# A New Approach For Smart Soil Erosion Modelling: Integration of Empirical And Machine Learning Models

**Mohammadtaghi Avand** ( mt.avand70@gmail.com )
Tarbiat Modares University https://orcid.org/0000-0001-7196-5051
**Maziar Mohammadi**
Tarbiat Modares University
**Fahimeh Mirchooli**
Tarbiat Modares University
**Ataollah Kavian**
Sari Agricultural Sciences and Natural Resources University
**John P Tiefenbacher**
Texas State University

**Research Article**

**Keywords:** Machine learning, RUSLE, Soil erosion, Spatial modelling, Talar watershed.

**Posted Date:** September 29th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-809330/v1

# A New Approach for Smart Soil Erosion Modelling: Integration of Empirical and Machine Learning Models

**Mohammadtaghi Avand[*1a,b], Maziar Mohammadi[2], Fahimeh Mirchooli[3], Ataollah Kavian[4], and John P. Tiefenbacher[5]**

[1a] Department of Watershed Management and Engineering, Faculty of Natural Resources, Tarbiat Modares University, Tehran, Iran

[1b] Department of Forests, Rangelands, and Watershed Management Engineering, Kohgiluyeh & Boyer Ahmad Agricultural and Natural Resources Research and Education Centre, AREEO, Yasouj, Iran.

[2] Department of Watershed Management and Engineering, Faculty of Natural Resources, Tarbiat Modares University, Tehran, Iran

[3] Department of Watershed Management and Engineering, Faculty of Natural Resources, Tarbiat Modares University, Tehran, Iran

[4] Faculty of Natural Resources, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

[5] Department of Geography, Texas State University, San Marcos, TX, USA

Corresponding author: mt.avand70@gmail.com

## Abstract

Despite advances in artificial intelligence modelling, the lack of soil erosion data and other watershed information is still one of the important factors limiting soil-erosion modelling. Additionally, the limited number of parameters and the lack of evaluation criteria are major disadvantages of empirical soil-erosion models. To overcome these limitations, we introduce a new approach that integrates empirical and artificial intelligence models. Erosion-prone locations (erosion ≥16 tons/ha/year) are identified using RUSLE model and a soil-erosion map is prepared

using random forest (RF), artificial neural network (ANN), classification tree analysis (CTA), and generalized linear model (GLM). This study uses 13 factors affecting soil erosion in the Talar watershed, Iran, to increase prediction accuracy. The results reveal that the RF model has the highest prediction performance (AUC=0.95, Kappa=0.87, Accuracy=0.93, and Bias=0.88), outperforming the three machine-learning models. The results show that slope angle, land use/land cover, elevation, and rainfall erosivity are the factors that contribute the most to soil erosion propensity in the watershed. Curvature and topography position index (TPI) were removed from the analysis due to multicollinearity with other factors. The results can be used to improve the identification of hot spots of soil erosion, especially in watersheds for which soil-erosion data are limited.

**Keywords:** Machine learning, RUSLE, Soil erosion, Spatial modelling, Talar watershed.

## 1. Introduction

Human-induced soil erosion is considered to be the most widespread form of (Pournader et al. 2018) and main contributor (> 85 %) to land degradation (Tang et al. 2015). The United Nations Food and Agriculture Organization has determined that soil erosion is one of the 10 main threats to soils globally world (Chalise et al. 2019) and that it contributes to food insecurity (Phinzi et al. 2020). Several studies have highlighted the on- and off-site consequences of soil erosion, including water quality, agricultural productivity, and sedimentation of rivers, which cause numerous social and economic issues (Sharma et al., 2011; Pournader et al., 2018). The quantification of soil erosion, identification of the most important factors contributing to soil-erosion inducement, and

the mapping of areas that are most prone to soil erosion are required for successful design and implementation of soil and water conservation projects.

Numerous quantitative models designed for different uses and requiring diverse data have been proposed for to estimate soil erosion. Some, like Water Erosion Prediction Project Model (WEPP) (Nearing et al. 1989), Areal Non-point Source Watershed Environment Response Simulation (ANSWERS) (Beasley et al. 1980), and the European Soil Erosion Model (EuroSEM), are physical models. Some, like the Universal Soil Loss Equation (USLE) (Wischmeier and Smith 1978), Revised Universal Soil Loss Equation (RUSLE) (Renard et al. 1991), have been used to assess soil erosion empirically. RUSLE is the model most commonly used to assess erosion as it is simple to use, it is relatively inexpensive, and it is reliable (Pham et al. 2018). This model predicts soil loss based on rainfall erosivity, topography, soil erodibility, vegetation cover, and land use management techniques (Vaezi et al. 2008). Koirala et al. (2019) and Atoma et al. (2020) are examples of its applications. Combining RUSLE with remote sensing data and a geographical information system (GIS) can make erosion assessments time-efficient and cost-effective, particularly for large watersheds. Numerous studies have used this combination to map soil erosion (Nyesheja et al. 2019; Mohammed et al. 2020), and to assess the influences of different factors like soil particle size (Wang et al. 2008), crop types (Ruysschaert et al. 2007), and land use (Vanacker et al. 2019).

RUSLE is challenged by the impacts of other soil characteristics and environmental features (Gayen et al. 2019). To resolve these issues, machine learning (ML) algorithms have been tested and are becoming more popular because they improve accuracy, performance, and have other strengths (Mousavi et al. 2017). Models using artificial neural-networks (ANN), logistic regression (LR) models, random forest models, generalized linear models (GLMs), and classification tree

analysis (CTA) have been developed to assess the spatial relationships between conditioning factors and soil erosion rates.

This study integrates RUSLE with ML to analyze soil erosion at the watershed scale. Soil erosion is assessed in the Talar watershed, Mazandaran Province, Iran to determine the locations of erosion hot spots. The identified hot spots are used to increase soil-erosion prediction accuracy with ML models. The ML models tested include RF, ANN, GLM, and CTA and they are used to explore the relationships between erosion and several conditioning factors. The most important innovation in this research is the use of RUSLE to improve hot spot identification to increase ML accuracy.

## 2. Methods

### 2.1. Description of study area

The 210000 ha Talar watershed extends from mountains to the Caspian Sea. The south-to-north flowing Talar River passes through the city of Ghaemshahr and across a lowland plain before emptying into the Caspian (Fig. 1). Much of the watershed is rangeland and forest, especially in the highlands. Irrigated agriculture, dry farming, and urban/residential are the other primary land uses in the study area. The highest elevation is 3910 m above sea level (asl) at Shaljmar Zardin in the southwestern part of the watershed. The lowest point is at the outlet of the Talar river at 217 m asl (Fig. 1). Due to the topography and proximity to the sea, precipitation is orographic and averages 752.7 mm annually. The rainfall gradient increases linearly with elevation, especially in the southern and southwestern parts of the watershed, but at elevations above 2400 m, rainfall decreases, particularly in the southeastern part of the basin due to reduced moisture availability. There are 58 types of soil with varying physical and chemical properties in the region. In addition,

the hydrological soil groups A, B, and C cover 60%, 21%, and 19% of the watershed, respectively (Mohammadi et al. 2020).

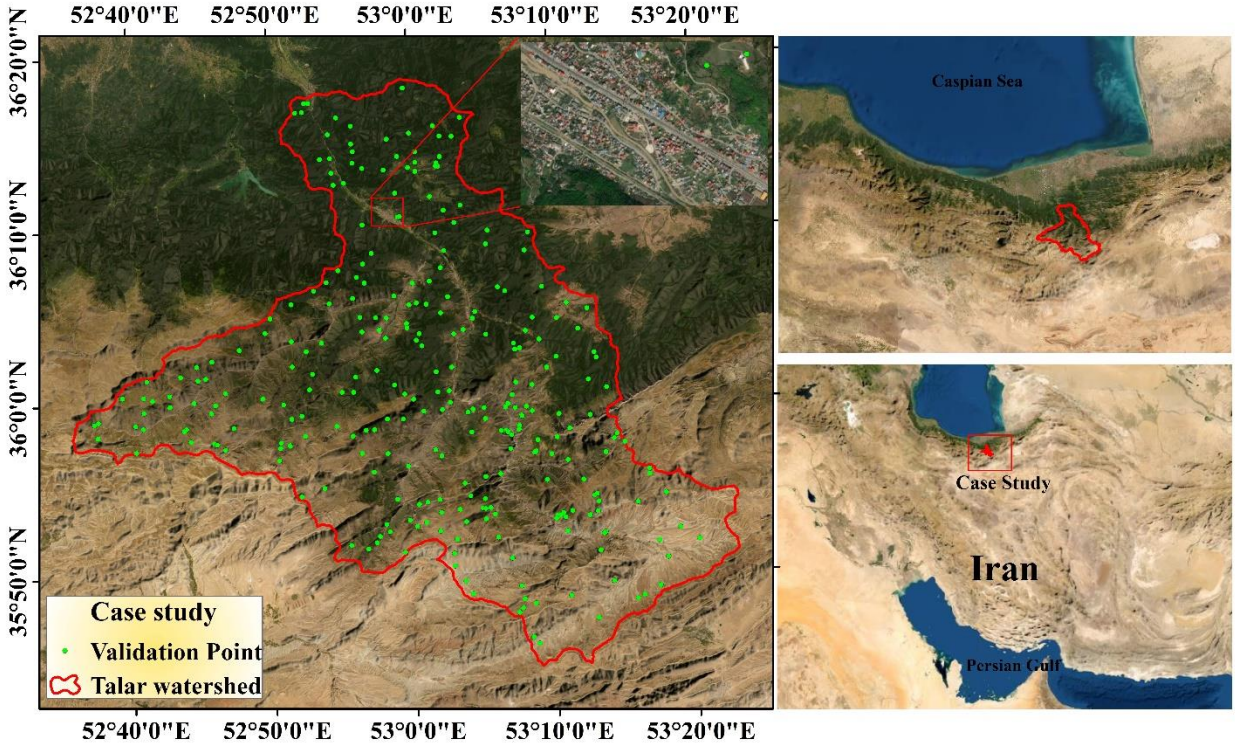



**Fig. 1** The location of study area in the Mazandaran province, northern Iran

## 2.2. Methods and software

Spatial modelling of soil erosion in the Talar watershed was performed by integrating RUSLE with ML models in 5 main steps: (1) Preparation of parameters (R, K, LS, C and P) and erosion map using RUSLE in GIS; (2) Extraction of locations of severe erosion locations (> 16 tons/ha/yr) for model training; (3) Preparation of the parameters that affect erosion and the ML models using training data in R software; (4) Removal of parameters with similar performance through

multicollinearity analysis; (5) Introduction of the relative importance of the parameters to cause soil erosion; and (6) Comparison of the results from the ML models to select the best model (Fig.2).

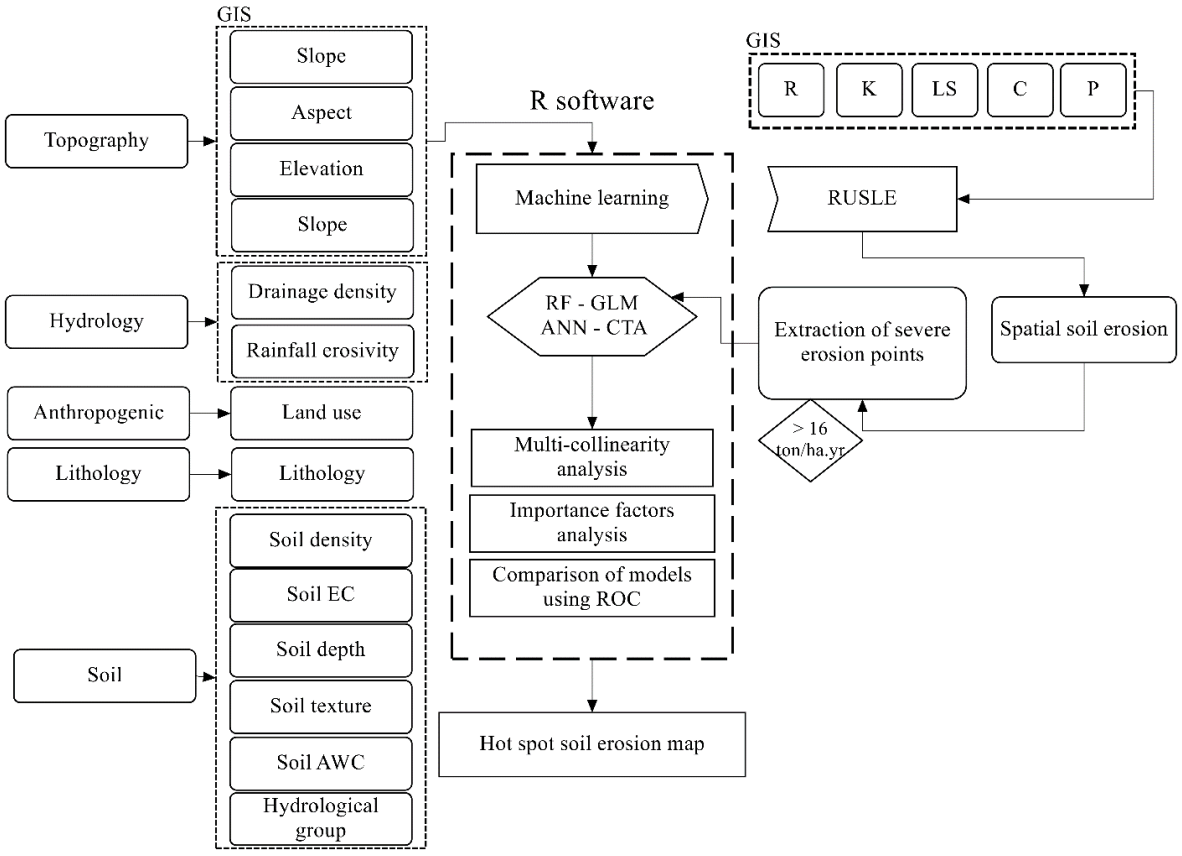



**Fig. 2** A flowchart of this study

## 2.3. Spatial data-sets

The watershed boundary was delineated using a digital elevation model (DEM) having a resolution of 30 m with the Arc-SWAT extension in ArcGIS (Table 1). The rainfall records were obtained from seven rainfall monitoring stations in the watershed to calculate rainfall erosivity in RUSLE. Land uses were classified from a satellite image (OLI) using supervised classification in ENVI version 5.3. RUSLE was mapped (Fig. 3a). The RUSLE map was classified by erosion rates and

areas >16 ton/ha/yr are considered zones of severe erosion. These areas were extracted and 500 locations were randomly selected as training points for the ML models (Fig. 3b). The average soil erosion rate in Iran is 15-20 tons/ha/yr. This is 3-4 times higher than the global average due to climatic, topographic and anthropogenic conditions (Karamidehkordi, 2010; Arekhi et al., 2012).

**Table 1.** Variables, types, and sources of data used in study

| Data | Data item | Type | Period | Source |
|---|---|---|---|---|
| DEM | Raster map (30m resolution) | Raster | | Iran National Cartographic Center |
| Geology | Geology map and information | Shape file | 1997 | Geological Survey & Mineral Explorations of Iran (GSI) |
| Land use | Land use classes | Raster | 2018 | USGS (https://earthexplorer.usgs.gov/) |
| Weather data | Observed rainfall | Daily | 2000-2017 | Iran Meteorological Organization |
| Land cover | Vegetation cover | Shape file | 2017 | (Mohammadi et al., 2017) |
| Soil | Soil physical and chemical properties | Shape file | 2002 | (Engineering Services Company 2002) |
| River | Names and tributaries | Shape file | | Iran National Cartographic Center |

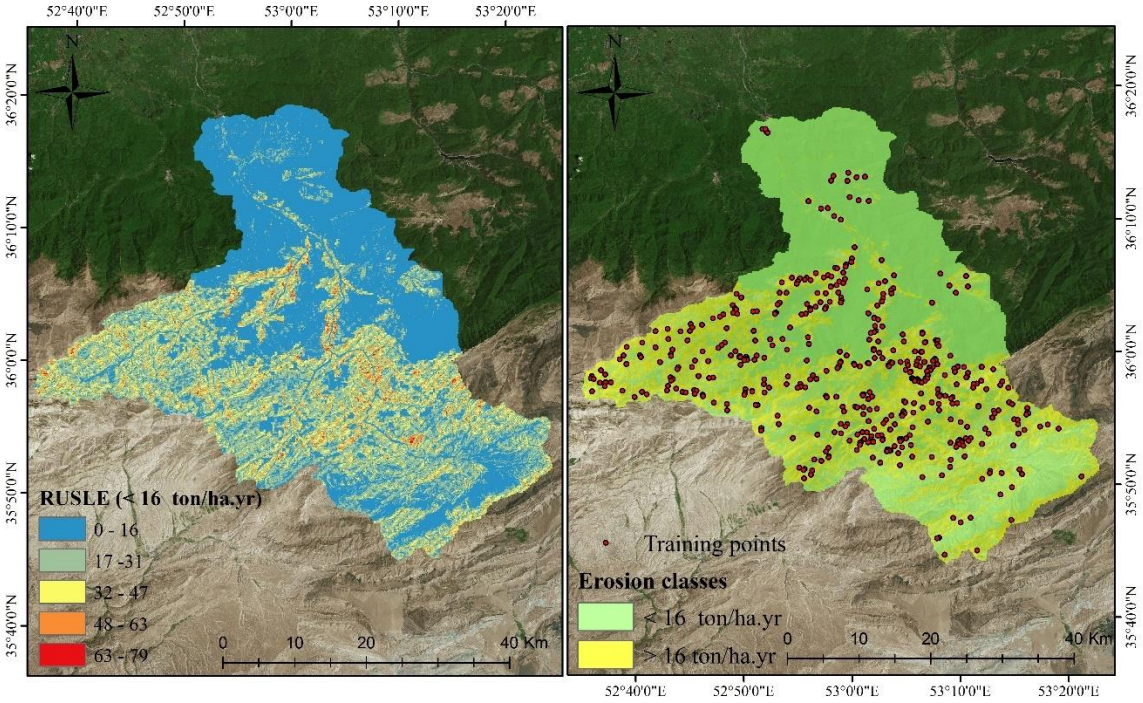



**Fig. 3** RUSLE map (a) and training point selection (b)

## 2.4. Building Models for Predicting Soil Erosion

### Selection of Predictors

### Topographical factors

Three topographical factors – slope length, gradient, and shape – are important drivers of erosion due to their influences on the speed and volume of runoff (Bagio et al. 2017). The DEM provided basic land surface data regarding slope, aspect, curvature, drainage area, drainage networks, and topographic indices (Mukherjee et al. 2013). Slope angle, aspect and, elevation were extracted from the DEM and classified in ArcGIS 10.3. Slope angles ranged from 0 to 72 and there are 5 aspect classes from -1 (flat area) and 0 to 359 degrees. Topography controls flow velocity and erosion and deposition rates, and influences erosion through conditions that determine flow convergence and divergence on the land surface (Pourghasemi et al. 2014).

## Lithology and soil factors

Lithology and soil properties also affect soil and land degradation processes like piping and gullying (Bouchnak et al. 2009, Amine et al. 2012, Faulkner 2013, Lei et al. 2020). Data that describe lithology (Table 2) and soil properties were entered for electrical conductivity (EC), soil available water (SAW), texture, density, depth, and soil hydrological group into the GIS.

**Table 2.** Geological units of the Talar watershed

| | Unit | Description | Area covered (%) |
|---|---|---|---|
| 1 | Murm | Marl and gypsiferous marl with sandstone intercalations | 0.03 |
| 2 | Mc | Conglomerate and sandstone | 0.07 |
| 3 | Olc,s | Conglomerate and sandstone | 0.13 |
| 4 | Pgkc | Coarse grained, polygenic conglomerate with sandstone intercalations | 0.26 |
| 5 | Mur | Marl, gypsiferous marl, sandstone and conglomerate | 0.26 |
| 6 | Qft2 | Low level pediment fan and valley terrace deposits | 0.34 |
| 7 | K1bvt | Basaltic volcanic tuff | 0.36 |
| 8 | Jd | Argillaceous limestone with intercalations of calcareous shale | 0.37 |
| 9 | $K_2l_2$ | Thick - bedded to massive limestone | 0.56 |
| 10 | Db-sh | Undifferentiated limestone, shale and marl | 0.58 |
| 11 | Pr | Medium - bedded to massive limestone | 0.62 |
| 12 | Qft1 | High level piedmont fan and valley terrace deposits | 0.68 |
| 13 | Czl | Undifferentiated unit, composed of dark red micaceous siltstone and sandstone | 0.70 |
| 14 | E1l | Nummulitic limestone | 0.71 |
| 15 | K2l1 | Hyporite bearing limestone | 0.78 |
| 16 | Jk | Conglomerate, sandstone and shale with plant remains and coal seams | 0.82 |
| 17 | Pz | Undifferentiated lower Paleozoic rocks | 1.12 |
| 18 | Ktzl | Thick bedded to massive, white to pinkish orbitolina bearing limestone | 1.79 |
| 19 | Cb | Alternation of dolomite, limestone and variegated shale | 2.14 |
| 20 | TRe | Thick bedded grey o'olitic limestone | 2.63 |
| 21 | Ku | Upper cretaceous, undifferentiated rocks | 3.34 |
| 22 | Ek | Well bedded green tuff and tophaceous shale | 3.76 |
| 23 | Qm | Swamp and marsh | 4.63 |
| 24 | Jl | Light grey, thin - bedded to massive limestone | 5.94 |
| 25 | Plc | Polymictic conglomerate and sandstone | 7.48 |
| 26 | E1m | Marl, gypsiferous marl and limestone | 8.36 |
| 27 | Mm,s,l | Marl, calcareous sandstone, sandy limestone and minor conglomerate | 10.92 |
| 28 | TRJs | Shale and sandstone (Shemshak formation) | 40.62 |

147 **Hydrologic factors**

148 Hydrological factors like drainage density, river buffer, and precipitation also influence erosion
149 and affect sediment-transport processes (Wulf et al. 2010, Ali et al. 2014). Drainage density was
150 determined from the DEM in the GIS. Rainfall erosivity was determined from precipitation data
151 recorded at seven stations.

152

153 **Anthropogenic factors**

154 Land use is one of the most important anthropogenic factors in soil degradation. Development of
155 agricultural and residential areas, and unscientific and unsustainable agricultural operations for
156 livelihoods, over grazing and deforestation have led to increasing degradation and pressure on soil
157 resources, especially in developing countries (Tadesse et al. 2017, Tang et al. 2020). The major
158 land use types in the Talar watershed are rangeland, forestland, dry or rain-fed agriculture, irrigated
159 agriculture, and urban/residential land use.

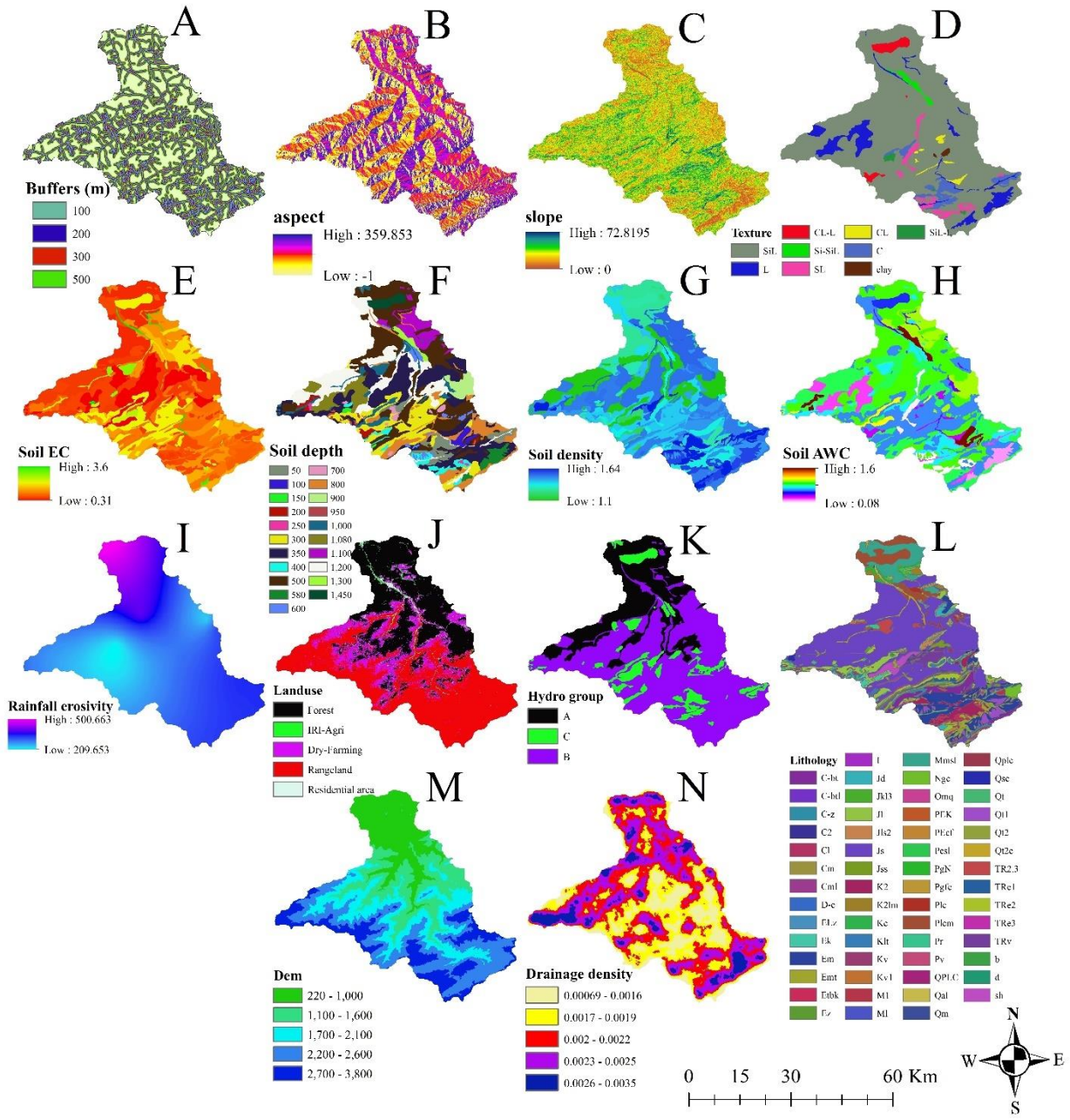



**Fig. 4** The distribution of variables that influence soil erosion in Talar Watershed – A: river buffer, B: aspect, C: slope, D: texture, E: EC, F: soil depth, G: soil density, H: soil available water, I: rainfall erosivity, J: land use, K: hydrological group, L: lithology M: elevation, N: drainage density.

## 2.5. Application of the Models

166 **2.5.1 RUSLE model**

167 The RUSLE model shows how rills are affected by soils, climate, topography, and land use.
168 Raindrop impacts and surface runoff cause inter-rill soil erosion (Renard et al. 1991). This equation
169 is a function of five factors that were mapped as rasters: slope length and gradient, rainfall
170 erosivity, soil erodibility, land cover and vegetation management, and land use practices. These
171 parameters vary in time and space and depend on other variables. The average annual soil loss (*A*)
172 (ton/hectare/year) at any location can be estimated using RUSLE (Eq. 1) (Renard et al. 1991):
173

$$A = R \times K \times LS \times C \times P \quad (1)$$

174

175 where *R* is rainfall erosivity (MJ/hectare/mm); *K* is the soil erodibility (ton/hectare/year); *LS* is the
176 slope length and slope gradient (dimensionless); *C* is the cover management practice
177 (dimensionless); and *P* is the conservation support or erosion control practices (dimensionless).
178 Wischmeier and Smith's equation was used to calculate the soil erodibility factor (Eq.2)
179 (Wischmeier 1976):
180

$$100K = 2.1M^{1.14} \times 10^{-4} \times (12 - \%OM) + 3.25(S - 2) + 2.5(P - 3) \quad (2)$$

181

182 where *K* is the soil erodibility, *M* is particle size (% silt + %very fine sand) * (100 - % clay), *OM*
183 is the organic matter content (%), and *S* and *P* are the soil structure and permeability classes,
184 respectively. The number of meteorological stations in the study area is low. Therefore, to increase
185 accuracy, rainfall gauges located outside the watershed but near the watershed boundaries were

186 employed as well. Renard and Freimund's (1994) method is used to calculate annual rainfall
187 erosivity from monthly and annual average of rainfall data (Eqs. 3-5):
188

$$R = 0.7397F^{1.847},\ F < 5\ mm \quad (3)$$

$$R = 95.77 - 6.081F + 0.477F^{2},\ F > 55\ mm \quad (4)$$

$$F = \sum_{i=1}^{12} P_i^2 / \bar{P} \quad (5)$$

189
190 where $F$ is the modified index value, $P_i$ is average monthly precipitation (mm), $\bar{P}$ is average annual
191 precipitation (mm), and $R$ is rainfall erosivity (MJ/ha/mm). Some have estimated C using NDVI
192 (normalized different vegetation index) (De Jong 1994, Lin et al. 2002, Wang et al. 2002). The
193 vegetation cover was determined from Landsat satellite images from the USGS website
194 (https://earthexplorer.usgs.gov/). Vegetation cover maps were created using supervised
195 classification of imagery based on the maximum likelihood algorithm in the ENVI 5.1 by applying
196 overall accuracy and the κ coefficient to determine classification accuracy. NDVI is derived from
197 the red and near-infrared bands using the following (Eq. 6) (Zhang et al. 2010):
198

$$NDVI = (\rho_{NIR} - \rho_{Red}) / (\rho_{NIR} + \rho_{Red}) \quad (6)$$

199
200 where $\rho_{Red}$ and $\rho_{NIR}$ are the spectral reflectance of the red and near-infrared bands in each pixel of
201 the image. $C$ was obtained for each pixel with (Eq. 7):
202

$$C\ Factor = 0.407 - 0.5953 \times NDVI \quad (7)$$

203

*LS* incorporates gradient, slope length and shape to determine sediment yield (Pradhan et al. 2012). A program written in Arc Macro Language was acquired from http://www.iamg.org and was used to calculate *LS* (Hickey 2000). It was updated in 2004 with C++ programming language. The program automatically processes DEM data to calculate *LS* (Van Remortel et al. 2004). A value for *P* was determined for each land use (Troeh et al., 1980). A map of soil erosion in the Talar watershed was produced by integrating the distribution of the five RUSLE factors and then reclassifying their product into three erosion classes of low, moderate, and high (Fig. 5).

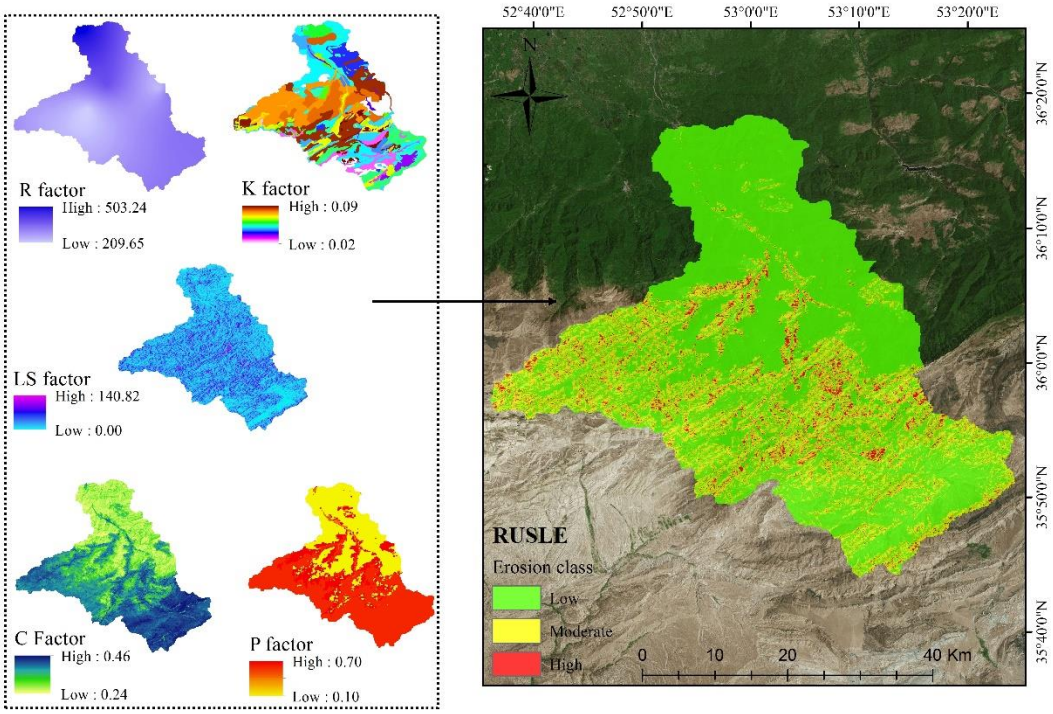



**Fig. 5** The RUSLE factors and erosion classes

### 2.5.2 Application of random forest (RF)

RF is a popular ML algorithm that is used for both classification and regression. This model includes decision tree that are used for determine the relationships between soil erosion intensity and the contributing conditioning factors. RF combines numerous trees. It handles random trees that implement subsets of the observations by bootstrapping. A random selection of training data is used to create a model, and the data that are not used to calibrate k-trees in the bagging process are regarded as out-of-bag (OOB) data. To use RF, it is necessary to define two parameters: the number of factors to be used in each tree-building process (mtry) and the number of trees to be built in the forest (ntree). RF can determine the importance of each conditioning factor based on the amount of increase in the error of prediction as the OBB for the factor are permuted while all others remain fixed (Naghibi et al. 2017). The RF modelling was performed using the "Random Forest" package of R. The output map was input into ArcGIS to create the soil-erosion map of the Talar watershed.

### 2.5.3 Application of artificial neural network (ANN)

The design of ANNs is inspired by a biological neural system. They mimic human brain performance. Generally, ANNs consist of interconnected processing elements (PEs) called nodes. These nodes are organized into three or more layers that include an input layer, one or more hidden layers, and an output layer. The number of neurons in the input and output layers depends on the problem, and the number of neurons in hidden layers are determined through trial and error, a process that is fixed for the particular application (Isazadeh et al. 2017). Many types of architectures, like the multilayer feed-forward and back-propagation algorithms, have been proposed. In this study, ANN was structured on a multilayer-perceptron architecture and the back-

238 propagation algorithm for the training. The number of hidden layers in the ANN was set at 5 and
239 the learning rate of the model was set at 9.
240

241 **2.5.4 Application of generalized linear model (GLM)**

242 GLM statistical models are increasingly used to study the relationships between dependent and
243 independent variables. GLMs can be used with categorical and continuous data, or a combination
244 thereof (Atkinson et al. 1998). This model does not force data into unnatural scales. The data can
245 be linear, non-linear, or have non-constant variance structures (Mirchooli et al. 2019). The
246 relationships between combinations of predictors and the mean of the response variable is
247 represented as a link function. GLMs can be based on Poisson, Binomial, Gaussian, or Gamma
248 distributions. The general form of this model is:
249 $l = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k$
250 where l is the linear predictor and is related to the expectation, u. The link function form is:
251 $g(u) = l$
252

253 **2.5.5 Application of classification-tree analysis (CTA)**

254 CTA is a rule-based process that generates a binary tree by dividing a node into yes or no predictor
255 values. It is called the binary recursive partitioning process. Each node is based on a single factor,
256 and the rule generated at each step minimizes the variability within each resulting subset, splitting
257 them further based on each relationship. The result of CTA is a hierarchical binary tree. The
258 response to each factor is premised on the values of prior inputs higher in the tree so that the
259 relationships and interactions between factors will be modelled automatically. This model is easy
260 and straight forward to interpret. It produces results that are intuitive and easy to visualize.

## 2.6. Model Calibration and Validation

The 30% of the soil-erosion hot spots that were not employed in the modeling process, were used to validate the results of these models (Rahmati et al. 2016). The accuracies of the models were determined using the receiver operating characteristic (ROC) curve and associated area under the curve (AUC). ROC indicates how well a model predicts an outcome (i.e., soil erosion). It is widely used to measure performance. ROC ranges from 0 to 1, but values below 0.5 indicate the model is random and not meaningful, but values above 0.5 indicate the model has some predictive value (Pourghasemi et al. 2013).

# 3. Results

## 3.1. Multi-collinearity analysis

The variance inflation factor (VIF) and tolerance (TOL) analyses were used to eliminate factors are high correlated to others. Highly correlated variables will have very similar effects on soil erosion, so removal can improve the accuracies of predictive models. The threshold value of VIF for removal is 5. The analysis revealed that all factors except curvature (VIF=5.43) and TPI (VIF=6.47) had VIF values below 5. Curvature and TPI were removed from the modeling process (Table 3).

**Table 3.** Multi-collinearity analysis between the factors used

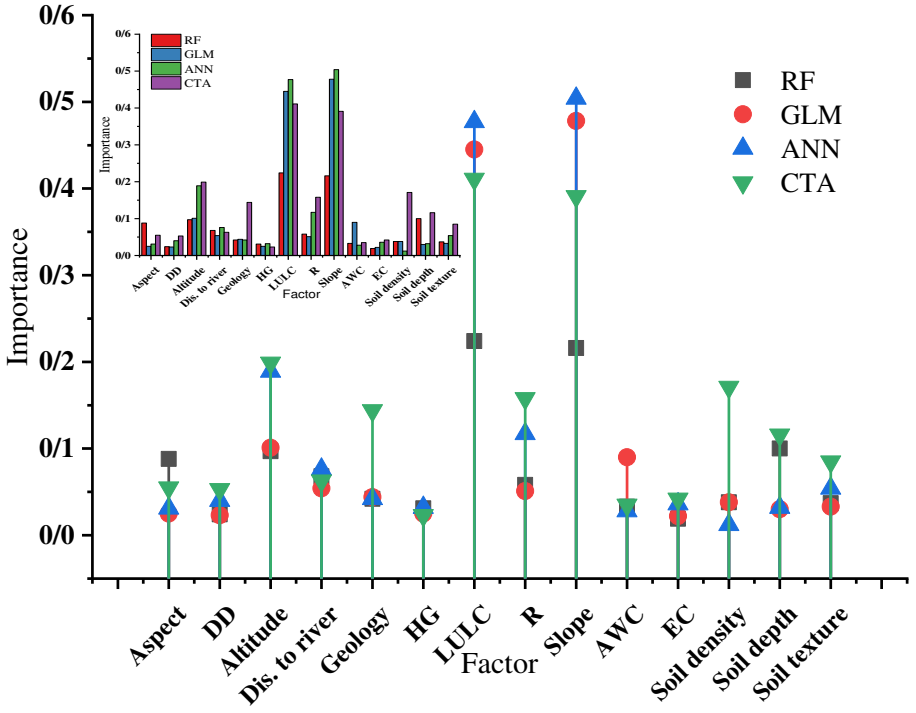
| Number | Variables | VIF (th≤5) | TOL |
|---|---|---|---|
| 1 | Aspect | 1.13 | 0.88 |
| 2 | Drainage density | 1.20 | 0.83 |
| 3 | Elevation | 2.83 | 0.35 |
| 4 | Distance to river | 1.35 | 0.74 |
| 5 | Geology | 1.29 | 0.77 |
| 6 | Hydro-group | 2.30 | 0.43 |
| 7 | Land use | 1.84 | 0.54 |

| | | | |
|---|---|---|---|
| 8 | Rainfall erosivity | 1.88 | 0.53 |
| 9 | Slope | 1.05 | 0.95 |
| 10 | Soil AWC | 1.10 | 0.90 |
| 11 | Soil EC | 1.16 | 0.86 |
| 12 | Soil density | 1.55 | 0.64 |
| 13 | Soil depth | 1.67 | 0.59 |
| 14 | Soil texture | 1.12 | 0.89 |
| 15 | Curvature | 5.43 | 0.18 |
| 16 | TPI | 6.47 | 0.15 |

279

280 **3.2. Importance factors analysis**

281 The relative importance of the conditioning factors in the RF, ANN, GLM, and CTA models was
282 determined (Fig. 6). Slope angle, land use/land cover, elevation, and rainfall erosivity are the most
283 important soil erosion conditioning factors in all four models of the Talar watershed. These factors
284 are also reflected in the previous spatial analysis of soil erosion events (i.e., observations) (Fig. 3)
285 which demonstrated that there were locations where several factors contributed to soil erosion.
286 However, there were significant differences in the contribution of soil erosion factors among the
287 assorted models.




288

289 **Fig. 6** Importance soil erosion factors affecting

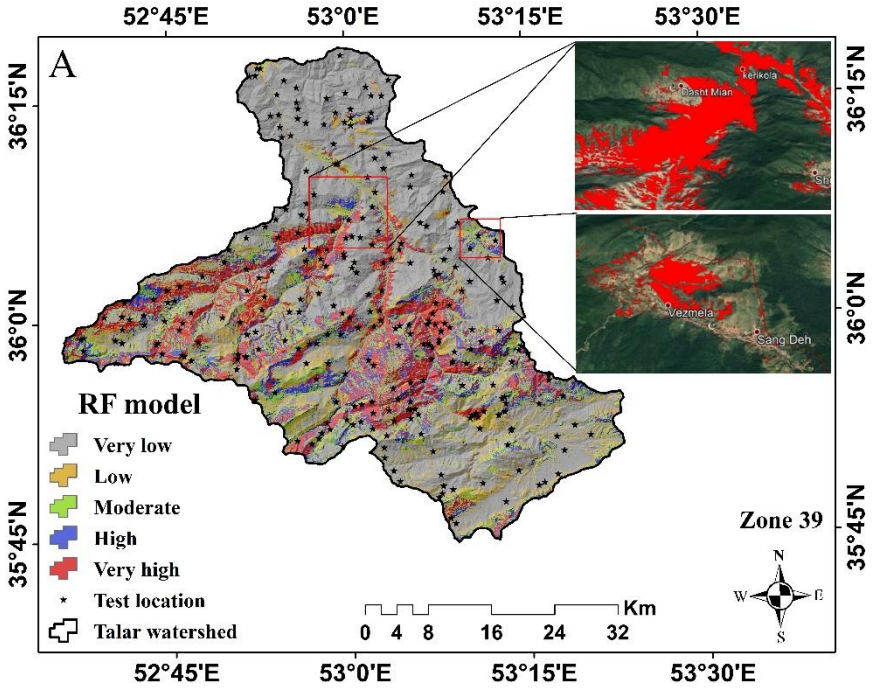
290

291 **3.3. Spatial Prediction of Water Erosion Susceptibility**

292 To better visualize and prioritize areas of erosion, the classes of the models' predictions were
293 mapped and classified as very low, low, moderate, high and very high (Fig. 7). These were divided
294 using the natural break algorithm in Arc Map software (Talukdar et al. 2020, Yousefi et al. 2020).
295 Soil erosion susceptibility is an estimation of risk that an area experiences erosion in any future
296 year. In these maps, red indicates areas with very high potential for soil erosion and gray indicates
297 that, comparatively, there is very low potential for soil erosion (though all of these zones would
298 likely generate more than 16 tons/ha/yr. In all four maps, it can be seen that the highest potential
299 for erosion is found in upstream and middle portions of the watershed. There is, however, a
300 significant difference between the proportions of the watershed classified at particular levels of
301 susceptibility by the four models.

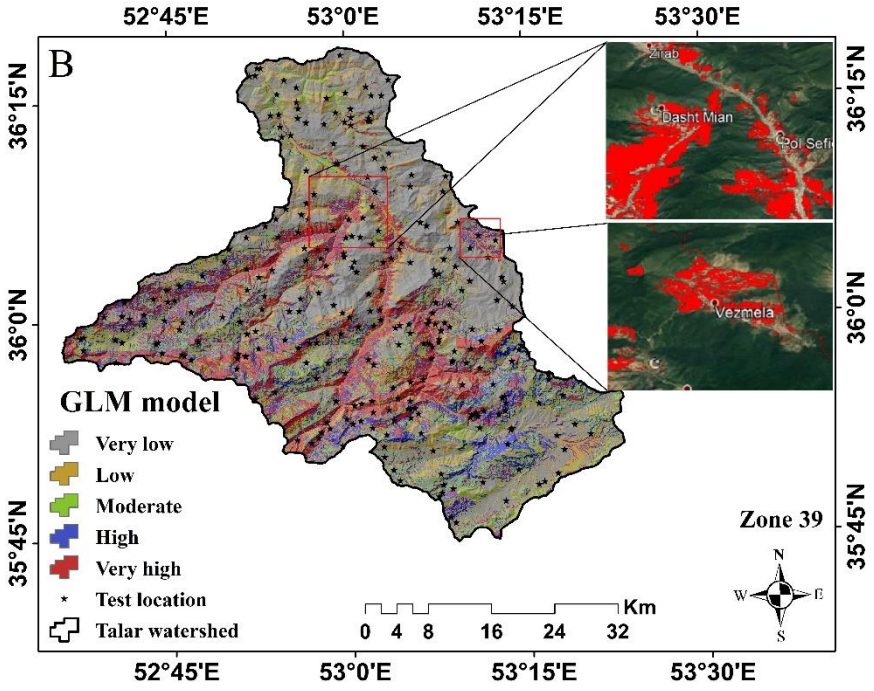
302

303

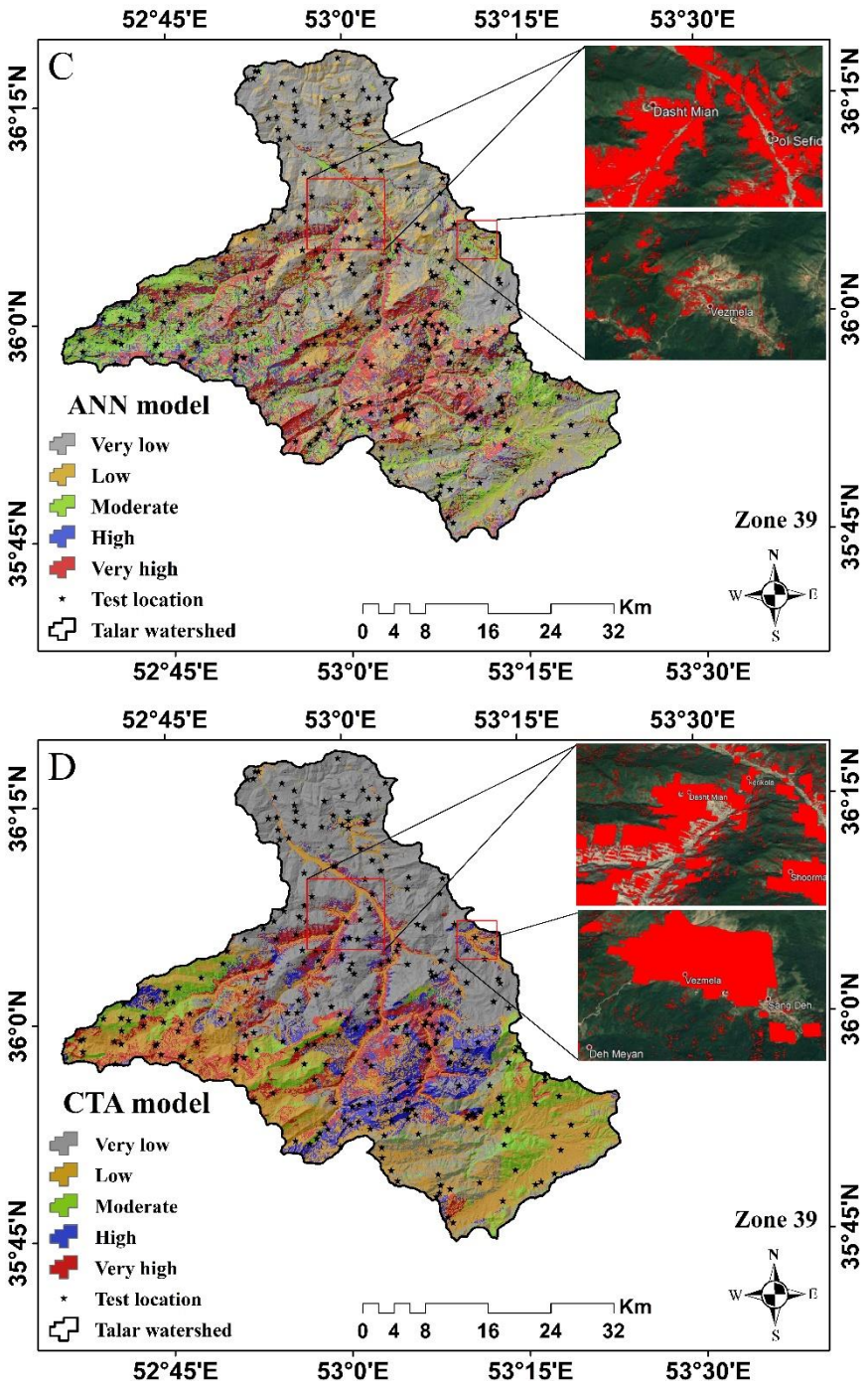

A
52°45'E
53°0'E
53°15'E
53°30'E
36°15'N
36°0'N
35°45'N
Kenkola
Dasht Mian
Vezmela
Sang Deh
RF model
Very low
Low
Moderate
High
Very high
Test location
Talar watershed
Km
0 4 8 16 24 32
Zone 39
N
W
E
S

304

305


B
52°45'E
53°0'E
53°15'E
53°30'E
36°15'N
36°0'N
35°45'N
Zirab
Dasht Mian
Vezmela
GLM model
Very low
Low
Moderate
High
Very high
Test location
Talar watershed
Km
0 4 8 16 24 32
Zone 39
N
W
E
S

306




307

308

309 **Fig. 7** Predictions of soil erosion susceptibility using the four models: A: RF; B: GLM; C: ANN;
310 D: CTA

311 The proportions of the watershed classified into the five soil-erosion susceptibility classes (very
312 low, low, moderate, high, and very high) were determined for each of the four methods (Fig. 8).
313 The RF model predicted 32% of the study area has very low susceptibility, 16% has low

susceptibility, 18% moderate, 22% high, and 24% very high. The ANN predicted 22% very low, 20% low, 39% moderate, 20% high, and 32% very high. GLM predicted 23% very low, 21% low, 26% moderate, 35% high, 28% very high. And CTA predicted 23% very low, 43% low, 16% moderate, 23% high, and 16% very high. In comparison, ANN predicted 32% of the area was very highly susceptible, GLM predicted 28%, CTA predicted 24%, and RF only 16%. All five classes produce more than 16 tons of erosion per ha per year; more erosion than allowed in Iran and the world, so all of these classes are of concern, but very high erosion rates are significantly bad.

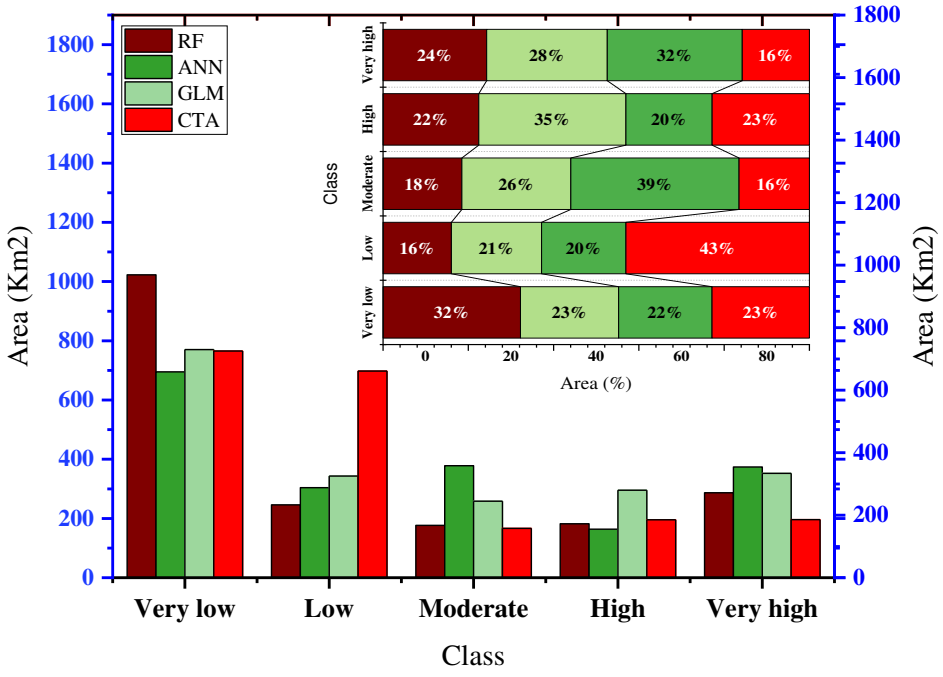



**Fig.** 8 Area of the soil erosion classes found using each model

### 4.3 Assessment of soil erosion risk zone and management strategies

The predictive performances (i.e., accuracies assessed in the validation step) of the four models were determined based on AUC values (Fig. 9). To objectively evaluate the predictive capabilities of these models, we also reviewed recent ML extreme-event studies (i.e., floods, landslides, and erosion) (Avand et al. 2020, Pham et al. 2020, Yariyan et al. 2020). The RF model had the highest accuracy (AUC = 0.973). This was followed by the ANN (AUC = 0.90), GLM (AUC = 0.894) and CTA (AUC = 0.866) models. Also, the binary comparison diagram of the models using the two criteria of sensitivity and precision is shown in Figure 10.

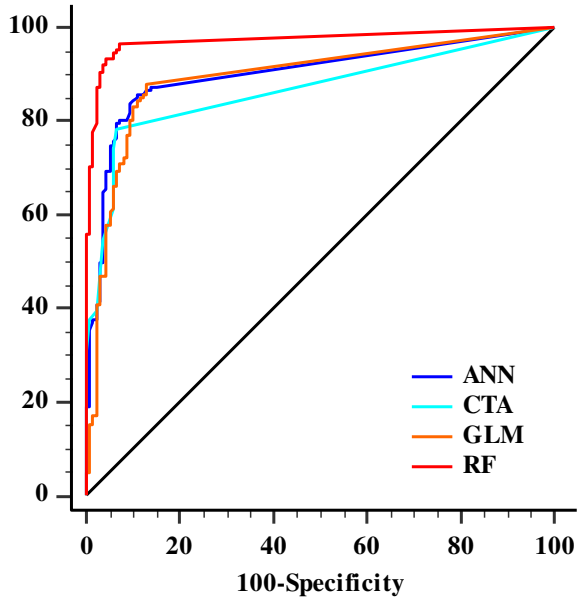



**Fig. 9** ROC curves for all the methods using validation dataset.

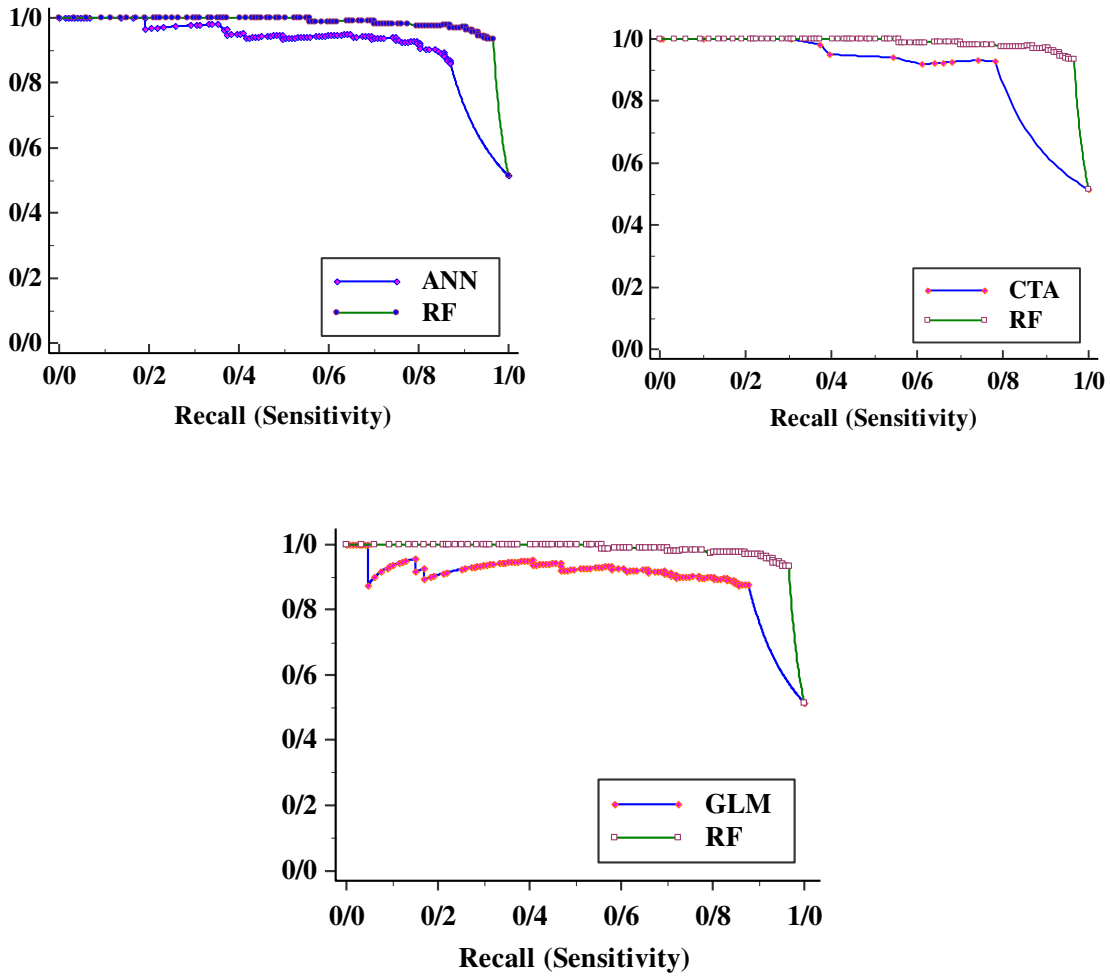



**Fig. 10** Sensitivity analysis of GLM, ANN and CTA models compared to the RF model

RF had superior performance by all metrics: SE = 0.008, Accuracy = 0.92, Sensitivity = 0.966, Specificity = 0.928, and Kappa = 0.78. ANN was second best: SE = 0.018, Accuracy = 0.84, Sensitivity = 0.851, Specificity = 0.893, and Kappa = 0.68. GLM ranked third: SE = 0.019, Accuracy = 0.878, Sensitivity = 0.871, Specificity = 0.835, and Kappa = 0.64. And the CTA was weakest: SE = 0.019, Accuracy = 0.77, Sensitivity = 0.782, Specificity = 0.935, and Kappa = 62 (Table 4).

347 **Table 4.** The performances of the models using the testing dataset.

| Criteria / Model | AUC | SE | 95% CI | Accuracy | Sensitivity | Specificity | TSS | Bias | Kappa |
|---|---|---|---|---|---|---|---|---|---|
| RF | 0.973 | 0.008 | 0.94 to 0.98 | 0.93 | 0.966 | 0.928 | 0.73 | 0.92 | 0.78 |
| GLM | 0.894 | 0.019 | 0.85 to 0.92 | 0.878 | 0.871 | 0.835 | 0.62 | 0.95 | 0.64 |
| ANN | 0.90 | 0.018 | 0.86 to 0.93 | 0.84 | 0.851 | 0.893 | 0.65 | 0.94 | 0.68 |
| CTA | 0.866 | 0.019 | 0.82 to 0.93 | 0.77 | 0.782 | 0.935 | 0.61 | 0.92 | 0.62 |

348

349

350 ## 4. Discussion

351 Identification of hot spots of soil erosion can help to identify ways to reduce erosion. This study
352 integrated RUSLE with four ML models – RF, ANN, GLM and CTA – to predict the areas in the
353 Talar watershed of Mazandaran Province, Iran, with the greatest potential for soil erosion. To
354 operationalize this research, data for 14 independent variables believed to contribute to soil erosion
355 were compiled. The ML models were trained and validated with 488 soil erosion locations
356 (locations with observed soil erosion). These points were extracted from the RUSLE output map
357 and were randomly separated into two groups for training (70% = 342) and validation (30% =
358 146). A multicollinearity analysis indicated that profile curvature and TPI were highly correlated
359 with elevation and slope, respectively, and were removed. All other factors were independent of
360 each other. Therefore, 12 independent factors were used in the model training and validation
361 processes.

362 ### Driving factors of soil erosion

363 Analysis of the twelve factors' contributions to erosion showed that four – slope, land use,
364 elevation, and rain erosivity were the most important. A large portion of the Talar watershed is

365 mountainous, especially the southern part of the watershed. Decreasing soil depth, diminishing
366 vegetation, and increasing slope angle in mountains accelerates erosion (Mohammadi et al. 2017).
367 Slope is usually the factor deemed most important promoter of soil erosion in most studies (Arekhi
368 et al. 2012, Samanta et al. 2016, Lei et al. 2020). Slope angle changes erosion rates by influencing
369 runoff velocity, the abundance of vegetation, and soil type. Slopes greater than 30 degrees, though
370 having increased erosion potential from faster moving water and less vegetation, eventually
371 become rocky with erosion-resistant formations and erosion is reduced due to lack of soil.
372 Increasing the slope length, also increases erosion rates due to flow conditions (flow velocity,
373 volume, and shear stress). These findings are consistent with studies by Atoma et al. (2020) and
374 Tadesse and Tefera (2020).

375 As slopes decrease in flatter topography, different soil types will form, and vegetation will usually
376 provide more dense cover and protection from the impact of rainfall. Rain intensity and the kinetic
377 energy of raindrops greatly affect soil erosion potential. Changing climates in Iran, particularly in
378 areas experiencing increased precipitation rates in the northern portions of the country, have
379 prompted greater soil erosion in many watersheds as discussed in Kavian et al. (2011) and Sadeghi
380 et al. (2017).

381

382 The type and severity of erosion depends upon land use as well. Subsurface water flow in forested
383 areas can produce gully erosion. Forests in the Talar watershed influence soil erosion on steep
384 slopes and produce debris that can augment mass wasting. Agricultural tillage enables greater
385 infiltration into soil and can cause mass erosion. The use of machinery and frequent plowing affects
386 surface runoff and erosion rates by changing the physical properties of the soils. Such mechanisms
387 for soil erosion in agricultural lands have been discussed (Silva et al. 2009, Filoso et al. 2015).

388 And rill erosion is more prevalent in rangelands (Vaezi and Sadeghi 2011, Zare et al. 2017).
389 Deforestation, mining, and the development of residential land uses and roads has also intensified
390 erosion in this region.

391 To understand the patterns of soil erosion susceptibility in the landscapes of the Talar watershed ,
392 the ML models' predictions were mapped and classified using the natural break method in
393 ArcMap. The classifications were summed, and a proportion of the watershed area covered by
394 each of the 5 classes (very low to very high) were calculated for each of the four ML models. All
395 five classes had erosion rates above the original base rate of 16 ton/ha/yr. The results reveal some
396 variation in the areal extents of the classes produced by each model, but the visual pattern indicates
397 that the regions of high and very high susceptibility are near stream paths and in upstream areas of
398 steep slopes, areas that typically have little to no vegetation.

399 **Performance of MLMs on hot spot soil erosion mapping**

400 RF is the ML modeling approach that produced the greatest prediction power. Ranking next was
401 ANN, then GLM, and CTA. An advantage of the RF model is that it employs the best randomly
402 selected variables (or dividing points) in sub-groups to produce a growth tree and in so doing it
403 reduces the strength of each individual regression tree. This reduction in the matching rate reduces
404 model error (Breiman 2001). To improve the learning of classification machines and regression
405 models and to reduce variance and prevent overfitting, a complex RF model is used (Immitzer et
406 al. 2012, Breiman 2017), which is consistent with Towfiqul Islam et al. (2020) and Chen et al.
407 (2020) who state that the RF is very accurate when predicting extreme events (like floods, erosion,
408 landslides, et al.).

Soil erosion and sediment deposition, especially in agricultural lands, lead to soil loss, reduced soil depth, soil structure decomposition, reduced fertility, and reduced organic matter and nutrients. These losses may lead to abandonment. ML models can help decision makers identify areas with high erosion potential and can focus management efforts. Based on the results of this study, RUSLE in combination with RF could be used by natural resource managers to guide zoning of land use practices and to identify hot spots of soil-erosion susceptibility in regions similar to the Talar watershed.

## Author Declarations


### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.


### Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Availability of data and material

Data not available due to [ethical/legal/commercial] restrictions.

### Code availability

Not applicable

### Authors' contributions

Data curation, **M.A** and **M.M;** formal analysis, **M.A** and **M.M**; funding acquisition, **M.A** and **A.K**; investigation, **M.A** and **M.M;** methodology, **A.K., M.A.,** and **M.M**; supervision, **M.A**; validation, **F.M** and **M.M;** visualization, **M.A, M.M,** and **F.M;** writing—original draft, **M.A.**

and **M.M.;** writing—review and editing, **J.P.T.,** and **A.K;** All authors have read and agreed to the published version of the manuscript.

## References


Ali, G., C. Birkel, D. Tetzlaff, C. Soulsby, J. J. McDonnell, and P. Tarolli. 2014. A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions. Earth Surface Processes and Landforms 39(3):399–413.

Amine, M., E. Maaoui, M. Sfar, M. Rached, and M. Habib. 2012. Catena Sediment yield from irregularly shaped gullies located on the Fortuna lithologic formation in semi-arid area of Tunisia. Catena 93 (2012):97–104.

Arekhi, S., A. Darvishi, A. Shabani, H. Fathizad, and S. Ahmadai Abchin. 2012. Mapping soil erosion and sediment yield susceptibility using RUSLE, remote sensing and GIS (Case study: Cham Gardalan Watershed, Iran). J. Adv. Environ. Biol 6(1):109–124.

Atoma, H., K. V. Suryabhagavan, and M. Balakrishnan. 2020. Soil erosion assessment using RUSLE model and GIS in Huluka watershed, Central Ethiopia. Sustainable Water Resources Management 6(1):1–17.

Avand, M., H. Moradi, and M. Ramazanzadeh. 2020. Using machine learning models , remote sensing , and GIS to investigate the effects of changing climates and land uses on flood probability. Journal of Hydrology(October):125663.

Bagio, B., I. Bertol, N. H. Wolschick, D. Schneiders, and M. A. do N. dos Santos. 2017. Water Erosion in Different Slope Lengths on Bare Soil. Revista Brasileira de Ciência do Solo 41 (2017).

Bouchnak, H., M. S. Felfoul, M. R. Boussema, and M. H. Snane. 2009. Slope and rainfall effects on the volume of sediment yield by gully erosion in the Souar lithologic formation (Tunisia). Catena

78(2):170–177.

Breiman, L. 2001. Random forests. Machine Learning 45(1):5–32.

Breiman, L. 2017. Classification and regression trees. Routledge.

Chalise, D., L. Kumar, V. Spalevic, and G. Skataric. 2019. Estimation of sediment yield and maximum outflow using the IntErO model in the Sarada River Basin of Nepal. Water (Switzerland) 11(5).

Chen, W., Y. Li, W. Xue, H. Shahabi, S. Li, H. Hong, X. Wang, H. Bian, S. Zhang, B. Pradhan, and B. Bin Ahmad. 2020. Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. Science of the Total Environment 701 (January 2020):134979.

Engineering Services Company. 2002. Comprehensive study: Talar watershed. Mazandaran.

Faulkner, H. 2013. Badlands in marl lithologies: a field guide to soil dispersion, subsurface erosion and piping-origin gullies. Catena 106 (2013):42–53.

Filoso, S., J. B. do Carmo, S. F. Mardegan, S. R. M. Lins, T. F. Gomes, and L. A. Martinelli. 2015. Reassessing the environmental impacts of sugarcane ethanol production in Brazil to help meet sustainability goals. Renewable and Sustainable Energy Reviews 52 (2015):1847–1856.

Gayen, A., H. Reza, S. Saha, S. Keesstra, and S. Bai. 2019. Science of the Total Environment Gully erosion susceptibility assessment and management of hazard- prone areas in India using different machine learning algorithms. Science of the Total Environment 668 (2019):124–138.

Hickey, R. 2000. Slope angle and slope length solutions for GIS. Cartography 29(1):1–8.

Immitzer, M., C. Atzberger, and T. Koukal. 2012. Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. Remote Sensing 4(9):2661–2693.

Isazadeh, M., S. M. Biazar, and A. Ashrafzadeh. 2017. Support vector machines and feed-forward neural

networks for spatial modeling of groundwater qualitative parameters. Environmental Earth Sciences 76(17):1–14.

De Jong, S. M. 1994. Application of Reflective Remote Sensing for Land Degradation Studies in a Mediterranean Environment (Utrecht: Netherlands Geographical Studies, University of Utrecht) (1994).

Karamidehkordi, E. 2010. A country report: Challenges facing Iranian agriculture and natural resource management in the twenty-first century. Human Ecology 38(2):295–303.

Kavian, A., Y. Fathollah Nejad, M. Habibnejad, and K. Soleimani. 2011. Modeling seasonal rainfall erosivity on a regional scale: a case study from Northeastern Iran. International Journal of Environmental Research 5(4):939–950.

Koirala, P., S. Thakuri, S. Joshi, and R. Chauhan. 2019. Estimation of Soil Erosion in Nepal Using a RUSLE Modeling and Geospatial Tool (2019).

Lei, X., W. Chen, M. Avand, S. Janizadeh, N. Kariminejad, H. Shahabi, R. Costache, H. Shahabi, A. Shirzadi, and A. Mosavi. 2020. GIS-based machine learning algorithms for gully erosion susceptibility mapping in a semi-arid region of Iran. Remote Sensing 12(15):2478.

Lin, C.-Y., W.-T. Lin, and W.-C. Chou. 2002. Soil erosion prediction and sediment yield estimation: the Taiwan experience. Soil and Tillage Research 68(2):143–152.

Mirchooli, F., A. Motevalli, H. R. Pourghasemi, M. Mohammadi, P. Bhattacharya, F. F. Maghsood, and J. P. Tiefenbacher. 2019. How do data-mining models consider arsenic contamination in sediments and variables importance? Environmental Monitoring and Assessment 191(12).

Mohammadi, M., H. Darabi, F. Mirchooli, A. Bakhshaee, and A. Torabi Haghighi. 2020. Flood risk mapping and crop-water loss modeling using water footprint analysis in agricultural watershed, northern Iran. Natural Hazards(0123456789).

MOHAMMADI, M., M. FALLAH, A. KAVIAN, L. GHOLAMI, and E. OMIDVAR. 2017. The Application of RUSLE Model in Spatial DistributionDetermination of Soil loss Hazard (2017).

Mohammed, S., K. Alsafadi, S. Talukdar, S. Kiwan, S. Hennawi, O. Alshihabi, M. Sharaf, and E. Harsanyie. 2020. Estimation of soil erosion risk in southern part of Syria by using RUSLE integrating geo informatics approach. Remote Sensing Applications: Society and Environment 20(July):100375.

Mousavi, S. M., A. Golkarian, S. A. Naghibi, B. Kalantar, and B. Pradhan. 2017. GIS-based Groundwater Spring Potential Mapping Using Data Mining Boosted Regression Tree and Probabilistic Frequency Ratio Models in Iran. AIMS Geosciences 3(1):91–115.

Mukherjee, S., P. K. Joshi, S. Mukherjee, A. Ghosh, R. D. Garg, and A. Mukhopadhyay. 2013. Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). International Journal of Applied Earth Observation and Geoinformation 21 (2013):205–217.

Naghibi, S. A., K. Ahmadi, and A. Daneshi. 2017. Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping. Water Resources Management 31(9):2761–2775.

Nyesheja, E. M., X. Chen, A. M. El-Tantawi, F. Karamage, C. Mupenzi, and J. B. Nsengiyumva. 2019. Soil erosion assessment using RUSLE model in the Congo Nile Ridge region of Rwanda. Physical Geography 40(4):339–360.

Pham, B. T., T. Van Phong, M. Avand, N. Al-Ansari, S. K. Singh, H. Van Le, and I. Prakash. 2020. Improving Voting Feature Intervals for Spatial Prediction of Landslides. Mathematical Problems in Engineering 2020 (2020).

Pham, T. G., J. Degener, and M. Kappas. 2018. Integrated universal soil loss equation (USLE) and Geographical Information System (GIS) for soil erosion estimation in A Sap basin: Central

Vietnam. International Soil and Water Conservation Research 6(2):99–110.

Phinzi, K., N. S. Ngetar, and O. Ebhuoma. 2020. Soil erosion risk assessment in the Umzintlava catchment (T32E), Eastern Cape, South Africa, using RUSLE and random forest algorithm. South African Geographical Journal 00(00):1–24.

Pourghasemi, H. R., A. G. Jirandeh, B. Pradhan, C. Xu, and C. Gokceoglu. 2013. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. J. Earth Syst. Sci. Indian Academy of Sciences 122(2):349–369.

Pourghasemi, H. R., H. R. Moradi, S. M. Fatemi Aghda, C. Gokceoglu, and B. Pradhan. 2014. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). Arabian Journal of Geosciences 7(5):1857–1878.

Pournader, M., H. Ahmadi, S. Feiznia, H. Karimi, and H. R. Peirovan. 2018. Spatial prediction of soil erosion susceptibility: an evaluation of the maximum entropy model. Earth Science Informatics 11(3):389–401.

Pradhan, B., A. Chaudhari, J. Adinarayana, and M. F. Buchroithner. 2012. Soil erosion assessment and its correlation with landslide events using remote sensing data and GIS: a case study at Penang Island, Malaysia. Environmental monitoring and assessment 184(2):715–727.

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

Van Remortel, R. D., R. W. Maichle, and R. J. Hickey. 2004. Computing the LS factor for the Revised Universal Soil Loss Equation through array-based slope processing of digital elevation data using a C++ executable. Computers & Geosciences 30(9–10):1043–1053.

Renard, K. G., G. R. Foster, G. A. Weesies, and J. P. Porter. 1991. RUSLE: Revised universal soil loss

546 equation. Journal of soil and Water Conservation 46(1):30–33.

547 Renard, K. G., and J. R. Freimund. 1994. Using monthly precipitation data to estimate the R-factor in the
548 revised USLE. Journal of hydrology 157(1–4):287–306.

549 Ruysschaert, G., J. Poesen, G. Verstraeten, and G. Govers. 2007. Soil loss due to harvesting of various
550 crop types in contrasting agro-ecological environments. Agriculture, ecosystems & environment 120
551 (2007):153–165.

552 Sadeghi, S. H., M. Zabihi, M. Vafakhah, and Z. Hazbavi. 2017. Spatiotemporal mapping of rainfall
553 erosivity index for different return periods in Iran. Natural Hazards 87(1):35–56.

554 Samanta, R. K., G. S. Bhunia, and P. K. shit. 2016. Spatial modelling of soil erosion susceptibility
555 mapping in lower basin of Subarnarekha river (India) based on geospatial techniques. Modeling
556 Earth Systems and Environment 2(2):1–13.

557 Sharma, A., K. N. Tiwari, and P. B. S. Bhadoria. 2011. Effect of land use land cover change on soil
558 erosion potential in an agricultural watershed (2011):789–801.

559 Silva, R. B. da, K. P. Lanças, E. E. V Miranda, F. A. de M. Silva, and F. H. R. Baio. 2009. Estimation and
560 evaluation of dynamic properties as indicators of changes on soil structure in sugarcane fields of Sao
561 Paulo State--Brazil. Soil and Tillage Research 103(2):265–270.

562 Tadesse, L., K. V Suryabhagavan, G. Sridhar, and G. Legesse. 2017. Land use and land cover changes
563 and Soil erosion in Yezat Watershed, North Western Ethiopia. International soil and water
564 conservation research 5(2):85–94.

565 Tadesse, T. B., and S. A. Tefera. 2020. Comparing potential risk of soil erosion using RUSLE and
566 MCDA techniques in Central Ethiopia. Modeling Earth Systems and Environment(Lal 1993).

567 Talukdar, S., B. Ghose, R. Salam, S. Mahato, Q. B. Pham, N. T. T. Linh, R. Costache, M. Avand, and
568 others. 2020. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel

569 ensembles of bagging algorithms. Stochastic Environmental Research and Risk Assessment
570 (2020):1–24.

571 Tang, B., J. Jiao, Y. Zhang, Y. Chen, N. Wang, and L. Bai. 2020. The magnitude of soil erosion on
572 hillslopes with different land use patterns under an extreme rainstorm on the Northern Loess
573 Plateau, China. Soil and Tillage Research 204 (2020):104716.

574 Tang, Q., Y. Xu, and S. J. Bennett. 2015. Assessment of soil erosion using RUSLE and GIS : a case study
575 of the Yangou watershed in the Loess Plateau , China. Environmental Earth Sciences(73):1715–
576 1724.

577 Towfiqul Islam, A. R. M., S. Talukdar, S. Mahato, S. Kundu, K. U. Eibek, Q. B. Pham, A. Kuriqi, and N.
578 T. T. Linh. 2020. Flood susceptibility modelling using advanced ensemble machine learning models.
579 Geoscience Frontiers (2020).

580 Troeh, F. R., J. A. Hobbs, R. L. Donahue, and others. 1980. Soil and water conservation for productivity
581 and environmental protection. Prentice-Hall, Inc.

582 Vaezi, A. R., H. A. BAHRAMI, S. H. R. Sadeghi, and M. H. Mahdian. 2008. Evaluating Erosivity
583 Indices of the USLE, MUSLE, RUSLE and USLE-M Models in Soils of a Semi-Arid Region in
584 Northwest of Iran (2008).

585 Vaezi, A. R., and S. H. R. Sadeghi. 2011. Evaluating the RUSLE model and developing an empirical
586 equation for estimating soil erodibility factor in a semi-arid region. Spanish journal of agricultural
587 research(3):912–923.

588 Vanacker, V., Y. Ameijeiras-mariño, J. Schoonejans, J. Cornélis, J. P. G. Minella, F. Lamouline, M.
589 Vermeire, B. Campforts, J. Robinet, M. Van De Broek, P. Delmelle, and S. Opfergelt. 2019. Land
590 use impacts on soil erosion and rejuvenation in Southern Brazil. Catena 178(March):256–266.

591 Wang, D., B. Fu, W. Zhao, H. Hu, and Y. Wang. 2008. Multifractal characteristics of soil particle size

592 distribution under different land-use types on the Loess Plateau , China 72 (2008):29–36.

593 Wang, G., S. Wente, G. Z. Gertner, and A. Anderson. 2002. Improvement in mapping vegetation cover
594 factor for the universal soil loss equation by geostatistical methods with Landsat Thematic Mapper
595 images. International Journal of Remote Sensing 23(18):3649–3667.

596 Wischmeier, W. H. 1976. Use and misuse of the universal soil loss equation. Journal of soil and water
597 conservation (1976).

598 Wulf, H., B. Bookhagen, and D. Scherler. 2010. Seasonal precipitation gradients and their impact on
599 fluvial sediment flux in the Northwest Himalaya. Geomorphology 118(1–2):13–21.

600 Yariyan, P., M. Avand, R. A. Abbaspour, M. Karami, and J. P. Tiefenbacher. 2020. GIS-based spatial
601 modeling of snow avalanches using four novel ensemble models. Science of the Total Environment
602 745 (2020):141008.

603 Yousefi, S., M. Avand, P. Yariyan, H. R. Pourghasemi, S. Keesstra, S. Tavangar, and S. Tabibian. 2020.
604 A novel GIS-based ensemble technique for rangeland downward trend mapping as an ecological
605 indicator change. Ecological Indicators 117 (2020):106591.

606 Zare, M., T. Panagopoulos, and L. Loures. 2017. Simulating the impacts of future land use change on soil
607 erosion in the Kasilian watershed , Iran. Land Use Policy 67(June):558–572.

608 Zhang, X., B. Wu, F. Ling, Y. Zeng, N. Yan, and C. Yuan. 2010. Identification of priority areas for
609 controlling soil erosion. Catena 83(1):76–86.

610