0.72

According to the 10-fold results, the selected factors should be between the minimum and maximum selected features and should be mostly equal to the mean selected factors across all folds. However, the percentage of selected factors in all folds can be a good criterion for selecting the final variables [48–50]. Figure 7 shows the percentage of selected factors in all folds. Twelve variables had a frequency of at least 50% across all folds. As can be seen, the NDVI with a 100% frequency (F) was selected in all folds. However, the important role of the vegetation and NDVI is obvious and shown in previous studies [8,9,46]. Followed the NDVI (F = 100%), the variables of lithology (F = 80%), R (F = 80%), SL (F = 80%), FD (F = 70%), FA (F = 70%), soil texture (F = 60%), DFS (F = 60%), aspect (F = 50%), curvature (F = 50%), HSG (F = 50%), and DD (F = 50%) were selected.

Although the feature selection has largely not been done in studies on the water erosion of soil, the importance of these selected variables in the water erosion of soil is demonstrated by previous studies, such as those of De Baets et al. [46], Md. Rejaun et al. [9], Sajedi-Hosseini et al. [8], Di Stefano et al. [42], Arabameri et al. [43], Lin et al. [37], Choubin et al. [41], Auzet et al. [44], and Nekhay et al. [35].

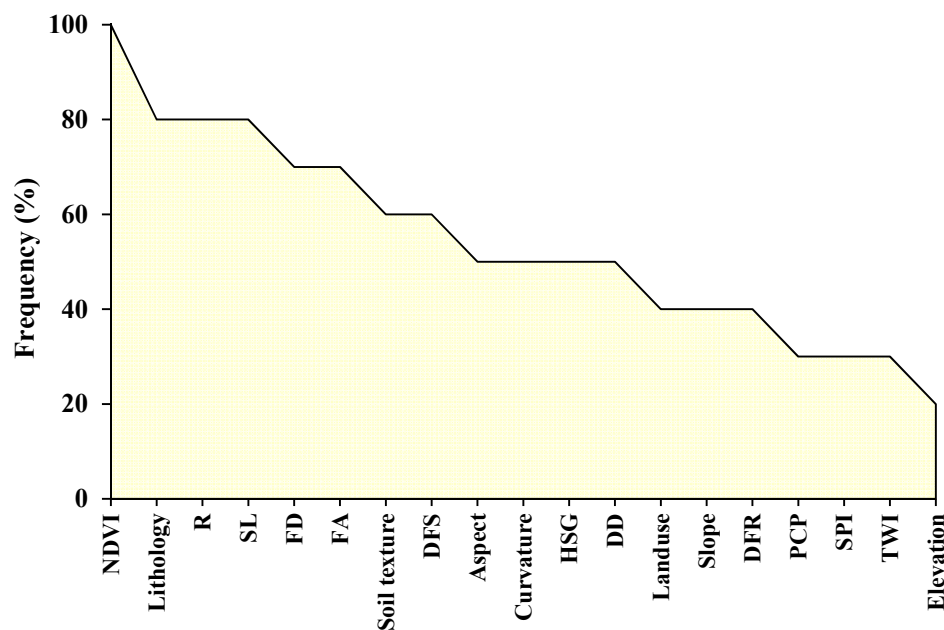


Figure 7. The percentage of selected factors in all folds.

3.2. Results of Water Erosion Modeling

The calibration of models was conducted using the “tunelength” function in the Caret R package [51]. The performance results of the three models (WSRF, Gaussprradial, and NB) were evaluated using the three statistics of accuracy, kappa, and the probability of detection (POD), which are presented in Table 3.

Table 3. The performances of the models using the testing dataset.

Statistic	WSRF	Gaussprradial	NB
Accuracy	0.91	0.88	0.85
Kappa	0.82	0.76	0.71
POD	0.94	0.94	0.94

WSRF: Weighted Subspace Random Forest, NB: Naive Bayes.

As can be seen from Table 3, the evaluation of the models’ performance indicated that the WSRF model had a higher accuracy (accuracy = 0.91), followed by the Gaussprradial (accuracy = 0.88) and NB (accuracy = 0.85) models. According to Monserud and Leemans [65], the kappa values indicated that all three models were in the “very good” degree of agreement (i.e., $0.70 < \text{Kappa} < 0.85$) (Table 3). However, like the accuracy, the kappa statistic for the WSRF model (Kappa = 0.82) was more than the Gaussprradial (Kappa = 0.76) and NB (Kappa = 0.71) models. Regarding the POD, the WSRF, Gaussprradial, and NB models showed an equal performance (POD = 0.94) (Table 3).

Generally, the evaluation of the applied machine learning (ML) models in this study indicated an acceptable performance for all the ML models. However, regarding the accuracy and kappa values, the models’ performances were ranked as follows: WSRF > Gaussprradial > NB. A direct comparison between the results of this study and previous ones is not possible because the application of the WSRF and Gaussprradial models was undertaken for the water erosion of soil for the first time. However, two novel ML models (WSRF and Gaussprradial) applied in this study indicated a better performance than the NB model that has previously been used in this field. Previous studies have indicated the accurate performance of the NB model in the assessment of soil erosion, such as Weihua et al. [66] and Nhu et al. [67].

3.3. Spatial Prediction of Water Erosion Susceptibility

After the calibration and validation of the models, the maps of the soil erosion probability were predicted using the values of the pixels throughout the study area. Then, the probability maps were classified into five susceptibility classes of very low, low, moderate, high, and very high based on the classification method of natural breaks through the ArcGIS software (Figure 8).

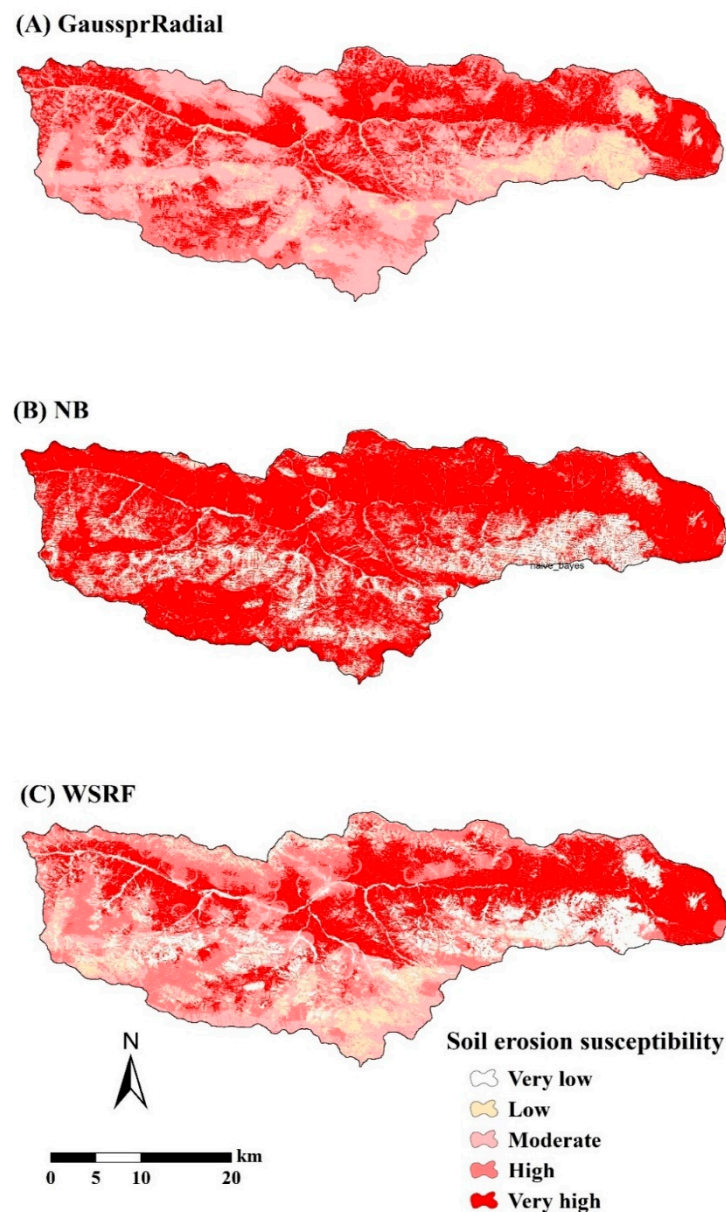


Figure 8. Spatial prediction of water erosion using various methods: (A) Gaussprradial, (B) NB, and (C) WSRF.

The area of the susceptibility classes found using each model is presented in Table 4. As can be seen, the Gaussprradial model predicted most of the area in the moderate class (about 450 km², 34.69% of the study area). The sum of the areas for low and very low classes was less than 7% (about 85 km²) (Figure 8A). According to the NB model, more than 65% of the study area (about 850 km²) was located in very high susceptibility zones (Figure 8B). Results of the WSRF model indicated that the classes of the very low, low, moderate, high, and very high susceptibilities covered 11.91% (154.52 km²),

9.76% (126.6 km²), 20.25% (262.64 km²), 28.66% (371.87 km²), and 29.42% (381.7 km²) of the study area, respectively (Figure 8C).

Table 4. Area of the susceptibility classes found using each model.

Susceptibility Class	Gaussprradial		NB		WSRF	
	Area (km ²)	Area (%)	Area (km ²)	Area (%)	Area (km ²)	Area (%)
Very low	0.08	0.01	115.73	8.92	154.52	11.91
Low	84.87	6.54	75.49	5.82	126.60	9.76
Moderate	450.02	34.69	93.00	7.17	262.64	20.25
High	386.76	29.81	163.39	12.59	371.87	28.66
Very high	375.60	28.95	849.72	65.50	381.70	29.42

Although the predicted models indicated different areas for each class, there was something in common for all predicted maps. By comparing the predicated maps (Figure 8) with the NDVI map (Figure 5a), it was clear that the susceptibility maps were approximately matched with the NDVI and lithology maps. For example, the green areas in the east of the region on the NDVI map (Figure 6a) had higher values of NDVI, which correspond to lower values on the water erosion susceptibility maps (Figure 7). Furthermore, the higher susceptibility values (Figure 8) corresponded to the TRJ lithology (Figure 5b). TRJs include dark grey shale, claystone, siltstone, and sandstone of the Shemshak formation. In this formation, various kinds of water erosion, such as rill, riverbank, gully, and badland erosions can be seen. This agrees with the SAFS results, which indicated that the NDVI and lithology were the most important variables during the feature selection.

4. Conclusions

This study focused on the probability of water erosion occurring in the Nur-Rood watershed. Using the SAFS model, the most important factors were selected among nineteen parameters, namely, NDVI, lithology, R, SL, FD, FA, soil texture, DFS, aspect, curvature, HSG, and DD. Based on the performance analysis of the machine learning (ML) models, the two novel applied ML models of WSRF (accuracy = 0.91, Kappa = 0.82) and Gaussprradial (accuracy = 0.88, Kappa = 0.76) displayed better performances than the NB (accuracy = 0.85, Kappa = 0.71) model that has previously been used in this field. The predicted maps created using the ML models indicated the different areas for each susceptibility class but it was obvious that the susceptibility maps were approximately matched with the NDVI and lithology maps (which were identified as the most important variables). One of the main limitations in this study that also occurs in other spatial modeling studies is that different scales are used for the input variables all over the world. Although all of the input variables were resampled into the same spatial resolution, the data collection and sampling of them were not on the same scale; this is an inevitable limitation for the time being. It may be the case that the data availability of the NDVI (30 m resolution) helped this variable to be the most important variable during the soil erosion modeling compared with the other variables (such as the soil dataset and lithology with scales of 1:100,000 or more). Despite these limitations, producing the water erosion susceptibility maps in developing countries can be a useful tool for sustainable management, the conservation of watersheds, the reduction of soil degradation, and alleviating water quality decline.

Author Contributions: Conceptualization, A.M.; data curation, F.S.-H. and B.C.; formal analysis, F.S.-H., B.C., and G.R.; investigation, F.S.-H., B.C., F.T., and G.R.; methodology, F.S.-H. and B.C.; project administration, F.S.-H. and F.T.; resources, F.T. and G.R.; software, B.C. and G.R.; supervision, A.M. and A.A.D.; validation, F.T.; visualization, G.R. and A.A.D.; writing—original draft, A.M.; writing—review and editing, F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Morgan, R.P.C. *Soil Erosion and Conservation*; John Wiley & Sons: Hoboken, NJ, USA, 2005; ISBN 9788578110796.
2. Vanmaercke, M.; Poesen, J.; Verstraeten, G.; de Vente, J.; Ocakoglu, F. Sediment yield in Europe: Spatial patterns and scale dependency. *Geomorphology* **2011**, *130*, 142–161. [\[CrossRef\]](#)
3. Panagos, P.; Borrelli, P.; Meusburger, K.; Alewell, C.; Lugato, E.; Montanarella, L. Estimating the soil erosion cover-management factor at the European scale. *Land Use Policy* **2015**, *48*, 38–50. [\[CrossRef\]](#)
4. García-Ruiz, J.M.; Beguería, S.; Lana-Renault, N.; Nadal-Romero, E.; Cerdà, A. Ongoing and Emerging Questions in Water Erosion Studies. *Land Degrad. Dev.* **2017**, *28*, 5–21. [\[CrossRef\]](#)
5. Scott, A. *Water Erosion in the Murray-Darling Basin: Learning from the Past*; Elsevier: New York, NY, USA, 2001.
6. Sharma, A.; Tiwari, K.N.; Bhadoria, P.B.S. Effect of land use land cover change on soil erosion potential in an agricultural watershed. *Environ. Monit. Assess.* **2011**, *173*, 789–801. [\[CrossRef\]](#)
7. Alkharabsheh, M.M.; Alexandridis, T.K.; Bilas, G.; Misopolinos, N.; Silleos, N. Impact of Land Cover Change on Soil Erosion Hazard in Northern Jordan Using Remote Sensing and GIS. *Procedia Environ. Sci.* **2013**, *19*, 912–921. [\[CrossRef\]](#)
8. Sajedi-Hosseini, F.; Choubin, B.; Solaimani, K.; Cerdà, A.; Kavian, A. Spatial prediction of soil erosion susceptibility using a fuzzy analytical network process: Application of the fuzzy decision making trial and evaluation laboratory approach. *Land Degrad. Dev.* **2018**, *29*, 3092–3103. [\[CrossRef\]](#)
9. Md. Rejaur, R.; Shi, Z.H.; Chongfa, C. Land use/land cover change analysis using geo- information technology: Two case studies in Bangladesh and China. *Int. J. Geoinformatics* **2009**, *5*, 25–37.
10. Vaezi, A.R.; Abbasi, M.; Bussi, G.; Keesstra, S. Modeling Sediment Yield in Semi-Arid Pasture Micro-Catchments, NW Iran. *Land Degrad. Dev.* **2017**, *28*, 1274–1286. [\[CrossRef\]](#)
11. Afshar, F.A.; Ayoubi, S.; Jalalian, A. Soil redistribution rate and its relationship with soil organic carbon and total nitrogen using ¹³⁷Cs technique in a cultivated complex hillslope in western Iran. *J. Environ. Radioact.* **2010**, *101*, 606–614. [\[CrossRef\]](#)
12. Khalili Moghadam, B.; Jabarifar, M.; Bagheri, M.; Shahbazi, E. Effects of land use change on soil splash erosion in the semi-arid region of Iran. *Geoderma* **2015**, *241*, 210–220. [\[CrossRef\]](#)
13. Pradeep, G.S.; Krishnan, M.V.N.; Vijith, H. Identification of critical soil erosion prone areas and annual average soil loss in an upland agricultural watershed of Western Ghats, using analytical hierarchy process (AHP) and RUSLE techniques. *Arab. J. Geosci.* **2015**, *8*, 3697–3711. [\[CrossRef\]](#)
14. Choubin, B.; Rahmati, O.; Tahmasebipour, N.; Feizizadeh, B.; Pourghasemi, H.R. Application of fuzzy analytical network process model for analyzing the gully erosion susceptibility. *Adv. Nat. Technol. Hazards Res.* **2019**, *48*, 105–125. [\[CrossRef\]](#)
15. Angileri, S.E.; Conoscenti, C.; Hochschild, V.; Märker, M.; Rotigliano, E.; Agnesi, V. Water erosion susceptibility mapping by applying Stochastic Gradient Treeboost to the Imera Meridionale River Basin (Sicily, Italy). *Geomorphology* **2016**, *262*, 61–76. [\[CrossRef\]](#)
16. Svoray, T.; Michailov, E.; Cohen, A.; Rokah, L.; Sturm, A. Predicting gully initiation: Comparing data mining techniques, analytical hierarchy processes and the topographic threshold. *Earth Surf. Process. Landf.* **2012**, *37*, 607–619. [\[CrossRef\]](#)
17. Lee, S.; Ryu, J.H.; Lee, M.J.; Won, J.S. Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. *Environ. Geol.* **2003**, *44*, 820–833. [\[CrossRef\]](#)
18. Pourghasemi, H.R.; Rossi, M. Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: A comparison between GLM, GAM, MARS, and M-AHP methods. *Theor. Appl. Climatol.* **2017**, *130*, 609–633. [\[CrossRef\]](#)
19. Pradhan, B. A 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.* **2013**, *51*, 350–365. [\[CrossRef\]](#)
20. Taner San, B. An evaluation of SVM using polygon-based random sampling in landslide susceptibility mapping: The Candir catchment area (western Antalya, Turkey). *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 399–412. [\[CrossRef\]](#)
21. Dou, J.; Yamagishi, H.; Pourghasemi, H.R.; Yunus, A.P.; Song, X.; Xu, Y.; Zhu, Z. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. *Nat. Hazards* **2015**, *78*, 1749–1776. [\[CrossRef\]](#)

22. Gorsevski, P.V.; Brown, M.K.; Panter, K.; Onasch, C.M.; Simic, A.; Snyder, J. Landslide detection and susceptibility mapping using LiDAR and an artificial neural network approach: A case study in the Cuyahoga Valley National Park, Ohio. *Landslides* **2016**, *13*, 467–484. [[CrossRef](#)]
23. Hong, H.; Pourghasemi, H.R.; Pourtaghi, Z.S. Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* **2016**, *259*, 105–118. [[CrossRef](#)]
24. Yuan, L.; Zhang, Q.; Li, W.; Zou, L. Debris Flow Hazard Assessment Based on Support Vector Machine. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*; IEEE: Piscataway, NJ, USA, 2006; Volume 11, pp. 4221–4224.
25. Chang, T.C. Risk degree of debris flow applying neural networks. *Nat. Hazards* **2007**, *42*, 209–224. [[CrossRef](#)]
26. Chang, T.C.; Chao, R.J. Application of back-propagation networks in debris flow prediction. *Eng. Geol.* **2006**, *85*, 270–280. [[CrossRef](#)]
27. Eustace, A.H.; Pringle, M.J.; Denham, R.J. A risk map for gully locations in central Queensland, Australia. *Eur. J. Soil Sci.* **2011**, *62*, 431–441. [[CrossRef](#)]
28. Rahmati, O.; Tahmasebipour, N.; Haghizadeh, A.; Pourghasemi, H.R.; Feizizadeh, B. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* **2017**, *298*, 118–137. [[CrossRef](#)]
29. Arabameri, A.; Rezaei, K.; Cerda, A.; Lombardo, L.; Rodrigo-Comino, J. GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *Sci. Total Environ.* **2019**, *658*, 160–177. [[CrossRef](#)] [[PubMed](#)]
30. Pourghasemi, H.R.; Yousefi, S.; Kornejady, A.; Cerdà, A. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.* **2017**, *609*, 764–775. [[CrossRef](#)]
31. Garosi, Y.; Sheklabadi, M.; Conoscenti, C.; Pourghasemi, H.R.; Van Oost, K. Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. *Sci. Total Environ.* **2019**, *664*, 1117–1132. [[CrossRef](#)]
32. Mao, D.; Zeng, Z.; Wang, C.; Lin, W. Support Vector Machines with PSO Algorithm for Soil Erosion Evaluation and Prediction. In *Proceedings of the Third International Conference on Natural Computation (ICNC 2007)*, Haikou, China, 24–27 August 2007; IEEE: Piscataway, NJ, USA, 2007; Volume 1, pp. 656–660.
33. Azareh, A.; Rafiei Sardooi, E.; Choubin, B.; Barkhori, S.; Shahdadi, A.; Adamowski, J.; Shamshirband, S. Incorporating multi-criteria decision-making and fuzzy-value functions for flood susceptibility assessment. *Geocarto Int.* **2019**. [[CrossRef](#)]
34. Salcedo-Sanz, S.; Ghamisi, P.; Piles, M.; Werner, M.; Cuadra, L.; Moreno-Martínez, A.; Izquierdo-Verdiguier, E.; Muñoz-Mari, J.; Mosavi, A.; Camps-Valls, G. Machine Learning Information Fusion in Earth Observation: A Comprehensive Review of Methods, Applications and Data Sources. *Inf. Fusion.* **2020**, *22*, 480–545. [[CrossRef](#)]
35. Nekhay, O.; Arriaza, M.; Boerboom, L. Evaluation of soil erosion risk using Analytic Network Process and GIS: A case study from Spanish mountain olive plantations. *J. Environ. Manag.* **2009**, *90*, 3091–3104. [[CrossRef](#)]
36. Emadi, M.; Taghizadeh-Mehrjardi, R.; Cherati, A.; Danesh, M.; Mosavi, A.; Scholten, T. Predicting and Mapping of Soil Organic Carbon Using Machine Learning Algorithms in Northern Iran. *Remote Sens.* **2020**, *12*, 2234.
37. Lin, B.S.; Thomas, K.; Chen, C.K.; Ho, H.C. Evaluation of soil erosion risk for watershed management in Shenmu watershed, central Taiwan using USLE model parameters. *Paddy Water Environ.* **2016**, *14*, 19–43. [[CrossRef](#)]
38. Yu, B.; Rosewell, C.J. A Robust estimator of the R-factor for the universal soil loss equation. *Trans. Am. Soc. Agric. Eng.* **1996**, *39*, 559–561. [[CrossRef](#)]
39. Food and Agriculture Organization of the United Nations. *Digital Soil Map of the World and Derived Soil Properties*; Food and Agriculture Organization of the United Nations: Bangkok, Thailand, 2003.
40. Kadam, A.K.; Kale, S.S.; Pande, N.N.; Pawar, N.J.; Sankhua, R.N. Identifying Potential Rainwater Harvesting Sites of a Semi-arid, Basaltic Region of Western India, Using SCS-CN Method. *Water Resour. Manag.* **2012**, *26*, 2537–2554. [[CrossRef](#)]
41. Choubin, B.; Borji, M.; Mosavi, A.; Sajedi-Hosseini, F.; Singh, V.P.; Shamshirband, S. Snow avalanche hazard prediction using machine learning methods. *J. Hydrol.* **2019**, *577*, 123929. [[CrossRef](#)]

42. Di Stefano, C.; Ferro, V.; Porto, P.; Tusa, G. Slope curvature influence on soil erosion and deposition processes. *Water Resour. Res.* **2000**, *36*, 607–617. [\[CrossRef\]](#)
43. Arabameri, A.; Pourghasemi, H.R. Spatial Modeling of Gully Erosion Using Linear and Quadratic Discriminant Analyses in GIS and R. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 299–321. [\[CrossRef\]](#)
44. Auzet, A.V.; Poesen, J.; Valentin, C. Soil patterns as a key controlling factor of soil erosion by water. *Catena* **2002**, *46*, 85–87. [\[CrossRef\]](#)
45. Crippen, R.E. Calculating the vegetation index faster. *Remote Sens. Environ.* **1990**, *34*, 71–73. [\[CrossRef\]](#)
46. De Baets, S.; Poesen, J.; Reubens, B.; Wemans, K.; De Baerdemaeker, J.; Muys, B. Root tensile strength and root distribution of typical Mediterranean plant species and their contribution to soil shear strength. *Plant Soil* **2008**, *305*, 207–226. [\[CrossRef\]](#)
47. Bertsimas, D.; Tsitsiklis, J. Simulated annealing. *Stat. Sci.* **1993**. [\[CrossRef\]](#)
48. Hosseini, F.S.; Choubin, B.; Mosavi, A.; Nabipour, N.; Shamshirband, S.; Darabi, H.; Haghighi, A.T. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Sci. Total Environ.* **2020**, *711*. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Choubin, B.; Abdolshahnejad, M.; Moradi, E.; Querol, X.; Mosavi, A.; Shamshirband, S.; Ghamisi, P. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Sci. Total Environ.* **2020**, *701*. [\[CrossRef\]](#)
50. Choubin, B.; Mosavi, A.; Alamdarloo, E.H.; Hosseini, F.S.; Shamshirband, S.; Dashtekian, K.; Ghamisi, P. Earth fissure hazard prediction using machine learning models. *Environ. Res.* **2019**, *179*. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Max, K.; Weston, S.; Wing, J.; Williams, A.; Keefer, C.; Engelhardt, A. Package ‘caret’. Classification and Regression Training. *Adsbabs Harv. Edu* **2011**, *138*, 454–466.
52. Xu, B.; Huang, J.Z.; Williams, G.; Wang, Q.; Ye, Y. Classifying very high-dimensional data with random forests built from small subspaces. *Int. J. Data Warehous. Min.* **2012**, *8*, 44–63. [\[CrossRef\]](#)
53. Zhao, H.; Williams, G.J.; Huang, J.Z. Wsr: An R package for classification with scalable weighted subspace random forests. *J. Stat. Softw.* **2017**, *77*. [\[CrossRef\]](#)
54. Rennie, J.D.M.; Shih, L.; Teevan, J.; Karger, D. Tackling the Poor Assumptions of Naive Bayes Text Classifiers. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; Volume 2, pp. 616–623.
55. Schneider, K.-M. A Comparison of Event Models for Naive Bayes Anti-Spam E-Mail Filtering. In Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary, 12–17 April 2003.
56. Webb, G.I.; Boughton, J.R.; Wang, Z. Not so naive Bayes: Aggregating one-dependence estimators. *Mach. Learn.* **2005**, *58*, 5–24. [\[CrossRef\]](#)
57. Roever, C.; Raabe, N.; Luebke, K.; Ligges, U.; Szepannek, G. Package “klaR”; Springer: Berlin, Germany, 2020.
58. Weihs, C.; Ligges, U.; Luebke, K.; Raabe, N. klaR Analyzing German Business Cycles. In *Data Analysis and Decision Support*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 335–343.
59. Schulz, E.; Speekenbrink, M.; Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **2018**, *85*, 1–16. [\[CrossRef\]](#)
60. Gershman, S.J.; Blei, D.M. A tutorial on Bayesian nonparametric models. *J. Math. Psychol.* **2012**, *56*, 1–12. [\[CrossRef\]](#)
61. Williams, C.K.I. Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond. In *Learning in Graphical Models*; Springer: Dordrecht, The Netherlands, 1998; pp. 599–621.
62. Alexandros, A.; Smola, A.; Hornik, K. Kernel-Based Machine Learning Lab-Package ‘kernlab’; Springer: Berlin, Germany, 2019.
63. Darabi, H.; Choubin, B.; Rahmati, O.; Torabi Haghighi, A.; Pradhan, B.; Kløve, B. Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *J. Hydrol.* **2019**, *569*, 142–154. [\[CrossRef\]](#)
64. Beucher, A.; Møller, A.B.; Greve, M.H. Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark. *Geoderma* **2019**, *352*, 351–359. [\[CrossRef\]](#)
65. Monserud, R.A.; Leemans, R. Comparing global vegetation maps with the Kappa statistic. *Ecol. Modell.* **1992**, *62*, 275–293. [\[CrossRef\]](#)

66. Weihua, L.; Yonggang, W.; Dianhui, M.; Yan, Y. Region Assessment of Soil Erosion Based on Naive Bayes. In Proceedings of the 2007 International Conference on Computational Intelligence and Security, CIS 2007, Harbin, China, 15–19 December 2007.
67. Nhu, V.-H.; Shirzadi, A.; Shahabi, H.; Singh, S.K.; Al-Ansari, N.; Clague, J.J.; Jaafari, A.; Chen, W.; Miraki, S.; Dou, J.; et al. Shallow Landslide Susceptibility Mapping: A Comparison between Logistic Model Tree, Logistic Regression, Naïve Bayes Tree, Artificial Neural Network, and Support Vector Machine Algorithms. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2749. [[CrossRef](#)] [[PubMed](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).