



Article

Novel Ensemble Approaches of Machine Learning Techniques in Modeling the Gully Erosion Susceptibility

Alireza Arabameri ^{1,*}, Omid Asadi Nalivan ², Sunil Saha ³ , Jagabandhu Roy ³, Biswajeet Pradhan ^{4,5} , John P. Tiefenbacher ⁶ and Phuong Thao Thi Ngo ⁷

¹ Department of Geomorphology, Tarbiat Modares University, Tehran 14117-13116, Iran

² Department of Watershed Management, Gorgan University of Agricultural Sciences and Natural Resources (GUASNR), Gorgan 3184761174, Iran; omid.asadi@alumni.ut.ac.ir

³ Department of Geography, University of Gour Banga, Malda, West Bengal 732101, India; sunilgeog@ugb.ac.in (S.S.); jagabandhuroy1991@gmail.com (J.R.)

⁴ Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, Sydney 2007, Australia; Biswajeet.Pradhan@uts.edu.au

⁵ Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea

⁶ Department of Geography, Texas State University, San Marcos, TX 78666, USA; tief@txstate.edu

⁷ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; ngotphuongthao5@duytan.edu.vn

* Correspondence: a.arabameri@modares.ac.ir

Received: 17 May 2020; Accepted: 7 June 2020; Published: 11 June 2020



Abstract: Gully erosion has become one of the major environmental issues, due to the severity of its impact in many parts of the world. Gully erosion directly and indirectly affects agriculture and infrastructural development. The Golestan Dam basin, where soil erosion and degradation are very severe problems, was selected as the study area. This research maps gully erosion susceptibility (GES) by integrating four models: maximum entropy (MaxEnt), artificial neural network (ANN), support vector machine (SVM), and general linear model (GLM). Of 1042 gully locations, 729 (70%) and 313 (30%) gully locations were used for modeling and validation purposes, respectively. Fourteen effective gully erosion conditioning factors (GECFs) were selected for spatial gully erosion modeling. Tolerance and variance inflation factors (VIFs) were used to examine the collinearity among the GECFs. The random forest (RF) model was used to assess factors' effectiveness and significance in gully erosion modeling. An ensemble of techniques can provide more accurate results than can single, standalone models. Therefore, we compared two-, three-, and four-model ensembles (ANN-SVM, GLM-ANN, GLM-MaxEnt, GLM-SVM, MaxEnt-ANN, MaxEnt-SVM, ANN-SVM-GLM, GLM-MaxEnt-ANN, GLM-MaxEnt-SVM, MaxEnt-ANN-SVM and GLM-ANN-SVM-MaxEnt) for GES modeling. The susceptibility zones of the GESMs were classified as very-low, low, medium, high, and very-high using Jenks' natural break classification method (NBM). Subsequently, the receiver operating characteristics (ROC) curve and the seed cell area index (SCAI) methods measured the reliability of the models. The success rate curve (SRC) and predication rate curve (PRC) and their area under the curve (AUC) values were obtained from the GES maps. The results show that the ANN model combined with two and three models are more accurate than the other combinations, but the ANN-SVM model had the highest accuracy. The rank of the others from best to worst accuracy is GLM, MaxEnt, SVM, GLM-ANN, GLM-MaxEnt, GLM-SVM, MaxEnt-ANN, MaxEnt-SVM, GLM-ANN-SVM-MaxEnt, GLM-MaxEnt-ANN, GLM-MaxEnt-SVM and MaxEnt-ANN-SVM. The resulting gully erosion susceptibility models (GESMs) are efficient and powerful and could be used to improve soil and water conservation and management.

Keywords: gully erosion susceptibility; artificial neural network; support vector machine; general linear model; GIS; ensemble models; Golestan Dambasin

1. Introduction

Human livelihood depends on soil [1]. Soil erosion is the removal and deposition of particles from the soil surface to other locations [2]. Soil erosion threatens both the quality and quantity of soil and water resources [3]. Approximately 58% of land degradation worldwide stems from soil erosion [4]. Soil erosion is a natural process, but it can be exacerbated by human activities [5]. The long-term impacts of soil erosion become evident, but short-term impacts tend to be less visible [6]. Moreover, soil loss can cause both flooding and desertification, and can damage both urban and rural infrastructure [7]. Rapid global population growth intensifies pressure on both the land and natural resources for living space and food [8]. In addition to this, soil loss rates are increasing [9,10]. Both the conservation of natural resources and the planning for sustainability are critical for environmental protection. The studies of soil erosion and erosion susceptibility are fundamental to successful soil conservation [11]. Over the last 40 years, one-third of the world's cultivated lands have been impacted by soil erosion. Today, approximately 10 million hectares are affected annually [12]. Therefore, measures should be taken to mitigate the effects of soil erosion and to reduce erosion rates [13].

Gully erosion is the most visible form of erosion among the types of soil erosion [14]. It is a major concern due to the damages that gullying has on environmental quality and resources, including the sedimentation of reservoirs, constriction of river flows, devastation to arable lands downstream, impacts on wildlife habitats, and the squelching of regional economic growth [15].

Soil erosion in Iran is a significant problem, affecting agricultural development, the supplies of natural resources, and environmental quality. Approximately 125 million hectares of Iran's 165 million hectares of land suffer water erosion [16]. In arid and semi-arid regions, gullies form due to overexploitation of soils and water and unscientific land and environmental management [17]. Ecosystems, soils, and aquifers are in decline due to gully erosion [18]. Some of the areas that are most vulnerable to gully erosion are the Ardib watershed [19], the Biarjamand watershed [20], the Bayazeh watershed [21], the northeastern Semnan Province [22], the Bastam watershed [23], the Mahabia watershed [24], and the Najafabad watershed [25]. The Golestan Dam basin in the Golestan Province is also experiencing significant soil erosion problems. This is the primary reason it was selected as a study area.

The geo-environmental factors that seem to affect gully erosion are topographical (elevation, slope, aspect, plan and profile curvature, convergence index, topographic position index, topographic ruggedness index), hydrological (drainage density, length of the overland flow, distance to stream, topographical wetness index), lithological (geology, distance to lineament), and environmental (soil type, land use/land cover (LULC), normalized difference vegetation index (NDVI), distance to stream) [15,19,23,25]. Gully erosion conditioning factors (GECFs) are the important conditions that must be monitored for sustainable soil and water conservation planning [26]. Gully erosion inventory maps (GEIMs) are fundamental to the creation of GES maps. Several statistical techniques, like the revised universal soil loss equation (RUSELE) [27] and the modified universal soil loss equation (MUSLE) [28], are used to estimate soil loss rates in regions. These techniques are well known, but calculating soil loss and sediment discharge is time-consuming. Several studies, however, have used an array of knowledge-driven and probabilistic models in their place. Subsequently, machine learning models have been developed to map and assess numerous extreme natural processes, like gullying. The combination of probabilistic and machine learning methods is often used to model extreme events.

Popular and effective models have been used to predict patterns of gully erosion susceptibility, landslide susceptibility, groundwater potential, and flood susceptibility mapping in regions around

the world. These methods have included: analytic hierarchy process [29], frequency ratio (FR) [30], information value [31], conditional probability [32], certainty factor (CF) [33], index of entropy (IOE) [34], logistic regression [35], weight of evidence [36], maximum entropy (MaxEnt) [37], multivariate adaptive regression spline [38], artificial neural network [39], adaptive neuro-fuzzy inference system [40], boosted regression tree [41], random forest (RF) [42], linear discriminate analysis (LDA) [5], support vector machine (SVM), bagging best-first decision tree [43], and classification and regression trees [42].

Ensemble machine-learning models are new powerful tools in the preparation of predictive maps of gully erosion. Several studies have demonstrated ensemble techniques for gully erosion, landslides, land subsidence, groundwater potential, and flood risk mapping. Weighted and predicted values of bivariate statistical methods, multi-criteria decision approaches, and machine learning methods can be integrated to create a single-layer ensemble model. For instance, fuzzy logic with frequency ratio, logistic regression and cosine amplitude [44], and artificial neural network and support vector machine (ANN-SVM) [45] combined with maximum entropy (MaxEnt-ANN-SVM) [46], a multilayer perceptron neural network (MLP) ensemble with dagging, rotation forest, Bagging, random subspace and AdaBoost models [47], complex proportional assessment of alternatives with random forest (COPRAS-RF), COPRAS-FR-boosted regression tree (BRT), and COPRAS-FR-logistic regression (LR) [48], ensemble of density area (DA)-RF and information value (IV)-RF [49], weight-of-evidence (WofE) with LR and Decision Tree ensemble [50], ensemble decision tree [43], generalized linear model (GLM)-fisher's discrimination analysis (FDA), FR-RF and statistical index (SI)-RF [51], RF-alternative decision tree (ADTree) and Bagging-ADTree [52], random subspace (RS)-Fisher's linear discriminant analysis (FLDA), RS-logistic model tree (LMT) and RS-naïve bayes tree (NBTree) [53], FR-RF [54], Vlse Kriterijumska Optimizacija I Kompromisno Resenje (VIKOR)-index of entropy (IoE) [55], WoE-IoE [56] and MaxEnt-ANN, SVM-ANN, SVM-MaxEnt [45] have been used. These ensemble models have proven to achieve greater prediction success than do single models. Machine learning-based approaches are better able to determine accurate relationships between the spatial and temporal data, and to predict future conditions. From this array of models, four individual models (ANN, SVM, MaxEnt, and GLM) were combined in stepwise fashion (i.e., combinations of two then three and then four) to create ensembles GES models. The three-fold ensembles (i.e., two-, three-, and four-model combinations) are quite different from those tested in previous gully erosion research. As natural resources are being degraded by human activities, information technology, modeling, and other approaches are needed to protect and revive ecosystems and environments.

Remote sensing (RS) and geographical information systems (GISs) have been vital to spatial mapping, gully detection, and analysis of the contribution of effective factors, but they also enable the preparation of datasets for analysis. RS- and GIS-based studies are more accurate and provide more meaningful results than by use of more primitive methods to map risk [57]. Over the last two decades, machine learning approaches have developed rapidly. Combination of machine learning, RS, and GIS produces more accurate results. For instance, Arabameri et al. [19–24] used machine learning ensembles to prepare a predictive model in a GIS. A high-resolution ALOS PALSAR digital elevation model (DEM) was combined with Landsat imagery to detect and map gully erosion, providing very accurate results [49].

The main objective of this study is to map the most severe gully-erosion zones with machine learning ensemble models. A MaxEnt, a GLM, an SVM, and an ANN were tested for preparation of a gully erosion susceptibility model (GESM) for the Golestan Dam basin. Furthermore, this study compares the individual, two-, three-, and four-model ensembles to determine the best gully erosion assessment model (or model combination).

2. Material and Methods

2.1. Study Area

The Golestan Dam basin occupies 219,846 hectares (between $37^{\circ}24'00''$ to $37^{\circ}45'00''$ N and $55^{\circ}20'00''$ to $56^{\circ}00'00''$ E) in northeastern Golestan Province. The elevation ranges between 58 and 2168 meters above sea level (m.a.s.l.) (Figure 1). A map of elevation of the study area was prepared from an advanced land observing satellite (ALOS) phased-array type L-band synthetic aperture radar (PALSAR) digital elevation model (DEM) acquired from the Alaska Satellite Facility. More than half of the basin is a gently sloped portion of the Alborz Mountains. The average annual rainfall varies from 224 to 736 mm, and rainfall is highest in the southeastern section of the basin; meteorological data were acquired from the Islamic Republic of Iran Metrological Organization [58]. Based on a land use map (Table 1) prepared at the 1:100,000 scale by the Natural Resources Department of Golestan Province, moderate quality rangeland comprises the largest portion of the area (25.9%). Poor rangelands account for 19.09%, garden and dry land-farming for 14.25%, dry land-farming for 11.2%, and dense forests for 11.19%. The remaining area (18.34%) is shared among gardens, moderate quality forest land, good rangelands, flood crossings, low quality forests, and residential areas. Based on a lithology map (Table 2) acquired from the Geological Society of Iran, the region is comprised of wetlands (60.56%), swamp and marsh (13.4%), grey shale, siltstone and sandstone (8.03%), and shale and calcareous (5.57%), and the remaining (12.44%) is comprised of other formations (Table 1). Rock outcrops, entisols, entisols/inceptisols, and mollisols are the most common soils in the study area [59].

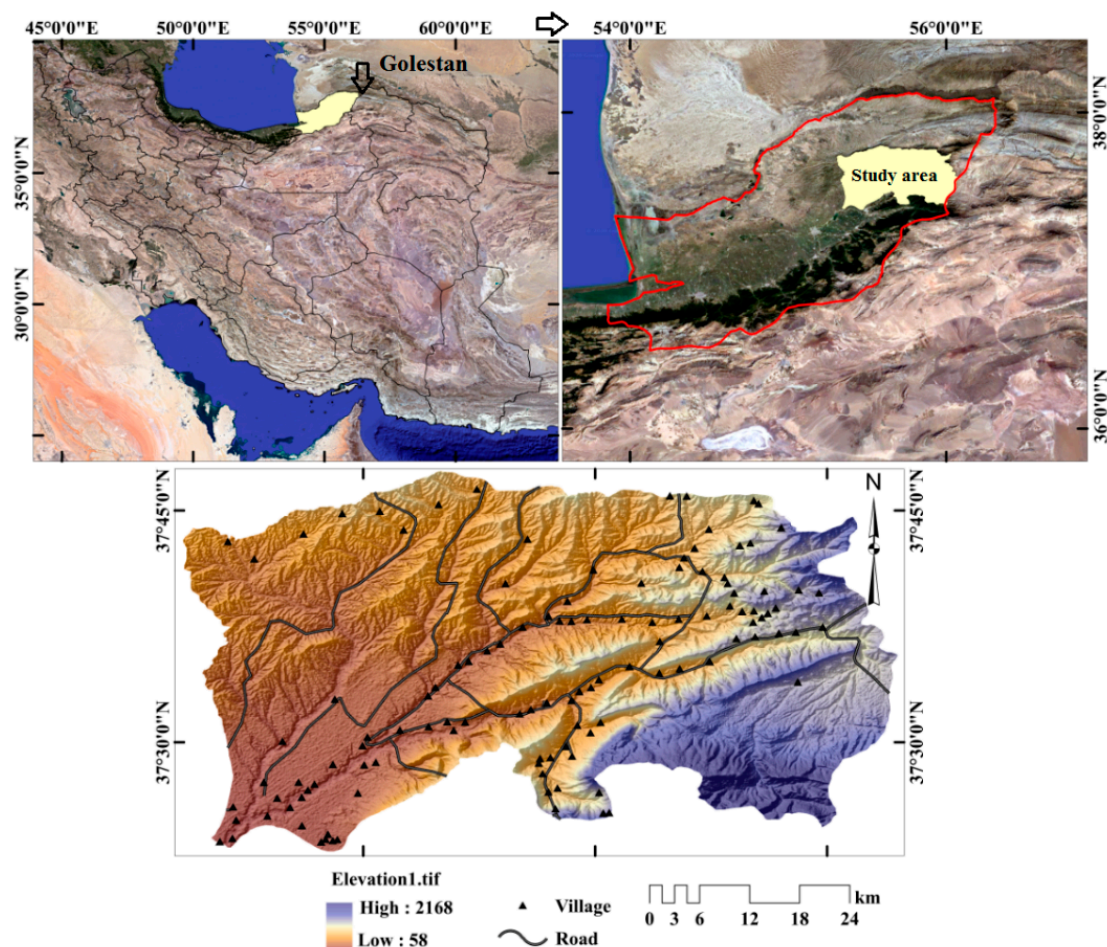


Figure 1. Location of study area.

Table 1. Land use classes in the study area.

Land Use	Area (ha)	Area (%)
Moderate Range	56,971.77	25.91
Poor Range	41,968.52	19.09
Dry farming-Garden	31,329.69	14.25
Dry Farming	24,647.25	11.21
Dense Forest	24,595.63	11.19
Orchard	12,381.7	5.63
Moderate Forest	8679.79	3.95
Good Range	6471.76	2.94
Low Forest	4486.76	2.04
Flood Crossing	3824.86	1.74
Agriculture	3666.75	1.67
Residential Areas	819.21	0.37

Table 2. Lithology of the study area.

Geo Unit	Description	Area (ha)	Area (%)
Kat	Olive green glauconitic sandstone and shale	11,516.28	5.24
Qsw	Swamp	133,117.1	60.56
Ksn	Grey to block shale and thin layers of siltstone and sandstone	17,645.92	8.03
Ekh	Olive—green shale and sandstone	2644.15	1.2
Qm	Swamp and marsh	29,445.41	13.4
Ksr	Ammonite bearing shale with interaction of orbitolin limestone	12,242.8	5.57
Jmz	Grey thick—bedded limestone and dolomite	3253.9	1.48
Jl	Light grey, thin—bedded to massive limestone	9945.08	4.52

2.2. Methodology

This study involves several steps (Figure 2) and processes for creating the gully erosion susceptibility (GES) maps which includes:

- (i) To prepare the gully erosion inventory map and the GECFs dataset, 1042 gully head cut locations were identified using high-resolution images, field investigation, and global positioning system (GPS). Data for fourteen environmental factors identified from a literature review were compiled (data sources are described below).
- (ii) Multi-collinearity analysis among the GECFs using tolerance and variance inflation factor (VIF) techniques was done.
- (iii) The significance and effectiveness of GECFs was determined using the random forest (RF) model.
- (iv) GES maps were prepared with MaxEnt, ANN, SVM and GLM models. The ensemble models were prepared by combining sets of two, three and four models.
- (v) The performances of gully erosion susceptibility models (GESMs) were validated with the area under receiver operating characteristic curve (AUROC) and seed cell area index (SCAI) methods.

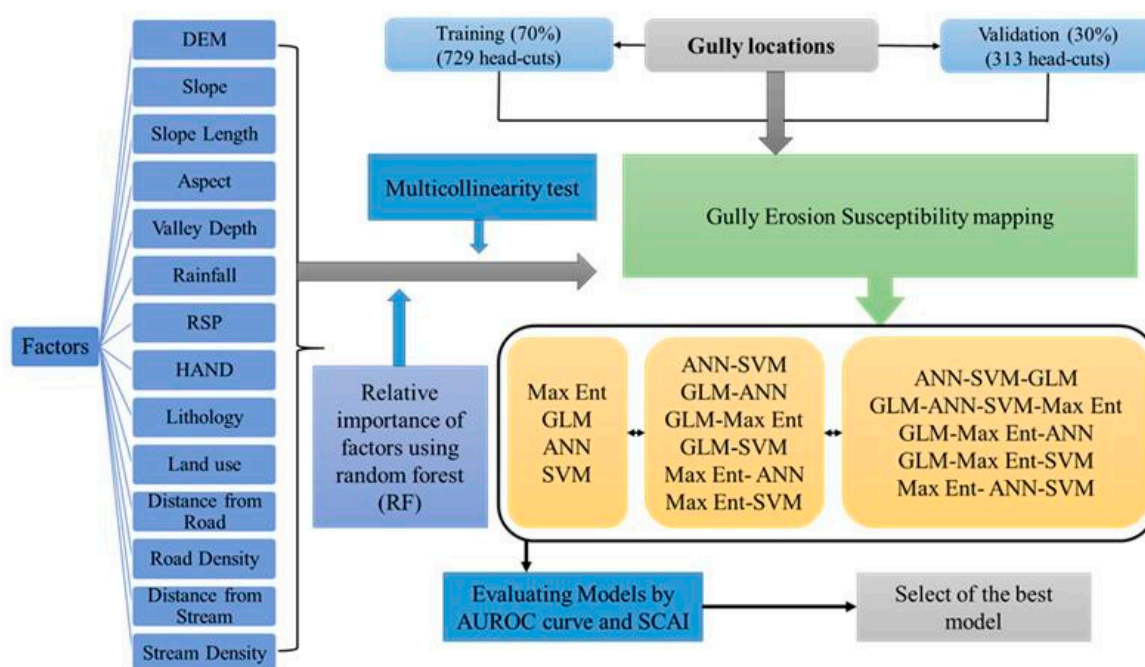


Figure 2. Flowchart of research in the study area.

2.3. Gully Erosion Inventory Map (GEIM)

The GEIM is the basic tool for preparing the GESMs. It consists of spatial datasets with several dependent factors. GEIM represents the spatial distribution of gullies and their geometries (size, shape, length, and width). Through historical and present-day distributions of gullies one can predict future probabilities of gully erosion in the region. Several studies have used GEIM to analyze gully erosion susceptibility [8,24,25,44,60]. ArcGIS 10.3 software was used to generate the GEIM of the Golestan Dam basin. High-resolution images from the Google Earth engine and the ALOS PALSAR DEM at a 12.5×12.5 m resolution were used to prepare the GEIM. Reports of 1042 historical gully formations were acquired from the Agricultural and Natural Resources Research Centre of Golestan Province and they were verified by extensive field surveys with handheld GPS. The inventoried sites were divided randomly into 729 training (70%) and 313 validation (30%) sites [19–25]. The dependent variable for gully formation was the presence or absence of gully and each pixel was coded binarily 1 or 0, respectively. An equal number of locations (pixels) without gullies (1042) were also selected for modeling [61]. The geometric characteristics of the gullies were also measured and recorded in the field [60] (Figure 3). The longest gully was 2.41 km and the shortest was 0.003 km. The deepest gully was 3.15 m and shallowest gully was 0.04 m. The widest gully was 21 m and the narrowest was 1.3 m.



Figure 3. Examples of some gullies in the study area. (a) Lat: 377007.7; Long 4183089. (b) Lat: 392843.2; Long 4176928.2. (c) Lat: 389282.7; Long 4173468.7. (d) Lat: 388929.7; Long 4169277.2.

2.4. Preparing the Gully Erosion Conditioning Factors (GECFs)

Several factors affect gully erosion [62–64]. Rainfall measurements like rainfall intensity, period, and spatial distribution are GECFs [65]. Topographic GECFs included slope, slope aspect, plan and profile curvature, slope length, distance from ridge, convergence index (CI), topographical position index (TPI), topographical rugged index (TRI) [66]. Rock type, rock arrangement, and fault locations are lithological GECFs, as are soil content, soil texture, sub-surface flow, erosivity features [67]. In addition, land use/land cover (LULC) is also an important GECF [68]. Normalized difference vegetation index (NDVI) is a GECF as erosion is more likely to occur on denuded surfaces more than on thickly vegetated surfaces [51,58]. Some studies illustrated that numerous environmental conditions (elevation, slope, aspect, plan and profile curvature, CI, TPI, TRI, drainage density, length of the overland flow, distance to stream (DtS), topographical wetness index (TWI), soil type, LULC, NDVI, and distance to road) influencing gully formation [15,19,23,25].

GECFs' reciprocal relationships were assessed with a multi-collinearity test. The 14 GECFs were mapped in GIS (Figure 4a–n). Spatial resolution of each factor, however, was not the same. The base resolution of these maps was set by the PALSAR DEM resolution (12.5×12.5 m). All factors

were converted to this resolution in ArcGIS 10.3. The preparation procedure for each factor is described below.

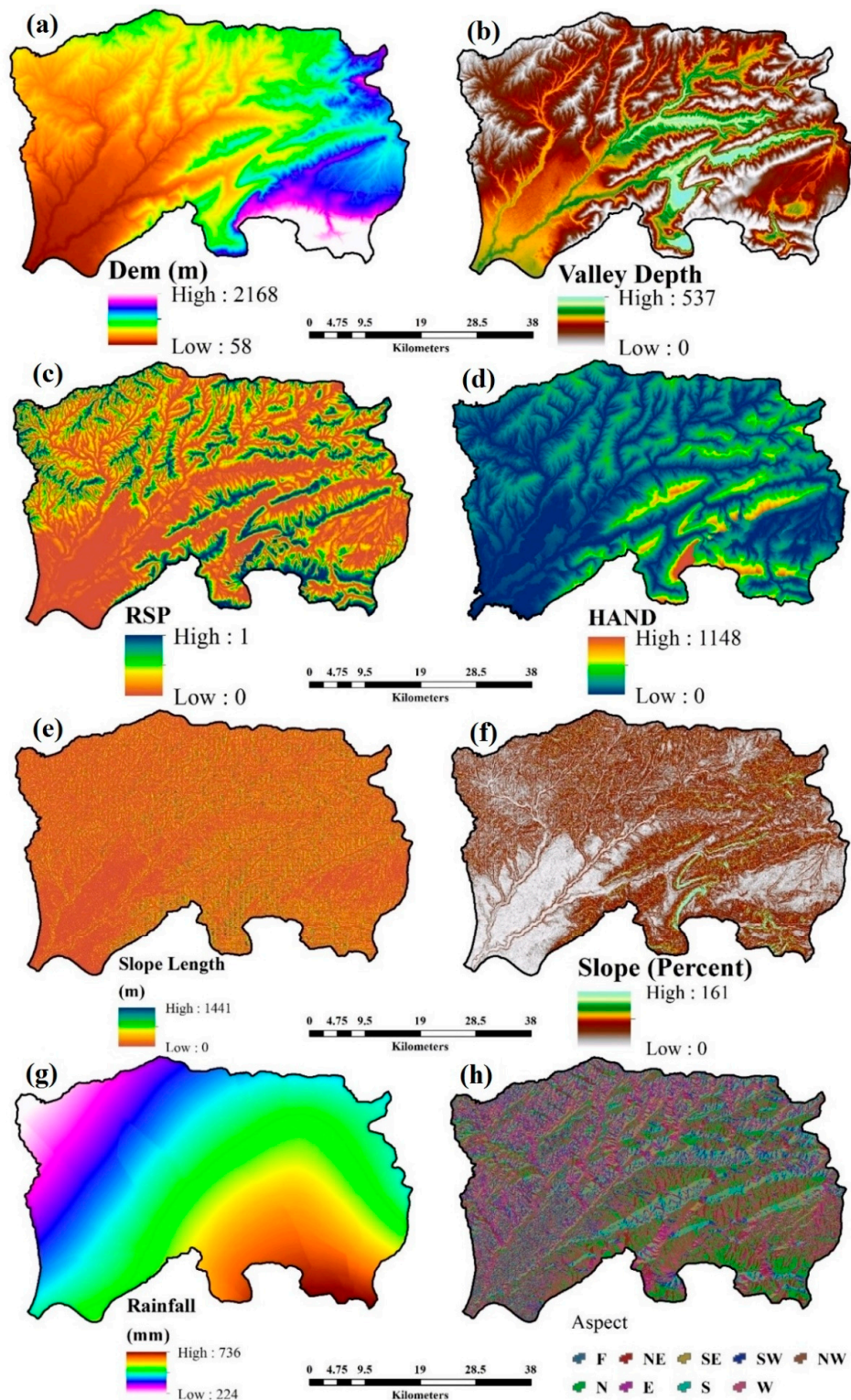


Figure 4. Cont.

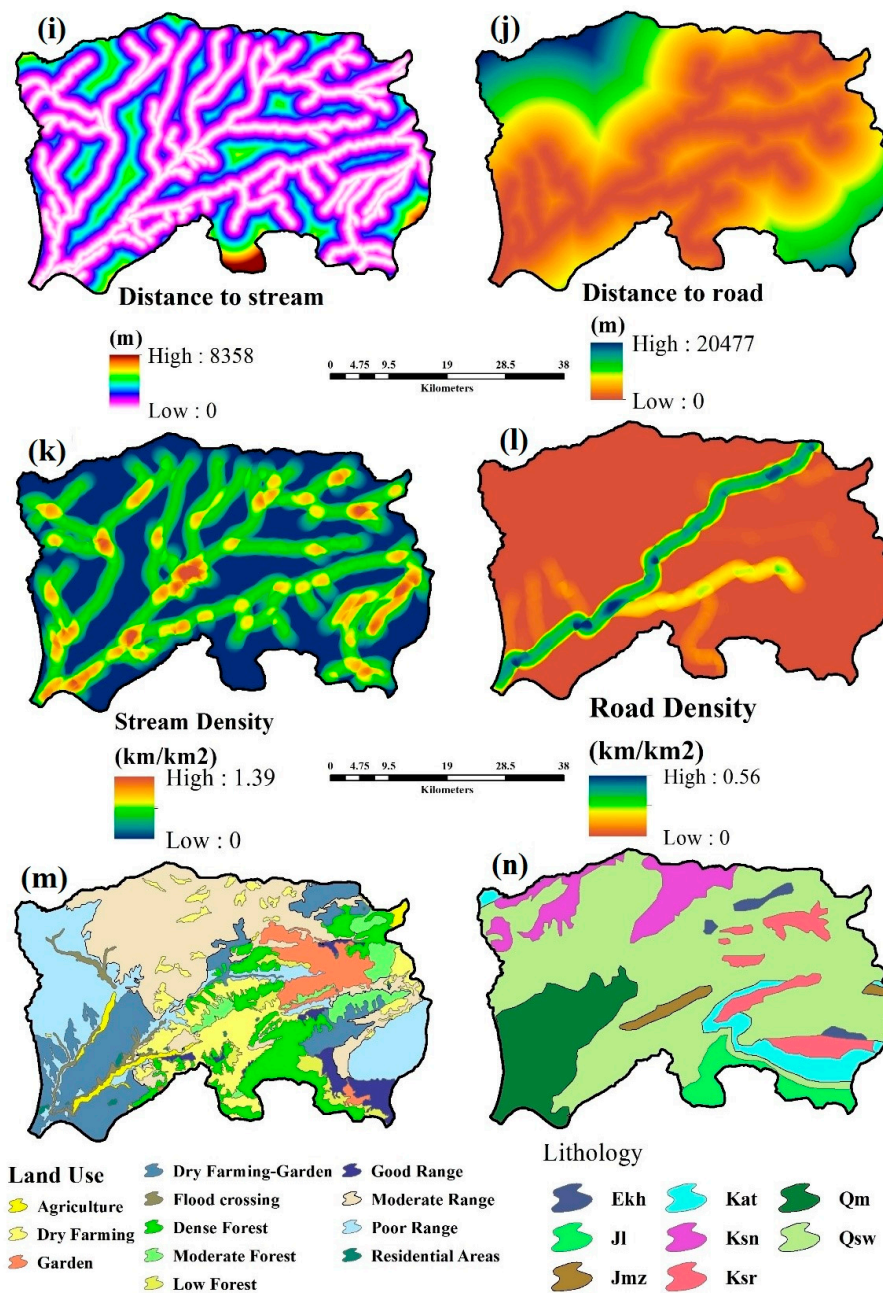


Figure 4. Gully erosion conditioning factors: (a) elevation, (b) valley depth, (c) relative slope position (RSP), (d) height above nearest drainage (HAND), (e) slope length, (f) slope, (g) rainfall, (h) slope aspect, (i) distance to stream (DtS), (j) distance to road (DtR), (k) stream density, (l) road density, (m) LULC, and (n) lithology.

2.4.1. DEM Derived Factors

The topographical and hydrological factors (elevation, slope, slope aspect, slope length, CI, TPI, TRI, topographical wetness index (TWI), stream power index (SPI), STI, distance to stream (DtS), and drainage density) were extracted from the ALOS PALSAR DEM following several steps and processes. The characteristic and morphology of the topographical and hydrological factors depends upon the accuracy and gridding techniques employed [69]. The efficiency of the gridding procedure determines the quality of several of the topographical and hydrological GECFs. The vertical accuracy of the DEM was 0.3 m and was used for all of these variables (a similar accuracy assessment was used

in Gesch et al. [70]). Vertical accuracy was assessed by comparing elevation on the DEM with ground control points (GCPs). At each point, GCP elevations were subtracted from the DEM elevations at the locations in the GIS. The differences were identified as errors. Positive errors occurred where the DEM value was greater than the GCP, negative errors were for DEM values below GCP elevations. The mean error, root means square error, and standard deviation of the mean error were computed from these measured errors [70]. Interferometric synthetic aperture radar (InSAR) was used to produce the ALOS DEM, and this was analyzed by Zhou et al. [71] and Zhang et al. [72]. Phase measurement is the main step in InSAR for DEM creation and this was followed by the transformation of phase to height [71].

After processing the ALOS PALSAR DEM, the topographical and hydrological GECFs were created and measured in GIS with the SAGA GIS tool. The topographical factors influence erosive power, velocity of flow, and surface runoff [73]. Slope controls surface runoff and drainage, influencing gully formation [74]. Slope aspect is also an important GECF [75].

Nubre et al. [76] created the height above nearest drainage (HAND) as a new terrain model. The HAND model indicates topography according to the local relative heights in the drainage network. Kornejady et al. [77] used the HAND tool as a topo-hydrological factor for landslide susceptibility mapping. The ALOS PALSAR DEM was used in this study to create a DEM raster with a resolution of 10 m. For a series of computations, the DEM layer was used to construct a hydrologically consistent DEM, to define flow directions, and to delineate drainage channels. The mapped stream network (MSN) layers were acquired from the Iranian Department of Water Resources Management IDWRM [77]. The MSN was field survey assessed. The automated process for generating the HAND map was conducted in this case study, thus, the spatial data used in the HAND Tool derived from both the DEM and the MSN. When the user chooses the automated process, the “drainage network” area should become operational. The design of the Hand tool’s user interface is valuable, as it allowed fast entry of layers and the choice of either the manual or automatic procedure. The relative slope position (RSP) was determined from the DEM using the SAGA GIS tool. The RSP value ranges from 0 to 1 (Figure 4c). Rahamati et al. [78] used the RSP factor as a conditioning factor for groundwater potential mapping.

2.4.2. Hydrological Factors

As rainfall impacts drainage processes and collapses gully material [79], precipitation data were collected from the Metrological Department of Iran. The thin plate spline method, first formulated by Wahba, is an effective and accurate interpolation method. The modified orthogonal least squares of thin plate spline (TPS-M) was used to prepared the raster layer of the rainfall data collected from the different weather stations. This is more accurate method than the classical interpolation method. The detailed of the methodology could be found in the work of Boer et al. [80]. The rainfall amounts ranged from 224 to 736 mm (Figure 4). DtS and SD are essential hydrological components. DtS and SD determine the width of erosion in rills and gullies [81]. The drainage network was derived from the DEM with the automatic ArcHydro tool in the GIS. Distances from locations (pixels) to streams was measured with the Euclidean distance tool. The drainage density was determined by Horton’s formula on a 1 km² grid. The highest drainage density was measured as 1.39 km/km² (Figure 4k).

2.4.3. Environmental Factors

The distance to road (DtR) is important because roads affect runoff and influences the formation of gullies [82]. A road map at a scale of 1:100,000 was acquired from the National Geographic Organization of Iran [19–25]. Euclidean distance buffering was used to measure distances to roads. Distances ranged from 0 to 20,477 m (Figure 4j). The road density map was created using the line density method. Road densities in the basin range from 0 to 0.56 km/km² (Figure 4l). Geomorphology, LULC, NDVI, and lithology are regarded as significant to GES [83]. The geological map of the study area was acquired from the Geological Society of Iran [84]. The lithology map was digitized in GIS from the map at a scale of 1:100,000. Geologically, the study area is composed of eight geological segments and units: Kat, Qsw, Ksn, Ekh, Qm, Ksr, Jmz, and JI (Figure 4n). LULC contributes directly or indirectly

by influencing evapotranspiration, infiltration, run-off, and sediment dynamics [85]. By comparison, many agricultural practices promote gully erosion and genesis [86]. The land use map was created from the combination of a map of 1:100,000 scale created by the Golestan Province Department of Natural Resources and images acquired from Google Earth. A total of 346 ground control points (GCPs) were selected for validation using the kappa index. The obtained kappa coefficient for the prepared map was 0.924, indicating its high accuracy. The land use types include poor range, dry farming-garden, good range, moderate forest, low forest, residential areas, moderate range, dense forest, agriculture, dry farming, flood crossings, and gardens (Table 1 and Figure 4).

2.5. Multi Collinearity Analysis

The GECFs were tested for the collinearity using tolerance (TOI) and the variance inflation factor (VIF) [19–24,62]. The TOI and VIF methods indicate whether there are linear relationships between GECF pairs. The linear relationship depends on the threshold range value of TOI and VIF. The values <10 for VIF and >0.1 for TOI, suggest no collinearity problems among the GECFs [62]. A multi-collinearity problem exists when two and more factors are highly correlated. Arabameri et al. [19], Saha et al. [58], and Roy and Saha [44] used this method for GES assessment. It is a vital test for subsequent modeling phases.

2.6. Methods

One statistical probability method (GLM), one artificial intelligence method (ANN), and two machine learning models (SVM and MaxEnt) were chosen to prepare GESMs. A number of ensemble models were created by combining these models in stepwise combinations (ensembles of two-, three-, and four-models) to determine which was best.

2.6.1. Maximum Entropy (MaxEnt)

MaxEnt is a presence-only feature and machine-learning model [87]. The presence-only feature is important for modeling as it is more reliable for remote areas [88]. On the other hand, the model is biased when exposed to data from nearby accessible areas [89]. MaxEnt identifies the existence probabilities of locations where a phenomenon is present based on the information theory and statistics [90]. To evaluate the uncertain distribution of probability it consists of set geo-environmental variables. MaxEnt assumes that the probability of occurrence for a phenomenon (here gully erosion) is equal at the all pixels so that it selects the uniform probability distribution function (PDF) as the target distribution. However, certain restrictions compel this PDF to circulate the true target. These constraints are determined by the GECF data that were used to map GES. An interchangeable equation of those factors, called the features was used in the MaxEnt [91]. Specific variables form the features that force MaxEnt's first conjecture to have those constraints. A continuous layer, such as distance to streams for example, creates a linear function that imposes a limit on the numerical average of factor values in the locations where the phenomenon is present. If the mean value of the factor at the presence points is, for instance, equal to 102 m, MaxEnt will consider the locations close to this number as highly susceptible areas. Specific types of features are contained in continuous variables, specifically quadratic, product, threshold, categorical, k binary variables [92]. The maximum entropy would be the final best estimation that satisfies all constraints [93,94]. All the mathematical details of the MaxEnt model can be found in Phillips et al. [87,92] and Elith et al. [91].

2.6.2. Artificial Neural Network (ANN)

The human brain is the inspiration for ANN [95]. ANN has several algorithms that can analyze and predict the nonlinear properties of a phenomenon [96]. Multi-layer perceptron (MLP) is one of the popular algorithms of the ANN model [97]. An ANN consists of the nodes of three layers: the input layer, hidden layers, and the output layer. The nodes in hidden layers measure the information inside the data when the input layer is not sufficiently involved and responsive to do so. GECFs and gully

erosion training results connect the input layer to the output layer. The input and hidden layers then systematically strive to simplify and predict the structures by embracing the knowledge from the input nodes and working with dynamic functions [98].

The number of input and output nodes is defined by a structured code, where nodes are equal to GECFs and output nodes are equal to a Boolean value (0 and 1) for each pixel: 1 indicates probable gully erosion and 0 probably not gully erosion. The trials and errors are determined by the number of hidden layers or nodes [99]. The application of and detailed information about the ANN model is discussed in Arora et al. [100].

2.6.3. Support Vector Machine (SVM)

SVM is a supervised machine learning classifier introduced by Vapnik and Chervonenkis in 1963, that depends on the principle of statistical learning [101]. SVM can be used for classification and regression [102]. It consists of several types of classification functions and can be used to analyze errors and to generalize information that requires the least amount of model tuning [103]. To limit the inherent complexity in the behaviors of phenomena, these processes use stored information inside factors in many iterations [104]. Typically, the hyper-plane with the highest margin should provide the strongest classification performance [105]. A strong hyper-plane between objects is to be much greater than the problem's theoretical vision, when facing actual noisy artifacts. To assign training data around the margin an appropriate methodological error (total gap between the margin and training points), a soft hyper-plane is configured [106]. Subsequently, the rise in the uncertainty of the model contributes to a growth in the complexity of the model and to become the least suited model, thus leading to a decline in generalization, and vice versa. A standardized n-dimensional hyper-plane solution with a maximum distance that is rather complicated and that also has few training errors should thus be found [107–110].

2.6.4. General Linear Model (GLM)

The GLM was derived from the classic linear-regression model platform extension [111]. The GLM applies a common regression structure to non-normal distributions [112]. The statistical function that is known as LOGIT for the GLM model was computed by the Equation (1) [113,114]. GLM is used when there are n independent observations, Y is a function of explanatory variables, X_1, X_2, \dots, X_n (these can be either continuous or categorical), and observations derive from any exponential family distribution.

$$Y = \Pr(Y = 1) = \frac{e^{C_0 + C_1 X_1 + \dots + C_n X_n}}{1 + e^{C_0 + C_1 X_1 + \dots + C_n X_n}} \quad (1)$$

2.7. Measuring the Importance of GECFs by RF

RF is a non-parametric multivariate model [115]. RF models generate thousands of trees, developing 'forests' based on decision rules. Each tree in the RF model depends on a sample of bootstrapped data, applying a CART process with a random subset of variables used at each node. The final decision of class membership and output is made according to majority voting among all decision trees [116]. The RF model was used in this study to evaluate the relative importance of the variables. RF is a machine learning model developed by Breiman [115]. More details of this model can be found in Breiman et al. [116] and Calle et al. [117,118].

2.8. Validation Techniques

2.8.1. Receiver Operating Characteristics (ROC)

The ROC curve is a popular validation method for assessing a model's performance [119]. The area under curve (AUC) of ROC measures the discriminatory capacity of a classification model [120]. This method was widely applied for accuracy assessment of numerous assessments of natural

hazards [121–125]. The ROC curve is used to test the efficiency of a classifier system over the spectrum of cut-off points [126]. The ROC curve is two-dimensional because there are two terms (events and non-events) in the modeling of precision assessment, and therefore, two types of possible accuracy [127]. The first aspect is the success rate of event identification (shown along the Y- or vertical axis). The ROC is a graph which plots the sensitivity (Y-axis) against the 1-specificity (X-axis) using the following Equations (2) and (3):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (3)$$

$$\text{AUC} = \frac{(\sum TP + \sum TN)}{(P + N)} \quad (4)$$

where TP is true positive, FN is false negative, TN is true negative, FP is false positive. P is the total number of gullies and N is the total number of non-gullies.

The area under curve (AUC) indicates the predictive capacity of the model. The threshold values of AUC range from 0.5 to 1 (1 indicates the best performance and 0.5 indicates poor performance) [128]. AUC values have been classified into four levels of accuracy: poor (AUC = 0.6 to 0.7), fair (AUC = 0.7 to 0.8), good (AUC = 0.8 to 0.9) and excellent (AUC = 0.9 to 1) [129].

ROC curves were plotted with both the training gullies points and validation gullies points. The sizes of the training and validation gully sets are proportional and relevant. A 70:30% split is common and widely used, for instance in gully erosion susceptibility [19–25], landslide susceptibility [121], flood susceptibility [122], groundwater potential [130], and land subsidence susceptibility modeling [123–125]. Other sample ratios have been tested: 80:20 [131], 70:30 [132], and 50:50 [133]. There are no absolute rules or standard methods for partitioning of samples, but the decision reflects the user's and experts' opinions and goals. A model's performance depends on the partitions of the samples in the inventory datasets. The sample partitions make a suitable result when the larger percentage of the training data are selected for modeling and the smaller percentage of validation data are used to test the accuracy of the models. Therefore, a 70:30 ratio of the inventory datasets was used to train the GES models and to validate them.

2.8.2. Seed Cell Area Index (SCAI)

The SCAI is the ratio of gullies' pixels to each model's classes of susceptibility. Arabameri et al. [19] states that high SCAI values indicate good model accuracy and low values indicate poor performance. This method has been used to assess the predictive accuracies of models' susceptibility maps.

2.9. Creating Ensemble Models

Ensemble modeling combines two or more single models into a composite model to enhance the reliability of predictive power [134]. This technique has received significant interest among the experts, especially those who use data-mining and machine-learning models [135]. In all ensemble strategies (e.g., bagging and boosting), individual models are weighted. However, there is a diversity of methods to calculate the weights. This study used the heterogeneous group, an integration system that includes addition, subtraction, multiplication, and division when an expanded method was developed for further analysis. The weighted mean function was used to construct the ensemble structures of the three models using Equation (5):

$$EM = \frac{\sum_{i=1}^n (AUSRC_i \times M_i)}{\sum_{i=1}^n AUSRC} \quad (5)$$

where EM is the resulting ensemble model, $AUSRC_i$ is the $AUSRC$ value of the i th model (M_i).

3. Results

3.1. Multi-Collinearity Analysis (MA)

The multi-collinearity test among GECFs was performed in SPSS 17 statistical software. The GECFs in this study achieved tolerance and variance inflation factor (VIF) values below the upper and lower limits, respectively (Table 3). There are no multicollinearity problems among the GECFs. Therefore, all GECFs were suitable for gully erosion modelling.

Table 3. Multi-collinearity analysis of the gully conditioning factors.

Conditioning Factors	Collinearity Statistics	
	Tolerance	VIF
LULC	0.923	1.083
Drainage density	0.911	1.098
Distance to road	0.906	1.104
Valley depth	0.854	1.171
Relative Slope Position (RSP)	0.765	1.307
Geology	0.743	1.346
Rainfall	0.654	1.529
Road density	0.645	1.550
Slope length	0.518	1.931
Aspect	0.465	2.151
Distance to stream	0.456	2.193
Slope	0.423	2.364
Height Above the Nearest Drainage (HAND)	0.384	2.604
Elevation	0.355	2.817

3.2. Gully Erosion Modeling with Individual Models

GESMs were produced using the MaxEnt, ANN, SVM, and GLM models. The GESMs were categorized into very low, low, moderate, high, and very-high susceptibility zones based on Jenks' natural break classification method. The SVM model categorized 16.01% of the basin as very-high GES. The MaxEnt assigned this class to 14.32%, ANN assigned 14.15%, and GLM 14.47% (Figure 5). The classifications of very low susceptibility areas were 33.11% by MaxEnt, 56.30% by ANN, 39.99% by SVM, and 33.86% by GLM.

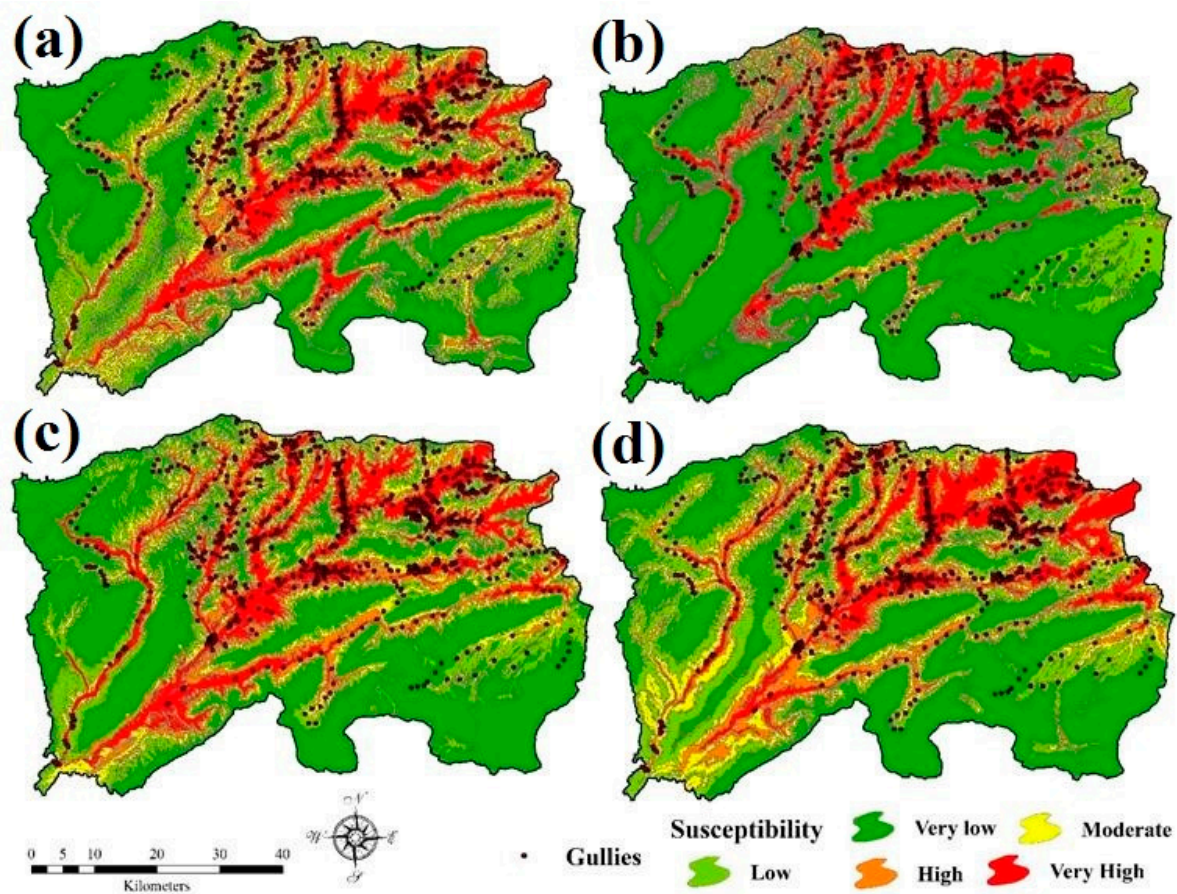


Figure 5. Gully erosion susceptibility mapping using individual models: (a) maximum entropy (MaxEnt), (b) artificial neural network (ANN), (c) support vector machine (SVM), and (d) general linear model (GLM).

3.3. Gully Erosion Modeling by Ensemble of Two Models

Combining two models provides better performance for GES mapping. Pourghasemi, et al. [45] used ensemble machine learning algorithms GESM. Their results demonstrated that the ensemble model achieved the highest predictive performance. We used six two-model ensembles: GLM-MaxEnt, GLM-ANN, GLM-SVM, MaxEnt-ANN, MaxEnt-SVM, and ANN-SVM. The GESMs produced by these ensemble models were also classified into the five susceptibility classes. The GLM-MaxEnt model classified 13.52% of the basin as very-high GES. MaxEnt-ANN assigned very-high to 23.76%, the highest among these ensemble models, and GLM-ANN assigned 12.91%, the lowest among these models (Figure 6). All ensemble models indicated that the central and northeastern parts of the basin have very-high GES.

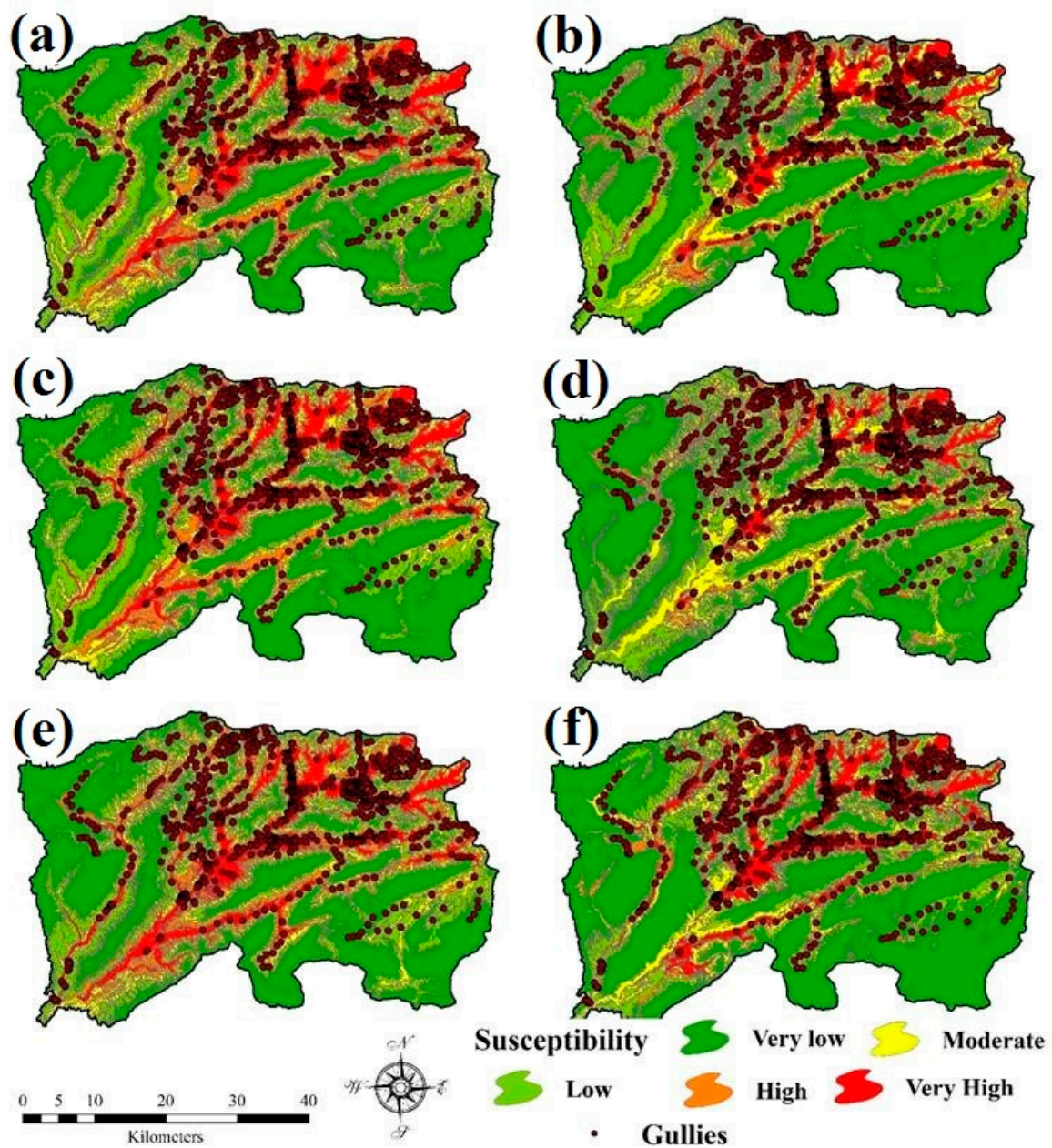


Figure 6. Gully erosion susceptibility maps using two-model ensembles: (a) GLM-MaxEnt, (b) GLM-ANN, (c) GLM-SVM, (d) MaxEnt-ANN, (e) MaxEnt-SVM, and (f) ANN-SVM.

3.4. Gully Erosion Modeling by Ensemble of Three and Four Models

Two-model ensembles are the models most typically used [19–25,45]. The application of three- or four-model ensembles has been rare. Arabameri et al. [49] applied three-model ensembles of COPRAS-FR-RF, COPRAS-FR-BRT, and COPRAS-FR-LR to map GES. These ensembles included a multi-criteria decision approach model (COPRAS), two probabilistic models (FR and LR), and two machine learning models (RF and BRT). In this study, three- and four-model ensembles were used (Figure 7). The ensembles comprised of two machine learning models have been found to be effective and accurate for the prediction and mapping of natural hazards. The GLM-MaxEnt-ANN model classified 13.10% of basin as very-high susceptibility, 10.62% as high, 15.85% as moderate, 22.73% as low, and 37.71% as very low. The GLM-MaxEnt-SVM ensemble assigned the classes (in the same order) to 14.32%, 14.30%, 15.09%, 22.15%, and 34.14%. The MaxEnt-ANN-SVM ensemble applied

these classes to 12.92%, 9.04%, 15.18%, 22.33%, and 40.53%. The ANN-SVM-GLM model classified 14.27% of the basin as very-high, 9.32% high, 15.17% moderate, 20.67% low, and 40.57% very low. The classification by the four-model ensemble GLM-ANN-SVM-MaxEnt was 13.48%, 11.32%, 14.87%, 22.11%, and 38.21%.

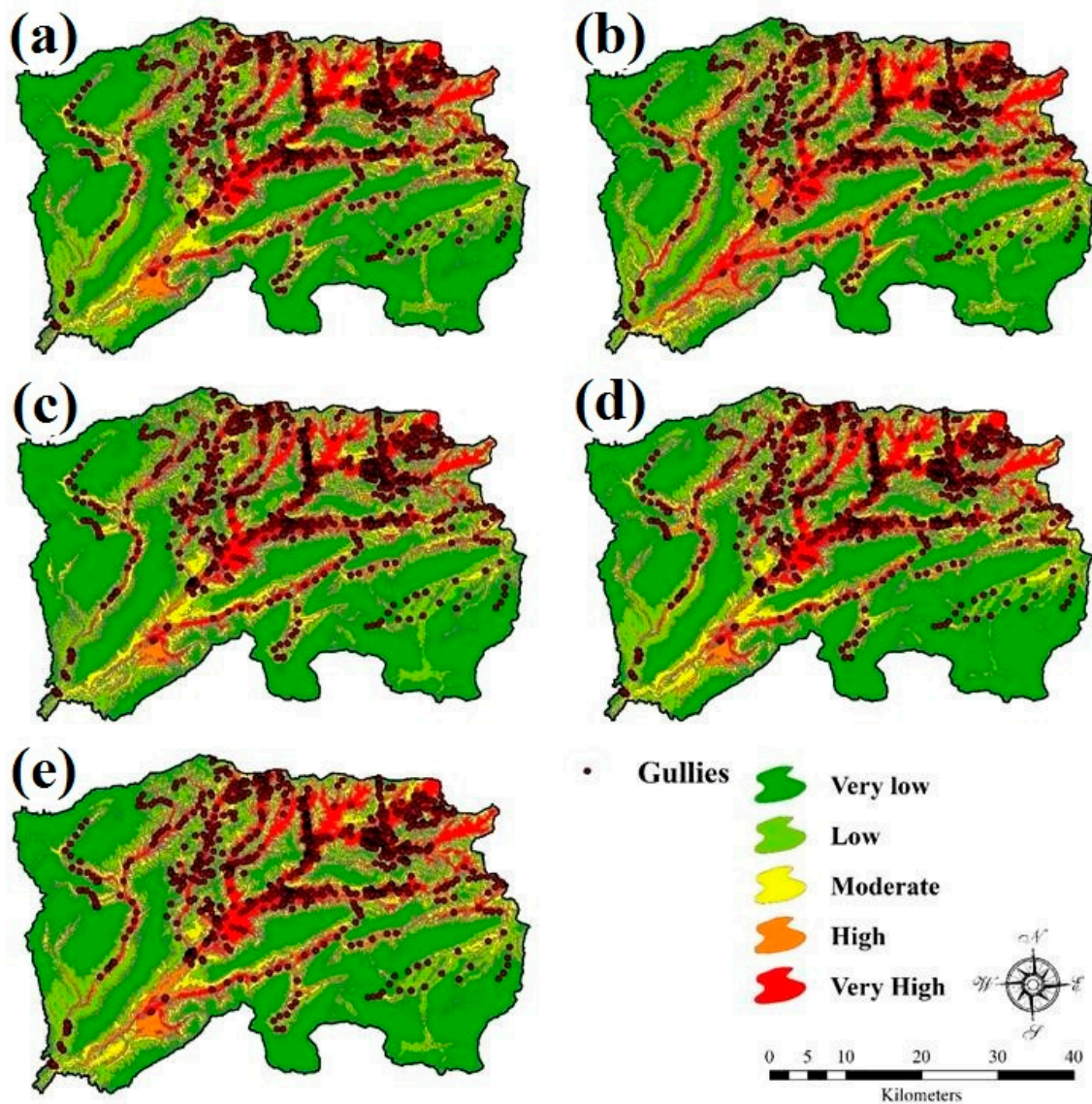


Figure 7. Gully erosion susceptibility mapping using ensemble of three and four models: (a) GLM-MaxEnt-ANN, (b) GLM-MaxEnt-SVM, (c) MaxEnt-ANN-SVM, (d) ANN-SVM-GLM, and (e) GLM-ANN-SVM.

3.5. Assessing the Importance of the Factors

Evaluation of the significance of each GECF provides planners important information that can guide sustainable natural resource use [10,20]. The spatial correlations of the GECFs to gullies were calculated with the RF model in this study (Table 4). Others have used RF techniques for factor-importance analysis [20,58]. The results indicate that of greatest importance was distance to road (RF = 19.23). This factor was followed by LULC, HAND, rainfall, valley depth, distance to stream, slope length, stream density, aspect, elevation, geology, slope, and RSP. GES was higher on rangeland than on non-agricultural lands and on areas nearer roads. Higher gully erosion was also associated with high stream density, HAND, and lower distances to streams. Elevation has direct and

indirect effects on hydrological properties and soil moisture and these may influence gully erosion potential. Overall, gully erosion susceptibility was highest on rangelands that had higher densities of roads and streams.

Table 4. Relative influence of effective conditioning factors in the random forest (RF) model.

Factor	Weight
Distance to road	19.23
LULC	18.60
Height Above the Nearest Drainage (HAND)	17.03
Rainfall	16.13
Valley depth	15.34
Distance to stream	14.65
Slope length	14.42
Stream density	13.54
Aspect	11.85
Elevation	9.34
Geology	9.08
Slope	4.43
Relative Slope Position (RSP)	2.76
Road density	1.65

3.6. Validation of the Models

The GES model's performances were evaluated with the AUC of ROC and SCAI methods applied to the training and validation datasets. The training dataset produces the success rate curve (SRC) and the testing dataset the prediction rate curve (PRC). According to AUC of the SRC the ANN-SVM ensemble achieved the highest accuracy (0.948) (Figure 8). In order of success were models using ANN, ANN-SVM-GLM, GLM-ANN, MaxEnt-ANN-SVM, GLM-MaxEnt-ANN, ANN-SVM-GLM-MaxEnt, MaxEnt-ANN, SVM, GLM-SVM, GLM-MaxEnt-SVM, MaxEnt-SVM, GLM, GLM-MaxEnt, and MaxEnt (Figure 9a–c). Based on the AUC of the PRC, the ANN-SVM model again had the highest accuracy (0.923). The MaxEnt model was least accurate (0.858) (Figure 9d–f). The AUCs of both the SRC and PRC indicate that all models achieved very good to excellent performance for GESM. Yet, though the ANN-SVM ensemble was superior in both assessments, the success and prediction rates of the other were less consistent. The AUC of SRC value of the ANN model was 0.946 (second best among the models), but the AUC of the PRC was 0.906 (seventh best).

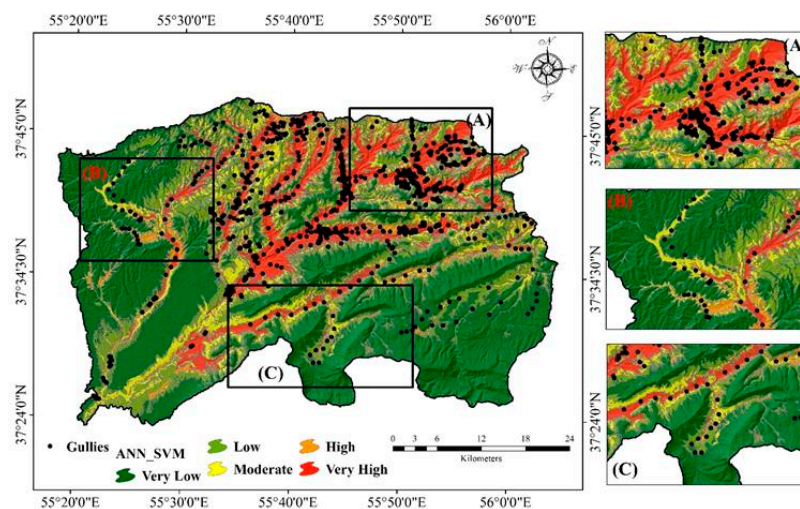


Figure 8. Gully erosion susceptibility mapping using the best model (ANN-SVM ensemble model). (A–C) are zoomed areas in the study area.

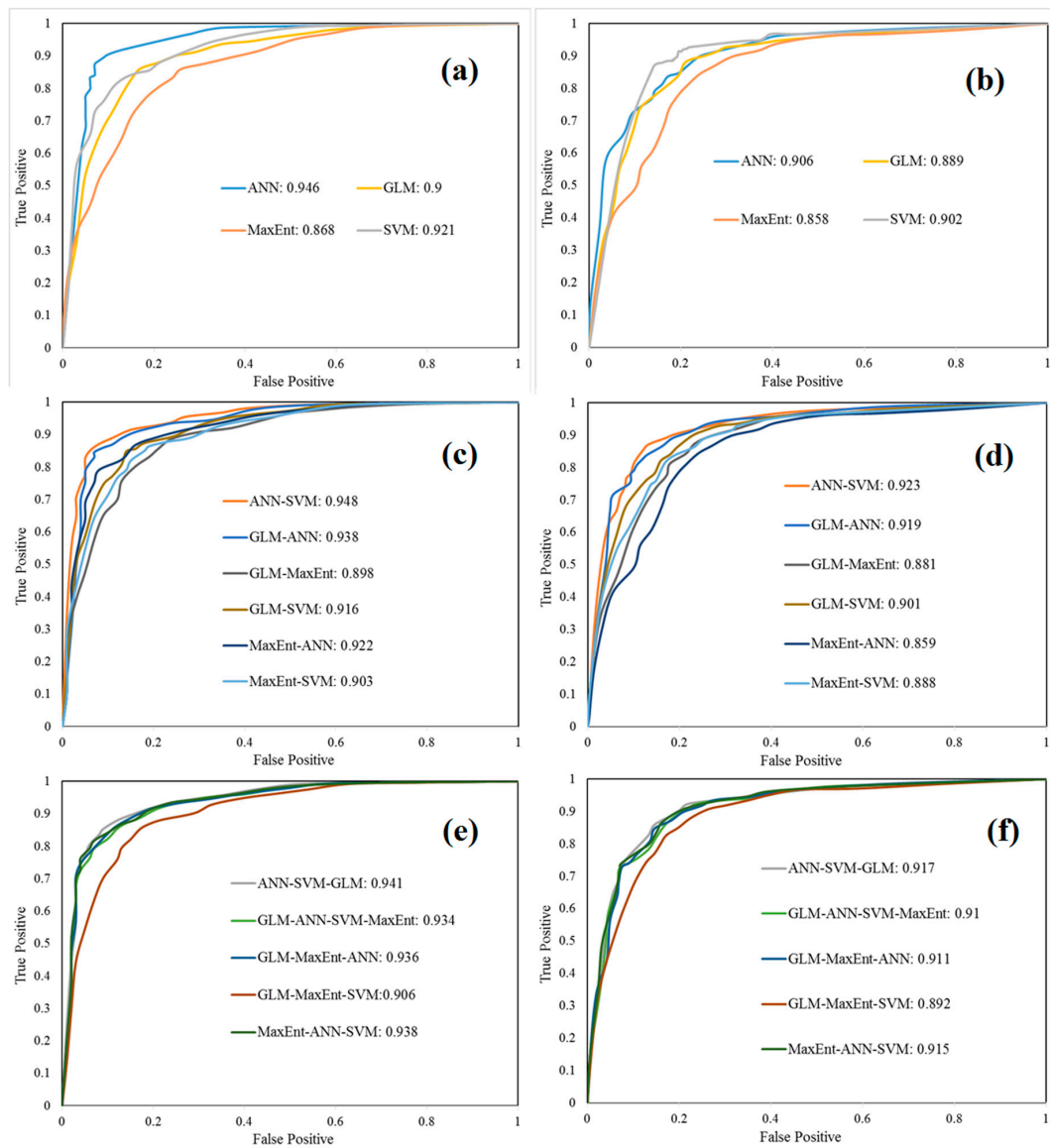


Figure 9. Area under the curves based on training datasets (success rate curve): (a) individual models, (b) ensemble of two models, (c) ensemble of three or four ensemble models based on validation datasets (prediction rate curve), (d) individual models, (e) ensemble of two models, and (f) ensemble of three or four ensemble models.

SCAI was also used to evaluate prediction performance. The SCAI for all the individual and ensemble models indicated that all models achieved at least very good accuracy (Figure 10). The susceptibility classifications of all 15 models were evaluated. The SCAI values decreased from as the susceptibility class increased (Figure 11).

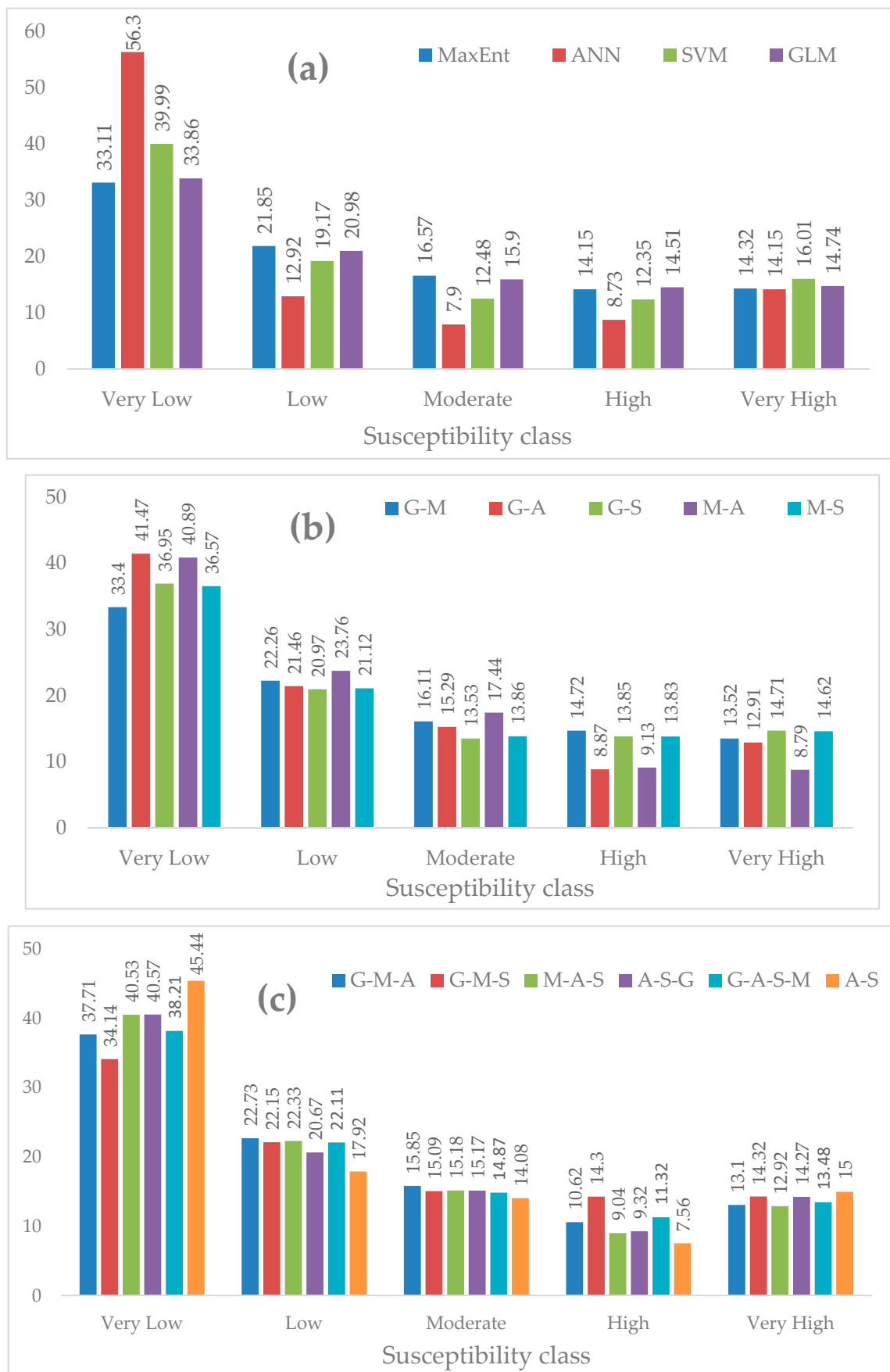


Figure 10. Percentage of each susceptibility class. (a) individual models, (b) ensemble of two models, and (c) ensemble of three and four ensemble models.

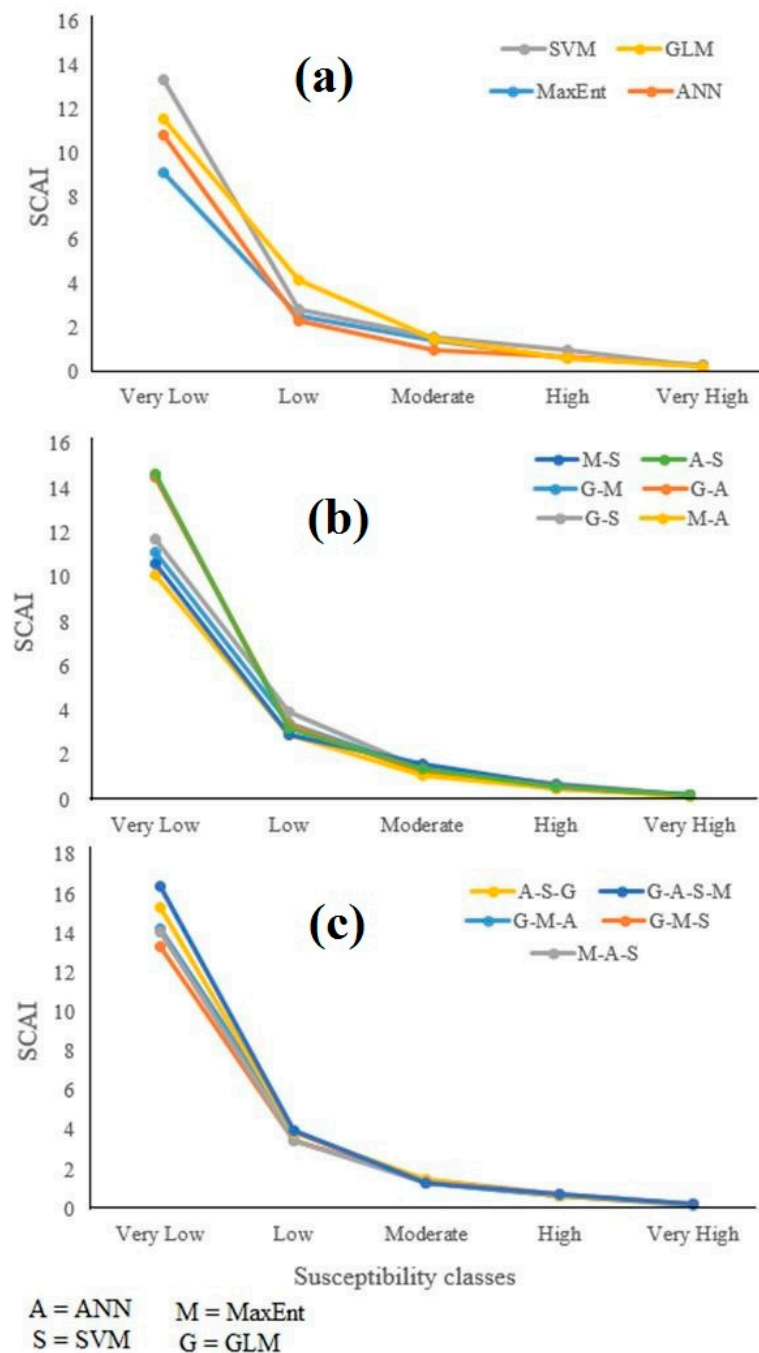


Figure 11. Seed cell area index (SCAI): (a) individual models, (b) two-model ensembles, and (c) three- and four-model ensembles.

4. Discussion

Gully erosion is a geomorphic process but anthropogenic activities accelerate erosion and create imbalances in environmental mechanisms. Borrelli et al. [136] and Azareh et al. [33] concluded that gully erosion is a complex hazard that can interrupt economic growth, cause infrastructural collapse, and disrupt ecosystems. This study used novel machine learning ensembles comprised of MaxEnt, ANN, SVM, and GLM models to produce gully erosion susceptibility models (GESMs). These were evaluated using ROC curves and the SCAI. GESMs can be useful to governmental institutions, especially as they can improve decision making for gully erosion mitigation [19–22]. Several GESMs

have been developed over the last few years and scholars have been searching for the best combinations of models and methods [19–25,45].

The Golestan Dam basin, encompassing most of Golestan Province, Iran, was chosen as the study site in which severe gully erosion can be assessed with individual and ensemble predictive models. Unique among GESM studies, this study tested and evaluated the results from two-, three-, and four-model ensembles. A gully erosion inventory data set was created and used in a now-common [19–25,44,58–60] 70:30 random partition to calibrate the models and evaluate the results, respectively. GECFs that are useful in one study area may not be suitable other regions, therefore, factors were selected based on a review of literatures describing GES models [45,58]. The literatures helped to identify fourteen geo-environmental factors—elevation, slope, slope aspect, rainfall, geology, relative slope position, distance to stream, distance to road, drainage density, road density, slope length, HAND—that were used to map GES in the study area [19–25,44,58–60,78]. After modeling, the contributions of each of the GECFs were assessed using the RF method [58]. The results indicated that the DtR and rainfall were the most important predictive variables among the GECFs. Using Jenks' natural break method, the GES maps were classified into five susceptibility classes. The GESMs classified 12.91%, to 16.01% of the study area as very-highly susceptible to gully erosion.

It is very difficult to avoid the problem of overfitting because non-gully pixels were chosen in a 1:1 ratio with the gully-present pixels, even though the actual number non-gully pixels is greater than the number of gully pixels. Furthermore, the training and testing datasets were based on a sampling ratio of 70:30 without testing the accuracy of the sampling ratio. Additionally, despite the multicollinearity test, noise may remain among the GECFs. The key benefit of machine learning algorithms is that they optimize the searches of several datasets from which valuable information is collected. Such methods may compensate for assumptions and can aid the formulation of strategies and the statistical processing of large datasets. Despite several drawbacks of machine learning, this study can help scholars create new models to improve performance not only for GES modelling, but for modelling of other hazards like floods, landslides, and others.

The distribution of GES in the Golestan Dam basin was successfully mapped by all individual and ensemble models. SVM-based models are very capable of dealing with non-linear and high-dimensional grouping problems by use of the kernel feature and the adding of slack parameters [137]. However, there may still be overestimation in SVM's GES model [138]. Some methods can be paired with SVM to boost the efficiency of GES assessments, which may achieve better results than other models [138]. The ANN-SVM ensemble achieved the highest accuracy, resembling the results in Chen et al. [46] and Pourghasemi et al. [45].

Models Prioritization

Comparing and appraising the machine learning models' results was a key objective of this study. The GESMs were constructed using the MaxEnt, ANN, GLM, and SVM models and eleven ensembles of these models. Model prioritization is an innovative and creative technique to evaluate model performance to select the best model. The prioritization process has been used in drainage morphometrics studies as sub-basin prioritization [139]. A similar approach was used here. The individual models, and the two-, three-, and four-model ensembles were used to produce fifteen GESMs. To rate the models, the AUC of SRC and the AUC of PRC were calculated and for each the models were ranked (i.e., prioritized) (Figure 12). The ANN-SVM (PR = 1) model ranked first. It was followed in rank by GLM, MaxEnt, ANN, SVM, GLM-MaxEnt, GLM-ANN, GLM-SVM, MaxEnt-ANN, MaxEnt-SVM, GLM-MaxEnt-ANN, GLM-MaxEnt-SVM, MaxEnt-ANN-SVM, ANN-SVM-GLM, and ANN-SVM-GLM-MaxEnt.

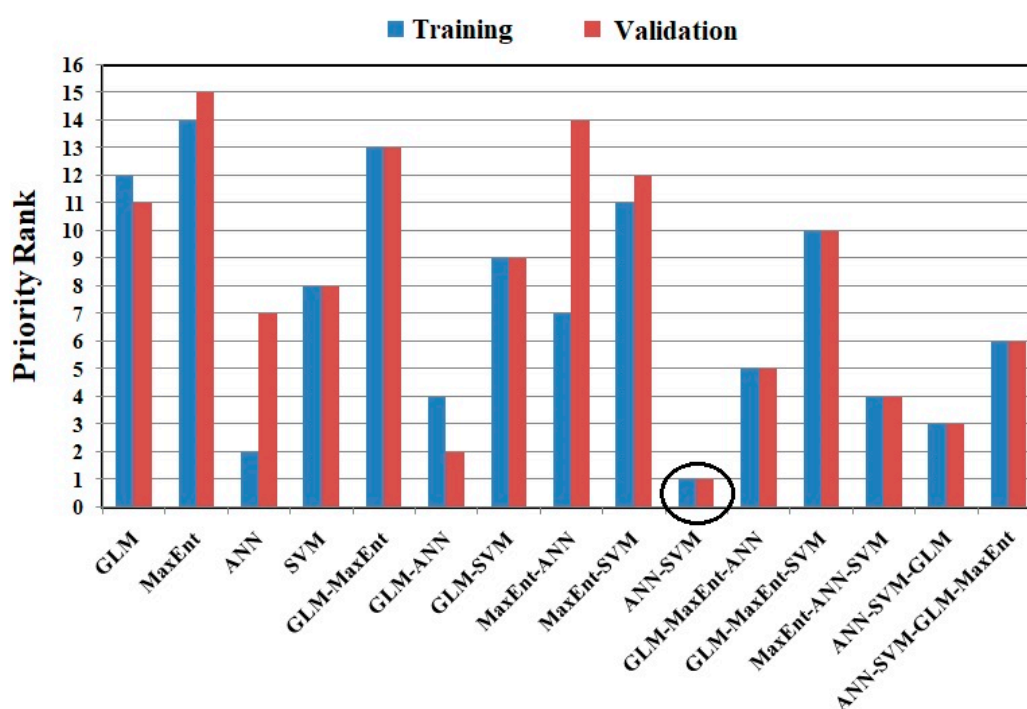


Figure 12. Prioritization of GESMs based on the AUC values of PRC and SRC.

5. Conclusions

The Golestan Dam basin is one of the most gully-prone regions in Golestan Province owing to the combined impacts of GECFs. Modelling of GES is essential for the management of this destructive phenomenon in the Golestan Dam catchment. The broad goal of models is to provide more accessible and meaningful soil information to decision-makers based on the cutting-edge knowledge in this field. Communities in the Golestan Dam basin are facing soil erosion rates that threaten agriculture and local infrastructure. GESMs were developed from four standalone models (MaxEnt, ANN, SVM, and GLM). They were integrated into two-, three-, and four-model ensembles (GLM-MaxEnt, GLM-ANN, GLM-SVM, MaxEnt-ANN, MaxEnt-SVM, GLM-MaxEnt-ANN, GLM-MaxEnt-SVM, MaxEnt-ANN-SVM, ANN-SVM-GLM and ANN-SVM-GLM-MaxEnt). The estimations of GES from all of the models were categorized into five levels. AUC of the ROC curve and SCAI methods were used to evaluate the ensemble accuracies of the models' mapped predictions of GES. All models achieved excellent results. Among the 15 GES models, ANN-SVM produced the highest predictive capability and was selected as the best model for these purposes in this basin. GESMs are efficient and effective measures for the potential of gully erosion in this area. The government institutions in Iran can employ the ANN-SVM GESM tool to prevent formation of gullies and to mitigate soil erosion. To prevent soil erosion in the areas of highest GES, more intensive measures need to be implemented. Community outreach and public education are vital to soil conservation and management. Thus, different scholars, engineers, and decision makers have important roles in the quest for sustainable management in these areas that are susceptible to soil erosion and particularly to the formation of gullies.

Author Contributions: Conceptualization, A.A.; methodology, A.A.; software, A.A. and O.A.N.; validation, A.A., and O.A.N.; formal analysis, A.A. and O.A.N.; investigation, A.A.; resources, A.A.; data curation, A.A.; writing—original draft preparation, A.A., S.S., and J.R.; writing—review and editing, A.A., S.S., J.R., B.P., J.P.T. and P.T.T.N.; visualization, A.A.; supervision, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We are grateful to the Editor, Vicky Huang, and three anonymous referees for their constructive comments which were valuable to improve our manuscript.

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

References

1. Ul Zaman, M.; Bhat, S.; Sharma, S.; Bhat, O. Methods to control soil erosion—a review. *Int. J. Pure Appl. Biosci.* **2018**, *6*, 1114–1121. [[CrossRef](#)]
2. Parlak, M. Determination of erosion risk according to CORINE methodology (a case study: Kurtboğazi Dam). *Int. Congr. River Basin Manag.* **2007**, *1*, 856.
3. Swarnkar, S.; Malini, A.; Tripathi, S.; Sinha, R. Assessment of uncertainties in soil erosion and sediment yield estimates at ungauged basins: An application to the Garra River basin, India. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 2471–2485. [[CrossRef](#)]
4. Arabameri, A.; Pradhan, B.; Pourghasemi, H.R.; Rezaei, K. Identification of erosion-prone areas using different multi-criteria decision-making techniques and GIS. *Geomat. Nat. Haz. Risk* **2018**, *9*, 1129–1155. [[CrossRef](#)]
5. 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*, 1st ed.; Pourghasemi, H.R., Gokceoglu, C., Eds.; Elsevier Publication: Amsterdam, The Netherlands, 2019. [[CrossRef](#)]
6. Singh, O.; Singh, J. Soil Erosion Susceptibility Assessment of the Lower Himachal Himalayan Watershed. *J. Geol. Soc. India* **2018**, *92*, 157–165. [[CrossRef](#)]
7. Odunuga, S.; Ajjola, A.; Igwet, N.; Adegun, O. Land susceptibility to soil erosion in Orashi Catchment, Nnewi South, Anambra State, Nigeria. *Proc. Int. Assoc. Hydrol. Sci.* **2018**, *376*, 87–95. [[CrossRef](#)]
8. Magliulo, P. Assessing the susceptibility to water-induced soil erosion using a geomorphological, bivariate statistics-based approach. *Environ. Earth Sci.* **2012**, *67*, s1801–s1820. [[CrossRef](#)]
9. Arabameri, A.; Pourghasemi, H.R.; Yamani, M. Applying different scenarios for landslide spatial modeling using computational intelligence methods. *Environ. Earth Sci.* **2017**, *76*. [[CrossRef](#)]
10. Arabameri, A.; Pourghasemi, H.R.; Cerda, A. Erodibility prioritization of subwatersheds using morphometric parameters analysis and its mapping: A comparison among TOPSIS, VIKOR, SAW, and CF multi-criteria decision making models. *Sci. Total Environ.* **2017**, *613–614*, 1385–1400. [[CrossRef](#)]
11. Arabameri, A.; Rezaei, K.; Pourghasemi, H.R.; Lee, S.; Yamani, M. GIS-based gully erosion susceptibility mapping: A comparison among three data-driven models and AHP knowledge-based technique. *Environ. Earth Sci.* **2018**, *77*, 628. [[CrossRef](#)]
12. Sun, W.; Shao, Q.; Liu, J.; Zhai, J. Assessing the effects of land use and topography on soil erosion on the Loess Plateau in China. *Catena* **2014**, *121*, 151–163. [[CrossRef](#)]
13. Alam, M.; Hussain, R.R.; Islam, A.S. Impact assessment of rainfall-vegetation on sedimentation and predicting erosion-prone region by GIS and RS. *Geomat. Nat. Hazards Risk* **2016**, *7*, 667–679. [[CrossRef](#)]
14. Li, H.; Cruse, R.M.; Bingner, R.L.; Gesch, K.R.; Zhang, X. Evaluating ephemeral gully erosion impact on Zea mays L. yield and economics using AnnAGNPS. *Soil Tillage Res.* **2016**, *155*, 157–165. [[CrossRef](#)]
15. Saha, S.; Gayen, A.; Pourghasemi, H.R.; Tiefenbacher, J.P. Identification of soil erosion-susceptible areas using fuzzy logic and analytical hierarchy process modeling in an agricultural watershed of Burdwan district, India. *Environ. Earth Sci.* **2019**, *78*, 649. [[CrossRef](#)]
16. Refahi, H. *Soil Erosion by Water & Conservation*; Tehran University Press: Tehran, Iran, 2009; pp. 10–202, (In Farsi with English Summary).
17. Ezechi, J.I. *The Influence of Runoff, Lithology and Water Table on the Dimensions and Rate of Gully Processes in Eastern, Nigeria*; Elsevier: Cremlingen, Germany, 2000.

18. Paolo, P.; Desmond, W.E.; Antonina, C. Using 137 CS and 210 Pbex measurements and conventional surveys to investigate the relative contributions of inter rill/rill and gully erosion to soil loss from a small cultivated catchment in Sicily. *Soil Tillage Res.* **2014**, *135*, 18–27. [[CrossRef](#)]
19. Arabameri, A.; Pradhan, B.; Bui, D.T. Spatial modelling of gully erosion in the Ardib River Watershed using three statistical-based techniques. *Catena* **2020**, *190*, 104545. [[CrossRef](#)]
20. Arabameri, A.; Cerda, A.; Pradhan, B.; Tiefenbacher, J.P.; Lombardo, L.; Bui, D.T. A methodological comparison of head-cut based gully erosion susceptibility models: Combined use of statistical and artificial intelligence. *Geomorphology* **2020**, 107136. [[CrossRef](#)]
21. Zabihi, M.; Mirchooli, F.; Motevalli, A.; Darvishan, A.K.; Pourghasemi, H.R.; Zakeri, M.A.; Sadighi, F. Spatial modelling of gully erosion in Mazandaran Province, northern Iran. *Catena* **2018**, *161*, 1–13. [[CrossRef](#)]
22. Arabameri, A.; Pradhan, B.; Rezaei, K.; Conoscenti, C. Gully erosion susceptibility mapping using GIS-based multi-criteria decision analysis techniques. *Catena* **2019**, *180*, 282–297. [[CrossRef](#)]
23. Arabameri, A.; Chen, W.; Loche, M.; Zhao, X.; Li, Y.; Lombardo, L.; Cerda, A.; Pradhan, B.; Bui, D.T. Comparison of machine learning models for gully erosion susceptibility mapping. *Geosci. Front.* **2019**. [[CrossRef](#)]
24. Arabameri, A.; Pradhan, B.; Rezaei, K. Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS. *J. Environ. Manag.* **2019**, *232*, 928–942. [[CrossRef](#)]
25. Zakerinejad, R.; Maerker, M. An integrated assessment of soil erosion dynamics with special emphasis on gully erosion in the Mazayjan basin, southwestern Iran. *Nat. Hazards* **2015**, *79*, 25–50. [[CrossRef](#)]
26. Shit, P.K.; Paira, R.; Bhunia, G.; Maiti, R. Modeling of potential gully erosion hazard using geo-spatial technology at Garbgheta block, West Bengal in India. *Modeling Earth Syst. Environ.* **2015**, *1*, 1–16. [[CrossRef](#)]
27. Ganasri, B.P.; Ramesh, H. Assessment of soil erosion by RUSLE model using remote sensing and GIS-A case study of Nethravathi Basin. *Geosci. Front.* **2016**, *7*, 953–961. [[CrossRef](#)]
28. Arekhi, S.; Niazi, Y. Assessment of GIS and RS applications to estimate soil erosion and sediment loading by using RUSLE model (Case Study: Upstream basin of Ilam dam). *J. Soil Water Conserv.* **2010**, *17*, 1–27.
29. Roy, J.; Saha, S. Assessment of land suitability for the paddy cultivation using analytical hierarchical process (AHP): A study on Hinglo river basin, Eastern India. *Modeling Earth Syst. Environ.* **2018**, *4*, 601–618. [[CrossRef](#)]
30. Rahmati, O.; Haghizadeh, A.; Pourghasemi, H.R.; Noormohamadi, F. Gully erosion susceptibility mapping: The role of GIS-based bivariate statistical models and their comparison. *Nat. Hazards* **2016**, *82*, 1231–1258. [[CrossRef](#)]
31. Conforti, M.; Aucelli, P.P.; Robustelli, G.; Scarciglia, F. Geomorphology and GIS analysis formapping gully erosion susceptibility in the Turbolo streamcatchment (Northern Calabria, Italy). *Nat. Hazards* **2011**, *56*, 881–898. [[CrossRef](#)]
32. Mojjaddadi, H.; Pradhan, B.; Nampak, H.; Ahmad, N.; Ghazali, A.H. Ensemble machine-learning-based geospatial approach for flood risk assessment using multisensory remote-sensing data and GIS. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1080–1102. [[CrossRef](#)]
33. Azareh, A.; Rahmati, O.; Rafiei-Sardooi, E.; Sankey, J.B.; Lee, S.; Shahabi, H.; Bin Ahmad, B. Modelling gully-erosion susceptibility in a semi-arid region, Iran: Investigation of applicability of certainty factor and maximum entropy models. *Sci. Total Environ.* **2019**, *655*, 684–696. [[CrossRef](#)]
34. Aghdam, I.N.; Varzandeh, M.H.M.; Pradhan, B. Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran). *Environ. Earth Sci.* **2016**, *75*, 553. [[CrossRef](#)]
35. Kornejady, A.; Heidari, K.; Nakhavali, M. Assessment of landslide susceptibility, semi-quantitative risk and management in the Ilam dam basin Ilam, Iran. *Environ. Resour. Res.* **2015**, *3*, 85–109. [[CrossRef](#)]
36. Dube, F.; Nhapi, I.; Murwira, A.; Gumindoga, W.; Goldin, J.; Mashauri, D.A. Potential of weight of evidence modelling for gully erosion hazard assessment in Mbire District—Zimbabwe. *Phys. Chem. Earth* **2014**, *67*, 145–152. [[CrossRef](#)]
37. Zakerinejad, R.; Maerker, M. Prediction of gully erosion susceptibilities using detailed terrain analysis and maximum entropy modeling: A case study in the Mazayejan Plain, Southwest Iran. *Geogr. Fis. Din. Quat.* **2014**, *37*, 67–76. [[CrossRef](#)]

38. Gómez-Gutiérrez, Á.; Conoscenti, C.; Angileri, S.E.; Rotigliano, E.; Schnabel, S. Using topographical attributes to evaluate gully erosion proneness (susceptibility) in two Mediterranean basins: Advantages and limitations. *Nat. Hazards* **2015**, *79*, 291–314. [[CrossRef](#)]
39. Pradhan, B.; Lee, S. Regional landslide susceptibility analysis using backpropagation neural network model at Cameron Highland, Malaysia. *Landslides* **2010**, *7*, 13–30. [[CrossRef](#)]
40. Dehnavi, A.; Aghdam, I.N.; Pradhan, B.; Varzandeh, M.H.M. A new hybrid model using step-wise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *Catena* **2015**, *135*, 122–148. [[CrossRef](#)]
41. Amiri, M.; Pourghasemi, H.R.; Ghanbarian, G.A.; Afzali, S.F. Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma* **2019**, *340*, 55–69. [[CrossRef](#)]
42. Arabameri, A.; Pradhan, B.; Rezaei, K.; Yamani, M.; Pourghasemi, H.R.; Lombardo, L. Spatial modelling of gully erosion using evidential belief function, logistic regression, and a new ensemble of evidential belief function–logistic regression algorithm. *Land Degrad. Dev.* **2018**, *29*, 4035–4049. [[CrossRef](#)]
43. Hosseinalizadeh, M.; Kariminejad, N.; Chen, W.; Pourghasemi, H.R.; Alinejad, M.; Behbahani, A.M.; Tiefenbacher, J.P. Spatial modelling of gully headcuts using UAV data and four best-first decision classifier ensembles (BFTree, Bag-BFTree, RS-BFTree, and RF-BFTree). *Geomorphology* **2019**, *329*, 184–193. [[CrossRef](#)]
44. Roy, J.; Saha, S. GIS-based Gully Erosion Susceptibility Evaluation Using Frequency Ratio, Cosine Amplitude and Logistic Regression Ensembled with fuzzy logic in Hinglo River Basin, India. *Remote Sens. Appl. Soc. Environ.* **2019**, *15*, 100247. [[CrossRef](#)]
45. Pourghasemi, H.R.; Yourself, 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](#)]
46. Chen, W.; Pourghasemi, H.R.; Kornejady, A.; Zhang, N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* **2017**, *305*, 314–327. [[CrossRef](#)]
47. Pham, B.T.; Bui, D.T.; Prakash, I.; Dholakia, M.B. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **2017**, *149*, 52–63. [[CrossRef](#)]
48. Arabameri, A.; Yamani, M.; Pradhan, B.; Melesse, A.; Shirani, K.; Bui, D.T. Novel ensembles of COPRAS multi-criteria decision-making with logistic regression, boosted regression tree, and random forest for spatial prediction of gully erosion susceptibility. *Sci. Total Environ.* **2019**, *688*, 903–916. [[CrossRef](#)] [[PubMed](#)]
49. Arabameri, A.; Pradhan, B.; Rezaei, K. Spatial prediction of gully erosion using ALOS PALSAR data and ensemble bivariate and data mining models. *Geosci. J.* **2019**, *23*, 669–686. [[CrossRef](#)]
50. Chen, W.; Li, H.; Hou, E.; Wang, S.; Wang, G.; Panahi, M.; Li, T.; Peng, T.; Guo, C.; Niu, C.; et al. GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci. Total Environ.* **2018**, *634*, 853–867. [[CrossRef](#)]
51. Arabameri, A.; Pradhan, B.; Lombardo, L. Comparative assessment using boosted regression trees, binary logistic regression, frequency ratio and numerical risk factor for gully erosion susceptibility modelling. *Catena* **2019**, *183*, 104223. [[CrossRef](#)]
52. Arabameri, A.; Chen, W.; Blaschke, T.; Tiefenbacher, J.P.; Pradhan, B.; Tien Bui, D. Gully Head-Cut Distribution Modeling Using Machine Learning Methods—A Case Study of NW Iran. *Water* **2020**, *12*, 16. [[CrossRef](#)]
53. Arabameri, A.; Chen, W.; Lombardo, L.; Blaschke, T.; Tien Bui, D. Hybrid computational intelligence models for improvement gully erosion assessment. *Remote Sens.* **2020**, *12*, 140. [[CrossRef](#)]
54. Arabameri, A.; Pradhan, B.; Rezaei, K.; Lee, C.W. Assessment of landslide susceptibility using statistical-and artificial intelligence-based FR–RF integrated model and multiresolution DEMs. *Remote Sens.* **2019**, *11*, 999. [[CrossRef](#)]
55. Arabameri, A.; Cerda, A.; Rodrigo-Comino, J.; Pradhan, B.; Sohrabi, M.; Blaschke, T.; Tien Bui, D. Proposing a novel predictive technique for gully erosion susceptibility mapping in arid and semi-arid regions (Iran). *Remote Sens.* **2019**, *11*, 2577. [[CrossRef](#)]
56. Arabameri, A.; Cerda, A.; Tiefenbacher, J.P. Spatial pattern analysis and prediction of gully erosion using novel hybrid model of entropy-weight of evidence. *Water* **2019**, *11*, 1129. [[CrossRef](#)]
57. Zerihun, M.; Mohammedyasin, M.S.; Sewnet, D.; Adem, A.A.; Lakew, M. Assessment of soil erosion using RUSLE, GIS and remote sensing in NW Ethiopia. *Geoderma Reg.* **2018**, *12*, 83–90. [[CrossRef](#)]

58. Islamic Republic of Iran Meteorological Organization (IRIMO). 2012. Available online: <http://www.mazandaranmet.ir> (accessed on 12 August 2018).
59. IUSS Working Group WRB. *World Reference Base for Soil Resources 2014; World Soil Resources Report*; FAO: Rome, Italy, 2014.
60. Saha, S.; Roy, J.; Arabameri, A.; Blaschke, T.; Tien Bui, D. Machine learning-based gully erosion susceptibility mapping: A case study of Eastern India. *Sensors* **2020**, *20*, 1313. [[CrossRef](#)]
61. Conoscenti, C.; Angileri, S.; Cappadonia, C.; Rotigliano, E.; Agnesi, V.; Marker, M. Gully erosion susceptibility assessment by means of GIS-based logistic regression: A case of Sicily (Italy). *Geomorphology* **2014**, *204*, 399–411. [[CrossRef](#)]
62. Cama, M.; Lombardo, L.; Conoscenti, C.; Rotigliano, E. Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). *Geomorphology* **2017**, *288*, 52–65. [[CrossRef](#)]
63. Garosi, Y.; Sheklabadi, M.; Pourghasemi, H.R.; Besalatpour, A.A.; Conoscenti, C.; Van Oost, K. Comparison of differences in resolution and sources of controlling factors for gully erosion susceptibility mapping. *Geoderma* **2018**, *330*, 65–78. [[CrossRef](#)]
64. Ollobarren Del Barrio, P.; Campo-Bescós, M.A.; Giménez, R.; Casalí, J. Assessment of soil factors controlling ephemeral gully erosion on agricultural fields. *Earth Surf. Process. Landf.* **2018**, *43*, 1993–2008. [[CrossRef](#)]
65. Kheir, R.B.; Wilson, J.; Deng, Y. Use of terrain variables for mapping gully erosion susceptibility in Lebanon. *Earth Surf. Process. Landf.* **2007**, *32*, 1770–1782. [[CrossRef](#)]
66. Montgomery, D.R.; Dietrich, W.E. Channel initiation and the problem of landscape scale. *Science* **1992**, *255*, 826. [[CrossRef](#)] [[PubMed](#)]
67. Parras-Alcántara, L.; Lozano-García, B.; Keesstra, S.; Cerdà, A.; Brevik, E.C. Long-term effects of soil management on ecosystem services and soil loss estimation in olive grove top soils. *Sci. Total Environ.* **2016**, *571*, 498–506. [[CrossRef](#)] [[PubMed](#)]
68. Poesen, J.; Nachtergaele, J.; Verstraeten, G.; Valentin, C. Gully erosion and environmental change: Importance and research needs. *Catena* **2003**, *50*, 91–133. [[CrossRef](#)]
69. Boreggio, M.; Bernard, M.; Gregoretti, C. Evaluating the influence of gridding techniques for Digital Elevation Models generation on the debris flow routing modeling: A case study from Rovina di Cancia basin (North-eastern Italian Alps). *Front. Earth Sci.* **2018**, *6*, 89. [[CrossRef](#)]
70. Gesch, D.; Oimoen, M.; Zhang, Z.; Meyer, D.; Danielson, J. Validation of the ASTER global digital elevation model version 2 over the conterminous United States. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *39*, 281–286. [[CrossRef](#)]
71. Zhou, C.; Ge, L.E.D.; Chang, H.C. A case study of using external DEM in InSAR DEM generation. *Geo-Spat. Inf. Sci.* **2005**, *8*, 14–18. [[CrossRef](#)]
72. Zhang, W.; Wang, W.; Chen, L. Constructing DEM based on InSAR and the relationship between InSAR DEM's precision and terrain factors. *Energy Procedia* **2012**, *16*, 184–189. [[CrossRef](#)]
73. Conoscenti, C.; Agnesi, V.; Cama, M.; Caraballo-Arias, N.A.; Rotigliano, E. Assessment of gully erosion susceptibility using multivariate adaptive regression splines and accounting for terrain connectivity. *Land Degrad. Dev.* **2018**, *29*, 724–736. [[CrossRef](#)]
74. Ghorbani Nejad, S.; Falah, F.; Daneshfar, M.; Haghizadeh, A.; Rahmati, O. Delineation of groundwater potential zones using remote sensing and GIS-based data-driven models. *Geocarto Int.* **2016**. [[CrossRef](#)]
75. Jaafari, A.; Najafi, A.; Pourghasemi, H.R.; Rezaeian, J.; Sattarian, A. GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *Int. J. Environ. Sci. Technol.* **2014**, *11*, 909–926. [[CrossRef](#)]
76. Nobre, A.D.; Cuartas, L.A.; Hodnett, M.; Rennó, C.D.; Rodrigues, G.; Silveira, A.; Waterloo, M.; Saleska, S. Height above the nearest drainage—A hydrologically relevant new terrain model. *J. Hydrol.* **2011**, *404*, 13–29. [[CrossRef](#)]
77. Kornejady, A.; Ownegh, M.; Bahremand, A. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena* **2017**, *152*, 144–162. [[CrossRef](#)]
78. Rahmati, O.; Moghaddam, D.D.; Moosavi, V.; Kalantari, Z.; Samadi, M.; Lee, S.; Tien Bui, D. An automated python language-based tool for creating absence samples in groundwater potential mapping. *Remote Sens.* **2019**, *11*, 1375. [[CrossRef](#)]

79. Liao, Y.; Yuan, Z.; Zhuo, M.; Huang, B.; Nie, X.; Xie, Z.; Tang, C.; Li, D. Coupling effects of erosion and surface roughness on colluvial deposits under continuous rainfall. *Soil Tillage Res.* **2019**, *191*, 98–107. [[CrossRef](#)]
80. Boer, E.P.; de Beurs, K.M.; Hartkamp, A.D. Kriging and thin plate splines for mapping climate variables. *Int. J. Appl. Earth Obs. Geoinf.* **2001**, *3*, 146–154. [[CrossRef](#)]
81. Beach, T. The fate of eroded soil: Sediment sinks and sediment budgets of agrarian landscapes in Southern Minnesota, 1851–1988. *Ann. Assoc. Am. Geogr.* **1994**, *84*, 5–28. [[CrossRef](#)]
82. Nyssen, J.; Poesen, J.; Moeyersons, J.; Luyten, E.; Veyret-Picot, M.; Deckers, J.; Haile, M.; Govers, G. Impact of road building on gully erosion risk: A case study from the northern Ethiopian highlands. *Earth Surf. Process. Landf. J. Br. Geomorphol. Res. Group* **2002**, *27*, 1267–1283. [[CrossRef](#)]
83. Bean, T.; Sumner, P.; Boojhawon, R.; Tatayah, V.; Khadun, A.; Hedding, D.W.; Rughooputh, S.D.D.V.; Nel, W. Bedrock-incised gully erosion phenomena on Round Island, Mauritius. *Catena* **2017**, *151*, 107–117. [[CrossRef](#)]
84. Geology Survey of Iran (GSI). 1997. Available online: http://www.gsi.ir/Main/Lang_en/index.html (accessed on 5 June 2020).
85. Maestre, F.T.; Cortina, J. Spatial patterns of surface soil properties and vegetation in a Mediterranean semiarid steppe. *Plant Soil* **2002**, *241*, 279–291. [[CrossRef](#)]
86. Zucca, C.; Canu, A.; Della Peruta, R. Effects of land use and landscape on spatial distribution and morphological features of gullies in an agropastoral area in Sardinia (Italy). *Catena* **2006**, *68*, 87–95. [[CrossRef](#)]
87. Phillips, S.J.; Dudík, M.; Elith, J.; Graham, C.H.; Lehmann, A.; Leathwick, J.; Ferrier, S. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol. Appl.* **2009**, *19*, 181–197. [[CrossRef](#)]
88. Keesstra, S.; Pereira, P.; Novara, A.; Brevik, E.C.; Azorin-Molina, C.; Parras-Alcántara, L.; Jordán, A.; Cerdà, A. Effects of soil management techniques on soil water erosion in apricot orchards. *Sci. Total Environ.* **2016**, *551*, 357–366. [[CrossRef](#)] [[PubMed](#)]
89. Reddy, S.; Dávalos, L.M. Geographical sampling bias and its implications for conservation priorities in Africa. *J. Biogeogr.* **2003**, *30*, 1719–1727. [[CrossRef](#)]
90. Shannon, C.E. Amathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
91. Elith, J.; Phillips, S.J.; Hastie, T.; Dudík, M.; Chee, Y.E.; Yates, C.J. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **2011**, *17*, 43–57. [[CrossRef](#)]
92. Phillips, S.J.; Dudík, M.; Schapire, R.E. A maximum entropy approach to species distribution modeling. In Proceedings of the Twenty-First International Conference on Machine Learning, London, UK, 24 July 2004; p. 83.
93. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620. [[CrossRef](#)]
94. Jaynes, E.T. Information theory and statistical mechanics. II. *Phys. Rev.* **1957**, *108*, 171. [[CrossRef](#)]
95. Cherkassky, V.; Krasnopolsky, V.; Solomatine, D.P.; Valdes, J. Computational intelligence in earth sciences and environmental applications: Issues and challenges. *Neural Netw.* **2006**, *19*, 113–121. [[CrossRef](#)]
96. Peddle, D.R.; Foody, G.M.; Zhang, A.; Franklin, S.E.; LeDrew, E.F. Multi-source image classification II: An empirical comparison of evidential reasoning and neural network approaches. *Can. J. Remote Sens.* **1994**, *20*, 396–407. [[CrossRef](#)]
97. Kosko, B. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence/Book and Disk*; Prentice Hall: Upper Saddle River, NJ, USA, 1992.
98. Falaschi, F.; Giacomelli, F.; Federici, P.R.; Puccinelli, A.; Avanzi, G.A.; Pochini, A.; Ribolini, A. Logistic regression versus artificial neural networks: Landslide susceptibility evaluation in a sample area of the Serchio River valley, Italy. *Nat. Hazards* **2009**, *50*, 551–569. [[CrossRef](#)]
99. Gong, P.; Pu, R.; Chen, J. Elevation and forest-cover data using neural networks. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 1249–1260.
100. Arora, M.K.; Das Gupta, A.S.; Gupta, R.P. An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. *Int. J. Remote Sens.* **2004**, *25*, 559–572. [[CrossRef](#)]
101. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
102. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
103. Joachims, T. Text categorization with support vector machines: Learning with many relevant features. In *European Conference on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 137–142.

104. Burges, C.J. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]
105. Kanevski, M.; Pozdnoukhov, A.; Timonin, V. *Machine Learning for Spatial Environmental Data: Theory, Applications, and Software*; EPFL Press: Lausanne, Switzerland, 2009.
106. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženilek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [[CrossRef](#)]
107. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2001.
108. Tax, D.M.; Duin, R.P. Support vector domain description. *Pattern Recognit. Lett.* **1999**, *20*, 1191–1199. [[CrossRef](#)]
109. Pourghasemi, H.R.; Jirandeh, A.G.; Pradhan, B.; Xu, C.; Gokceoglu, C. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *J. Earth Syst. Sci.* **2013**, *122*, 349–369. [[CrossRef](#)]
110. Poeppl, R.E.; Keesstra, S.D.; Maroulis, J. A conceptual connectivity framework for understanding geomorphic change in human-impacted fluvial systems. *Geomorphology* **2017**, *277*, 237–250. [[CrossRef](#)]
111. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 1989; p. 532.
112. Payne, R. *A Guide to Regression, Nonlinear and Generalized Linear Models in GenStat*; VSN International: Hemel Hempstead, UK, 2012; p. 88.
113. Bernknopf, R.L.; Brookshire, D.S.; Shapiro, C.D. A probabilistic approach to landslide hazard mapping in Cincinnati, Ohio, with applications for economic evaluation. *Bull. Assoc. Eng. Geol.* **1988**, *24*, 39–56. [[CrossRef](#)]
114. Nikita, E. The use of generalized linear models and generalized estimating equations in bioarchaeological studies. *Am. J. Phys. Anthropol.* **2014**, *153*, 473–483. [[CrossRef](#)]
115. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
116. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman & Hall: New York, NY, USA, 1984.
117. Micheletti, N.; Foresti, L.; Robert, S.; Leuenberger, M.; Pedrazzini, A.; Jaboyedoff, M.; Kanevski, M. Machine learning feature selection methods for landslide susceptibility mapping. *Math. Geosci.* **2014**, *46*, 33–57. [[CrossRef](#)]
118. Calle, M.L.; Urrea, V. Letter to the Editor: Stability of random forest importance measures. *Brief. Bioinform.* **2010**, *12*, 86–89. [[CrossRef](#)] [[PubMed](#)]
119. Kumar, R.; Indrayan, A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics* **2011**, *48*, 277–287. [[CrossRef](#)] [[PubMed](#)]
120. Pepe, M.S. Receiver operating characteristic methodology. *J. Am. Stat. Assoc.* **2000**, *95*, 308–311. [[CrossRef](#)]
121. Roy, J.; Saha, S.; Arabameri, A.; Blaschke, T.; Bui, D.T. A novel ensemble approach for Landslide Susceptibility Mapping (LSM) in Darjeeling and Kalimpong Districts, West Bengal, India. *Remote Sens.* **2019**, *11*, 2866. [[CrossRef](#)]
122. Paul, G.C.; Saha, S.; Hembram, T.K. Application of the GIS-based probabilistic models for mapping the flood susceptibility in Bansloi Sub-basin of Ganga-Bhagirathi River and their comparison. *Remote Sens. Earth Syst. Sci.* **2019**, *2*, 120–146. [[CrossRef](#)]
123. Saha, S. Groundwater potential mapping using analytical hierarchical process: A study on Md. Bazar Block of Birbhum District, West Bengal. *Spat. Inf. Res.* **2017**. [[CrossRef](#)]
124. Rahmati, O.; Golkarian, A.; Biggs, T.; Keesstra, S.; Mohammadi, F.; Daliakopoulos, I.N. Land subsidence hazard modeling: Machine learning to identify predictors and the role of human activities. *J. Environ. Manag.* **2019**, *236*, 466–480. [[CrossRef](#)]
125. Rahmati, O.; Falah, F.; Naghibi, S.A.; Biggs, T.; Soltani, M.; Deo, R.C.; Cerdà, A.; Mohammadi, F.; Bui, D.T. Land subsidence modelling using tree-based machine learning algorithms. *Sci. Total Environ.* **2019**, *672*, 239–252. [[CrossRef](#)]
126. Pourghasemi, H.R.; Mohammady, M.; Pradhan, B. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* **2012**, *97*, 71–84. [[CrossRef](#)]
127. Frattini, P.; Crosta, G.; Carrara, A. Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* **2010**, *111*, 62–72. [[CrossRef](#)]

128. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378. [[CrossRef](#)]
129. Fressard, M.; Thiery, Y.; Maquaire, O. Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the Pays d’Auge plateau hillslopes (Normandy, France). *Nat. Hazards Earth Syst. Sci. Discuss.* **2014**, *14*, 569–588. [[CrossRef](#)]
130. Arabameri, A.; Roy, J.; Saha, S.; Blaschke, T.; Ghorbanzadeh, O.; Tien Bui, D. Application of probabilistic and machine learning models for groundwater potentiality mapping in Damghan Sedimentary Plain, Iran. *Remote Sens.* **2019**, *11*, 3015. [[CrossRef](#)]
131. Lipovetsky, S. Pareto 80/20 law: Derivation via random partitioning. *Int. J. Math. Educ. Sci. Technol.* **2009**, *40*, 271–277. [[CrossRef](#)]
132. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F.; Mosavi, A. An ensemble prediction of flood susceptibility using multivariate discriminate analysis, classification and regression trees, and support vector machines. *Sci. Total Environ.* **2019**, *651*, 2087–2096. [[CrossRef](#)]
133. Deo, R.C.; Tiwari, M.K.; Adamowski, J.F.; Quilty, J.M. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. *Stoch. Environ. Res. Risk Assess.* **2017**, *31*, 1211–1240. [[CrossRef](#)]
134. Rokach, L. Ensemble-based classifiers. *Artif. Intell. Rev.* **2010**, *33*, 1–39. [[CrossRef](#)]
135. Tien Bui, D.; Pradhan, B.; Revhaug, I.; Tran, C.T. *A Comparative Assessment between the Application of Fuzzy Unordered Rules Induction Algorithm and J48 Decision Tree Models in Spatial Prediction of Shallow Landslides at Lang Son City*; Springer International Publishing: Cam, Swiss, 2014.
136. Borrelli, P.; Märker, M.; Panagos, P.; Schütt, B. Modeling soil erosion and river sediment yield for an intermountain drainage basin of the Central Apennines, Italy. *Catena* **2014**, *114*, 45–58. [[CrossRef](#)]
137. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **2018**, *165*, 520–529. [[CrossRef](#)]
138. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [[CrossRef](#)]
139. Ahirwar, R.; Malik, M.S.; Shukla, J.P. Prioritization of sub-watersheds for soil and water conservation in parts of Narmada River through morphometric analysis using remote sensing and GIS. *J. Geol. Soc. India* **2019**, *94*, 515–524. [[CrossRef](#)]



© 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/>).

© 2020. This work is licensed under <http://creativecommons.org/licenses/by/3.0/> (the “License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.