applied sciences

MDPI

Article

Groundwater Spring Potential Mapping Using Artificial Intelligence Approach Based on Kernel Logistic Regression, Random Forest, and Alternating Decision Tree Models

Wei Chen [1,2], Yang Li [1], Paraskevas Tsangaratos [3,*], Himan Shahabi [4], Ioanna Ilia [3], Weifeng Xue [5] and Huiyuan Bian [1]

1 College of Geology & Environment, Xi'an University of Science and Technology, Xi'an 710054, China; chenwei0930@xust.edu.cn (W.C.); 19209071008@stu.xust.edu.cn (Y.L.); bianhuiyuan@xust.edu.cn (H.B.)
2 Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Natural Resources, Xi'an 710021, China
3 Zografou Campus: Heroon Polytechniou 9, Laboratory of Engineering Geology and Hydrogeology, Department of Geological Sciences, National Technical University of Athens, School of Mining and Metallurgical Engineering, 15780 Zografou, Greece; gilia@metal.ntua.gr
4 Departments of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran; h.shahabi@uok.ac.ir
5 Shaanxi Coal and Chemical Technology Institute Co Ltd., Xi'an 710065, China; xuewf@sxccti.com
* Correspondence: ptsag@metal.ntua.gr

Received: 20 November 2019; Accepted: 3 January 2020; Published: 7 January 2020



Abstract: This study presents a methodology for constructing groundwater spring potential maps by kernel logistic regression, (KLR), random forest (RF), and alternating decision tree (ADTree) models. The analysis was based on data concerning groundwater springs and fourteen explanatory factors (elevation, slope, aspect, plan curvature, profile curvature, stream power index, sediment transport index, topographic wetness index, distance to streams, distance to roads, normalized difference vegetation index (NDVI), lithology, soil, and land use), which were divided into training and validation datasets. Ningtiaota region in the northern territory of Shaanxi Province, China, was considered as a test site. Frequency Ratio method was applied to provide to each factor's class a coefficient weight, whereas the linear support vector machine method was used as a feature selection method to determine the optimal set of factors. The Receiver Operating Characteristic curve and the area under the curve (AUC) were used to evaluate the performance of each model using the training dataset, with the RF model providing the highest AUC value (0.909) followed by the KLR (0.877) and ADTree (0.812) models. The same performance pattern was estimated based on the validation dataset, with the RF model providing the highest AUC value (0.811) followed by the KLR (0.797) and ADTree (0.773) models. This study highlights that the artificial intelligence approach could be considered as a valid and accurate approach for groundwater spring potential zoning.

Keywords: groundwater spring potential mapping; kernel logistic regression; random forest; alternating decision tree; China

1. Introduction

As pointed out by many researchers, one of the most important natural resource worldwide is groundwater, with one third of the world's population depending on it [1–4]. Several areas in the world are subject to overexploitation of groundwater, undergoing water shortages as a result of a difference between water supply and demand [5]. It is also well established that the demand for groundwater

will increase substantially in the following years, mainly due to the growing population and economic development [6–10]. According to Curran and de Sherbinin [11], even though the supply of water is mainly controlled by climatic parameters, the management and the followed practices significantly influence the availability of water. In the case of groundwater resources, inappropriate management may result in the deterioration of water resources but also a decrease in the quality of groundwater [12]. Similar with the rest of the world, China faces increasing consumption of groundwater, making imperative the application of accurate methods for assessing groundwater potential [13–16].

Groundwater spring potential mapping has been recognized as an investigation practice, the outcomes of which provide useful data inputs concerning groundwater management projects [17]. Identifying areas with high probability concerning the presence of groundwater springs assists in developing appropriate groundwater exploitation and groundwater resources conservation programs [18,19]. Over the past two decades, geographical information systems (GIS) and remote sensing techniques (RS) have been the main investigation tools concerning groundwater spring potential mapping [4,17,18,20–22]. Successful examples of studies concerning groundwater potential mapping involve bivariate and multivariate methods and specifically applications of frequency ratio (FR) [13,17,23], analytical hierarchy process (AHP) [20,24–27], weight of evidence (WofE) [18,23,28,29], evidential belief function (EBF) [19,30–33], and logistic regression (LR) [28,34,35].

Similar, machine learning methods have been introduced as an alternative option for groundwater potential, mapping mainly involving tree-based methods, such as classification and regression tree (CART) [36] and random forest (RF) [19,26,37], and neural network-based methods, such as artificial neural network (ANN) [18,38,39] and support vector machine (SVM) [40]. Other notable examples of machine learning methods that have been utilized in groundwater potential mapping assessments are the implementations of naive Bayes (NB) [41] and K-nearest neighbor (KNN) [42].

Quite recently, new hybrid and ensemble methods have been applied in groundwater mapping studies, showing in most cases enhanced performance than single predictive models [12]. Chen et al. [43] produced groundwater potential maps integrating WofE with LR and functional tree (FT) models, the validation of which clearly highlighted the efficacy of the integrated models. The authors reported that the integrated models provided better results, overcoming the drawbacks of bivariate statistics and machine learning. Khosravi et al. [44] proposed five hybrid artificial intelligence methods, integrating an adaptive neuro-fuzzy inference system (ANFIS) and meta-heuristic algorithms. The outcomes of their study illustrated that by applying the novel hybrid models they could produce more accurate groundwater potential models. Kordestani et al. [33] proposed an ensemble method, integrating EBF and boosted regression tree (BRT), reporting that the EBF–BRT model was capable of providing highly accurate results. The authors suggest that the produced model improves the weak points of each method, while taking advantage of the ability of the methods to analyze the relation of groundwater with each groundwater-related variable and with each class of the groundwater-related variable. In a similar study, Chen et al. [15] integrated an ANFIS model with a teaching–learning based optimization (TLBO) and a biogeography-based optimization (BBO) model. According to the authors, the two novel data mining methods could be useful in solving non-linear and high-dimensional problems and overall could be useful for groundwater management and exploration development projects.

In this context, the current study presents a novel hybrid integration approach of FR with artificial intelligence-based kernel logistic regression (KLR), alternating decision tree (ADTree), and RF models for groundwater spring potential mapping, having as a test site the Ningtiaota region, China. A hybrid integration approach of FR, KLR, ADTree, and RF is a relatively new contribution that has been seldom used for modeling of groundwater spring potential areas. It should also be mentioned that limited studies have been conducted concerning groundwater spring potential mapping in China, therefore, this research aims to fill this gap in the literature.

2. Study Area

Ningtiaota region is located in the northern territory of Shaanxi Province, China. The climate is characterized as dry throughout the year. The maximum and minimum temperatures are 38.9 °C and −29.0 °C, respectively, the average relative humidity is 56%, the average wind speed is 13.4 m/s, and the average annual rainfall is 434.1 mm. The study area, which is a portion of the Ningtiaota region, defined and limited to the area where data were available, is a geographical area of 119.77 km^2, located within latitudes 38°57′30′′ to 39°7′57′′ N and longitudes 110°9′36′′ to 110°16′20′′ E (Figure 1). According to the Soil Map produced by the Institute of Soil Science, Chinese Academy of Sciences [45], the typical soil types that cover the study region are Calcari-Gypsiric Arenosols (Arc), Haplic Arenosols (ARh), Calcareous Red Clay (CMe), and Luvi-Calcic Kastanozems (KSk)

Topographically, altitudes vary from 1118 to 1364 m above the sea level, and slope gradients vary from 0 to 37.88° based on a digital elevation model (DEM) with a 30 m regular grid. Approximately 75.38% of the area appears with less than 10° slope surface, whereas only 0.097% of the total study area have slopes greater than 30°. Areas with the slopes between 10 and 20° and 20 and 30° account for 21.77% and 2.74%, respectively.

Geologically, the strata of the study area belong to the Ordos Basin stratigraphic subarea in the North China stratigraphic area. Based on the geological map of China (http://www.cgs.gov.cn), the strata in the area from old to new are Yan'an formation (J_{2y}), Zhiluo formation (J_{2z}), Anding formation (J_{2a}), Baode formation (N_{2b}), Lishi formation (Q_{2l}), Salawusu formation (Q_{3s}), Malan formation (Q_{3m}), Alluvium (Q_{4al}), and Eolian deposit (Q_{4eol}), respectively (Table 1).

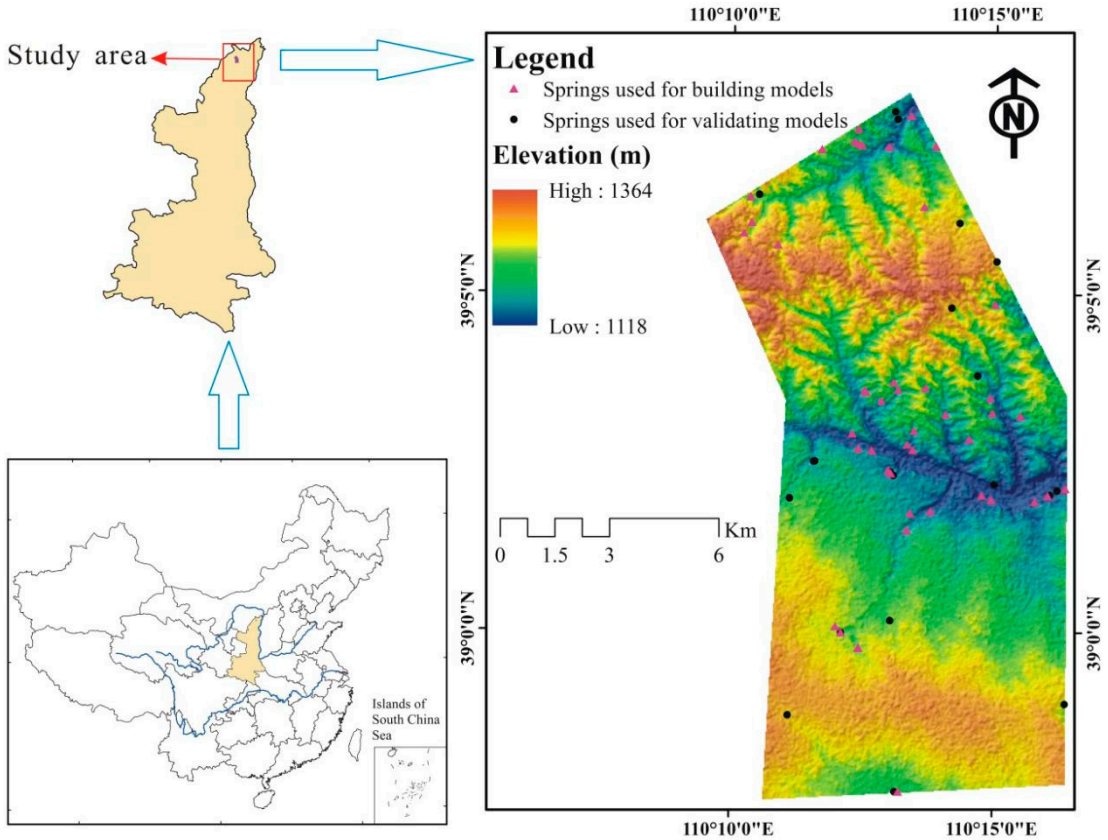


Figure 1. The study area.

Table 1. Lithology of the study area.

Categories	Codes	Lithologies
Group A	J_{2y}	Mudstone, sandy mudstone, arcose sandstone
Group B	J_{2z}	Mudstone, sandstone, glutenite
Group C	J_{2a}	Mudstone, sandstone
Group D	N_{2b}	Clay
Group E	Q_{2l}	Loess
Group F	Q_{3s}	Sand
Group G	Q_{3m}	Loess
Group H	Q_{4al}	Alluvium
Group I	Q_{4eol}	Eolian deposit

3. Methodology

The developed investigation approach followed in the present study was a four-step procedure: (i) data selection, generation of the spring inventory map and selection of nonspring areas, (ii) application of the Frequency Ration (FR) method and the linear support vector machine (LSVM) as a feature selection method so as to quantify the contribution of each explanatory factor and determine the optimal set of factors that have high predictive power, construction of the training and validation dataset, (iii) application of the Kernel logistic regression (KLR), Random Forest (RF), and Alternating Decision Tree (ADTree) models, and (iv) validation and comparison of the developed models. Figure 2 highlights the flowchart of the followed methodology, and each method used in our study will be briefly described in the following paragraphs.

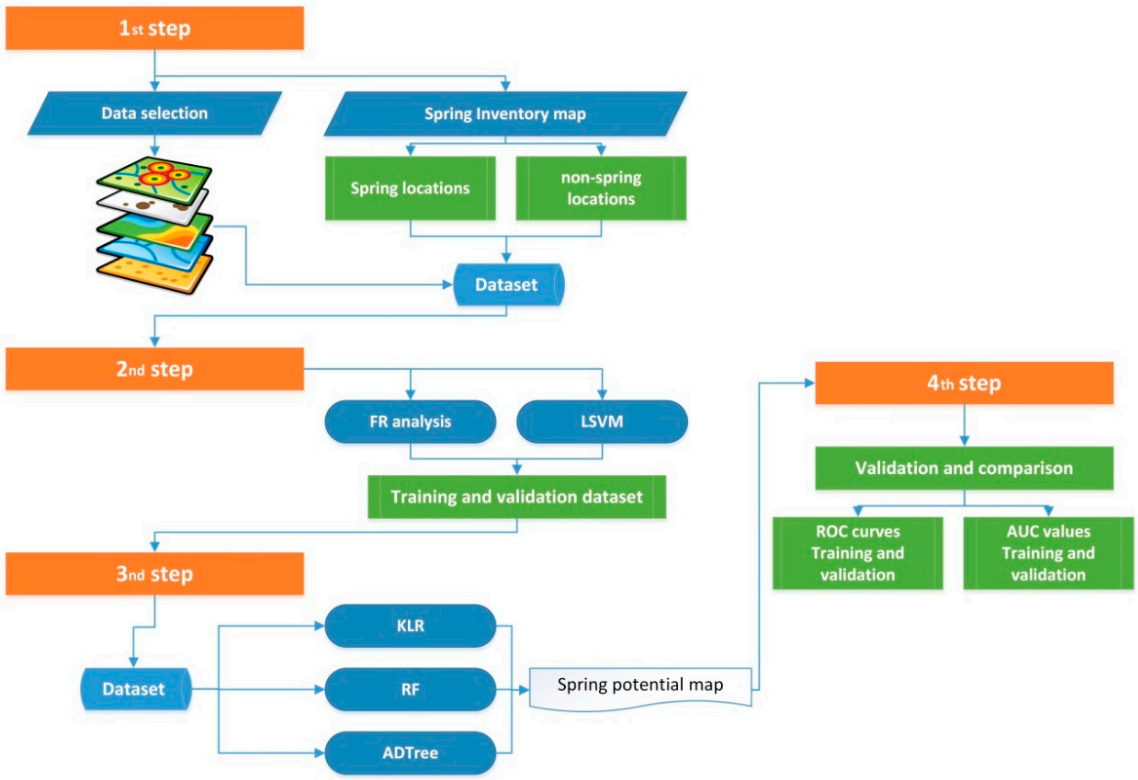


Figure 2. Flowchart of the developed methodology.

3.1. Frequency Ratio (FR)

The FR model, as a popular and efficient bivariate statistical techniques, is mainly used to estimate the potential probabilistic relation between dependent and independent factors but also for the potential

relation of multi-classified maps [35]. According to the FR model, the following formula (Equation (1)) was used to calculate the FR values for classes of the groundwater spring conditioning factors:

$$FR_{ij} = \frac{\frac{Spr_{ij}}{Sur_{ij}}}{\frac{Spr_T}{Sur_T}} \quad (1)$$

where FR_{ij} is the frequency ratio of a ith class for the jth factor, Spr_{ij} is the number of pixels with spring pixels in the ith class area of the jth factor, Spr_T is the total number of springs, Sur_{ij} is the number of pixels in the ith class area of the jth factor, and Sur_T is the total number of pixels.

3.2. Selection of Spring Explanatory Factors Using an SVM Classifier

The quality of groundwater spring potential mapping is influenced by the quality and quantity of the input data and also the predictive models that were used [15]. It is well known that groundwater spring explanatory factors may have unequal predictive capability in groundwater spring potential modeling. Therefore, groundwater spring explanatory factors that are characterized by negligible predictive capability should be not included in the analysis as they may produce less accurate results. Feature selection methods and specifically the gain ratio [46] and information gain ratio [47] are mainly used for estimating the predictive capability, however, in our case, the linear support vector machine (LSVM) method was used [48]. The determination of the contributions of the 14 groundwater spring explanatory factors was carried out as follows (Equation (2)) [49,50]:

$$g(x) = sgn\left(w^T m + n\right) \quad (2)$$

where $m = (m_1, m_2, m_3, \ldots, m_{12})$ is the input vector, w^T is the inverse matrix, and n is the offset from the origin of the hyper-plane.

3.3. Kernel Logistic Regression (KLR)

Kernel logistic regression is considered as a powerful discriminative method, described as the kernel version of logistic regression capable of transferring into a high-dimensional feature space the original input feature space by using kernel functions [51]. The following kernel function (Equation (3)) is the basic function in which φ is assumed to be unknown:

$$K(x, x') = \varphi(x)^T \varphi(x') \quad (3)$$

where T is the inner product in the Z space.

The training dataset has n vector input samples (x_i, Y_i) with xi belonging to R^n and Y_i belonging to $\{-1, 1\}$, where x_i is the ith input vector sample and Y is the target value. For $Y_i = 1$, the x_i is characterized as class 1, whereas for $Y_i = -1$, x_i is characterized as class 2. Let $Z_i = \varphi(x_i)$. Hence, the kernel-based method will solve the following optimization problem (Equation (4)):

$$\min_{k,b} E = \frac{1}{2}\|k\|^2 + C\sum_i g(-y_i(k * z_i - b) \quad (4)$$

where C corresponds to a regularization parameter, the optimal value of which is estimated by using techniques such as cross validation or a grid search technique. For the KLR function, the g is estimated by the following (Equation (5)):

$$g(\lambda) = log\left(1 + e^{\lambda}\right) \quad (5)$$

The goal of KLR is to estimate a discrimination function that distinguishes the two classes perfectly, in our case, spring from non-spring areas:

$$\log(p) = \log\left(\frac{p}{1-p}\right) = \sum_{i=1}^{n} a_i K\left(x_i, x_j\right) + b \quad (6)$$

where p is the logistic function with values ranging between −1 and 1, $K(x_i, x_j)$ is the kernel function that takes into consideration the Mercer's condition [52], α_i is a vector of dual parameters and b is the intercept.

Several kernel functions can be used, such as the linear kernel, the polynomial, and normalized polynomial [53]. However, in our case, the radial basis function (RDF) kernel was considered to be carried out:

$$K\left(x_i, x_j\right) = exp\frac{-x_i - x_j{}^2}{2\sigma^2} \quad (7)$$

where σ is a tuning parameter.

3.4. Random Forest (RF)

RF is an ensemble method of binary decision trees that are trained separately, and it is appropriate for classification and regression problems [54]. The fundamental approach used for classification problems by RF is based on training separately each decision tree, whereas the final outcome is estimated by taking into account the results obtained by each decision tree [55].

RF models have the ability to generalize and minimize the risk of over fitting, without having to undergo any pruning process. The training involves creating a number of different bootstrap samples from the original dataset, with one-third being left out of the process to act as test cases and based on this test cases to estimate an unbiased test error, referred to as the out-of-bag-error, that expresses the predictive ability of the RF model [56].

3.5. Alternating Decision Tree (ADTree)

ADTree is a combination of a decision tree and boosting techniques that generates classification rules with less nodes, is easier to explain, and provides a measure of confidence that is called the classification margin [57].

ADTree are similar to the option trees first described by Buntine [58] and further developed by Kohavi [59]. Compared to a single decision tree, option trees achieve a significant improvement in classification error. The ADTree's structure is similar to option trees since they also use a boosting technique and achieve better performance levels [60]. Because of the boosting iteration process, which adds three more nodes (one splitter node and two prediction nodes) to the tree, more boosting iterations will produce larger and more accurate trees. Different from original decision trees, ADTrees perform classification for a sample by mapping all possible paths for which all decision nodes are true, while summing up any prediction nodes that are traversed. In the case of unknown feature values, the ADTree algorithm only considers the reachable decision nodes. That is the reason why the ADTree algorithm can be applied widely in classification.

3.6. Validation and Comparison of the Results Obtained by the Models

The validation of the success and predictive performance of the three models was performed based on the receiver operating characteristic (ROC) curves [61–65]. The estimated AUC values range between 0.50 and 1.00 and can be classified based on a quantitative–qualitative classification scheme as follows: 0.5–0.6 (poor), 0.6–0.7 (average), 0.7–0.8 (good), 0.8–0.9 (very good), and 0.9–1 (excellent) [66]. In addition to the AUC values, two evaluation statistics, namely standard error (SE) and confidence interval (CI) at 95%, were also estimated. The best model has the smallest standard error, and the narrowest CI [67,68].

4. Data Used

A crucial aspect in groundwater spring potential mapping process is to identify spring locations. Based on extensive field surveys conducted during 2006–2017, 66 springs were detected in the study area (Figures 1 and 3a,b). An equal number of 66 nonspring locations were randomly selected from the free of spring's space by applying the Create Random Points function found in the Data Management Tools in the ArcGIS platform [69]. The spring and nonspring locations were randomly divided into two subsets, by using the Subset tool in the Geo statistical extension package of the ArcGIS platform [69]. The first subsets consisted of 46 spring and 46 nonspring locations, 70% of the total number of springs and nonspring areas and were used for training, whereas the second subset consisted of the remaining 30% (20 spring and 20 nonspring locations) and were used for validation.



Figure 3. (**a**,**b**) Process of identifying spring locations in the field.

Generally, several spring explanatory factors may influence spring occurrence, however, there are no guidelines for the selection of spring explanatory factors. Therefore, in the present study and based on the experienced gained from previous studies, 14 spring explanatory factors, including slope aspect, slope angle, plan curvature, profile curvature, elevation, stream power index (SPI), sediment transport index (STI), topographic wetness index (TWI), distance to streams, distance to roads, normalized difference vegetation index (NDVI), lithology, soil, and land use, were selected and prepared for further analysis within a GIS environment [43,70,71]. Eight geomorphometric factors, including slope aspect, slope angle, plan curvature, profile curvature, elevation, SPI, STI, and TWI, were extracted from the ASTER GDEM version 2 sensor (http://www.jspacesystems.or.jp/ersdac/GDEM/E/index.html) with a resolution of 30 m. These spring explanatory factors were reclassified into categories (Table 2) based on the outcomes of frequency analysis concerning spring occurrence and also characteristics of the study area. The distance-to-streams and distance-to-roads maps were produced using the topographic maps at 1:10,000-scale. The NVDI was calculated using Landsat 8 OLI (path/row 126–33) obtained on 4 November 2017 (available at http://www.gscloud.cn). A lithological map was extracted from the geological map at a scale of 1:10,000 and constructed with nine classes based on lithological similarities [43,72]. The soil types were extracted from soil maps at 1:1,000,000-scale in the study area and were classified into four classes [43,73]. In addition, the land use map was extracted from land use maps at 1:100,000-scale with six land use types based on the supervised classification method and maximum likelihood algorithm [19]. All the spring explanatory factors were finally converted into the same spatial resolution of 30×30 m^2 (Figure 4).

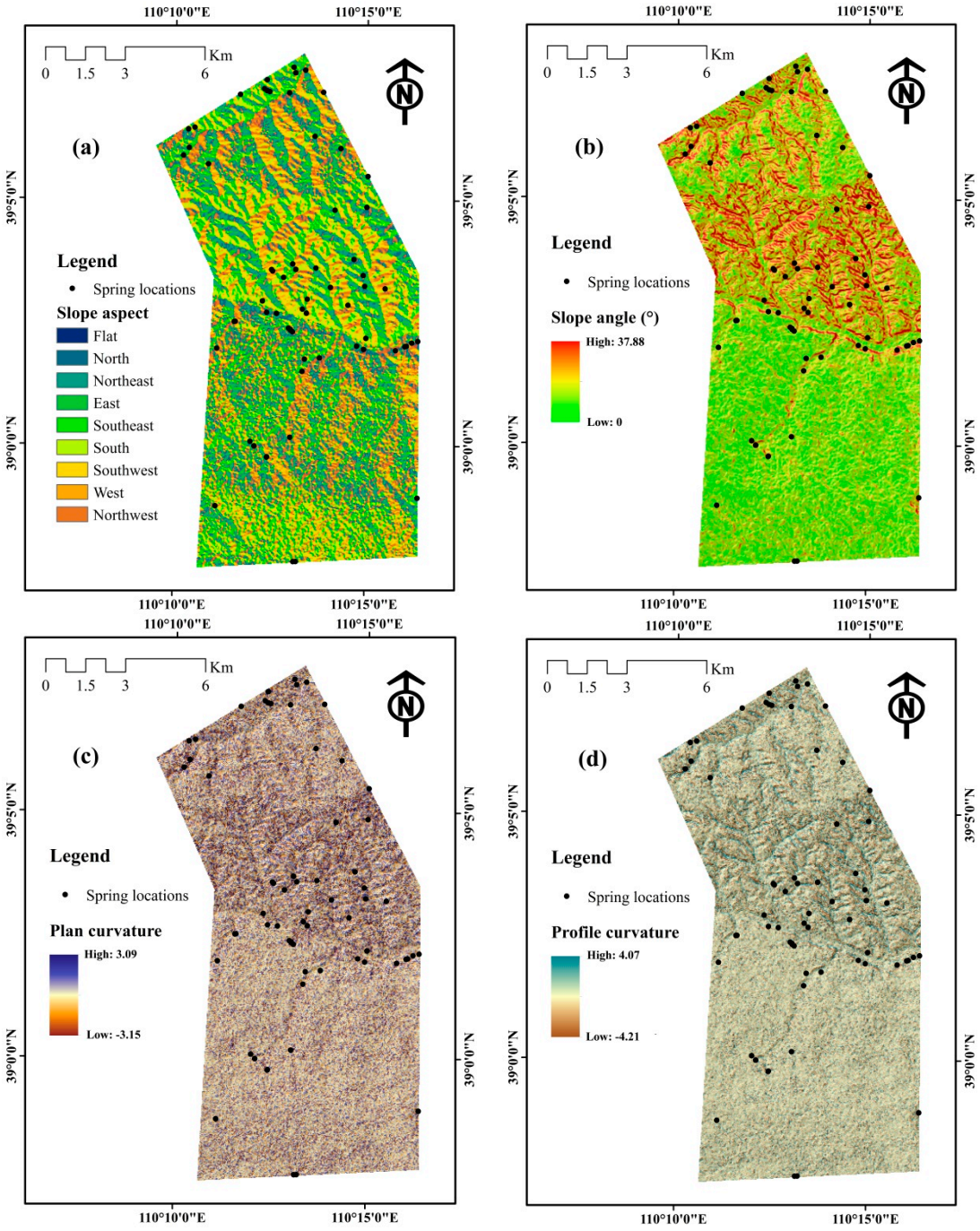


Figure 4. *Cont.*

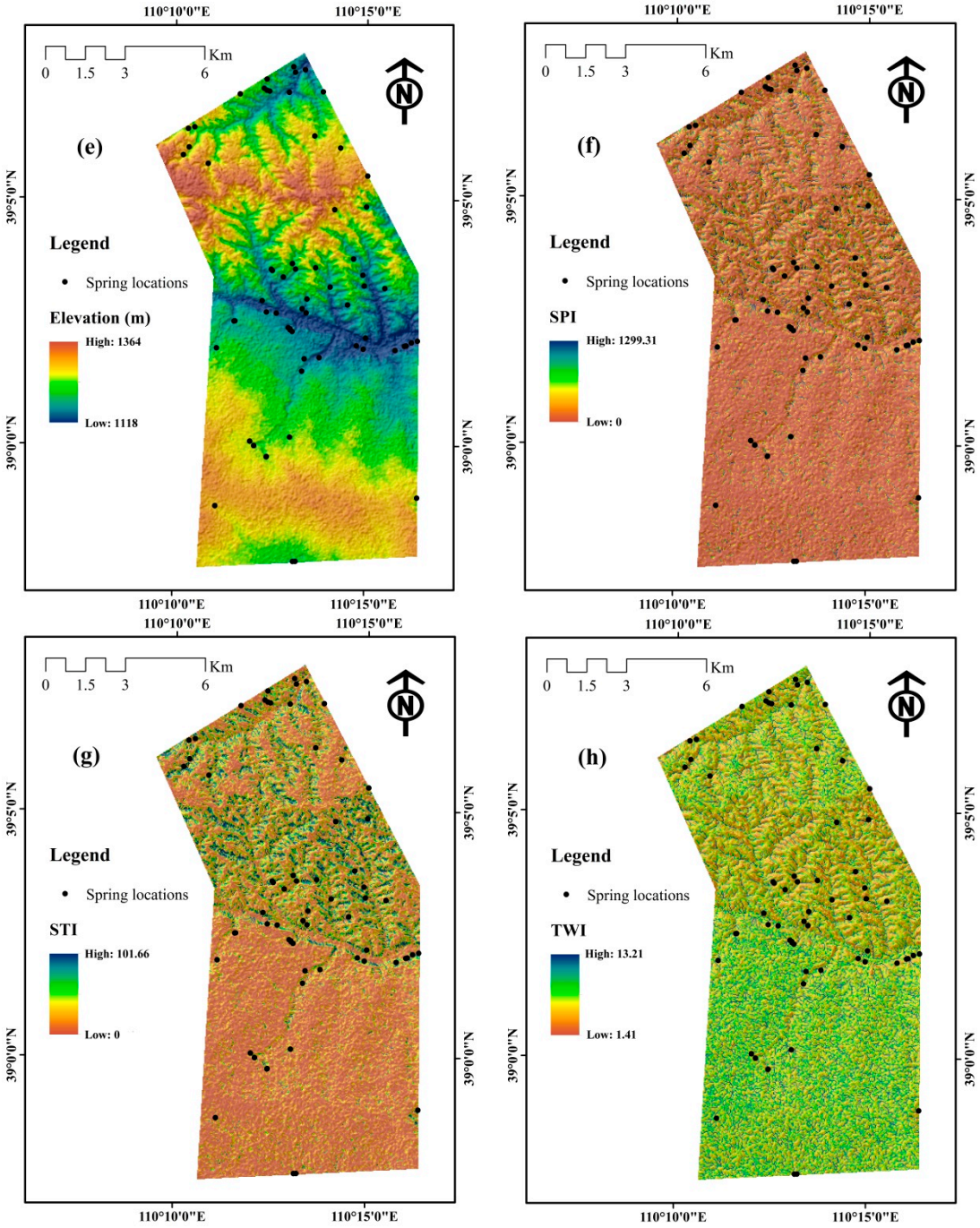


Figure 4. *Cont.*

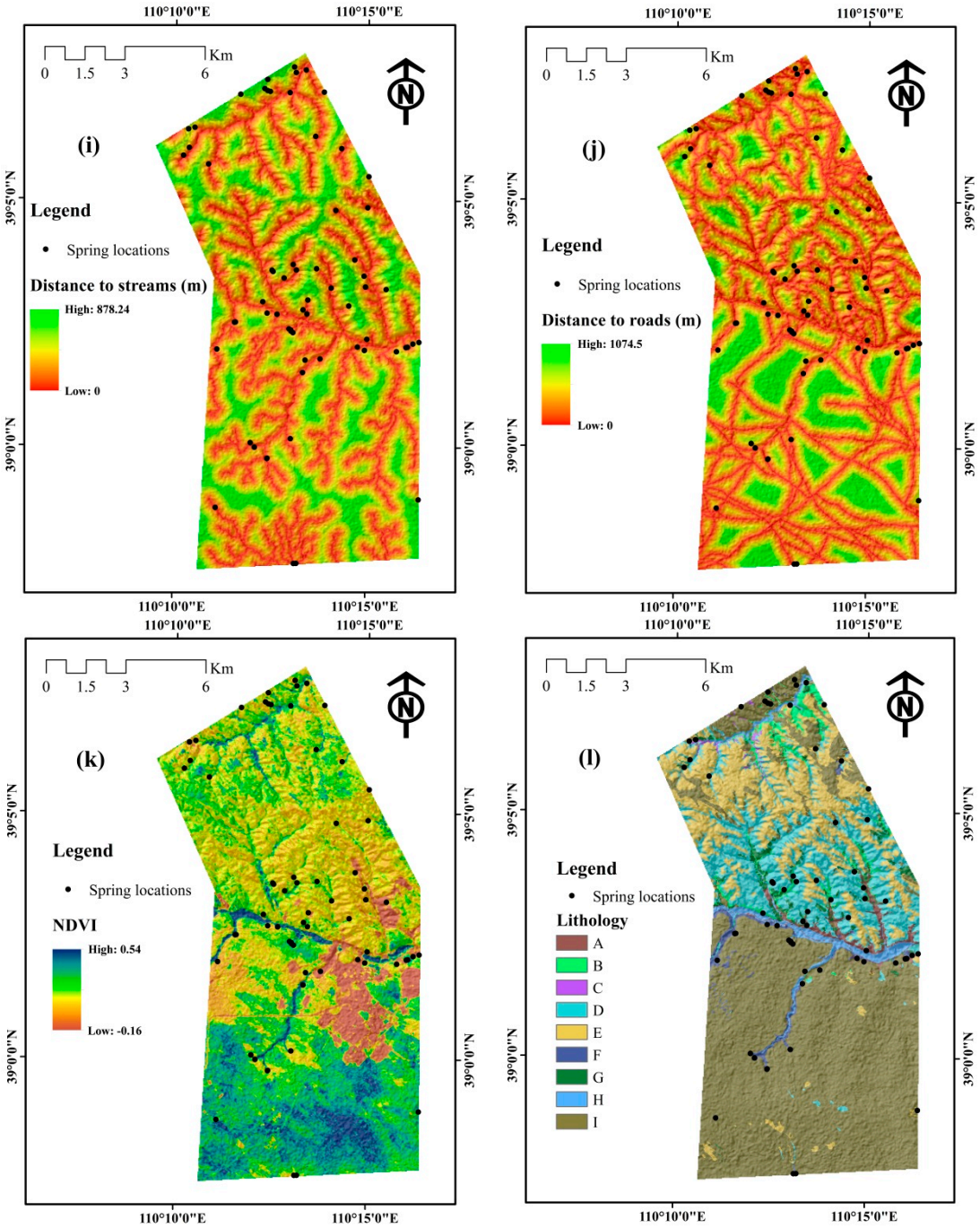


Figure 4. *Cont.*

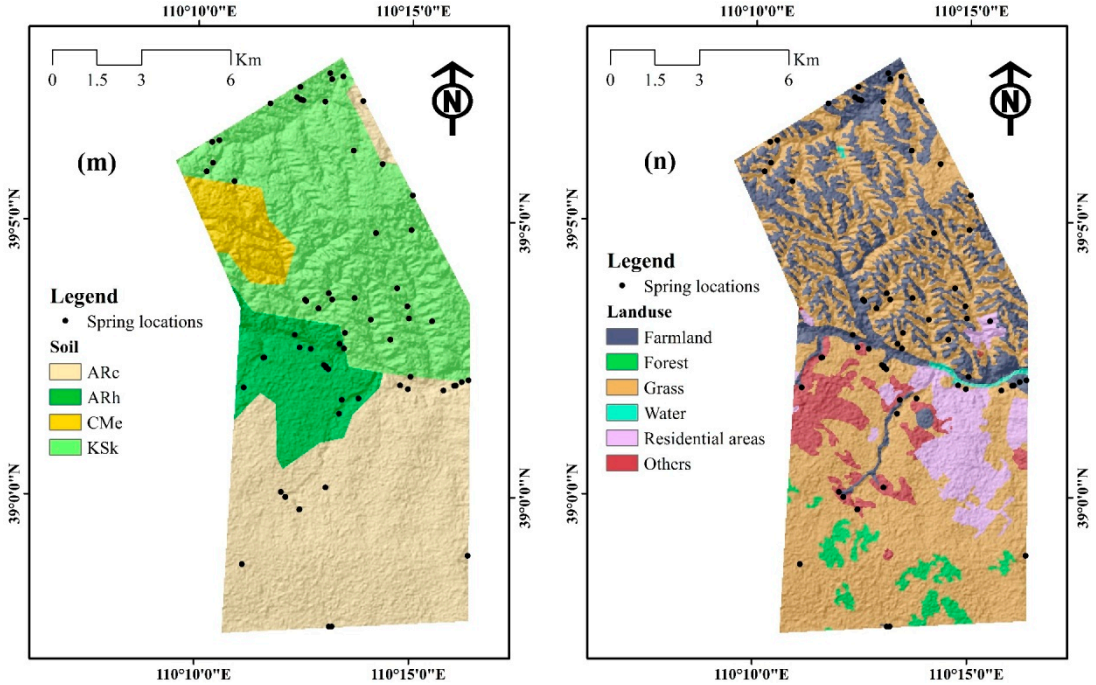


Figure 4. Thematic maps: (**a**) slope aspect; (**b**) slope angle; (**c**) plan curvature; (**d**) profile curvature; (**e**) elevation; (**f**) stream power index (SPI); (**g**) sediment transport index (STI); (**h**) topographic wetness index (TWI); (**i**) distance to streams; (**j**) distance to roads; (**k**) normalized difference vegetation index (NDVI); (**l**) lithology; (**m**) soil; (**n**) land use.

5. Results

5.1. Results of Explanatory Factors Selection

Table 2 illustrates the average merit (AM) values of the 14 spring explanatory factors based on the LSVM algorithm classifier using a 10-fold cross-validation method [53]. The results of the performed LSVM analysis revealed that lithology had the highest predictive power (14.0), followed by elevation (12.8), SPI (12.2), and soil cover (10.6), thus being the most significant factors that contribute to the predictive performance of a model. Since, all spring explanatory factors appear to have a positive predictive value, none of them were excluded from the analysis that followed.

Table 2. Predictive capabilities of spring explanatory factors using the LSVM algorithm. AM: average merit.

No.	Explanatory Factors	AM	Standard Deviation
1	Lithology	14.0	±0.000
2	Elevation	12.8	±0.400
3	SPI	12.2	±0.400
4	Soil cover	10.6	±0.490
5	Distance to roads	9.9	±0.831
6	Slope aspect	9.3	±0.900
7	TWI	7.4	±1.020
8	Slope angle	6.2	±1.327
9	STI	5.6	±1.685
10	Land use	4.9	±1.758
11	Distance to streams	4.7	±1.487
12	Profile curvature	3.0	±1.789
13	Plan curvature	2.5	±0.671
14	NDVI	1.9	±1.814

5.2. Correlation Analysis between Springs and Explanatory Factors Using FR

The correlation between groundwater springs and explanatory factors using the FR values is illustrated in Table 3. Based on the results, springs are found more frequently in southeast-facing (1.989) and south-facing (1.338) slopes. Flat slopes, with no springs occurrence, have the lowest FR value (0.000). For slope angle, FR values increase with the increasing slope angles and then decrease when slope angles are larger than 25° in the study area, and the class of 20–25 has the highest FR value (1.956), followed by the classes of 15–20 (1.955), 10–15 (1.296), 5–10 (1.248), and <5 (0.483). In the case of plan curvature, the FR value is the highest for the class of −0.77 to −0.28 (2.084), followed by the classes of 0.60–3.09 (1.200), 0.11–0.60 (0.989), −3.15 to −0.77 (0.856), and −0.28–0.11 (0.419). The FR values concerning the profile curvature indicate a decrease with the increasing profile curvature values. However, when the profile curvature values are higher than -0.22, the FR values show an increase. The class of 0.72–4.07 has the highest FR value (2.778). For elevation, the FR values decrease with increasing elevation. The class that is characterized by elevation greater than 1150 m has the highest FR value (6.914), followed by the class of 1150–1200 (2.592). Concerning SPI, the FR values show an increasing trend with the increasing SPI values. The class of 30–40 shows the highest FR value (4.524), whereas the class of 10–20 shows the lowest FR value (0.531). For STI, the class of >8 exhibits the highest FR value (2.246). In the case of TWI, the FR values decrease and then increase with the increasing TWI values. The class of <2 shows the highest FR value (2.429), followed by the class of >3.5 (1.539). For the factor distance to rivers, the class 150–200 m was found to have the highest FR value (1.759). For the factor distance to roads, the FR values indicate that the class of 150–200 m (2.226) has the strong influence of road proximity to spring occurrence. For NDVI, FR values are relatively equal for different NDVI classes except for the class of 0.19–0.26. The highest FR values are estimated for the classes of −0.16–0.04 (1.296) and 0.26–0.54 (1.263). Concerning lithology, the FR values are estimated to be the highest for the lithology groups of F (10.155), B (7.227), C (6.656), and A (5.793). In the case of soil, the ARh soil exhibits the highest FR value (1.902), followed by KSk soil (1.473). Finally, for the land use factor, the highest FR value (0.330) is estimated for the land use type of others (1.373), grassland (1.104), and farmland (0.984).

Table 3. Spatial relationship between springs and factors by FR model.

Explanatory Factors	Classes	No. of Pixels in Domain	No. of Springs	FR
Slope aspect	Flat	160	0	0.000
	North	20,781	7	1.110
	Northeast	23,407	3	0.422
	East	18,904	5	0.871
	Southeast	16,567	10	1.989
	South	19,695	8	1.338
	Southwest	19,789	6	0.999
	West	15,924	3	0.621
	Northwest	16,327	4	0.807
Slope angle (°)	<5	61,434	9	0.483
	5–10	52,808	20	1.248
	10–15	22,886	9	1.296
	15–20	10,114	6	1.955
	20–25	3368	2	1.956
	25–30	797	0	0.000
	>30	147	0	0.000
Plan curvature	−3.15 to −0.77	7701	2	0.856
	−0.77 to −0.28	28,451	18	2.084
	−0.28–0.11	55,026	7	0.419
	0.11–0.60	46,644	14	0.989
	0.60–3.09	13,732	5	1.200

Table 3. *Cont.*

Explanatory Factors	Classes	No. of Pixels in Domain	No. of Springs	FR
Profile curvature	−4.21 to −0.74	9855	4	1.337
	−0.74 to −0.22	37,272	8	0.707
	−0.22–0.20	56,394	14	0.818
	0.20–0.72	37,360	11	0.970
	0.72–4.07	10,673	9	2.778
Elevation (m)	<1150	1906	4	6.914
	1150–1200	13,983	11	2.592
	1200–1250	46,890	22	1.546
	1250–1300	61,507	9	0.482
	>1300	27,268	0	0.000
SPI	<10	112,248	29	0.851
	10–20	18,612	3	0.531
	20–30	7174	2	0.918
	30–40	3641	5	4.524
	>40	9879	7	2.335
STI	<2	83,093	16	0.634
	2–4	31,339	12	1.262
	4–6	14,344	4	0.919
	6–8	8108	4	1.625
	>8	14,670	10	2.246
TWI	<2	13,564	10	2.429
	2–2.5	62,484	10	0.527
	2.5–3	43,231	13	0.991
	3–3.5	17,294	6	1.143
	>3.5	14,981	7	1.539
Distance to streams (m)	<50	22,376	10	1.472
	50–100	19,583	10	1.682
	100–150	18,591	6	1.063
	150–200	13,111	7	1.759
	>200	77,893	13	0.550
Distance to roads (m)	<50	36,659	14	1.258
	50–100	26,995	11	1.343
	100–150	22,234	6	0.889
	150–200	13,323	9	2.226
	>200	52,343	6	0.378
NDVI	−0.16–0.04	10,165	4	1.296
	0.04–0.13	50,750	18	1.169
	0.13–0.19	46,711	17	1.199
	0.19–0.26	33,496	3	0.295
	0.26–0.54	10,432	4	1.263
Lithology	A	2275	4	5.793
	B	4103	9	7.227
	C	495	1	6.656
	D	24,980	8	1.055
	E	26,024	4	0.506
	F	2271	7	10.155
	G	2895	3	3.414
	H	3653	1	0.902
	I	84,858	9	0.349

Table 3. *Cont.*

Explanatory Factors	Classes	No. of Pixels in Domain	No. of Springs	FR
Soil	Calcari-Gypsiric Arenosols (Arc)	67,657	10	0.487
	Haplic Arenosols (ARh)	17,325	10	1.902
	Calcareous red clay (CMe)	8418	0	0.000
	Luvi-Calcic Kastanozems (KSk)	58,154	26	1.473
Land use	Farmland	33,485	10	0.984
	Forest	5148	0	0.000
	Grass	89,545	30	1.104
	Water	743	0	0.000
	Residential areas	13,036	2	0.505
	Others	9,597	4	1.373

5.3. Application of KLR, RF, and ADTree Models

Figure 5 illustrates the spring potential map constructed by the KLR method. Based on the visual inspection of the produced spring potential maps, the occurrence of spring appears to follow the spatial distribution of elevation and the factor distance to streams. The high and very high potential groundwater spring zones cover mainly the central and north areas, whereas the south area exhibits low to very low values. The high spring potential class was estimated to cover 5.02% of the study area, whereas low and very low spring potential classes cover 77.71% of the area (Table 4).

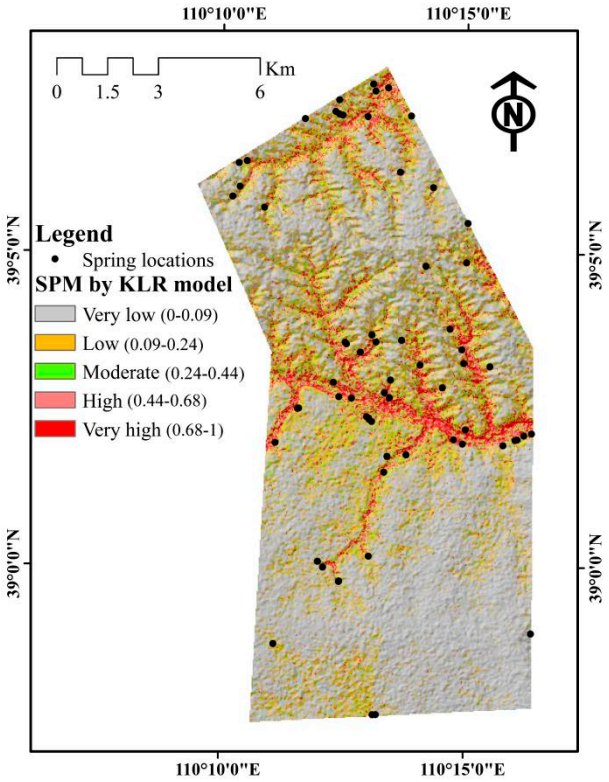


Figure 5. Groundwater spring potential map by the KLR model.

Table 4. Percentages of groundwater spring-potential classes.

Classes	KLR (%)	RF (%)	ADTree (%)
Very low	58.80	46.90	76.66
Low	18.91	21.57	5.90
Moderate	10.10	13.77	4.04
High	7.18	10.76	4.25
Very High	5.02	7.01	9.15

To enhance the performance of the RF method, a tuning process that is based on the grid search method was necessary [74]. The results of the tuning process indicated the optimal parameters to be for ntree 1500 trees and for the mtry parameter of 11. The implementation of RF also provided some extra information concerning the importance of the spring explanatory factors on the overall spring potential mapping. This was achieved by calculating the mean decrease accuracy and the mean decrease Gini [75] (Figure 6). Higher values for both measures indicate that the factor is relatively more significant [76]. According, to those two metrics, the most important factor was lithology, followed by elevation.

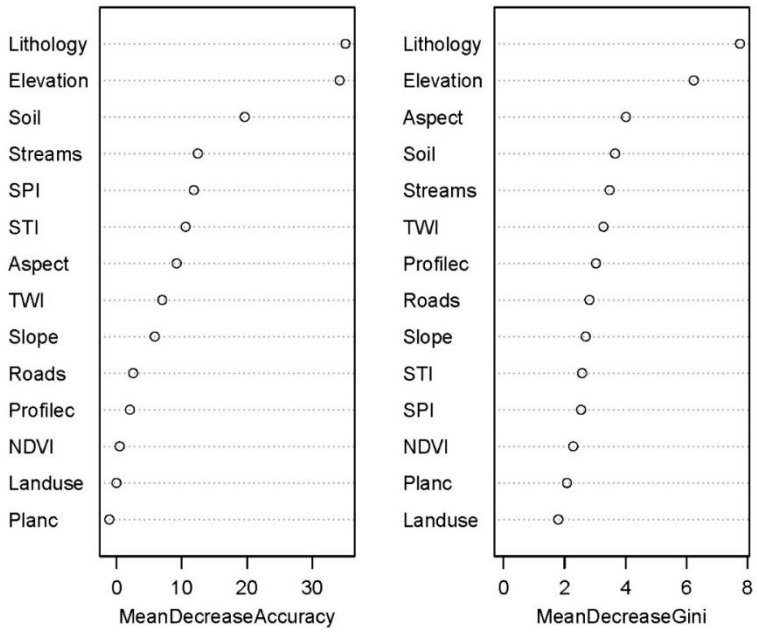


Figure 6. Mean decrease accuracy and mean decrease Gini (Streams: Distance to streams; Roads: Distance to roads; Planc: Plan curvature; Profilec: Profile curvature).

Figure 7 illustrates the groundwater spring potential map constructed by the RF model. Based on the visual inspection of the spring potential map, it could be concluded that spring occurrence follows in this case the spatial pattern of the stream network, with high and very high potential zones covering mainly the central area, whereas the south area illustrates low to very low values. The high spring potential class covers 7.01% of the area, whereas low and very low spring potential classes cover 68.47% of the area (Table 4).

Figure 8 shows the spring potential map constructed by the ADTree model based on the natural break method [77,78]. Compared to the previous methods, the ADTree method provides a rather different spatial distribution. The high spring potential class covers 9.15% of the area, whereas low and very low spring potential classes cover 82.56% of the area. It seems that the ADTree method distinguishes with clarity the potential nonspring and spring areas compared to the other two methods.

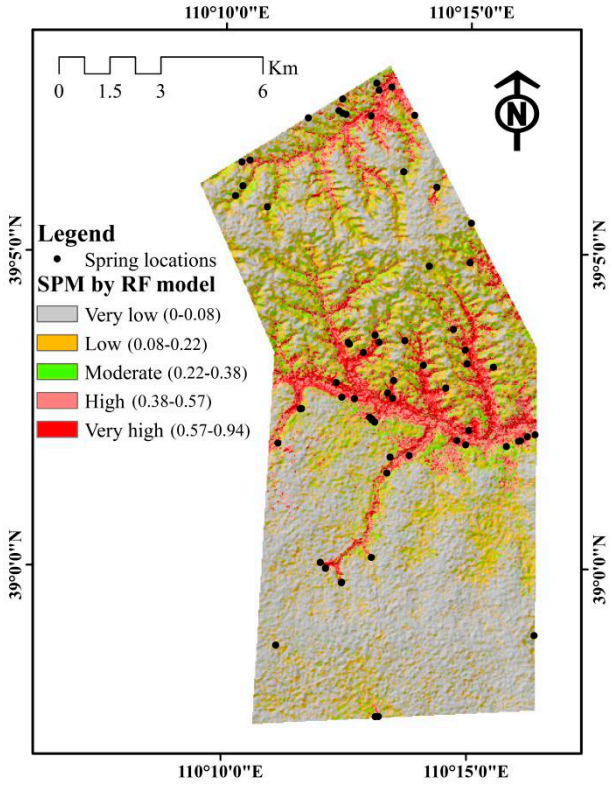


Figure 7. Groundwater spring potential map by the RF model.

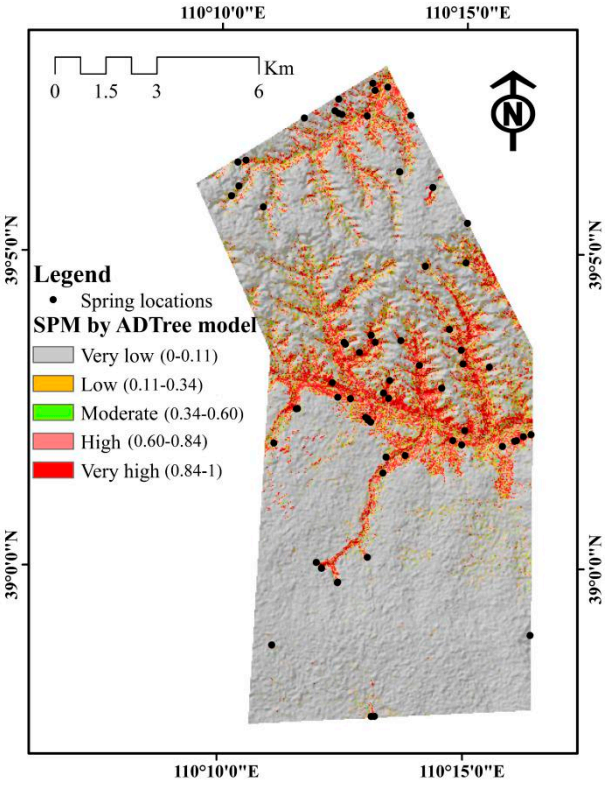


Figure 8. Groundwater spring potential map by the ADTree model.

5.4. Validation and Comparison

Figure 9a,b illustrates the ROC plot assessment results based on the training and validation subsets. The AUC value for the success rate curve using the RF model was estimated to be 0.909, which corresponds to a prediction accuracy of 90.90%, followed by the KLR (0.877) and ADTree (0.812) models.

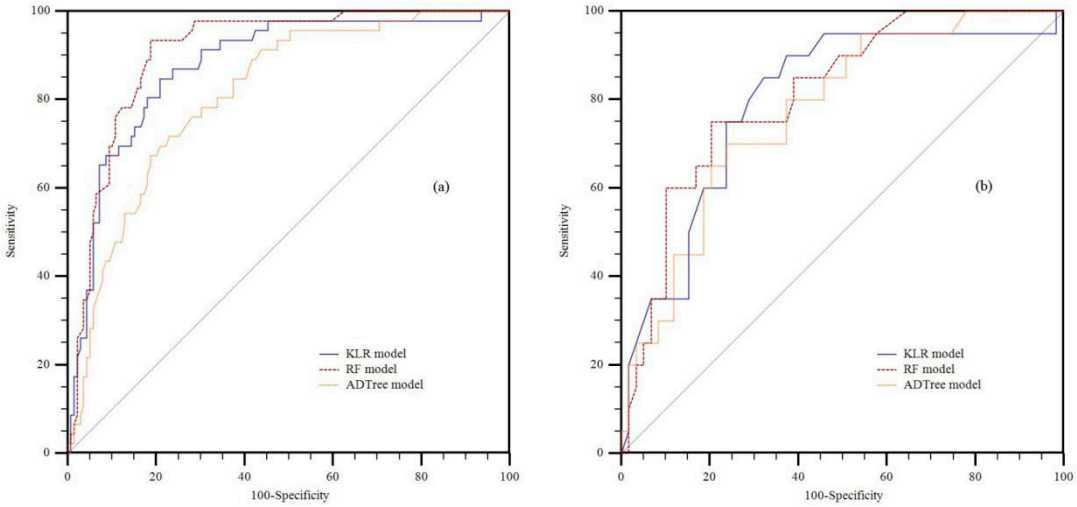


Figure 9. ROC curves of the three models using a (**a**) training dataset and (**b**) validation dataset.

RF showed the lowest SE value (0.0225), followed by KLR (0.0294) and ADTree (0.0341), and also the shorter CI value (0.088) followed by KLR (0.1) and ADTree (0.118) (Table 5). Similar performance patterns were estimated when using the validation subset, with the AUC value for the predictive rate curve using the RF model estimated at 0.811, followed by the KLR (0.797) and ADTree (0.773) models. Again, the RF model showed the lowest SE (0.0526), followed by ADTree (0.0578) and KLR (0.0591). As for the CI values, RF showed the shortest interval (0.183), followed by KLR (0.188) and ADTree (0.195) (Table 6). Based on the validation analysis, all three models appear to provide good accuracy, with the RF model producing slightly better results in term of AUC values, low SE, and short CI values for both the training and validation subsets. Concerning the performance of KLR based on the training subset, it was found that it provides results relatively close to the RF model AUC, SE, and CI values.

Table 5. Parameters of AUC values using training dataset. SE: standard error; CI: confidence interval.

Models	AUC	SE	95% CI
KLR	0.877	0.0294	0.821 to 0.921
RF	0.909	0.0225	0.858 to 0.946
ADTree	0.812	0.0341	0.748 to 0.866

Table 6. Parameters of AUC values using validation dataset.

Models	AUC	SE	95% CI
KLR	0.797	0.0591	0.691 to 0.879
RF	0.811	0.0526	0.707 to 0.890
ADTree	0.773	0.0578	0.665 to 0.860

6. Discussion

As several studies report, the significance and predictive power of spring related factors that are used in groundwater spring potential assessments are controlled by the geological, morphological, hydrological, and climatic settings of the area [19,22,79–81]. According to Ozdemir [35], topographic

features, such as elevation and slope, have a negative influence with groundwater spring potential, and on the other hand TWI and drainage density have a positive influence. Similar studies, report that topographic features along with the characteristics of the soil cover, tectonic features (fault density and distance to faults), and also hydrological features (drainage density) influence the rainfall-runoff rate and also the infiltration rate, thus possibly affecting the groundwater spring potential occurrence [19,35]. Chen et al. [43] reported that lithology, elevation, and distance to streams had a greater influence, whereas land use, NDVI, plan, and profile curvature appear to have the least influence.

During the present study, the implementation of LSVM revealed that lithology had the highest predictive power, followed by elevation, SPI, and soil cover. Concerning the lithology factor, lithological and structural differences lead to variations in the durability and permeability of rock and soil formations, and thus the presence of springs [35]. Based on the FR analysis, groundwater springs are more probable to be found in southeast facing slopes, in areas with slopes angles ranging between 15 to 25 degrees and elevation lower than 1150 m. Concerning slope angles, the outcomes of the study are persistent with previous studies that report that areas with slopes greater than 35° are considered unfavorable since as the slope increases so too does runoff, having as a result reduced infiltration rates [82,83]. Moreover, the most spring-probable areas are covered by Haplic Arenosols (ARh) soils, which are coarsely textured sandy soils, permeable to water, and Calcic Kastanozems (KSk) soils, which are characterized by a rather restricted water transmission with higher portions of clay particles. According to Srivastava and Bhattacharya [84], sandy soils and coarse sandy clays appear as potential favorable storage bodies due to their light texture and excellent rate of infiltration, which is persistent with the findings of our study.

Within the research area, sand, mudstone, and sandstone formations appear to be more likely to contain springs. Similar findings were found by the authors in a previous study concerning the area of research [43]. Mudstone layers, which could be defined as formations with very low infiltration capacity, form an impermeable layer while sand and sandstone formations act as permeable layers allowing the concentration of surface water within their mass. The alternation of these layers permits the formation of groundwater springs as can be found in the area of research.

An interesting point that should be mentioned is the high predictive value of the factor distance to roads. The distance-from-road network is considered to have an influence on the occurrence of groundwater springs since its presence can cause local hydrological and erosion issues while affecting indirectly the groundwater table [85]. Also, the presence of a road may influence the amount of soil moisture but also the infiltration rate as a result of the removal of geological formations and the disturbance of the surface during of the construction phase [43,85].

Concerning the validation and comparison of the three models (KLR, RF, and ADTree), the RF model appears to provide slightly higher AUC values, lower SE values, and shorter CI intervals than the other two methods. Several studies have indicated that RF models have higher accuracy, compared to other models. According to Naghibi et al. [12], who applied support vector machine (SVM), RF, and genetic algorithm optimized RF (RFGA) methods to assess groundwater potential by spring locations, RF and optimized RF models outperformed SVM models. According to Golkarian et al. (2018) [86], this could be attributed to the methodological approach they followed, which involves aggregating the outcomes of many decision trees in order to limit overfitting effects as well as to limit error due to bias and error due to variance, thus producing more accurate predictions. However, other studies report that the performance of RF models could be influenced by the presence of datasets with noisy data and by the presence of data that includes categorical variables with different numbers of levels where, in such a case, RF models are biased in favor of those variables that appear with more levels [36]. In the present study, KLR gave more accurate results than those from the ADTree model. In similar studies concerning landslide susceptibility assessments, which implemented KLR and ADTree methods, it was found that KLR produced more balanced results for the training and validation datasets in terms of the statistical index, while the ADTree models showed significant variance [74]. Finally, although the presented models appear to have satisfactory predictive performance, it must be kept in mind that their results are influenced by the quality and quantity of the available input,

and also the identification of nonspring areas. Concerning future work, the presented approach could be applied to an area with different geo-environmental settings or include in the analysis dynamic variables, such as precipitation and temperature, that may vary over short timeframes, so as to estimate the efficiency of the proposed models.

7. Conclusions

In the present study, three artificial intelligence methods (KLR, RF, and ADTree) were utilized for the generation of a groundwater spring potential map for the Ningtiaota region, which is located in the northern territory of Shaanxi Province, China. A linear support vector machine method was used as a feature selection method so as to determine the optimal set of factors, which included fourteen explanatory factors (elevation, slope, aspect, plan curvature, profile curvature, stream power index, sediment transport index, topographic wetness index, distance to streams, distance to roads, NDVI, lithology, soil, and land use). The performed analysis highlighted the higher predictive power of the spring explanatory factors lithology, elevation, SPI, and soil cover. These four factors significantly influence the prediction accuracy. The comparison between the performances of KLR, RF, and ADTree models revealed that the RF model had higher prediction accuracy than the other two models, based on the results of higher values of AUC metric, lower SE values, and shorter CI intervals. The RF model's ability to limit overfitting effects may be the reason for its higher predictive performance. While remembering that the results obtained by tree-based artificial intelligence approaches could be influenced by the quality and quantity of data, overall they could be appreciated as accurate and reliable investigation tools in groundwater spring potential assessments.

Author Contributions: W.C., Y.L., P.T., H.S., I.I., W.X., and H.B. contributed equally to the work. W.C., Y.L., W.X., and H.B. collected field data and conducted the analysis. W.C., Y.L., W.X., and H.B. wrote the manuscript. P.T., H.S., and I.I., edited the manuscript. All the authors discussed the results and revised the manuscript. The authors specially wish to thank Enke Hou for useful information provided. All authors have read and agreed to the published version of the manuscript.

Funding: This study is jointly supported by the National Natural Science Foundation of China (Grant No. 41807192), Natural Science Basic Research Program of Shaanxi (Program No. 2019JLM-7, Program No. 2019JQ-094), China Postdoctoral Science Foundation (Grant No. 2018T111084, 2017M613168), Project funded by Shaanxi Province Postdoctoral Science Foundation (Grant No. 2017BSHYDZZ07), and the National Major Science and Technology Project (Grant No. 2017ZX05030-002).

Acknowledgments: The authors specially wish to thank Enke Hou for useful information provided.

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

References

1. Ayazi, M.H.; Pirasteh, S.; Arvin, A.; Pradhan, B.; Nikouravan, B.; Mansor, S. Disasters and risk reduction in groundwater: Zagros mountain southwest Iran using geoinformatics techniques. *Disaster Adv.* **2010**, *3*, 51–57.
2. Neshat, A.; Pradhan, B.; Pirasteh, S.; Shafri, H.Z.M. Estimating groundwater vulnerability to pollution using a modified drastic model in the Kerman agricultural area, Iran. *Environ. Earth Sci.* **2014**, *71*, 3119–3131. [CrossRef]
3. Gleeson, T.; Befus, K.M.; Jasechko, S.; Luijendijk, E.; Cardenas, M.B. The global volume and distribution of modern groundwater. *Nat. Geosci.* **2016**, *9*, 161. [CrossRef]
4. 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]
5. De Vries, J.J.; Simmers, I. Groundwater recharge: An overview of processes and challenges. *Hydrogeol. J.* **2002**, *10*, 5–17. [CrossRef]
6. Jackson, R.B.; Carpenter, S.R.; Dahm, C.N.; McKnight, D.M.; Naiman, R.J.; Postel, S.L.; Running, S.W. Water in a changing world. *Ecol. Appl.* **2001**, *11*, 1027–1045. [CrossRef]

7. Rosegrant, M.W.; Cai, X. Global water demand and supply projections: Part 2. Results and prospects to 2025. *Water Int.* **2002**, *27*, 170–182. [CrossRef]
8. Ercin, A.E.; Hoekstra, A.Y. Water footprint scenarios for 2050: A global analysis. *Environ. Int.* **2014**, *64*, 71–82. [CrossRef]
9. Kummu, M.; Guillaume, J.; de Moel, H.; Eisner, S.; Flörke, M.; Porkka, M.; Siebert, S.; Veldkamp, T.I.; Ward, P.J. The world's road to water scarcity: Shortage and stress in the 20th century and pathways towards sustainability. *Sci. Rep.* **2016**, *6*, 38495. [CrossRef] [PubMed]
10. Kaushal, S.; Gold, A.; Mayer, P. *Land Use, Climate, and Water Resources—Global Stages of Interaction*; Multidisciplinary Digital Publishing Institute: Basel, Switzerland, 2017.
11. Curran, S.R.; De Sherbinin, A. Completing the picture: The challenges of bringing "consumption" into the population–environment equation. *Popul. Environ.* **2004**, *26*, 107–131. [CrossRef]
12. Naghibi, S.A.; Dashtpagerdi, M.M. Evaluation of four supervised learning methods for groundwater spring potential mapping in Khalkhal region (Iran) using GIS-based features. *Hydrogeol. J.* **2017**, *25*, 169–189. [CrossRef]
13. Oh, H.-J.; Kim, Y.-S.; Choi, J.-K.; Park, E.; Lee, S. Gis mapping of regional probabilistic groundwater potential in the area of Pohang city, Korea. *J. Hydrol.* **2011**, *399*, 158–172. [CrossRef]
14. Udimal, T.B.; Jincai, Z.; Ayamba, E.C.; Owusu, S.M. China's water situation; the supply of water and the pattern of its usage. *Int. J. Sustain. Built Environ.* **2017**, *6*, 491–500. [CrossRef]
15. Chen, W.; Panahi, M.; Khosravi, K.; Pourghasemi, H.R.; Rezaie, F.; Parvinnezhad, D. Spatial prediction of groundwater potentiality using anfis ensembled with teaching-learning-based and biogeography-based optimization. *J. Hydrol.* **2019**, *572*, 435–448. [CrossRef]
16. 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] [PubMed]
17. Moghaddam, D.D.; Rezaei, M.; Pourghasemi, H.; Pourtaghie, Z.; Pradhan, B. Groundwater spring potential mapping using bivariate statistical model and GIS in the Taleghan watershed, Iran. *Arab. J. Geosci.* **2015**, *8*, 913–929. [CrossRef]
18. Corsini, A.; Cervi, F.; Ronchetti, F. Weight of evidence and artificial neural networks for potential groundwater spring mapping: An application to the Mt. Modino area (Northern Apennines, Italy). *Geomorphology* **2009**, *111*, 79–87. [CrossRef]
19. Naghibi, S.A.; Pourghasemi, H.R. A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. *Water Resour. Manag.* **2015**, *29*, 5217–5236. [CrossRef]
20. Chowdhury, A.; Jha, M.; Chowdary, V.; Mal, B. Integrated remote sensing and GIS-based approach for assessing groundwater potential in west Medinipur district, west Bengal, India. *Int. J. Remote Sens.* **2009**, *30*, 231–250. [CrossRef]
21. Jha, M.K.; Bongane, G.M.; Chowdary, V. Groundwater potential zoning by remote sensing, GIS and mcdm techniques: A case study of eastern India. In *Hydroinformatics in Hydrology, Hydrogeology and Water Resources, Proceedings of the Symposium JS. 4 at the Joint Convention of the International Association of Hydrological Sciences (IAHS) and the International Association of Hydrogeologists (IAH) held in Hyderabad, Hyderabad, India, 6–12 September 2009*; IAHS Press: Wallingford, UK, 2009; pp. 432–441.
22. Kumar, U.; Kumar, B.; Mallick, N. Groundwater prospects zonation based on RS and GIS using fuzzy algebra in Khoh river watershed, Pauri-Garhwal district, Uttarakhand, India. *Glob. Perspect. Geogr. (GPG)* **2013**, *1*, 37–45.
23. Pourtaghi, Z.S.; Pourghasemi, H.R. GIS-based groundwater spring potential assessment and mapping in the Birjand Township, Southern Khorasan province, Iran. *Hydrogeol. J.* **2014**, *22*, 643–662. [CrossRef]
24. Machiwal, D.; Jha, M.K.; Mal, B.C. Assessment of groundwater potential in a semi-arid region of India using remote sensing, GIS and MCDM techniques. *Water Resour. Manag.* **2011**, *25*, 1359–1386. [CrossRef]
25. Adiat, K.; Nawawi, M.; Abdullah, K. Assessing the accuracy of GIS-based elementary multi criteria decision analysis as a spatial prediction tool—A case of predicting potential zones of sustainable groundwater resources. *J. Hydrol.* **2012**, *440*, 75–89. [CrossRef]

26. Rahmati, O.; Samani, A.N.; Mahdavi, M.; Pourghasemi, H.R.; Zeinivand, H. Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arab. J. Geosci.* **2015**, *8*, 7059–7071. [CrossRef]
27. Shekhar, S.; Pandey, A.C. Delineation of groundwater potential zone in hard rock terrain of India using remote sensing, geographical information system (GIS) and analytic hierarchy process (AHP) techniques. *Geocarto Int.* **2015**, *30*, 402–421. [CrossRef]
28. Lee, S.; Kim, Y.-S.; Oh, H.-J. Application of a weights-of-evidence method and GIS to regional groundwater productivity potential mapping. *J. Environ. Manag.* **2012**, *96*, 91–105. [CrossRef]
29. Al-Abadi, A.M. Groundwater potential mapping at northeastern Wasit and Missan governorates, Iraq using a data-driven weights of evidence technique in framework of GIS. *Environ. Earth Sci.* **2015**, *74*, 1109–1124. [CrossRef]
30. Nampak, H.; Pradhan, B.; Manap, M.A. Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *J. Hydrol.* **2014**, *513*, 283–300. [CrossRef]
31. Mogaji, K.; Omosuyi, G.; Adelusi, A.; Lim, H. Application of GIS-based evidential belief function model to regional groundwater recharge potential zones mapping in Hardrock geologic terrain. *Environ. Process.* **2016**, *3*, 93–123. [CrossRef]
32. Tahmassebipoor, N.; Rahmati, O.; Noormohamadi, F.; Lee, S. Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. *Arab. J. Geosci.* **2016**, *9*, 79. [CrossRef]
33. Kordestani, M.D.; Naghibi, S.A.; Hashemi, H.; Ahmadi, K.; Kalantar, B.; Pradhan, B. Groundwater potential mapping using a novel data-mining ensemble model. *Hydrogeol. J.* **2019**, *27*, 211–224. [CrossRef]
34. Kim, K.-D.; Lee, S.; Oh, H.-J.; Choi, J.-K.; Won, J.-S. Assessment of ground subsidence hazard near an abandoned underground coal mine using GIS. *Environ. Geol.* **2006**, *50*, 1183–1191. [CrossRef]
35. Ozdemir, A. Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the sultan mountains (Aksehir, Turkey). *J. Hydrol.* **2011**, *405*, 123–136. [CrossRef]
36. Naghibi, S.A.; Pourghasemi, H.R.; Dixon, B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* **2016**, *188*, 44. [CrossRef] [PubMed]
37. Zabihi, M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Behzadfar, M. GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environ. Earth Sci.* **2016**, *75*, 665. [CrossRef]
38. Lee, S.; Song, K.-Y.; Kim, Y.; Park, I. Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model. *Hydrogeol. J.* **2012**, *20*, 1511–1527. [CrossRef]
39. Kim, J.-C.; Jung, H.-S.; Lee, S. Groundwater productivity potential mapping using frequency ratio and evidential belief function and artificial neural network models: Focus on topographic factors. *J. Hydroinform.* **2018**, *20*, 1436–1451. [CrossRef]
40. Lee, S.; Hong, S.-M.; Jung, H.-S. GIS-based groundwater potential mapping using artificial neural network and support vector machine models: The case of Boryeong city in Korea. *Geocarto Int.* **2018**, *33*, 847–861. [CrossRef]
41. Aguilera, P.A.; Fernández, A.; Ropero, R.F.; Molina, L. Groundwater quality assessment using data clustering based on hybrid Bayesian networks. *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 435–447. [CrossRef]
42. Naghibi, S.A.; Pourghasemi, H.R.; Abbaspour, K. A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theor. Appl. Climatol.* **2018**, *131*, 967–984. [CrossRef]
43. 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] [PubMed]
44. Khosravi, K.; Panahi, M.; Bui, D.T. Spatial prediction of groundwater spring potential mapping based on an adaptive neuro-fuzzy inference system and metaheuristic optimization. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 4771–4792. [CrossRef]
45. National Soil Survey Office. *Chinese Soil Types*; China Agricultural Press: Beijing, China, 1995.
46. Nithya, N.S.; Duraiswamy, K. Gain ratio based fuzzy weighted association rule mining classifier for medical diagnostic interface. *Sadhana* **2014**, *39*, 39–52. [CrossRef]

47. Bui, D.T.; Tuan, T.A.; Hoang, N.-D.; Thanh, N.Q.; Nguyen, D.B.; Van Liem, N.; Pradhan, B. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* **2017**, *14*, 447–458.
48. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
49. Pham, B.T.; Bui, D.T.; Pourghasemi, H.R.; Indra, P.; Dholakia, M.B. Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2015**, *128*, 1–19. [CrossRef]
50. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [CrossRef]
51. Tanaka, K.; Kurita, T.; Kawabe, T. Selection of import vectors via binary particle swarm optimization and cross-validation for kernel logistic regression. In Proceedings of the 2007 International Joint Conference on Neural Networks, Orlando, FL, USA, 12–17 August 2007; pp. 1037–1042.
52. Mercer, J. *Functions of Positive and Negative Type, and Their Connection with the Theory of Integral Equations*; Royal Society of London Philosophical Transactions: London, UK, 1909; Volume 209, pp. 415–446.
53. Bui, D.T.; 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.
54. Jo, T.; Cheng, J. Improving protein fold recognition by random forest. *BMC Bioinform.* **2014**, *15*, 1–7. [CrossRef]
55. Ravìa, D.; Bober, M.; Farinella, G.M.; Guarnera, M.; Battiato, S. Semantic segmentation of images exploiting dct based features and random forest. *J. Pain Palliat. Care Pharmacother.* **2015**, *24*, 429. [CrossRef]
56. Breiman, L. Random forest. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
57. Freund, Y.; Mason, L. The alternating decision tree learning algorithm. In Proceedings of the Sixteenth International Machine Learning Conference, Bled, Slovenia, 27–30 June 1999; Morgan Kaufmann: San Francisco, CA, USA, 2002; pp. 124–133.
58. Buntine, W. Learning classification trees. *Stat. Comput.* **1992**, *2*, 63–73. [CrossRef]
59. Kohavi, R.; Kunz, C. Option decision trees with majority votes. In Proceedings of the Fifteenth International Machine Learning Conference (ICML 1998), Madison, MI, USA, 24–27 July 1998.
60. Min, S.L.; Oh, S. Alternating decision tree algorithm for assessing protein interaction reliability. *Vietnam J. Comput. Sci.* **2014**, *1*, 169–178.
61. Hosseinalizadeh, M.; Kariminejad, N.; Chen, W.; Pourghasemi, H.R.; Alinejad, M.; Mohammadian Behbahani, A.; Tiefenbacher, J.P. Gully headcut susceptibility modeling using functional trees, naïve bayes tree, and random forest models. *Geoderma* **2019**, *342*, 1–11. [CrossRef]
62. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S.; et al. 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]
63. Chen, W.; Li, Y.; Xue, W.; Shahabi, H.; Li, S.; Hong, H.; Wang, X.; Bian, H.; Zhang, S.; Pradhan, B.; et al. 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]
64. Chen, W.; Fan, L.; Li, C.; Pham, B.T. Spatial prediction of landslides using hybrid integration of artificial intelligence algorithms with frequency ratio and index of entropy in Nanzheng county, China. *Appl. Sci.* **2020**, *10*, 29. [CrossRef]
65. 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]
66. Yesilnacar, E.; Topal, T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Eng. Geol.* **2005**, *79*, 251–266. [CrossRef]
67. 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]

68. Chen, W.; Sun, Z.; Han, J. Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. *Appl. Sci.* **2019**, *9*, 171. [CrossRef]
69. ESRI. *ArcGIS Desktop: Release 10.3.1*; Environmental Systems Research Institute: Redlands, CA, USA, 2015.
70. Cantonati, M.; Segadelli, S.; Ogata, K.; Tran, H.; Sanders, D.; Gerecke, R.; Rott, E.; Filippini, M.; Gargini, A.; Celico, F. A global review on ambient Limestone-Precipitating Springs (LPS): Hydrogeological setting, ecology, and conservation. *Sci. Total Environ.* **2016**, *568*, 624–637. [CrossRef] [PubMed]
71. Hou, E.; Wang, J.; Chen, W. A comparative study on groundwater spring potential analysis based on statistical index, index of entropy and certainty factors models. *Geocarto Int.* **2018**, *33*, 754–769. [CrossRef]
72. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [CrossRef]
73. Zeinivand, H.; Ghorbani Nejad, S. Application of GIS-based data-driven models for groundwater potential mapping in Kuhdasht region of Iran. *Geocarto Int.* **2018**, *33*, 651–666. [CrossRef]
74. Chen, W.; Xie, X.; Peng, J.; Wang, J.; Duan, Z.; Hong, H. GIS-based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models. *Geomat. Nat. Hazards Risk* **2017**, *8*, 950–973. [CrossRef]
75. Chen, W.; Pourghasemi, H.R.; Naghibi, S.A. Prioritization of landslide conditioning factors and its spatial modeling in Shangnan county, China using GIS-based data mining algorithms. *Bull. Eng. Geol. Environ.* **2018**, *77*, 611–629. [CrossRef]
76. Williams, G. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*; Springer Science & Business Media: Berlin, Germany, 2011.
77. Chen, W.; Hong, H.; Li, S.; Shahabi, H.; Wang, Y.; Wang, X.; Ahmad, B.B. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* **2019**, *575*, 864–873. [CrossRef]
78. Li, Y.; Chen, W. Landslide susceptibility evaluation using hybrid integration of evidential belief function and machine learning techniques. *Water* **2020**, *12*, 113. [CrossRef]
79. Nagarajan, M.; Singh, S. Assessment of groundwater potential zones using GIS technique. *J. Indian Soc. Remote Sens.* **2009**, *37*, 69–77. [CrossRef]
80. Ballukraya, P.; Kalimuthu, R. Quantitative hydrogeological and geomorphological analyses for groundwater potential assessment in hard rock terrains. *Curr. Sci.* **2010**, 253–259.
81. 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]
82. Jaiswal, R.; Mukherjee, S.; Krishnamurthy, J.; Saxena, R. Role of remote sensing and GIS techniques for generation of groundwater prospect zones towards rural development—An approach. *Int. J. Remote Sens.* **2003**, *24*, 993–1008. [CrossRef]
83. Madrucci, V.; Taioli, F.; de Araújo, C.C. Groundwater favorability map using GIS multicriteria data analysis on crystalline terrain, Sao Paulo state, Brazil. *J. Hydrol.* **2008**, *357*, 153–173. [CrossRef]
84. Srivastava, P.K.; Bhattacharya, A.K. Groundwater assessment through an integrated approach using remote sensing, GIS and resistivity techniques: A case study from a hard rock terrain. *Int. J. Remote Sens.* **2006**, *27*, 4599–4620. [CrossRef]
85. Cuo, L.; Giambelluca, T.W.; Ziegler, A.D.; Nullet, M.A. Use of the distributed hydrology soil vegetation model to study road effects on hydrological processes in Pang Khum Experimental Watershed, northern Thailand. *For. Ecol. Manag.* **2006**, *224*, 81–94. [CrossRef]
86. Golkarian, A.; Naghibi, S.A.; Kalantar, B.; Pradhan, B. Groundwater potential mapping using c5.0, random forest, and multivariate adaptive regression spline models in GIS. *Environ. Monit. Assess.* **2018**, *190*, 149. [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/).