








Article

# GIS-Based Gully Erosion Susceptibility Mapping: A Comparison of Computational Ensemble Data Mining Models

Viet-Ha Nhu <sup>1,2</sup>, Saeid Janizadeh <sup>3</sup>, Mohammadtaghi Avand <sup>3</sup> , Wei Chen <sup>4</sup>, Mohsen Farzin <sup>5</sup> , Ebrahim Omidvar <sup>6</sup>, Ataollah Shirzadi <sup>7</sup> , Himan Shahabi <sup>8,9</sup> , John J. Clague <sup>10</sup>, Abolfazl Jaafari <sup>11</sup> , Fatemeh Mansoorypoor <sup>12</sup>, Binh Thai Pham <sup>13,\*</sup> , Baharin Bin Ahmad <sup>14</sup> and Saro Lee <sup>15,16,\*</sup> 

- <sup>1</sup> Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City 758307, Vietnam; nhuvietha@tdtu.edu.vn
  - <sup>2</sup> Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City 758307, Vietnam
  - <sup>3</sup> Department of Watershed Management Engineering, College of Natural Resources, Tarbiat Modares University, Tehran, P.O. Box 14115-111, Iran; Janizadeh.saeed@gmail.com (S.J.); Mt.avand70@gmail.com (M.A.)
  - <sup>4</sup> College of Geology & Environment, Xi'an University of Science and Technology, Xi'an 710054, China; chenwei.0930@163.com
  - <sup>5</sup> Department of Forestry, Range and Watershed Management, Faculty of Agriculture and Natural Resources, Yasouj University, Yasouj 75918-74934, Iran; m.farzin@yu.ac.ir
  - <sup>6</sup> Department of Rangeland and Watershed Management, Faculty of Natural Resources and Earth Sciences, University of Kashan, Kashan 87317-53153, Iran; ebrahimomidvar@kashanu.ac.ir
  - <sup>7</sup> Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran; a.shirzadi@uok.ac.ir
  - <sup>8</sup> Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran; h.shahabi@uok.ac.ir
  - <sup>9</sup> Board Member of Department of Zrebar Lake Environmental Research, Kurdistan Studies Institute, University of Kurdistan, Sanandaj 66177-15175, Iran
  - <sup>10</sup> Department of Earth Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; john\_clague@sfu.ca
  - <sup>11</sup> Research Institute of Forests and Rangelands, Agricultural Research, Education, and Extension Organization (AREEO), Tehran P.O. Box 64414-356, Iran; jaafari@rifra.ac.ir
  - <sup>12</sup> Data Mining Laboratory, Department of Engineering, College of Farabi, University of Tehran, Tehran 37181-17469, Iran; Fatemehmansoorypoor@gmail.com
  - <sup>13</sup> Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
  - <sup>14</sup> Department of Geoinformation, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia; baharinahmad@utm.my
  - <sup>15</sup> Geoscience Platform Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124 Gwahak-ro, Yuseong-gu, Daejeon 34132, Korea
  - <sup>16</sup> Department of Geophysical Exploration, Korea University of Science and Technology, 217 Gajeong-ro, Yuseong-gu, Daejeon 34113, Korea
- \* Correspondence: phamthaibinh2@duytan.edu.vn (B.T.P.); leesaro@kigam.re.kr (S.L.); Tel.: +82-42-8683-057 (S.L.)

Received: 25 January 2020; Accepted: 12 March 2020; Published: 17 March 2020



**Abstract:** Gully erosion destroys agricultural and domestic grazing land in many countries, especially those with arid and semi-arid climates and easily eroded rocks and soils. It also generates large amounts of sediment that can adversely impact downstream river channels. The main objective of this research is to accurately detect and predict areas prone to gully erosion. In this paper, we couple hybrid models of a commonly used base classifier (reduced pruning error tree, REPtree) with AdaBoost (AB), bagging (Bag), and random subspace (RS) algorithms to create gully erosion susceptibility maps for a sub-basin of the Shoor River watershed in northwestern Iran. We compare

the performance of these models in terms of their ability to predict gully erosion and discuss their potential use in other arid and semi-arid areas. Our database comprises 242 gully erosion locations, which we randomly divided into training and testing sets with a ratio of 70/30. Based on expert knowledge and analysis of aerial photographs and satellite images, we selected 12 conditioning factors for gully erosion. We used multi-collinearity statistical techniques in the modeling process, and checked model performance using statistical indexes including precision, recall, F-measure, Matthew correlation coefficient (MCC), receiver operating characteristic curve (ROC), precision–recall graph (PRC), Kappa, root mean square error (RMSE), relative absolute error (PRSE), mean absolute error (MAE), and relative absolute error (RAE). Results show that rainfall, elevation, and river density are the most important factors for gully erosion susceptibility mapping in the study area. All three hybrid models that we tested significantly enhanced and improved the predictive power of REPTree (AUC=0.800), but the RS-REPTree (AUC= 0.860) ensemble model outperformed the Bag-REPTree (AUC= 0.841) and the AB-REPTree (AUC= 0.805) models. We suggest that decision makers, planners, and environmental engineers employ the RS-REPTree hybrid model to better manage gully erosion-prone areas in Iran.

**Keywords:** gully erosion; watershed management; machine learning; hybrid models; GIS; Iran

---

## 1. Introduction

A global problem that seriously threatens soil and water resources is soil erosion [1–3]. Gully erosion affects soil productivity, can trigger debris landslides and debris flows [4,5], and—if sufficiently severe—can cause an undesirable buildup of sediment in waterways, reservoirs, and ponds [6,7]. Gullies are deep erosional channels on slopes and are commonly a product of ephemeral runoff during periods of heavy rainfall. They provide pathways for water and sediment transport from the upper to lower parts of watersheds. In some catchments, as much as one-third to one-half of the total sediment output is a product of gully erosion [8,9], and gully erosion constitutes 10 to 94 percent of erosion at the watershed scale [9,10]. Gully networks also lower the water table in eroded areas, reducing soil moisture and potentially lowering crop yields on the damaged terrain.

Identifying areas that are susceptible to gully erosion can help land-use managers and planners maintain soil and water resources [10]. Dealing with gullies after they begin to form is difficult and expensive, thus it is better to plan and implement preventative and protective schemes before erosion begins [11].

Past attempts to identify slopes susceptible to gully erosion have focused on topographic thresholds. However, models that use only topographic thresholds typically fail to identify locations sensitive to gully erosion [12,13]. They de-emphasize or ignore land-use, hydrological, climatic, and other environmental factors that have key roles in gully erosion, and do not consider the rapid growth of gully systems once they have initiated [14–17].

Scientists have used a variety of computational data mining methods and models in natural hazard research, including studies of floods [18–28], wildfire [29], sinkholes [30], droughtiness [31,32], earthquakes [33,34], land/ground subsidence [35,36], groundwater [21,37–44], and landslides [22,45–72]. These methods extract related patterns in historical data to predict future events [73]. Data mining methods used to predict gully erosion include logistic regression (LR) [2,30,74–77], artificial neural network (ANN) [20,48,78–80], random subspace (RS) [48,62,81], maximum entropy (ME) [82], artificial neural fuzzy system (ANFIS) [56,83–86], support vector machine (SVM) [18,59,73], fuzzy analytical network (FAN) [37], multi-criteria decision analysis (MCDA) [87,88], evidential belief function (EBF) [88,89], classification and regression tree (CART) [90,91], random forest (RF) [39,52,92–94], rotation forest (RoF) [95], weights of evidence (WofE) [96], frequency ratio (FR) [28,97], BFTree for gully headcut [81], boosted regression [24], ADTree, RF-ADTree [73,76,98], and naive Bayes tree (NBTree) [67].

Accurate gully erosion susceptibility maps are required to predict, control, and mitigate gully formation. This need has led researchers to apply and test a wide variety of data mining methods in gully-prone areas. This study uses three hybrid models—Ada-REPTree, Bag-REPTree, and RS-REPTree—to prepare gully erosion hazard zoning maps for the Rabat Turk watershed in northwestern Iran and to compare the results with those obtained using other models. The study area has an arid to semi-arid climate, a limited vegetation cover, and easily eroded bedrock, all of which make it susceptible to gully erosion.

## 2. Materials and Methods

### 2.1. Study Area

The Rabat Turk watershed is located between Markazi and Isfahan provinces in northwestern Iran (Figure 1). It is one of the catchments of the Shoor River watershed and has an area of about 242 km<sup>2</sup>. The lowest elevation in the watershed is 1807 m above sea level (a.s.l.); its maximum elevation is 2723 m a.s.l. The climate is arid and semi-arid, with average annual rainfall of 213 mm. Precipitation is seasonal, with about 80% of the annual rainfall falling between December and early April [93]. Most of the catchment is bare land, although some areas support agriculture and domestic animals. Gullies are concentrated in the northern part of the watershed, and most are active [93] (Figure 2).

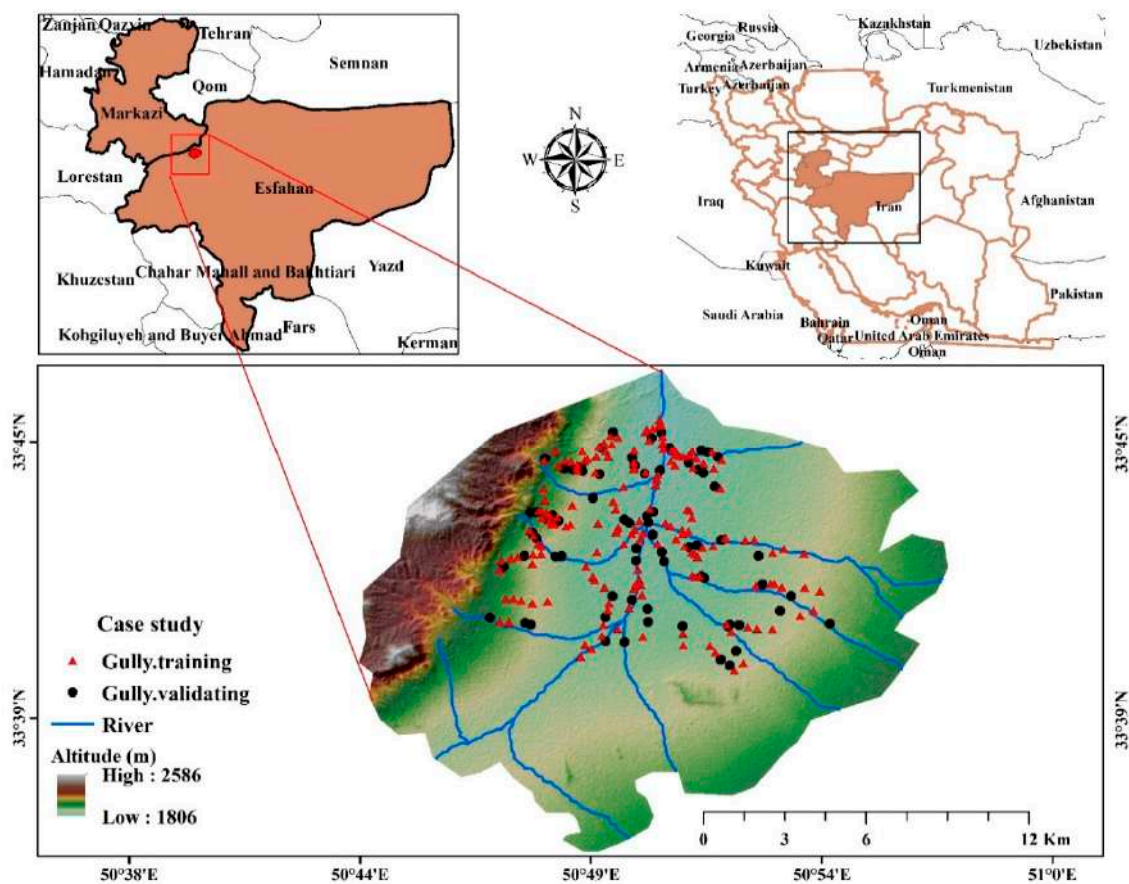


Figure 1. Location of the study area and model training and validating gullies.



Figure 2. Examples of gully erosion in the study area.

### 2.2. Methodology

A flowchart for the methodology used in this study is shown in Figure 3. The methodology involves the following steps: (1) preparing a gully erosion inventory map; (2) determining the appropriate gully erosion conditioning factors (factor ranking and selection); (3) modeling gully erosion susceptibility using REPTree and its ensembles—AdaBoost, bagging, and random subspace algorithms; (4) assessing the goodness-of-fit and prediction accuracy of the models, (5) generating flood susceptibility maps using a base classifier and its ensembles, and (6) assessing the goodness-of-fit and prediction accuracy of the maps.

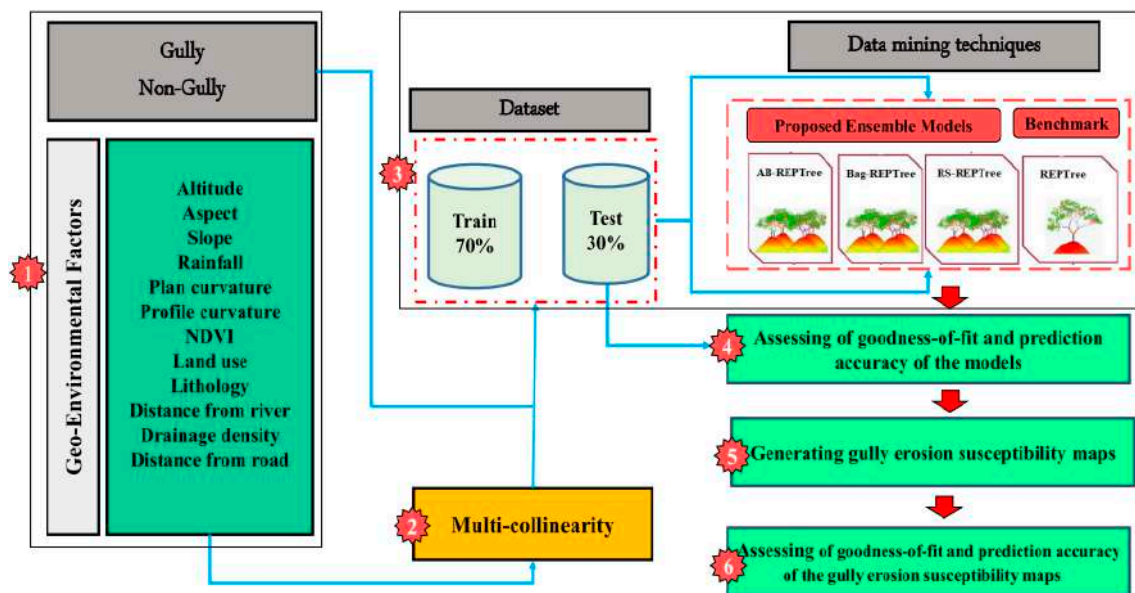


Figure 3. Flowchart of the study.

#### 2.2.1. Gully Inventory Map

Accurately predicting and modeling gully erosion susceptibility requires a high-quality gully erosion map is essential, which thus must be carefully prepared. We obtained an inventory map with 242 gully locations from the Administration of Natural Resources of Markazi Province. The gullies were mapped from aerial photographs and satellite images and were confirmed in the field. Typically, gullies in the study area have concave and vertical heads, indicating that they are active. Longitudinal

profiles are typically straight to convex, but gully widths differ greatly. Gullies on agricultural land commonly have V-shaped cross-sections, whereas those on rangeland more commonly are U-shaped.

Depending on map scale, a gully may be considered a point or a polygon. Most authors who have studied gully erosion consider the heads of gullies to be gully locations [76,99,100], because gully heads are the sources of much of the sediment carried by the gully channels and delivered to the fluvial system below [101,102]. However, some researchers have used grid cells to create gully polygons to prepare gully erosion susceptibility maps [92,103,104], whereas others have converted gully polygons to points using ‘feature to point’ tool in ArcGIS software [105]. However, an active gully is a dynamic landform, and its head moves landward over time as erosion proceeds. A gully consists of three parts: its head, the main channel, and its end point. For long gullies, we used these three points to define their locations. For short gullies, we considered only the head location point. For this study, we randomly selected 242 non-gully locations in the study area. We randomly chose 70% (169) of the mapped gullies to construct the model for gully erosion; the remaining 30% (73) were used to evaluate the predictive performance of model (Figure 3).

### 2.2.2. Gully Conditioning Factors

Gully erosion is a complex process that results from the interplay of numerous factors [106,107]. After reviewing gully erosion literature and considering local conditions and available data, we selected 12 topographic, hydrological, geological, and anthropogenic factors for inclusion in the modeling process.

The topographic factors chosen for this study are elevation, aspect, slope gradient, plan curvature, and profile curvature. The hydrological parameters are distance to rivers and drainage density. We extracted topographic and hydrological factors from a digital elevation model (DEM) obtained from ALOS PALSAR (Phased Array Type L-band Synthetic Aperture Radar) data, with a cell size  $12.5 \times 12.5$  m (<http://www.eorc.jaxa.jp/ALOS/en/aw3d30>) and prepared in ArcGIS 10.3 [93].

The elevation map has four classes (1800–2000, 2000–2200, 2200–2400, and >2400 m a.s.l) (Figure 4a). The highest gully frequency ratio (FR) is associated with the 1800–2000 m class (FR ratio = 1.16). The gully aspect map (Figure 4b) has nine classes, and the highest FR values are in the east, northeast, and southeast aspect classes, with values of, respectively, 1.30, 1.17, and 1.13. The slope gradient map has five classes: 0–5%, 5–10%, 10–20%, 20–30%, and >30% (Figure 4c). The 5–10% class has the highest FR value (1.23). Plan curvature was categorized as convex, flat, and concave forms (Figure 4d). Most gully erosion in the study area occurs in areas mapped as flat (FR = 1.09). There are three classes of profile curvature (< -0.35, -0.35–0.25, and >0.25) (Figure 4e). The <-0.35 class has the highest FR value (1.18).

Hydrological factors (distance from river, drainage density, and rainfall) were extracted from the stream network in the DEM using the Arc Hydro, Euclidean Distance, and Line Density in Spatial Analysis tools in ArcGIS 10.3 [108]. Distance-from-river classes are 0–500, 500–1000, 1000–2000, 2000–3000, and >3000 m (Figure 4f). Gully erosion in the study area is greatest near rivers, and thus the 0–500 m class has the highest FR (1.63). The drainage density map has five classes: 0–0.24, 0.24–0.64, 0.64–1.06, 1.06–1.62, and 1.62–2.46 km/km<sup>2</sup> (Figure 4g). Gully erosion and drainage density are positively correlated; therefore the 1.62–2.46 class has the highest FR value (4.32) and the 0–0.24 class has the lowest FR value (0.52). Annual rainfall data for the study area were obtained for the period 1984–2014 from Robat Turk watershed weather stations operated by the Iran Meteorological Organization. Based on previous related research [76], gully erosion and rainfall are inversely correlated. The rainfall data were interpolated using the inverse distance weighting (IDW) interpolation tool in ArcGIS 10.3 and placed into three classes: 148–159, 159–171, and 171–192 mm (Figure 4l). The largest and smallest number of gullies in the study area are in, respectively, the 148–159 mm (FR = 2.15) and 171–192 mm (FR = 0 classes).

Bedrock lithology is an important factor in gullying [8], and eight types were extracted from a 1:100,000-scale geological map using ArcGIS 10.3 (Figure 4h). The highest and lowest FR values belong to, respectively, the gypsum (Ekgy) class (4.43) and regional metamorphic rocks (pCmt2) (0.08).

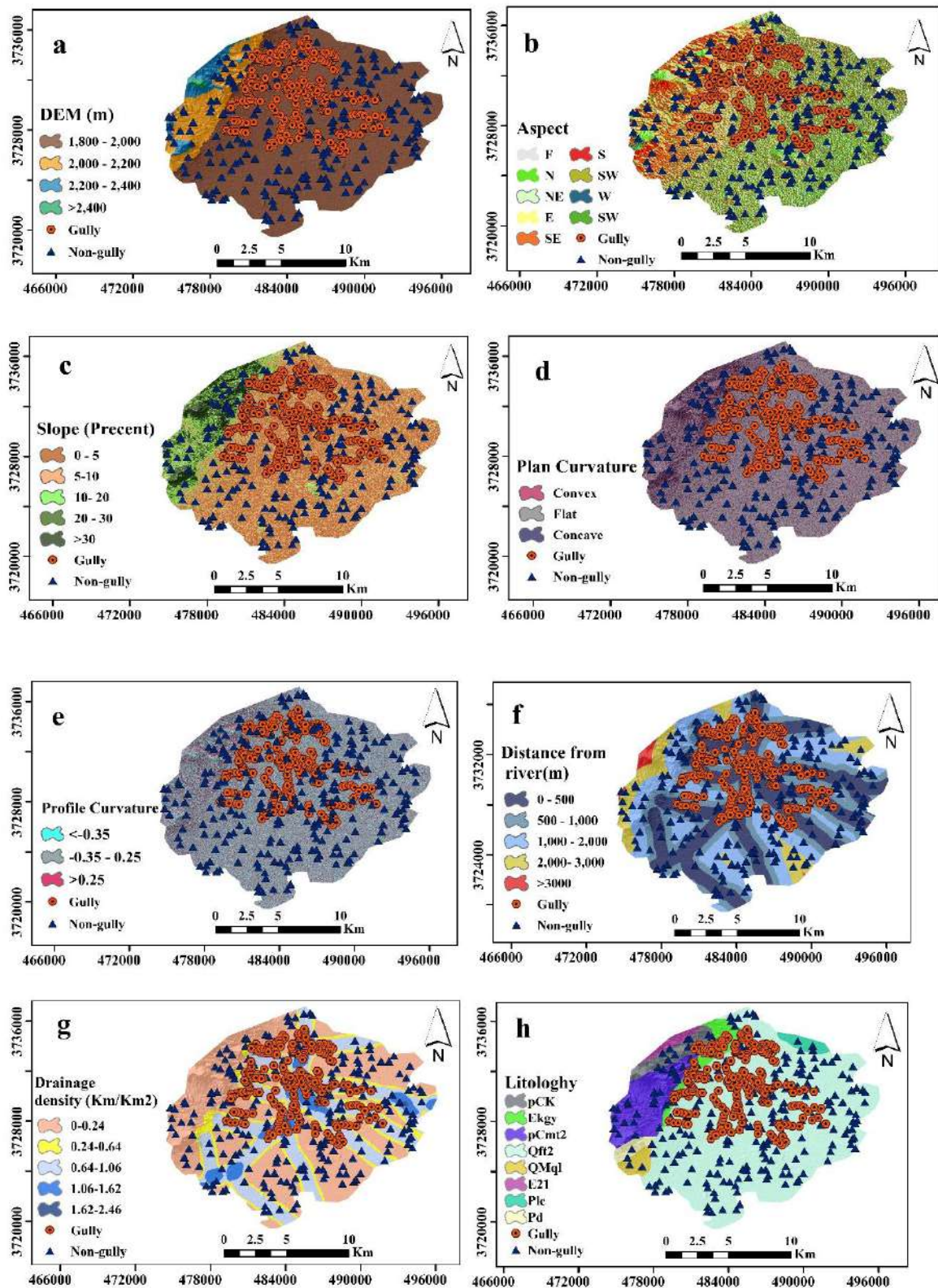
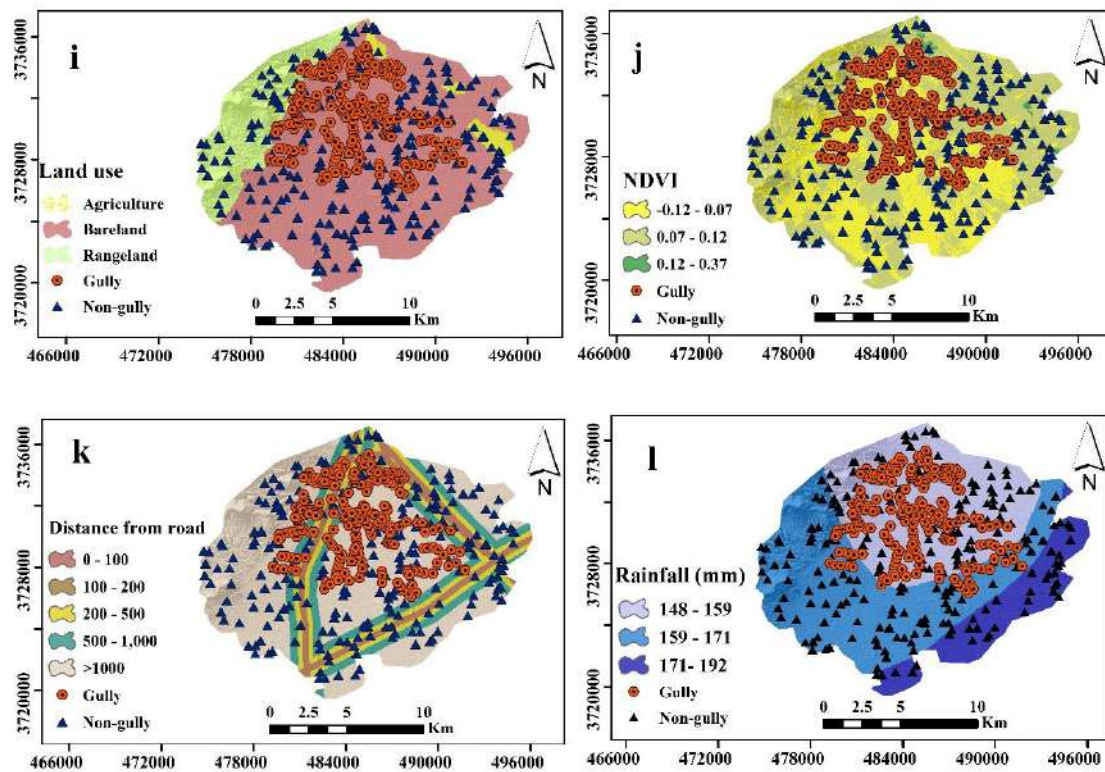


Figure 4. Cont.



**Figure 4.** Spatial database for gully susceptibility analysis. (a) Elevation, (b) aspect, (c) slope, (d) plan curvature, (e) profile curvature, (f) distance from river, (g) drainage density, (h) lithology, (i) land use, (j) NDVI, (k) distance from road, (l) rainfall.

Changing land use, for example deforestation and grazing, is an important cause of soil erosion [76]. For the current study, land use was inferred from Landsat 8 (OLI) satellite imagery and analyzed and processed with the ENVI 5.4 software. The land-use map includes three classes—agriculture, bare land, and rangeland (Figure 4i). Most gullies in the study area are found in the bare land class (FR = 1.21), and lowest number are in the rangeland class (FR = 0.62).

The incidence of gully erosion is greatest in areas with limited vegetation cover. A normalized difference vegetation index (NDVI) map of the study area was generated in ArcGIS 10.3 from Landsat 8 imagery acquired on 15 June 2017. This map is based on the formula  $(\text{NIR}-\text{Red})/(\text{NIR}+\text{Red})$ , where NIR (near-infrared) is band 5 and Red is band 4 of the Landsat 8 imagery. The map includes three NDVI classes:  $-0.12-0.07$ ,  $0.07-0.12$ , and  $0.12-0.37$  (Figure 4j). The  $0.12-0.37$  map class has the largest number of gullies.

Roads also affect gully erosion, as they intercept and concentrate overland flow [109,110]. This factor is represented by distances of gully and non-gully sites from roads, which were determined by vectorizing topographic maps and then transforming the data to a raster map using the Euclidean Distance tool in Arc GIS 10.3. Five classes were defined: 0–100, 100–200, 200–500, 500–1000, and >1000 m (Figure 4k). The largest and smallest number of gullies in the study area are, respectively, in the 200–500 m (FR = 1.10) and 0–100 m (FR = 0.73) classes.

### 2.2.3. Gully Erosion Susceptibility Modeling

In this study, we prepared gully susceptibility maps using REPTree as a base classifier and AdaBoost, bagging, and random subspace in ensemble models. The following subsections briefly describe the four ensemble models.

### AdaBoost (AB)

AdaBoost (adaptive boosting) was the first boosting algorithm used for binary classification [111] and is a starting point for understanding the concept of boosting. AdaBoost free users from the complexities involved in detecting and choosing parameters.

The steps of the AdaBoost algorithm can be summarized as follows:

First, each data point is calculated as

$$w(x_i, y_i) = \frac{1}{n}, i = 1, \dots, n \tag{1}$$

The obtained weights are updated after each step.

Second, a basic classifier  $C_b(X_i)$  is built from a training set and is applied to each training sample. The error of this classifier  $\epsilon_b$  is calculated as

$$\epsilon_b = \sum_{i=1}^n w_b(i) \xi_b(i) \text{ where } \xi_b(i) = \begin{cases} 0 & C_b(x_i) = y_i \\ 1 & C_b(x_i) \neq y_i \end{cases} \tag{2}$$

The new weight for each iteration is

$$w_{b+1}(i) = w_b(i) \cdot \exp(\alpha_b \xi_b(i)) \tag{3}$$

where  $\alpha_b$  is a constant that is calculated from the error of the classifier in each iteration

$$\alpha_b = \ln((1 - \epsilon_b) / \epsilon_b) \tag{4}$$

The calculated weights in each iteration are generally normalized, and their sum is one.

This process is repeated in every step for  $b = 1, 2, 3, \dots, B$ , and then the ensemble classifier is built as a linear combination of the single classifiers weighted by the corresponding constant  $\alpha_b$ :

$$C(x) = \text{sign} \left( \sum_{b=1}^B \alpha_b C_b(x) \right) \tag{5}$$

### Bagging (Bag)

Bagging is an ensemble learning method introduced by Breiman [112]. It creates parallel diverse classifiers that are then coupled. Specifically, each bootstrap sample dataset is generated by randomly drawing, with replacement,  $N$  instances ( $N$  is the size of the original training datasets). Then, a classifier  $C_i$  is built from each bootstrap sample  $B_i$ , and  $C^*$  is built from  $C_1, C_2, \dots, C_T$ . Bagging output is the class that is most often predicted by its sub-classifiers.

This algorithm can be summarized as follows:

Input: training set  $S$ , inducer  $T$ , integer  $T$  (number of bootstrap samples)

- (1) for  $i = 1$  to  $T$  {
- (2)  $S^i =$  bootstrap sample from  $S$  (sample with replacement)
- (3)  $C_i = T(S^i)$
- (4) }
- (5)  $C^*(x) = \underset{y \in Y}{\operatorname{argmax}} \sum_{i=C_i(x)=y} 1$
- (6) Output: classifier  $C^*$



### Random Subspace

The random subspace (RS) method [113] is an ensemble classifier technique in which each training sample is defined as a p-dimensional vector  $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  and  $r < p$  features are randomly selected from the p-dimensional dataset X in each iteration. Classifiers then are built into the random subspaces and aggregated through majority voting.

The RS algorithm can be summarized as follows:

- (1) Repeat for  $b = 1, 2, \dots, B$ :
  - (a) Select a r-dimensional random subspace  $\tilde{X}^b$  from the original p-dimensional feature space.
  - (b) Construct a classifier  $C^b(x)$  with a decision boundary  $C^b(x) = 0$  in  $\tilde{X}^b$ .
- (2) Combine classifiers  $C^b(x)$ ,  $b = 1, 2, \dots, B$  by simple majority voting to obtain a final decision rule:

$$\beta(x) = \operatorname{arimax}_{y \in [-1,1]} \sum_b \delta_{sgn}(C^b(x)) \cdot y \tag{6}$$

where  $\delta_{ij}$  is the Kronecker symbol and  $y \in [-1, 1]$  is a decision (class label) of the classifier.

### Reduced-Error Pruning Tree (REPTree)

Quinlan [114] introduced a method based on information gain or variance to build a decision tree that uses reduce-error pruning with back overfitting. The REPTree algorithm sorts values for numerical attributes once; missing values are created using an embedded method by C4.5 in fractional instances.

#### 2.2.4. Comparison and Validation of Gully Erosion Models and Susceptibility Maps

In this section, we introduce the evaluation metrics used in this study. We selected the most widely used metrics based on the machine learning literature, which include machine learning performance evaluation metrics and error metrics.

#### Machine Learning Evaluation Metrics

Machine learning evaluation metrics include true positive (TP), false positive (FP), precision, recall, F-measure, Matthews correlation coefficient (MCC), receiver operatic characteristic curve (ROC), and the precision recall (PRC) metric. All these metrics are obtained based on the four possibilities shown in Table 1: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). TP and TN are the number of gully erosion pixels that correctly classified as, respectively, gully erosion and non-gully erosion pixels. In contrast, FP and FN pixels are incorrectly classified as gully erosion and non-gully erosion pixels, respectively [76]. The above-monitored metrics can be formulated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 - \text{measure} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{9}$$

We used the Matthews correlation coefficient (MCC) [114] to check the quality of binary (two-class) classifications. This metric has a range from -1 (total disagreement between prediction and observation values) and +1 (perfect prediction). The MCC can be computed as

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{((TP + FP)(TP + FN)(TN + FP) - (TN + FN))}} \tag{10}$$

The receiver operating characteristic curve (ROC) is a popular and important metric to check the general performance of a model [115]. Recall and 1-specificity (FP / (FP + TN)) are plotted, respectively, on the *x* and *y*-axes of the ROC. A model with random performance has a straight diagonal line from (0, 0) to (1, 1) on the plot, which thus serves as a reference line. The area under the ROC curve (AUC) is a quantitative measure of the performance of the model. It ranges from 0 (inaccurate model) to 1 (perfect model) [21,116]. The PRC metric is a graph that provides a prediction of future classification performance [117]. The *x*- and *y*-axes are, respectively, recall and precision metrics. The higher the PRC line value, the better the performance of the model.

**Table 1.** Confusion matrix of machine learning models in this study.

		Predicted Target	
		Gully Erosion (+)	Non-Gully Erosion (-)
Actual target	Gully erosion (+)	TP	FP
	Non-gully erosion (-)	FN	TN

### Error-Based Evaluation Metrics

Error-based indexes are the second group of evaluation metrics used to check the performance of the gully erosion mapping. They include Kappa (K), root mean square error (RMSE), relative standard error of the prediction (PRSE), mean absolute error (MAE), and relative absolute error (RAE), which are formulated as

$$\text{Kappaindex (K)} = \frac{A - B}{1 - B} \tag{11}$$

$$A = (TP + TN) / (TP + TN + FN + FP) \tag{12}$$

$$B = ((TP + TN)(TP + FP) + (FP + TN)(FN + TN)) / \sqrt{(TP + TN + FN + FP)} \tag{13}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (p_i - a_i)^2}{n}} \tag{14}$$

$$\text{PRSE} = \frac{\sum_{i=1}^n (p_i - a_i)^2}{\sum_{i=1}^n (\bar{a} - a_i)^2} \tag{15}$$

$$\text{MAE} = \frac{\sum_{i=1}^n |p_i - a_i|}{n} \tag{16}$$

$$\text{RAE} = \frac{\sum_{i=1}^n |p_i - a_i|}{\sum_{i=1}^n |\bar{a} - a_i|} \tag{17}$$

### 2.2.5. Factor Ranking and Selection by the Information Gain Ratio Technique

Several techniques for factor ranking and selection have been proposed, but the relative advantages and weaknesses of these techniques are unknown [118]. Factor ranking techniques evaluate the relevance of each factor independently and eliminate factors determined to be irrelevant or redundant. They also search for the subset of factors that offers the largest reduction in dimensionality [118].

In this study, we used the information gain ratio (IGR) method to select and rank the most important factors for gully erosion modeling and susceptibility mapping. The IGR method is applied as follows [119]:

Let  $T$  be the total number of tuples in the training dataset;  $T_j$  as the total number of positive or negative tuples in the training dataset;  $v$  is the total number of classes in the dataset; and  $S$  is slope angle, which is one of the gully conditioning factors.

$$GainRatio(Slope) = \frac{Gain(Slope)}{SplitInfo(Slope)} \tag{18}$$

$$\text{where; } SplitInfo(T) = -\sum_{j=1}^v \frac{|T_j|}{|T|} \log_2\left(\frac{|T_j|}{|T|}\right) \tag{19}$$

$$Gain(Slope) = I(p, n) - E(Slope) \tag{20}$$

$$E(Slope) = -\sum_{i=1}^m \frac{p_i + n_i}{p + n} I(p_i, n) \tag{21}$$

$$I(p_i, n) = -\frac{p}{p + n} \log_2 \frac{p}{p + n} - \frac{n}{p + n} \log_2 \frac{n}{p + n} \tag{22}$$

$E(Slope)$  represents the entropy of the slope angle factor in the training dataset,  $I(p, n)$  denotes the information needed to satisfy a given training dataset,  $p$  is the total number of positive tuples in the training dataset,  $n$  is the total number of negative tuples in the training dataset, and  $m$  is the number of values for the slope angle factor.

### 3. Results

#### 3.1. Correlation between Conditioning Factors and Gully Occurrence Using the Frequency Ratio Method

We used the frequency ratio method to calculate the probabilistic relation between gullies as a dependent variable and conditioning factors as independent variables. Figure 5 presents FR values for the classes of each conditioning factor. In the case of rainfall, the 148–159 mm class has the highest FR value (2.15), followed by the 159–171 mm (0.29) and 171–192 mm (0) classes. The >1000 m distance-from-road class had the highest FR value (1.63), followed by the 200–500 m (1.10), 500–1000 m (0.96), 100–200 m (0.77), and 0–100 m (0.73) classes. In the case of NDVI, the 0.12–0.37 class has the highest FR value (3.48). Bare land areas have the highest FR values in the land-use class (1.21). In the case of drainage density, high FR values are associated high drainage density. The 1.62–2.46 km/km<sup>2</sup> class, for example has a value of 4.32. For lithology, the Ekgy has by far the largest FR (4.43), followed by Qft2 (1.02), PCK (0.34), and PCmt2 (0.08). No gullies are present on the other lithologies; therefore, their values are 0. Areas located <500 m from rivers have a FR value of 1.63; the more distant classes have 0 values. In the case of profile curvature, the highest FR value (1.18) is associated with the >0.25 class. Flat areas have a FR value of 1.09, which is higher than the values for convex and concave areas (0.72 and 0.53, respectively). The highest FR value for the slope factor is 1.23 (5–10% class). Values for the 0–5% and 10–20% classes are, respectively, 1.07 and 0.76; the 20–30% and >30% classes are 0. Slopes with an eastern aspect have the highest FR value (1.30), following by slopes with northeastern (1.17), southeastern (1.13), northern and flat (1.05), southwestern (0.94), northwestern (0.88), southern (0.70), and western (0.68) aspects. Finally, all gullies are located in areas with an elevation range of 1800–2000 m a.s.l. (FR = 1.16).

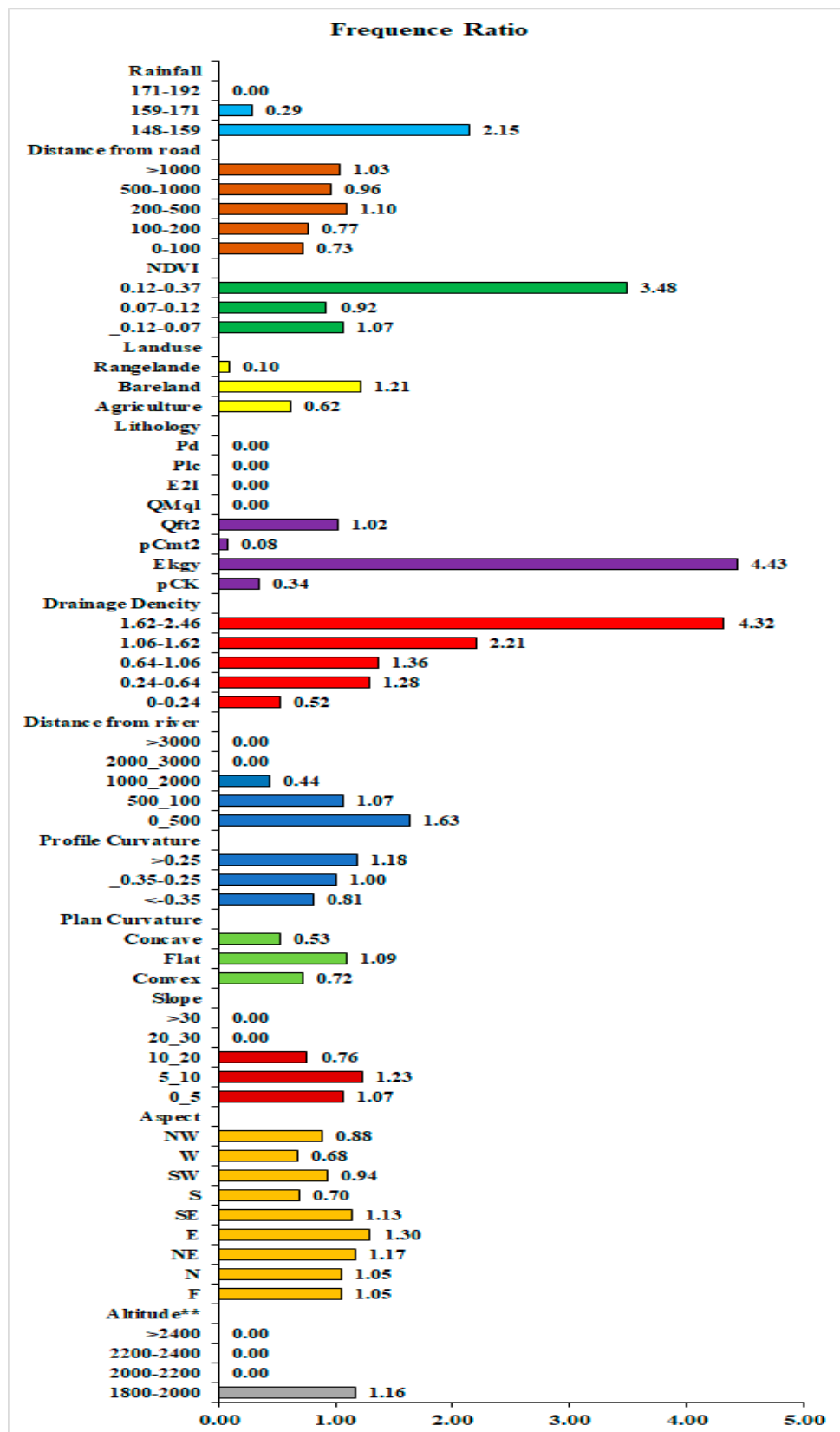


Figure 5. Frequency ratios for factors related to gully erosion.

### 3.2. Analysis of Factor Multi-Collinearity

We examined the multi-collinearity of gully erosion conditioning factors using the variance inflation factor (VIF) and tolerances (TOL). Values of VIF >10 and TOL <0.10 generally indicate a multi-collinearity problem [120]. VIF and TOL values for the conditioning factors used in this study are shown in Table 2. The highest VIF and the lowest TOL are, respectively, 2.673 and 0.184, which indicate that there is not a multi-collinearity problem among the conditioning factors and, hence, all factors can be used for gully erosion susceptibility mapping.

**Table 2.** Multi-collinearity statistics for the gully erosion affecting factors.

Parameters	Collinearity Statistics	
	Tolerance	VIF
Land use	0.184	1.525
Lithology	0.674	1.354
NDVI	0.628	2.047
Plan curvature	0.492	1.254
Profile curvature	0.398	2.673
Rainfall	0.712	1.951
River density	0.420	2.322
River distance	0.324	1.875
Road	0.583	1.840
Slope	0.809	1.245
Aspect	0.856	1.030
Altitude	0.198	2.329

### 3.3. The Most Important Factors for Gully Modeling

The average merit (AM) values calculated by information gain ratio (IGR) technique are summarized in Table 3. The results indicate that all factors can be included in gully erosion susceptibility modeling because their AM values are greater than zero. However, rainfall, with an AM value of 0.225, is the most effective factor for gully erosion susceptibility mapping in the study area. It is followed by elevation (AM = 0.186), river density (AM = 0.106), distance to river (AM = 0.093), land use (AM = 0.086), lithology (AM = 0.083), distance to road (AM = 0.038), profile curvature (AM = 0.031), aspect (AM = 0.028), NDVI (AM = 0.023), slope (AM = 0.020), and plan curvature (AM = 0.016).

**Table 3.** The most effective factors for gully erosion occurrence.

Rank	Conditioning Factor	Average Merit	Standard Deviation
1	Rainfall	0.225	± 0.012
2	Altitude	0.186	± 0.009
3	River density	0.106	± 0.011
4	River distance	0.093	± 0.015
5	Land use	0.086	± 0.007
6	Lithology	0.083	± 0.01
7	Profile curvature	0.031	± 0.017
8	Road	0.038	± 0.014
9	Aspect	0.028	± 0.021
10	NDVI	0.023	± 0.018
11	Slope	0.02	± 0.016
12	Plan curvature	0.016	± 0.018

### 3.4. Evaluation of Gully Erosion Susceptibility Models

We created four landslide susceptibility models (REPTree, AB-REPTree, Bag-REPTree, and RS-REPTree) using the training dataset. The 10-fold cross-validation method was used to prevent over-fitting and to decrease variability. Heuristic tests were used to find the best values for the parameters of the four models; these are shown in Table 4.

**Table 4.** Parameters of algorithms utilized in this study.

Methods	Algorithms	Parameters
Base classifier	Reduced-error pruning tree	Seed, 1; The minimum total weight of the instances in a leaf, 2; Number of folds, 10
Ensembles	Bagging	Seed, 1; The number of iterations, 10
	AdaBoost	Seed, 1; The number of iterations, 10
	Random subspace	Seed, 1; The number of iterations, 10

We validated gully erosion susceptibility models using error and machine learning comparison metrics (Tables 5 and 6). The highest values of the Kappa metric were obtained for the RS-REPTree model (0.61), followed by the Bag-REPTree (0.55), AB-REPTree (0.53), and REPTree (0.53) models. The RS-REPTree model has the highest value (0.33) for the MAE metric, followed by the Bag-REPTree (0.28), AB-REPTree (0.24), and REPTree (0.24) models. The RMSE, RAE, and RRSE metrics indicate that the Bag-REPTree model (RMSE = 0.37, RAE = 56.62, and RRSE = 77.75) has the lowest error. It is followed by the RS-REPTree model (RMSE = 0.38, RAE = 67.68, and RRSE = 77.57) and the AB-REPTree model (RMSE = 0.43, RAE = 49.76, and RRSE = 86.49). The REPTree model has the highest error (RMSE = 0.43, RAE = 79.76, and RRSE = 86.50).

**Table 5.** Evaluation of gully erosion susceptibility models using error metrics.

Models	Kappa	MAE	RMSE	RAE	PRSE
REPTree	0.53	0.24	0.43	79.76	86.50
AB-REPTree	0.53	0.24	0.43	49.76	86.49
Bag-REPTree	0.55	0.28	0.37	56.62	75.30
RS-REPTree	0.61	0.33	0.38	67.68	77.57

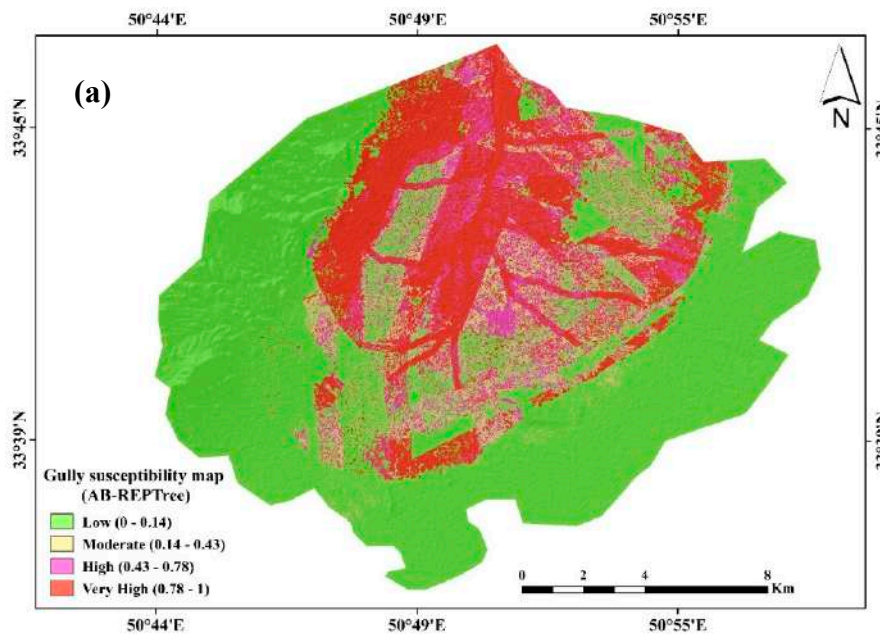
**Table 6.** Evaluation of gully erosion susceptibility models using machine learning metrics.

Models	TP	FP	Precision	Recall	F-Measure	MCC	AUC	PRSE
REPTree	0.774	0.226	0.776	0.774	0.773	0.549	0.819	0.782
AB-REPTree	0.768	0.232	0.77	0.768	0.767	0.537	0.844	0.838
Bag-REPTree	0.776	0.224	0.779	0.776	0.776	0.555	0.871	0.866
RS-REPTree	0.806	0.194	0.809	0.806	0.805	0.615	0.874	0.865

The machine learning comparison metrics shown in Table 6 indicate that the RS-REPTree model performed best based on TP, FP, precision, recall, F-measure, MCC, AUC, and PRSE values. It is followed by the Bag-REPTree, REPTree, and AB-REPTree models in terms of TP, FP, precision, recall, F-measure, and MCC metrics. The AB-REPTree model performed better than the REPTree model in term of the AUC and PRSE metrics.

### 3.5. Development of Gully Erosion Susceptibility Maps

We calculated gully erosion susceptibility indices for each cell based on the results of the ensemble models. We then constructed gully erosion susceptibility maps for the study area using the Ada-REPTree, Bag-REPTree, REPTree, and RS-REPTree models (Figure 6). Gully erosion susceptibility classes (low, moderate, high, and very high) were created using the natural breaks method. For example, in the case of the Ada-REPTree map, the four susceptibility classes have values of 0.00–0.13, 0.13–0.42, 0.42–0.78, and 0.78–1.00 (Figure 6a). Comparison of the four maps indicates that the REPTree model predicts a larger part of the watershed as having high and very high erosion susceptibilities. More generally, the maps show that most cells of low erosion susceptibility are located on steep slopes in the marginal parts of the watershed. The high and very high susceptibility classes cover the northern and central parts of the watershed where most of the observed gully sites are located.



**Figure 6.** Cont.

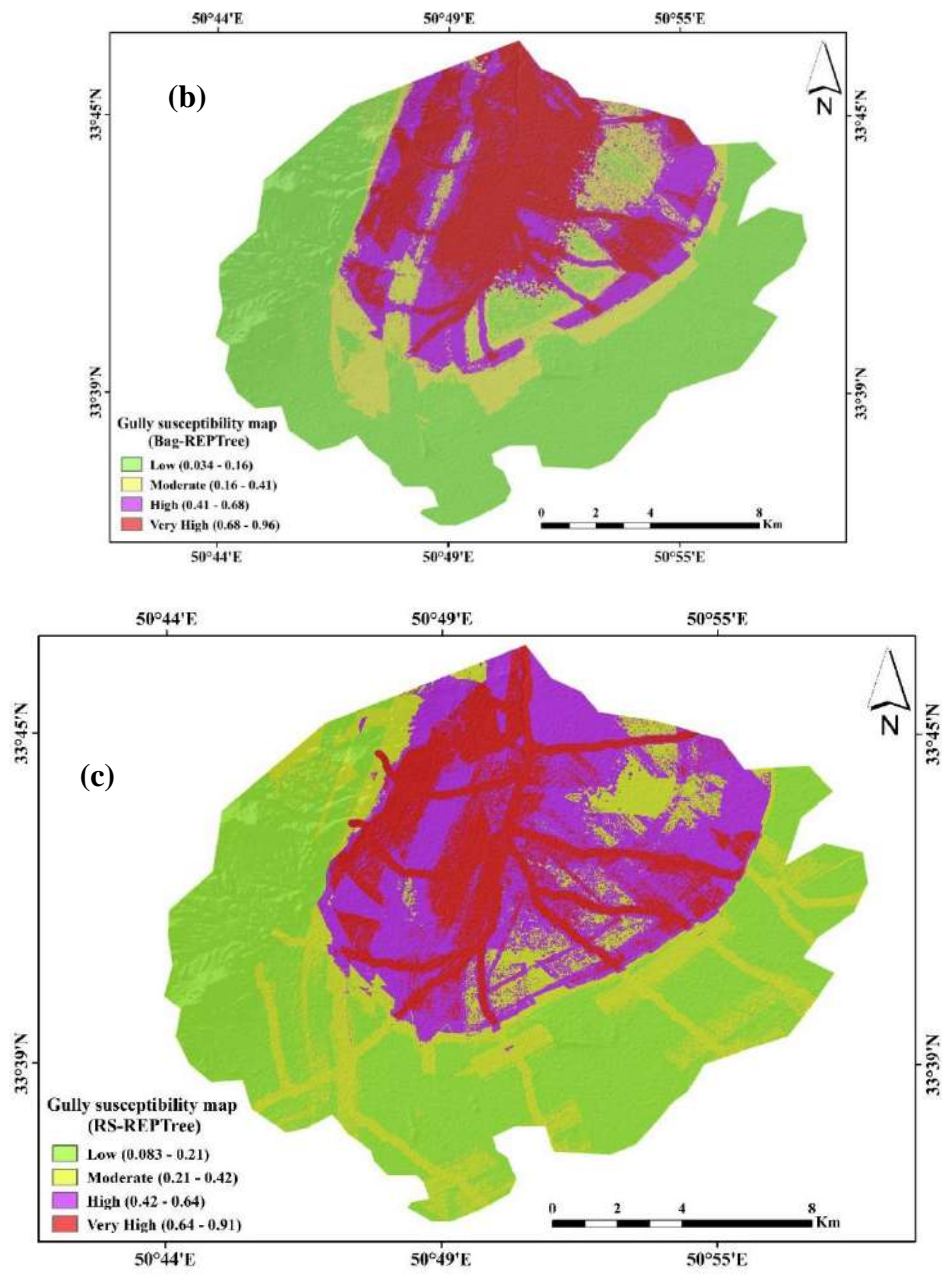
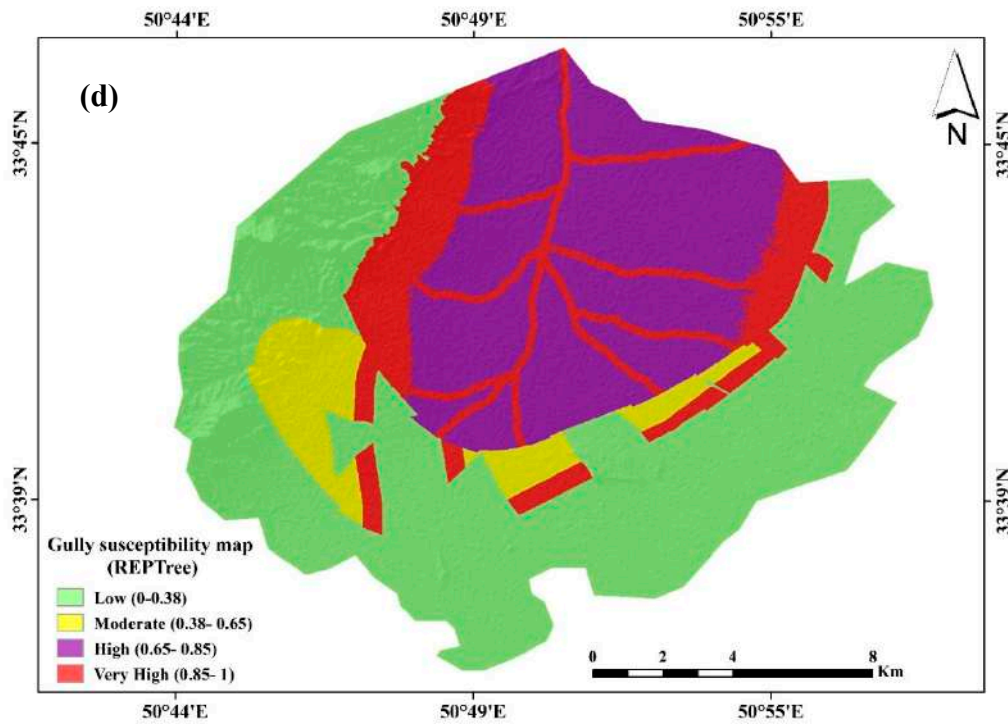


Figure 6. Cont.





**Figure 6.** Gully erosion susceptibility maps based on: (a) AB-REPTree, (b) Bag-REPTree, (c) RS-REPTree, and (d) REPTree.

3.6. Evaluation and Comparison of the Models

Evaluation of model performance is an important step in the spatial modeling process [88]. In this study, we evaluated the performance of the four ensemble models using the area under the ROC curve (AUC), standard error (SE), and 95% confidence interval for the training and testing datasets. The logistic regression (LR) model was used as a benchmark method. ROC curves for the training dataset are shown in Figure 7. The curves show that all tested ensemble models perform well in spatially predicting gully erosion susceptibility. However, the ROC curve for the REPTree model falls below the curves of the other models. Other results of the goodness-of-fit analysis of the training dataset are shown in Table 7. These results indicate that RS-REPTree model has the best performance with the highest AUC (0.874), lowest SE value (0.0191), and narrowest 95% CI (0.834–0.907). Sequentially, the Bag-REPTree, AB-REPTree, and REPTree models have slightly lower performances. Finally, the performances of three ensemble models are better than that of the benchmark LR model.

**Table 7.** ROC curve using the training dataset.

Variable	AUC	SE	95% CI	
			Lower Bound	Upper Bound
<b>REPTree</b>	<b>0.819</b>	0.0238	0.774	0.859
AB-REPTree	0.844	0.0210	0.801	0.881
Bag-REPTree	0.871	0.0191	0.830	0.905
RS-REPTree	0.874	0.0191	0.834	0.907
LR	0.825	0.0222	0.780	0.864

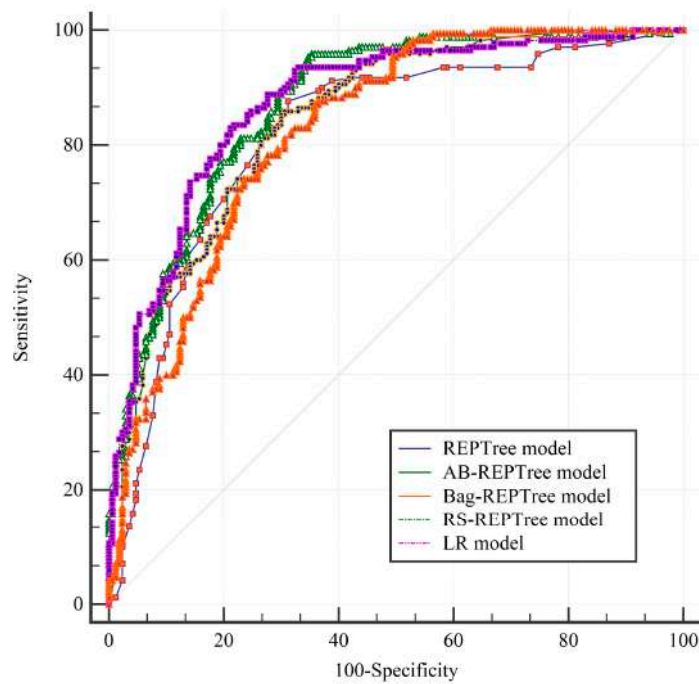


Figure 7. ROC curves related to the susceptibility models used in this study.

Model performances for the testing dataset based on the ROC curve, AUC, SE, and 95% CI values are shown in Figure 8 and summarized in Table 8. All models performed well, but the proposed new ensemble model, RS-REPTree, has the highest prediction capability based on its AUC (0.860), SE (0.0315), and 95% CI (0.793–0.912). It is followed by the Bag-REPTree (AUC = 0.841), LR (AUC = 0.824), AB-REPTree (AUC = 0.805), and REPTree (AUC = 0.800) models. Overall, our results show that the new ensemble models of REPTree outperform and outclass the standard REPTree model in gully erosion susceptibility mapping.

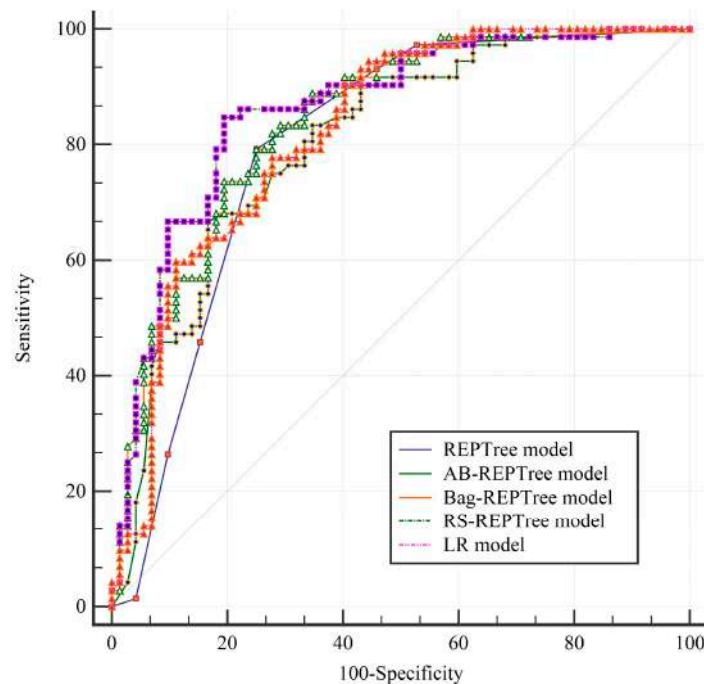


Figure 8. ROC curve and AUC of the models: (a) training dataset, and (b) validation dataset.

**Table 8.** ROC curve using the validation dataset.

Model	AUC	SE	95% CI	
			Lower Bound	Upper Bound
<b>REPTree</b>	<b>0.800</b>	0.0383	0.725	0.862
AB-REPTree	0.805	0.0368	0.731	0.866
Bag-REPTree	0.841	0.0329	0.771	0.896
RS-REPTree	0.860	0.0315	0.793	0.912
LR	0.824	0.0350	0.751	0.882

#### 4. Discussion

Obtaining reliable map of gully erosion susceptibility remains yet a challenge for managers, land use planners, and engineers. To address this challenge, researchers are proposing new models and testing them in different gully-prone regions around the world. In this paper, we propose and evaluate three ensembles of the REPTree model for gully erosion susceptibility mapping. The modeling process is based on an investigation of the relationships between spatial locations of gullies in the Rabat Turk watershed and a suite of different geo-environmental factors. We demonstrate that rainfall, elevation, river density, distance to rivers, land-use, and lithology are important factors for gully erosion in the study area. In contrast, plan curvature, slope, NDVI, aspect, and distance to roads are the less important.

An examination of the literature suggests that conditioning factors for gully erosion are area-specific and cannot be reliably extrapolated to other regions. For example, Amiri et al. [121] identified land-use as the most important factor in their study areas, whereas Rahmati et al. [122] and Garosi et al. [92] reported that distance from rivers is the most important factor in their studies. Furthermore, the slope factor, which we and Rahmati et al. [122] ranked as a relatively unimportant factor, was among the most effective factors identified by Rahmati et al. [97]. These differences call for further research on controls of gully erosion in different landscapes.

The ensemble learning techniques used in this study (AB, bagging, and RS) improved the goodness-of-fit and prediction performance of REPTree. Among these techniques, random subspace outperformed the other two techniques in improving both the training and validation of the base REPTree model. The RS ensemble learning technique performed better than the other techniques in decreasing the variance, bias, and noise of the modeling process, and protected the models from over-fitting. The superiority of the RS ensemble learning technique stems from the use of random subspaces for aggregating the base classifiers, which results in better performance compared to the original feature space [112]. In addition, the base classifier works better using smaller subspaces, as shown by Pham et al. [123]. The literature includes numerous successful applications of RS ensemble learning techniques for predicting different types of natural hazards. For example, Tien Bui et al. [76] showed that the naive Bayes tree performed better when used in combination with the RS technique for landslide modeling, and Shirzadi et al. [62] demonstrated that the RS technique improved the performance of the alternating decision tree base classifier.

Our results suggest that the Bagging technique is the second-best ensemble learning method for improving REPTree performance, which is in line with previous findings. For example, Hong et al. [124] reported that bagging, used in combination with the j48 decision tree, has higher predictive capacity than the single j48 and AB-j48 models alone. In another study, Bui et al. [58] reported that the functional tree (FT) model with bagging outperforming the AB-FT method.

Although our study is the first to use REPTree in combination with ensemble learning techniques for gully erosion modeling, this approach has been used by Pham et al. [123] for predicting landslides. They ranked the ensemble models in terms of prediction capability, from best to worst, to be: BA-REPTree (AUC = 0.872), rotation forest REPTree (AUC = 0.872), RSRETree (AUC = 0.864), and MultiBoost REPTree (AUC = 0.855). The differences in their results and ours suggest that the techniques are case-

and site-specific and that their performances depend heavily on the datasets that are trained and built upon.

Although it is difficult to directly compare the results of this study with those reported from other regions, we suggest that our ensemble models perform better than the generalized linear model (AUC = 0.71), boosted regression tree (AUC = 0.84), multivariate adaptive regression spline (AUC = 0.83), and ANN (AUC = 0.84) models used by Garosi et al. [104]; the certainty factor model (AUC = 0.82) used by Azareh et al. [82]; and the Fisher's linear discriminant analysis (AUC = 0.76), logistic model tree (AUC = 0.77), and NBT (AUC = 0.78) models of Arabameri et al. [125]. In contrast, however, our models were outperformed by the maximum entropy (AUC = 0.88, 0.90) models used by Azareh et al. (2019) and Kariminejad et al. [107]; BFTree and its ensembles (bagging and RS) (AUC = 0.92) used by Hosseinalizadeh et al. [81]; and the multivariate additive regression splines (AUC = 0.91), SVM (AUC = 0.88), and FR (AUC = 0.96) models employed by Gayen et al. [126]. Again, these different results are attributable to local differences in the environments in which the models were used.

Our field survey indicated that gullies in the study area are located along tributaries near the main river in the Rabat Turk study area. Erosion is initiated by focusing of runoff along these tributaries, gradual gully retrogression, and piping above gully heads. Gullies on the east side of the river have lower slopes than those on the west side of the river, perhaps because there is little vegetation in the former areas. There is also more upslope area for gully development on the west side of the river, allowing for more flow with the gully system. Our results are in agreement with the findings of Vandekerckhove et al. [127] and Bergonse and Reis [128], who argued that gullies are mainly formed through extreme runoff related to slope-area relations. The gully erosion susceptibility map of the study area obtained using the RS-REPTree ensemble model accurately predicts observed gullies along the main river and its tributaries.

Despite the improved prediction performance provided by ensemble models, the difficulty associated with proper parameter tuning still restricts their development and application. In this study, we manually tuned the parameters of the ensemble methods through a trial-and-error process [129,130]. There are, however, several optimization techniques (e.g., metaheuristic optimization algorithms) that can significantly speed up the process of model building [131,132]. Nevertheless, ensemble models are easy to develop within open-source WEKA software and do not require advanced programming knowledge. They can be applied to types of environmental research that involve datasets with a number of geo-environmental variables and a set of presence/absence locations of the phenomenon being modeled. Such datasets can be generated with automated GIS techniques from accessible geospatial data (e.g., DEM, soil, lithology, and meteorological records).

## 5. Conclusions

Gully erosion is an advanced stage of water erosion and sediment production that can transfer large volumes of sediment into stream channels, resulting in environmental damage. It is a common problem in arid and semi-arid landscapes, and therefore, prediction and mapping of areas susceptible to gully erosion are of interest to soil scientists, natural resource authorities, and land managers. Accordingly, researchers have used a variety of machine learning methods to understand the causes of gully erosion and to produce reliable erosion susceptibility maps [133].

We addressed this problem by studying gully erosion in a sub-basin of the Shoor River watershed in Isfahan Province (Iran), which has a semi-arid climate and a human-impacted landscape. We used 12 conditioning factors tested by the information gain ratio method, and REPTree coupled with the AB, BA, and RS ensemble learning methods to model gully erosion and produce gully erosion susceptibility maps. The following are key conclusions of our study:

(1) Rainfall, elevation, and river density are the most important factors for gully erosion in the study area. Most gully erosion sites are located in areas of lower rainfall and lower elevation.

(2) REPTree and all its ensembles yielded a high goodness-of-fit and prediction accuracy during the modeling process, but the ensemble RS-REPTree performed best. RS decreased over-fitting and noise in the training datasets, which resulted in better prediction. It successfully predicted gully erosion locations and allowed us to produce an accurate gully erosion susceptibility map of the study area.

(3) Modeling gully erosion is a complicated task, with many uncertainties. The proposed machine learning model is an easy-to-use, inexpensive decision-making tool that can supplement expensive field surveys. It also provides managers with guidance on what further information might be needed to provide a more accurate map of gully erosion.

(4) Gully erosion susceptibility maps are essential products for hazard analysis and management. We recommend our proposed ensemble RS-REPTree model for predicting gully erosion in other semi-arid and arid areas. However, the performance of this model depends on the quality of the data used.

(5) We recommend further research on other hybrid data mining methods, as well as ensemble boosting algorithms with REPTree. We also recommend further sensitivity analysis of gully erosion conditioning factors.

**Author Contributions:** V.-H.N., S.J., M.A., W.C., M.F., E.O., A.S., H.S., J.J.C., A.J., F.M., B.T.P., B.B.A., and S.L. contributed equally to the work. S.J., M.A., M.F., and F.M. collected field data and conducted the gully erosion mapping and analysis. S.J., M.A., W.C., M.F., E.O., A.S., H.S., A.J., and F.M. wrote the manuscript. V.-H.N., A.S., H.S., J.J.C., B.T.P., B.B.A., and S.L. provided critical comments in planning this paper and edited the manuscript. All the authors discussed the results and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Basic Research Project of the Korea Institute of Geoscience, Mineral Resources (KIGAM) funded by the Minister of Science and ICT.

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

## References

1. Morgan, R.P.C. *Soil Erosion and Conservation*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
2. Conoscenti, C.; Angileri, S.; Cappadonia, C.; Rotigliano, E.; Agnesi, V.; Märker, M. Gully erosion susceptibility assessment by means of gis-based logistic regression: A case of sicily (Italy). *Geomorphology* **2014**, *204*, 399–411. [[CrossRef](#)]
3. Moradi, H.; Avand, M.T.; Janizadeh, S. Landslide susceptibility survey using modeling methods. In *Spatial Modeling in Gis and R for Earth and Environmental Sciences*; Elsevier: New York, NY, USA, 2019; pp. 259–275.
4. Ionita, I.; Fullen, M.A.; Zgłobicki, W.; Poesen, J. Gully erosion as a natural and human-induced hazard. *Nat. Hazards* **2015**, *79*. [[CrossRef](#)]
5. Ni, H.; Li, Z.; Tie, Y.; Song, Z. Formation condition, disaster characteristics and developing trend analysis on debris flows in moxi river basin, sw China. In *Landslide Science for a Safer Geoenvironment*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 5–11.
6. Jurchescu, M.; Grecu, F. Modelling the occurrence of gullies at two spatial scales in the olteț drainage basin (Romania). *Nat. Hazards* **2015**, *79*, 255–289. [[CrossRef](#)]
7. Kirkby, M. *Thresholds and Instability in Stream Head Hollows: A Model of Magnitude and Frequency for Wash Processes*; School of Geography, University of Leeds: Leeds, UK, 1992.
8. Lucà, F.; Conforti, M.; Robustelli, G. Comparison of gis-based gullying susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. *Geomorphology* **2011**, *134*, 297–308. [[CrossRef](#)]
9. Poesen, J.; Vandekerckhove, L.; Nachtergaele, J.; Oostwoud Wijdenes, D.; Verstraeten, G.; van Wesemael, B. Gully erosion in dryland environments. In *Dryland Rivers: Hydrology and Geomorphology of Semi-Arid Channels*; Bull, L.J., Kirkby, M.J., Eds.; Wiley: Chichester, UK, 2002; pp. 229–262.
10. Valentin, C.; Poesen, J.; Li, Y. Gully erosion: Impacts, factors and control. *Catena* **2005**, *63*, 132–153. [[CrossRef](#)]
11. Istanbuluoğlu, E.; Tarboton, D.G.; Pack, R.T.; Luce, C. A probabilistic approach for channel initiation. *Water Resour. Res.* **2002**, *38*, 1–14. [[CrossRef](#)]

12. Shellberg, J.; Spencer, J.; Brooks, A.; Pietsch, T. Degradation of the mitchell river fluvial megafan by alluvial gully erosion increased by post-european land use change, queensland, australia. *Geomorphology* **2016**, *266*, 105–120. [[CrossRef](#)]
13. Burkard, M.; Kostaschuk, R. Patterns and controls of gully growth along the shoreline of lake huron. *Earth Surf. Process. Landf. J. Br. Geomorphol. Group* **1997**, *22*, 901–911. [[CrossRef](#)]
14. Heathwaite, A.L.; Burt, T.; Trudgill, S. Land-use controls on sediment production in a lowland catchment, south-west England. In *Soil Erosion on Agricultural Land, Proceedings of the Workshop Sponsored by the British Geomorphological Research Group, Coventry, UK, 17–19 January 1989*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 1990; pp. 69–86.
15. Nachtergaele, J.; Poesen, J.; Sidorchuk, A.; Torri, D. Prediction of concentrated flow width in ephemeral gully channels. *Hydrol. Process.* **2002**, *16*, 1935–1953. [[CrossRef](#)]
16. 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. Group* **2002**, *27*, 1267–1283. [[CrossRef](#)]
17. McCloskey, G.; Wasson, R.; Boggs, G.; Douglas, M. Timing and causes of gully erosion in the riparian zone of the semi-arid tropical victoria river, australia: Management implications. *Geomorphology* **2016**, *266*, 96–104. [[CrossRef](#)]
18. Wang, Y.; Hong, H.; Chen, W.; Li, S.; Panahi, M.; Khosravi, K.; Shirzadi, A.; Shahabi, H.; Panahi, S.; Costache, R. Flood susceptibility mapping in Dingnan county (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. *J. Environ. Manag.* **2019**, *247*, 712–729. [[CrossRef](#)] [[PubMed](#)]
19. Khosravi, K.; Shahabi, H.; Pham, B.T.; Adamowski, J.; Shirzadi, A.; Pradhan, B.; Dou, J.; Ly, H.-B.; Gróf, G.; Ho, H.L. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *J. Hydrol.* **2019**, *573*, 311–323. [[CrossRef](#)]
20. He, Q.; Shahabi, H.; Shirzadi, A.; Li, S.; Chen, W.; Wang, N.; Chai, H.; Bian, H.; Ma, J.; Chen, Y. Landslide spatial modelling using novel bivariate statistical based naïve bayes, rbf classifier, and rbf network machine learning algorithms. *Sci. Total Environ.* **2019**, *663*, 1–15. [[CrossRef](#)] [[PubMed](#)]
21. Tien Bui, D.; Shirzadi, A.; Chapi, K.; Shahabi, H.; Pradhan, B.; Pham, B.T.; Singh, V.P.; Chen, W.; Khosravi, K.; Bin Ahmad, B. A hybrid computational intelligence approach to groundwater spring potential mapping. *Water* **2019**, *11*, 2013. [[CrossRef](#)]
22. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Hoang, N.-D.; Pham, B.; Bui, Q.-T.; Tran, C.-T.; Panahi, M.; Bin Ahmad, B. A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides. *Remote Sens.* **2018**, *10*, 1538. [[CrossRef](#)]
23. Tien Bui, D.; Khosravi, K.; Li, S.; Shahabi, H.; Panahi, M.; Singh, V.; Chapi, K.; Shirzadi, A.; Panahi, S.; Chen, W. New hybrids of anfis with several optimization algorithms for flood susceptibility modeling. *Water* **2018**, *10*, 1210. [[CrossRef](#)]
24. Shafizadeh-Moghadam, H.; Valavi, R.; Shahabi, H.; Chapi, K.; Shirzadi, A. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *J. Environ. Manag.* **2018**, *217*, 1–11. [[CrossRef](#)]
25. Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Bui, D.T.; Pham, B.T.; Khosravi, K. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ. Model. Softw.* **2017**, *95*, 229–245. [[CrossRef](#)]
26. Chen, W.; Li, Y.; Xue, W.; Shahabi, H.; Li, S.; Hong, H.; Wang, X.; Bian, H.; Zhang, S.; Pradhan, B. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. *Sci. Total Environ.* **2020**, *701*, 134979. [[CrossRef](#)]
27. Shahabi, H.; Shirzadi, A.; Ghaderi, K.; Omidvar, E.; Al-Ansari, N.; Clague, J.J.; Geertsema, M.; Khosravi, K.; Amini, A.; Bahrami, S. Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on k-nearest neighbor classifier. *Remote Sens.* **2020**, *12*, 266. [[CrossRef](#)]
28. Khosravi, K.; Melesse, A.M.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Hong, H. Flood susceptibility mapping at Ningdu catchment, China using bivariate and data mining techniques. In *Extreme Hydrology and Climate Variability*; Elsevier: New York, NY, USA, 2019; pp. 419–434.

29. Jaafari, A.; Zenner, E.K.; Panahi, M.; Shahabi, H. Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agric. For. Meteorol.* **2019**, *266*, 198–207. [[CrossRef](#)]
30. Taheri, K.; Shahabi, H.; Chapi, K.; Shirzadi, A.; Gutiérrez, F.; Khosravi, K. Sinkhole susceptibility mapping: A comparison between bayes-based machine learning algorithms. *Land Degrad. Dev.* **2019**, *30*, 730–745. [[CrossRef](#)]
31. Roodposhti, M.S.; Safarrad, T.; Shahabi, H. Drought sensitivity mapping using two one-class support vector machine algorithms. *Atmos. Res.* **2017**, *193*, 73–82. [[CrossRef](#)]
32. Choubin, B.; Soleimani, F.; Pirnia, A.; Sajedi-Hosseini, F.; Alilou, H.; Rahmati, O.; Melesse, A.M.; Singh, V.P.; Shahabi, H. Effects of drought on vegetative cover changes: Investigating spatiotemporal patterns. In *Extreme Hydrology and Climate Variability*; Elsevier: New York, NY, USA, 2019; pp. 213–222.
33. Lee, S.; Panahi, M.; Pourghasemi, H.R.; Shahabi, H.; Alizadeh, M.; Shirzadi, A.; Khosravi, K.; Melesse, A.M.; Yekrangnia, M.; Rezaie, F. Sevucas: A novel gis-based machine learning software for seismic vulnerability assessment. *Appl. Sci.* **2019**, *9*, 3495. [[CrossRef](#)]
34. Alizadeh, M.; Alizadeh, E.; Asadollahpour Kotenaee, S.; Shahabi, H.; Beiranvand Pour, A.; Panahi, M.; Bin Ahmad, B.; Saro, L. Social vulnerability assessment using Artificial Neural Network (ANN) model for earthquake hazard in Tabriz city, Iran. *Sustainability* **2018**, *10*, 3376. [[CrossRef](#)]
35. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Pradhan, B.; Chen, W.; Khosravi, K.; Panahi, M.; Bin Ahmad, B.; Saro, L. Land subsidence susceptibility mapping in south korea using machine learning algorithms. *Sensors* **2018**, *18*, 2464. [[CrossRef](#)]
36. Rahmati, O.; Samadi, M.; Shahabi, H.; Azareh, A.; Rafiei-Sardooi, E.; Alilou, H.; Melesse, A.M.; Pradhan, B.; Chapi, K.; Shirzadi, A. Swpt: An automated gis-based tool for prioritization of sub-watersheds based on morphometric and topo-hydrological factors. *Geosci. Front.* **2019**, *10*, 2167–2175. [[CrossRef](#)]
37. Choubin, B.; Rahmati, O.; Tahmasebipour, N.; Feizizadeh, B.; Pourghasemi, H.R. Application of fuzzy analytical network process model for analyzing the gully erosion susceptibility. In *Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 105–125.
38. Chen, W.; Pradhan, B.; Li, S.; Shahabi, H.; Rizeei, H.M.; Hou, E.; Wang, S. Novel hybrid integration approach of bagging-based fisher’s linear discriminant function for groundwater potential analysis. *Nat. Resour. Res.* **2019**, *28*, 1239–1258. [[CrossRef](#)]
39. Miraki, S.; Zanganeh, S.H.; Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Pham, B.T. Mapping groundwater potential using a novel hybrid intelligence approach. *Water Resour. Manag.* **2019**, *33*, 281–302. [[CrossRef](#)]
40. Rahmati, O.; Naghibi, S.A.; Shahabi, H.; Bui, D.T.; Pradhan, B.; Azareh, A.; Rafiei-Sardooi, E.; Samani, A.N.; Melesse, A.M. Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. *J. Hydrol.* **2018**, *565*, 248–261. [[CrossRef](#)]
41. Chen, W.; Li, Y.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Xue, W.; Bian, H. Groundwater spring potential mapping using artificial intelligence approach based on kernel logistic regression, random forest, and alternating decision tree models. *Appl. Sci.* **2020**, *10*, 425. [[CrossRef](#)]
42. Avand, M.; Janizadeh, S.; Tien Bui, D.; Pham, V.H.; Ngo, P.T.T.; Nhu, V.H. A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *Int. J. Digit. Earth* **2020**, 1–22. [[CrossRef](#)]
43. Chen, W.; Tsangaratos, P.; Ilia, I.; Duan, Z.; Chen, X. Groundwater spring potential mapping using population-based evolutionary algorithms and data mining methods. *Sci. Total Environ.* **2019**, *684*, 31–49. [[CrossRef](#)]
44. Chen, W.; Zhao, X.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Xue, W.; Wang, X.; Ahmad, B.B. Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. *J. Hydrol.* **2020**, *583*, 124602. [[CrossRef](#)]
45. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Alizadeh, M.; Chen, W.; Mohammadi, A.; Ahmad, B.; Panahi, M.; Hong, H. Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, malaysia. *Remote Sens.* **2018**, *10*, 1527. [[CrossRef](#)]
46. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, A.-X.; Pei, X.; Duan, Z. Landslide susceptibility modelling using gis-based machine learning techniques for Chongren county, Jiangxi province, China. *Sci. Total Environ.* **2018**, *626*, 1121–1135. [[CrossRef](#)]

47. Pham, B.T.; Shirzadi, A.; Shahabi, H.; Omidvar, E.; Singh, S.K.; Sahana, M.; Asl, D.T.; Ahmad, B.B.; Quoc, N.K.; Lee, S. Landslide susceptibility assessment by novel hybrid machine learning algorithms. *Sustainability* **2019**, *11*, 4386. [[CrossRef](#)]
48. Shirzadi, A.; Bui, D.T.; Pham, B.T.; Solaimani, K.; Chapi, K.; Kavian, A.; Shahabi, H.; Revhaug, I. Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ. Earth Sci.* **2017**, *76*, 60. [[CrossRef](#)]
49. Pham, B.T.; Prakash, I.; Dou, J.; Singh, S.K.; Trinh, P.T.; Tran, H.T.; Le, T.M.; Van Phong, T.; Khoi, D.K.; Shirzadi, A. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto Int.* **2019**. [[CrossRef](#)]
50. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using gis. *Comput. Geosci.* **2013**, *51*, 350–365. [[CrossRef](#)]
51. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.; Akgun, A.; Tian, Y.; Liu, J.; Zhu, A.-X.; Li, S. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4397–4419. [[CrossRef](#)]
52. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Khosravi, K.; Yang, Y.; Pham, B.T. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci. Total Environ.* **2019**, *662*, 332–346. [[CrossRef](#)] [[PubMed](#)]
53. Jaafari, A.; Panahi, M.; Pham, B.T.; Shahabi, H.; Bui, D.T.; Rezaie, F.; Lee, S. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena* **2019**, *175*, 430–445. [[CrossRef](#)]
54. Hong, H.; Shahabi, H.; Shirzadi, A.; Chen, W.; Chapi, K.; Ahmad, B.B.; Roodposhti, M.S.; Hesar, A.Y.; Tian, Y.; Bui, D.T. Landslide susceptibility assessment at the Wuning area, China: A comparison between multi-criteria decision making, bivariate statistical and machine learning methods. *Nat. Hazards* **2019**, *96*, 173–212. [[CrossRef](#)]
55. Shafizadeh-Moghadam, H.; Minaei, M.; Shahabi, H.; Hagenauer, J. Big data in geohazard; pattern mining and large scale analysis of landslides in Iran. *Earth Sci. Inform.* **2019**, *12*, 1–17. [[CrossRef](#)]
56. Nguyen, V.V.; Pham, B.T.; Vu, B.T.; Prakash, I.; Jha, S.; Shahabi, H.; Shirzadi, A.; Ba, D.N.; Kumar, R.; Chatterjee, J.M. Hybrid machine learning approaches for landslide susceptibility modeling. *Forests* **2019**, *10*, 157. [[CrossRef](#)]
57. Nguyen, P.T.; Tuyen, T.T.; Shirzadi, A.; Pham, B.T.; Shahabi, H.; Omidvar, E.; Amini, A.; Entezami, H.; Prakash, I.; Phong, T.V. Development of a novel hybrid intelligence approach for landslide spatial prediction. *Appl. Sci.* **2019**, *9*, 2824. [[CrossRef](#)]
58. Tien Bui, D.; Shahabi, H.; Omidvar, E.; Shirzadi, A.; Geertsema, M.; Clague, J.J.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Chapi, K. Shallow landslide prediction using a novel hybrid functional machine learning algorithm. *Remote Sens.* **2019**, *11*, 931. [[CrossRef](#)]
59. Tien Bui, D.; Shirzadi, A.; Shahabi, H.; Geertsema, M.; Omidvar, E.; Clague, J.J.; Thai Pham, B.; Dou, J.; Talebpour Asl, D.; Bin Ahmad, B. New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed. *Forests* **2019**, *10*, 743. [[CrossRef](#)]
60. Chen, W.; Zhao, X.; Shahabi, H.; Shirzadi, A.; Khosravi, K.; Chai, H.; Zhang, S.; Zhang, L.; Ma, J.; Chen, Y. Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. *Geocarto Int.* **2019**, *34*, 1177–1201. [[CrossRef](#)]
61. Shirzadi, A.; Solaimani, K.; Roshan, M.H.; Kavian, A.; Chapi, K.; Shahabi, H.; Keesstra, S.; Ahmad, B.B.; Bui, D.T. Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. *Catena* **2019**, *178*, 172–188. [[CrossRef](#)]
62. Shirzadi, A.; Solaimani, K.; Habibnejhad, M.; Kavian, A.; Chapi, K.; Shahabi, H.; Chen, W.; Khosravi, K.; Thai Pham, B.; Pradhan, B. Novel gis based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* **2018**, *18*, 3777. [[CrossRef](#)] [[PubMed](#)]
63. Chen, W.; Shahabi, H.; Zhang, S.; Khosravi, K.; Shirzadi, A.; Chapi, K.; Pham, B.; Zhang, T.; Zhang, L.; Chai, H. Landslide susceptibility modeling based on gis and novel bagging-based kernel logistic regression. *Appl. Sci.* **2018**, *8*, 2540. [[CrossRef](#)]



64. Zhang, T.; Han, L.; Chen, W.; Shahabi, H. Hybrid integration approach of entropy with logistic regression and support vector machine for landslide susceptibility modeling. *Entropy* **2018**, *20*, 884. [[CrossRef](#)]
65. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Pham, B.T.; Bin Ahmad, B.; Tien Bui, D. A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment. *Geocarto Int.* **2019**, *34*, 1427–1457. [[CrossRef](#)]
66. Chen, W.; Xie, X.; Peng, J.; Shahabi, H.; Hong, H.; Bui, D.T.; Duan, Z.; Li, S.; Zhu, A.-X. Gis-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *Catena* **2018**, *164*, 135–149. [[CrossRef](#)]
67. Chen, W.; Shirzadi, A.; Shahabi, H.; Ahmad, B.B.; Zhang, S.; Hong, H.; Zhang, N. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve bayes tree classifiers for a landslide susceptibility assessment in Langao county, China. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1955–1977. [[CrossRef](#)]
68. Hong, H.; Liu, J.; Zhu, A.-X.; Shahabi, H.; Pham, B.T.; Chen, W.; Pradhan, B.; Bui, D.T. A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). *Environ. Earth Sci.* **2017**, *76*, 652. [[CrossRef](#)]
69. Shadman Roodposhti, M.; Aryal, J.; Shahabi, H.; Safarrad, T. Fuzzy shannon entropy: A hybrid gis-based landslide susceptibility mapping method. *Entropy* **2016**, *18*, 343. [[CrossRef](#)]
70. Shahabi, H.; Hashim, M.; Ahmad, B.B. Remote sensing and gis-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central zab basin, Iran. *Environ. Earth Sci.* **2015**, *73*, 8647–8668. [[CrossRef](#)]
71. Shahabi, H.; Khezri, S.; Ahmad, B.B.; Hashim, M. Landslide susceptibility mapping at central zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models. *Catena* **2014**, *115*, 55–70. [[CrossRef](#)]
72. Zhao, X.; Chen, W. Gis-based evaluation of landslide susceptibility models using certainty factors and functional trees-based ensemble techniques. *Appl. Sci.* **2020**, *10*, 16. [[CrossRef](#)]
73. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Bui, D.T. A comparative study of support vector machine and logistic model tree classifiers for shallow landslide susceptibility modeling. *Environ. Earth Sci.* **2019**, *78*, 560. [[CrossRef](#)]
74. Chaplot, V.; Le Brozec, E.C.; Silvera, N.; Valentin, C. Spatial and temporal assessment of linear erosion in catchments under sloping lands of northern laos. *Catena* **2005**, *63*, 167–184. [[CrossRef](#)]
75. 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](#)]
76. Tien Bui, D.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Omidavr, E.; Pham, B.T.; Talebpour Asl, D.; Khaledian, H.; Pradhan, B.; Panahi, M. A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (Iran). *Sensors* **2019**, *19*, 2444. [[CrossRef](#)]
77. Pham, B.T.; Avand, M.; Janizadeh, S.; Phong, T.V.; Al-Ansari, N.; Ho, L.S.; Jafari, F. GIS Based Hybrid Computational Approaches for Flash Flood Susceptibility Assessment. *Water* **2020**, *12*, 683. [[CrossRef](#)]
78. Pourghasemi, H.R.; Yousefi, S.; Kornejady, A.; Cerdà, A. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.* **2017**, *609*, 764–775. [[CrossRef](#)]
79. Yariyan, P.; Avand, M.; Soltani, F.; Ghorbanzadeh, O.; Blaschke, T. Earthquake Vulnerability Mapping Using Different Hybrid Models. *Symmetry* **2020**, *12*, 405. [[CrossRef](#)]
80. Nguyen, M.D.; Pham, B.T.; Tuyen, T.T.; Yen, H.; Phan, H.; Prakash, I.; Vu, T.T.; Chapi, K.; Shirzadi, A.; Shahabi, H. Development of an artificial intelligence approach for prediction of consolidation coefficient of soft soil: A sensitivity analysis. *Open Constr. Build. Technol. J.* **2019**, *13*. [[CrossRef](#)]
81. 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](#)]
82. Azareh, A.; Rahmati, O.; Rafiei-Sardooi, E.; Sankey, J.B.; Lee, S.; Shahabi, H.; Ahmad, B.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](#)] [[PubMed](#)]
83. Shahabi, H.; Jarhani, B.; Tavakkoli Piralilou, S.; Chittleborough, D.; Avand, M.; Ghorbanzadeh, O. A Semi-Automated Object-Based Gully Networks Detection Using Different Machine Learning Models: A Case Study of Bowen Catchment, Queensland, Australia. *Sensors* **2019**, *19*, 4893. [[CrossRef](#)]

84. Ahmadlou, M.; Karimi, M.; Alizadeh, S.; Shirzadi, A.; Parvinnejhad, D.; Shahabi, H.; Panahi, M. Flood susceptibility assessment using integration of Adaptive Network-Based Fuzzy Inference System (ANFIS) and Biogeography-Based Optimization (BBO) and Bat Algorithms (BA). *Geocarto Int.* **2019**, *34*, 1252–1272. [[CrossRef](#)]
85. Bui, D.T.; Panahi, M.; Shahabi, H.; Singh, V.P.; Shirzadi, A.; Chapi, K.; Khosravi, K.; Chen, W.; Panahi, S.; Li, S. Novel hybrid evolutionary algorithms for spatial prediction of floods. *Sci. Rep.* **2018**, *8*, 1–14. [[CrossRef](#)]
86. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S. Spatial prediction of landslide susceptibility using gis-based data mining techniques of anfis with Whale Optimization Algorithm (WOA) and Grey Wolf Optimizer (GWO). *Appl. Sci.* **2019**, *9*, 3755. [[CrossRef](#)]
87. Shirzadi, A.; Shahabi, H.; Chapi, K.; Bui, D.T.; Pham, B.T.; Shahedi, K.; Ahmad, B.B. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena* **2017**, *157*, 213–226. [[CrossRef](#)]
88. Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Bui, D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at haraz watershed, Northern Iran. *Sci. Total Environ.* **2018**, *627*, 744–755. [[CrossRef](#)]
89. Chen, W.; Shahabi, H.; Shirzadi, A.; Li, T.; Guo, C.; Hong, H.; Li, W.; Pan, D.; Hui, J.; Ma, M. A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. *Geocarto Int.* **2018**, *33*, 1398–1420. [[CrossRef](#)]
90. 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](#)]
91. Gayen, A.; Pourghasemi, H.R. Spatial modeling of gully erosion: A new ensemble of cart and glm data-mining algorithms. In *Spatial Modeling in Gis and R for Earth and Environmental Sciences*; Elsevier: New York, NY, USA, 2019; pp. 653–669.
92. Garosi, Y.; Sheklabadi, M.; Conoscenti, C.; Pourghasemi, H.R.; Van Oost, K. Assessing the performance of gis-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. *Sci. Total Environ.* **2019**, *664*, 1117–1132. [[CrossRef](#)] [[PubMed](#)]
93. Avand, M.; Janizadeh, S.; Naghibi, S.A.; Pourghasemi, H.R.; Khosrobeigi Bozchaloei, S.; Blaschke, T. A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. *Water* **2019**, *11*, 2076. [[CrossRef](#)]
94. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Bui, D.T. Landslide susceptibility modeling using reduced error pruning trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [[CrossRef](#)]
95. Pham, B.T.; Jaafari, A.; Prakash, I.; Singh, S.K.; Quoc, N.K.; Bui, D.T. Hybrid computational intelligence models for groundwater potential mapping. *Catena* **2019**, *182*, 104101. [[CrossRef](#)]
96. Dube, F.; Nhapi, I.; Murwira, A.; Gumindoga, W.; Goldin, J.; Mashauri, D. Potential of weight of evidence modelling for gully erosion hazard assessment in mbire district–zimbabwe. *Phys. Chem. Earth Parts A/B/C* **2014**, *67*, 145–152. [[CrossRef](#)]
97. Rahmati, O.; Pourghasemi, H.R.; Zeinivand, H. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan province, Iran. *Geocarto Int.* **2016**, *31*, 42–70. [[CrossRef](#)]
98. Janizadeh, S.; Avand, M.; Jaafari, A.; Phong, T.V.; Bayat, M.; Ahmadisharaf, E.; Prakash, I.; Pham, B.T.; Lee, S. Prediction success of machine learning methods for flash flood susceptibility mapping in the tafresh watershed, Iran. *Sustainability* **2019**, *11*, 5426. [[CrossRef](#)]
99. Rengers, F.K.; Tucker, G. Analysis and modeling of gully headcut dynamics, north american high plains. *J. Geophys. Res. Earth Surf.* **2014**, *119*, 983–1003. [[CrossRef](#)]
100. Hosseinalizadeh, M.; Kariminejad, N.; Chen, W.; Pourghasemi, H.R.; Alinejad, M.; Behbahani, A.M.; Tiefenbacher, J.P. Gully headcut susceptibility modeling using functional trees, naïve bayes tree, and random forest models. *Geoderma* **2019**, *342*, 1–11. [[CrossRef](#)]
101. Oostwoud Wijdenes, D.; BRYAN, R. The significance of gully headcuts as a source of sediment on low-angle slopes at baringo, kenya, and initial control measures. *Adv. Geoecol.* **1994**, *27*, 205–231.
102. Wijdenes, D.J.O.; Poesen, J.; Vandekerckhove, L.; Ghesquiere, M. Spatial distribution of gully head activity and sediment supply along an ephemeral channel in a mediterranean environment. *Catena* **2000**, *39*, 147–167. [[CrossRef](#)]

103. Kuhnert, P.M.; Henderson, A.K.; Bartley, R.; Herr, A. Incorporating uncertainty in gully erosion calculations using the random forests modelling approach. *Environmetrics* **2010**, *21*, 493–509. [[CrossRef](#)]
104. 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](#)]
105. 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](#)]
106. Di Stefano, C.; Ferro, V.; Palmeri, V.; Pampalona, V. Testing slope effect on flow resistance equation for mobile bed rills. *Hydrol. Process.* **2018**, *32*, 664–671. [[CrossRef](#)]
107. Kariminejad, N.; Hosseinalizadeh, M.; Pourghasemi, H.R.; Bernatek-Jakiel, A.; Campetella, G.; Ownegh, M. Evaluation of factors affecting gully headcut location using summary statistics and the maximum entropy model: Golestan province, ne Iran. *Sci. Total Environ.* **2019**, *677*, 281–298. [[CrossRef](#)]
108. 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](#)]
109. Frankl, A.; Poesen, J.; Deckers, J.; Haile, M.; Nyssen, J. Gully head retreat rates in the semi-arid highlands of northern ethiopia. *Geomorphology* **2012**, *173*, 185–195. [[CrossRef](#)]
110. Svoray, T.; Markovitch, H. Catchment scale analysis of the effect of topography, tillage direction and unpaved roads on ephemeral gully incision. *Earth Surf. Process. Landf. J. Br. Geomorphol. Res. Group* **2009**, *34*, 1970–1984. [[CrossRef](#)]
111. Freund, Y.; Schapire, R.E. Experiments with a new boosting algorithm. In Proceedings of the 13th International Conference on Machine Learning, Bari, Italy, 3–6 July 1996; Citeseer: Gaithersburg, MD, USA, 1996; pp. 148–156.
112. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
113. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1994.
114. Quinlan, J.R. Simplifying decision trees. *Int. J. Man-Mach. Stud.* **1987**, *27*, 221–234. [[CrossRef](#)]
115. Matthews, B.W. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Biophys. Acta (BBA)-Protein Struct.* **1975**, *405*, 442–451. [[CrossRef](#)]
116. Fawcett, T. An introduction to roc analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
117. Pham, B.T.; Bui, D.T.; Dholakia, M.; Prakash, I.; Pham, H.V. A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. *Geotech. Geol. Eng.* **2016**, *34*, 1807–1824. [[CrossRef](#)]
118. Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [[CrossRef](#)] [[PubMed](#)]
119. Duch, W.; Winiarski, T.; Biesiada, J.; Kachel, A. Feature selection and ranking filters. In Proceedings of the International Conference on Artificial Neural Networks (ICANN) and International Conference on Neural Information Processing (ICONIP), Istanbul, Turkey, 26–29 June 2003; p. 254.
120. Nandhini, M.; Sivanandam, S. An improved predictive association rule based classifier using gain ratio and t-test for health care data diagnosis. *Sadhana* **2015**, *40*, 1683–1699. [[CrossRef](#)]
121. 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](#)]
122. Rahmati, O.; Tahmasebipour, N.; Haghizadeh, A.; Pourghasemi, H.R.; Feizizadeh, B. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* **2017**, *298*, 118–137. [[CrossRef](#)]
123. Pham, B.T.; Prakash, I.; Bui, D.T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **2018**, *303*, 256–270. [[CrossRef](#)]
124. Hong, H.; Liu, J.; Bui, D.T.; Pradhan, B.; Acharya, T.D.; Pham, B.T.; Zhu, A.-X.; Chen, W.; Ahmad, B.B. Landslide susceptibility mapping using j48 decision tree with adaboost, bagging and rotation forest ensembles in the guangchang area (China). *Catena* **2018**, *163*, 399–413. [[CrossRef](#)]
125. 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](#)]

126. Gayen, A.; Pourghasemi, H.R.; Saha, S.; Keesstra, S.; Bai, S. Gully erosion susceptibility assessment and management of hazard-prone areas in india using different machine learning algorithms. *Sci. Total Environ.* **2019**, *668*, 124–138. [[CrossRef](#)]
127. Vandekerckhove, L.; Poesen, J.; Oostwoud Wijdenes, D.; Nachtergaele, J.; Kosmas, C.; Roxo, M.; De Figueiredo, T. Thresholds for gully initiation and sedimentation in mediterranean europe. *Earth Surf. Process. Landf.* **2000**, *25*, 1201–1220. [[CrossRef](#)]
128. Bergonse, R.; Reis, E. Controlling factors of the size and location of large gully systems: A regression-based exploration using reconstructed pre-erosion topography. *Catena* **2016**, *147*, 621–631. [[CrossRef](#)]
129. Bayat, M.; Ghorbanpour, M.; Zare, R.; Jaafari, A.; Pham, B.T. Application of artificial neural networks for predicting tree survival and mortality in the hyrcanian forest of Iran. *Comput. Electron. Agric.* **2019**, *164*, 104929. [[CrossRef](#)]
130. Van Dao, D.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayedi, H.; Van Phong, T.; Ly, H.-B.; Le, T.-T.; Trinh, P.T. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* **2020**, *188*, 104451.
131. Qiao, W.; Huang, K.; Azimi, M.; Han, S. A novel hybrid prediction model for hourly gas consumption in supply side based on improved whale optimization algorithm and relevance vector machine. *IEEE Access* **2019**, *7*, 88218–88230. [[CrossRef](#)]
132. Zhou, G.; Moayedi, H.; Bahiraei, M.; Lyu, Z. Employing artificial bee colony and particle swarm techniques for optimizing a neural network in prediction of heating and cooling loads of residential buildings. *J. Clean. Prod.* **2020**, 120082. [[CrossRef](#)]
133. Ujoh, F.; Igbawua, T.; Ogidi Paul, M. Suitability mapping for rice cultivation in Benue State, Nigeria using satellite data. *Geo. Spatial. Inform. Sci* **2019**, *22*, 332–344. [[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/>).