

Article

# Novel Ensemble of MCDM-Artificial Intelligence Techniques for Groundwater-Potential Mapping in Arid and Semi-Arid Regions (Iran)

Alireza Arabameri <sup>1</sup>, Saro Lee <sup>2,3,\*</sup>, John P. Tiefenbacher <sup>4</sup> and Phuong Thao Thi Ngo <sup>5,\*</sup>

<sup>1</sup> Department of Geomorphology, Tarbiat Modares University, Tehran 36581-17994, Iran; a.arabameri@modares.ac.ir

<sup>2</sup> Geoscience Platform Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124, Gwahak-ro Yuseong-gu, Daejeon 34132, Korea

<sup>3</sup> Korea University of Science and Technology, 217 Gajeong-ro Yuseong-gu, Daejeon 34113, Korea

<sup>4</sup> Department of Geography, Texas State University, San Marcos, TX 78666, USA; tief@txstate.edu

<sup>5</sup> Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

\* Correspondence: leesaro@kigam.re.kr (S.L.); Ngotphuongthao5@duytan.edu.vn (P.T.T.N.); Tel.: +82-10-5353-8179 (S.L.); +47-966-77-678 (P.T.T.N.)

Received: 25 December 2019; Accepted: 2 February 2020; Published: 4 February 2020



**Abstract:** The aim of this research is to introduce a novel ensemble approach using Vise Kriterijumska Optimizacija I Kompromisno Resenje (VIKOR), frequency ratio (FR), and random forest (RF) models for groundwater-potential mapping (GWPM) in Bastam watershed, Iran. This region suffers from freshwater shortages and the identification of new groundwater sites is a critical need. Remote sensing and geographic information system (GIS) were used to reduce time and financial costs of rapid assessment of groundwater resources. Seventeen physiographical, hydrological, and geological groundwater conditioning factors (GWCs) were derived from a spatial geo-database. Groundwater data were gathered in field surveys and well-yield data were acquired from the Iranian Department of Water Resources Management for 89 locations with high yield potential values  $\geq 11 \text{ m}^3 \text{ h}^{-1}$ . These data were mapped in a GIS. From these locations, 62 (70%) were randomly selected to be used for model training, and the remaining 27 (30%) were used for validation of the model. The relative weights of the GWCs were determined with an RF model. For GWPM, 220 randomly selected points in the study area and their final weights were determined with the VIKOR model. A groundwater potential map was created by interpolating the values at these points using Kriging in GIS. Finally, the area under receiver operating characteristic (AUROC) curve was plotted for the groundwater potential map. The success rate curve (SRC) was computed for the training dataset, and the prediction rate curve (PRC) was calculated for the validation dataset. Results of RF analysis show that land use and land cover, lithology, and elevation are the most significant determinants of groundwater occurrence. The validation results show that the ensemble model had excellent prediction performance (PRC = 0.934) and goodness-of-fit (SRC = 0.925) and reasonably high classification accuracy. The results of this study could aid management of groundwater resources and assist planners and decision makers in groundwater-investment planning to achieve sustainability.

**Keywords:** modeling; random forest; frequency ration; VIKOR model; Bastam watershed

## 1. Introduction

Increasing demands for water due to population growth and industrialization combined with anticipated changes in aquatic resources due to global warming and regional climate change highlight the urgent need for a quantitative methodology that can model groundwater production [1–7].

Groundwater-potential mapping (GPM) is the first and most important step in groundwater management [8]. Since the methods to identify the parameters that influence the spatial distribution of groundwater and the ways to acquire data are evolving, new approaches to develop accurate and useful information for decision makers are needed [9]. GPM is effectively achieved by combining field study, remote sensing (RS), and geographic information system (GIS) methods [10,11]. Over the past decade, the growing use of satellite data, thematic maps, and land cover and land use (LULC) data has made it easier to study groundwater potential. As the range of data and number of variables that must be gathered and compiled spatially to simultaneously examine the critical factors grows, GIS is increasingly useful because of its capacity to process diverse spatial data from diverse sources and in different formats. The combination of GIS with new modeling methods can be a powerful tool for decision makers [12,13].

Several statistical, multi-criteria decision-making (MCDM), and machine learning approaches have been developed for GPM with thematic data layers that describe soils, topography, hydrology, lithology, and LULC [1]. The most important advantages of these methods avoid the need for highly complex or expensive data and produce results that are simple to analyze and evaluate [14]. Data needed for these models are easily acquired by integrating RS data and field surveys and can be routinely integrated in a GIS [15,16]. The extraction of spatial information from RS data has been undertaken for many tasks, including prediction of the spatial occurrence of extreme processes that yield environmental hazards (such as landslides, soil erosion, and flooding) and prediction of the distribution and dimensions of natural resources (like groundwater) [17–19]. Using RS and GIS, it is possible to develop a GPM with limited costs and reasonable levels of accuracy over large areas [20]. GIS enables faster spatial analysis, combines a large array of data that describe diverse spatial characteristics from diverse sources, and makes information management easier [21–23]. GIS is used in environmental management because it offers several benefits. High-quality data analysis can be conducted with it and it is often adaptable to the use of new and advanced analytical methods. It enables easy and fast management and analyses of large datasets. Innovations occur constantly and there are improvements in methods to create and update maps. GIS can be used to link phenomena and relate vast amounts of information in data tables. Data can be rapidly updated and integrated from continuous and scattered data acquisition activities. GIS is methodologically easy to review, and such systems can be used to model, test hypotheses, and to make predictions [24].

Many powerful statistical, machine learning, and MCDM methods have been developed and tested by researchers. These can be classified into several groups:

- MCDM-based models, like the analytic hierarchy process (AHP) [25,26],
- Bivariate and multivariate statistical models, like weights of evidence (WoE) [27], evidential belief function (EBF) [28–30], Shannon’s entropy (SE) [31], maximum entropy (ME) [32], frequency ratio (FR) [33], certainty factor (CF) [34], and logistic regression (LR) [35,36],
- Data mining/machine learning models, like boosted regression tree (BRT) [37], classification and regression tree (CART) [38], multivariate adaptive regression splines (MARS) [37], artificial neural network (ANN) [39], random forest (RF) [40], fuzzy logic (FL) [41], Fisher’s linear discriminant function (FLDA) [42], support vector machine (SVM) [43], logistic model tree (LMT) [44], K-nearest neighbor (KNN) [45], and quadratic discriminant analysis [46].

Many approaches have been used to identify areas with high groundwater potential and the array of approaches has increased in recent years because of the growing pressure to clarify the distribution of groundwater and to identify new sources of water. High-precision maps are needed to better manage aquifers that are in use and to avoid costly exploration for new deposits, especially during droughts. Hybrid groundwater mapping techniques are a vital necessity [35]. Ensemble machine learning frameworks can be used to develop better prediction models and they easily cope with complex multi-dimensional data [40]. Kordestani et al. [30] hybridized an EBF and BRT (i.e., creating an EBF-BRT) algorithm and demonstrated a groundwater contamination-susceptibility model. Naghibi et al. [47]

hybridized four data mining models and FR into a novel ensemble model for a groundwater-related study. Their ensemble model exhibited better performance than stand-alone applications of each separate model. Chen et al. [35] developed a novel hybrid approach with WoE, LR, and functional tree (FT) models that individually outperformed the models. Miraki et al. [40] used a random subspace (RS) ensemble-RF (i.e., RS-RF) classifier to model groundwater susceptibility and found that the hybrid model outperformed individual models. Razavi-Termeh et al. [48] integrated three bivariate statistical models with RF and logistic tree models to map groundwater potential and revealed that the RF statistical models improved prediction accuracy. Though the idiographic studies demonstrated hybrid models that performed strong groundwater susceptibility modeling in given study areas, there is no evidence to show that they are universally appropriate or applicable. The applicability of the novel approaches for better prediction of groundwater susceptibility models needs further evaluation.

Therefore, the main purpose of this research is to evaluate the integration of MCDM with machine-learning approaches to map groundwater potential. FR models, based on simple statistical methods, are used to examine correlations between groundwater-related factors and groundwater potential. This ensemble approach is novel and has not yet been applied in groundwater potential mapping. This study evaluated the performance of the Vise Kriterijumska Optimizacija I Kompromisno Resenje (VIKOR)-RF-FR ensemble approach for groundwater potential mapping. Statistical evaluation metrics—area under the receiver operating characteristic curve (AUC-ROC), FR, and seed cell area index (SCAI)—were used to determine the goodness-of-fit and predictive performance of the ensemble algorithm. Data from the advanced land-observing satellite (ALOS) phased array type L-band synthetic aperture radar (PALSAR) digital elevation model (DEM) with a spatial resolution of 12.5 m were used to acquire topographical and hydrological factors. ALOS-PALSAR DEM has been determined to have a higher accuracy than advanced spaceborne thermal emission and reflection radiometer (ASTER) and shuttle radar topography mission (SRTM) DEMs [49].

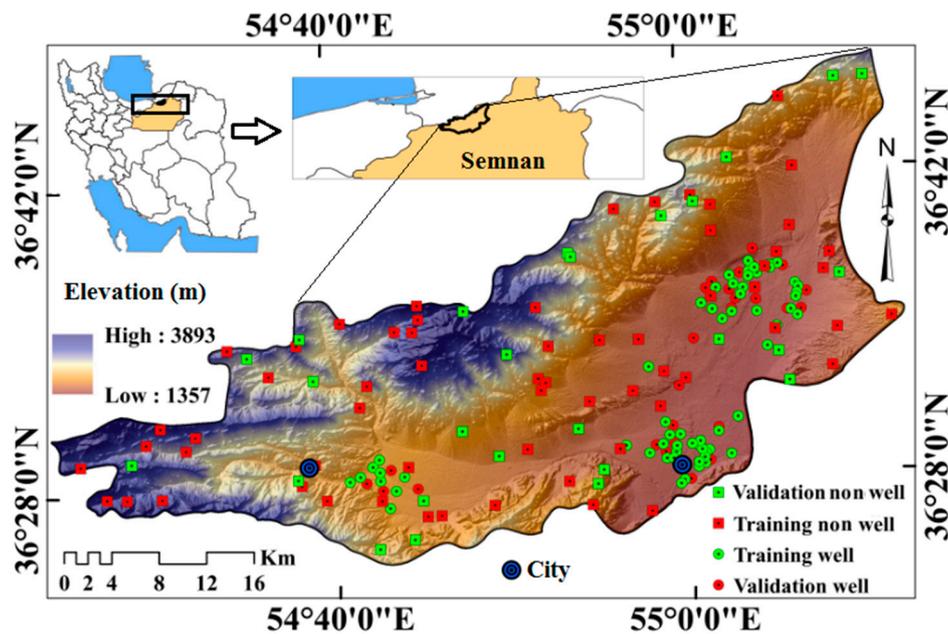
MCDM models are useful for solving complex problems because of their capacity to enable examination and selection of alternatives based on criteria [50–59]. MCDMs can analyze basic statistical trends in big data in various study areas [60,61]. These models are an important part of modern decision-making science which is decision-making in the face of varying decision criteria and selection from alternative outcomes. MCDM methods have been developed not only because of the diversity of processes and problems that require assessment of multiple factors, but because experts desire to improve decision making with better techniques [62–65]. The improvements are the outcomes of advances in mathematical science, computational science, and computer technology. Machine-learning models have also gained popularity among environmental managers and decision makers [66–68]. Machine-learning models compare favorably to other models because they do not require a defined and strict set of a priori assumptions as do many other statistical methods. These models are also used to assess the relative importance of each of the conditioning factors [66]. The FR model simplifies the probabilistic relationships between the dependent and independent variables to quantify the relationships between the reclassified factors and groundwater capacity. The relationships calculated using the FR model enabled easier interpretation of groundwater potential indicated by the rating of each factor.

## 2. Materials and Methods

### 2.1. Study Area

Bastam watershed, with an area of 1329.43 km<sup>2</sup>, is located between 36°27′02″ and 36°47′13″N and between 54°24′23″ and 55°11′08″E in Semnan Province, Iran (Figure 1). The region is topographically mountainous with an elevation range from 1357 to 3893 m.a.s.l. and an average elevation of 2078.3 m asl. The average slope in the study area is 13.5° and the maximum is 70.6°. The region's climates are arid and semi-arid: the average annual rainfall of the study area is 262.4 mm and the average annual temperature is 12.8 °C [69]. Valley terrace deposits, stream channels, braided channels, flood plain

deposits, low elevation piedmont fans, and valley terrace deposits are the most important lithological units of the study area [70]. Entisols (32.9%), entisols/inceptisols (32.4%), and mollisols (21.5%) are the most widespread soils in the study area [71].



**Figure 1.** Location of study area in Semnan Province and Iran (inset) and the locations of training and validation wells in the study area.

Geologically, Devonian-Eocene rocks, including andesitic and basaltic volcanic, conglomerate, and sandstone, are distributed as mountain ranges in the northern part of the study area extending from the eastern to the western boundaries. Cretaceous granitic and volcanic rocks are intruded as mountain ranges and isolated hills. They are found in the eastern part of the study area. Moreover, Quaternary sediments, fluvial deposits along streams, and collovium, are found in the central section of study area. Groundwater in this area is stored in these Quaternary sediments, both alluvial and sedimentary formations, in cracks, fractures, faults, bedding planes, and large cavities of consolidated rocks of Permo-Carboniferous meta-sediment and granitic aquifers.

Due to Iran's aridity, industrialization, and population growth, groundwater is a vital natural resource. More than 90% of groundwater is used for agriculture [72]. During the period 1971–2014, the number of water wells in the country increased dramatically, from 47,000 to more than 789,000. There are currently about 196,000 deep wells and 593,000 semi-deep wells in Iran and they extract an average of 0.17 and 0.02 million m<sup>3</sup>, respectively [72]. Groundwater consumption in Iran has gone from less than 18 billion m<sup>3</sup> in 1971 to more than 61 billion m<sup>3</sup> in 2014. In recent years, average groundwater levels in the country have dropped by about 51 cm, a decrease of 5.086 billion m<sup>3</sup>/year [48]. Approximately 20,000 people depend on the groundwater resources of the Bastam watershed for personal consumption and agriculture [73]. Mismanagement of water in this region has led to a 15 m drop of water tables over a 20-year period and to land-surface subsidence [73].

## 2.2. Methodology

This research involved six steps (Figure 2):

(i) Prepare the database—A groundwater-inventory map (GWIM) was created and 17 groundwater conditioning factors (GWCFs) were added as layers to a GIS (ArcGIS). The groundwater inventory locations were separated into two randomly selected groups for training and validation.

(ii) Assess the data—A multi-collinearity analysis was performed on the GWCFs using indices of tolerance (TOL) and the variance inflation factor (VIF). TOL < 0.2 and VIF > 5 indicate a multicollinearity

issue among the variables [74]. The linearity of the GWCFs was assessed, and increasing linearity between the factors will reduce the accuracy of the models [65].

(iii) Calculate relative importance of GWCFs—Because a VIKOR MCDM model cannot determine the relative importance of predictive variables, a RF model was used. To do this, the same number of non-well locations were randomly selected in the study area as there were wells yielding more than 11 m<sup>3</sup>/h and they were used to train the RF model. Next, the Extract Multi Values to Point tool in ArcGIS extracted the values of each GWCF at well (coded 1) and non-well (coded 0) locations. The data were formatted in Excel and imported into R 3.1.1 software. The Rattle package in R software was used to calculate the weight of each GWCF. Given a groundwater dataset,  $GWD = (x_i, y_i)$ , ( $i = 1, 2, 3, \dots, N$ ) with  $x_i \in R^M$  are the conditioning factors.  $N$  is the total number of samples, and  $R^M$  is the total number of conditioning factors,  $y_i \in (1,0)$  is the output which contains two classes, well and non-well locations, with the first coded “1” and the second coded “0”. The bootstrap aggregating algorithm generated  $n$  subsets, which consist of  $m$  factors ( $m \leq M$ ). A CART algorithm constructed a tree classifier for each subset. The tree classifiers were aggregated to an RF classifier. Two parameters of each the RF,  $n$ , and  $m$ , were determined when building the RF model. The parameter,  $n$ , should be large enough to ensure the robustness of the RF model (i.e.,  $n = 500$ ).

(iv) Prepare the decision matrix—Two hundred and twenty-two random points (Figure 3) (called alternatives) were selected in the study area and their GWCF values were extracted. A decision matrix of 17 columns (the GWCFs) and 220 rows (alternatives) was created. FR models were used to calculate the values of the alternatives’ pixels. In MCDM procedures, GWCF values should have either ascending or descending orders. The actual values did not follow this trend in this study, so FR was used to fit an ascending trend. Higher GWCF values indicated locations that were more likely to have groundwater.

(v) Run the hybrid model and map groundwater—The VIKOR was run to develop the groundwater potential model. The produced values indicated strength of association with the presence of groundwater at the 220 locations. Using the Ordinary Kriging tool in ArcGIS 10.5, the distribution of groundwater potential was mapped. Ordinary Kriging has been shown to efficiently map nonlinear relationships and has been used for a variety of real-world applications [75].

(vi) Assess performance—The receiver operating characteristics (ROC) curves and their area under the curve (AUROC), frequency ratio (FR), and seed cell area index (SCAI), were used to evaluate the accuracy of the model predictions.

### 2.3. Data Preparation

The most important step in groundwater-potential mapping (GWPM) is the creation of a well-inventory map (WIM) [64]. Extensive field surveys were conducted, and well-yield data were acquired from the Iranian Department of Water Resources Management to characterize the known spatial distribution of groundwater in the region. Actual pumping-test production measurements of wells expressed in m<sup>3</sup>h<sup>-1</sup> were used to represent yield. Eighty-nine wells with a high yield ( $\geq 11$  m<sup>3</sup>h<sup>-1</sup>) were selected for GWPM [1] and modeling (Figure 1). The data from these wells were randomly grouped into a training dataset of 62 (70%) wells and a validation dataset of 27 (30%) to be used to test the model [76].

Geo-environmental features were used as GWCFs to predict groundwater potential [1]. Based on a literature review [1,30,32,36,37,39,40], availability of data, environmental characteristics, scale of research, and a multicollinearity test, 17 RS-derived GWCFs that reflected topography, hydrology, and LULC were selected. Some GWCF measurements were derived from the ALOS PALSAR DEM with a spatial resolution of 12.5 m [1,52,53]. RS and thematic maps were also used to measure others. The GWCFs are elevation, slope, aspect, stream power index (SPI), topographic position index (TPI), topographic wetness index (TWI), terrain ruggedness index (TRI), convergence index (CI), distance to streams, distance to faults, distance to roads, drainage density (DD), rainfall, soil type, land use and land cover (LULC), lithology, and normalized difference vegetation index (NDVI) (Figure 4a–q).

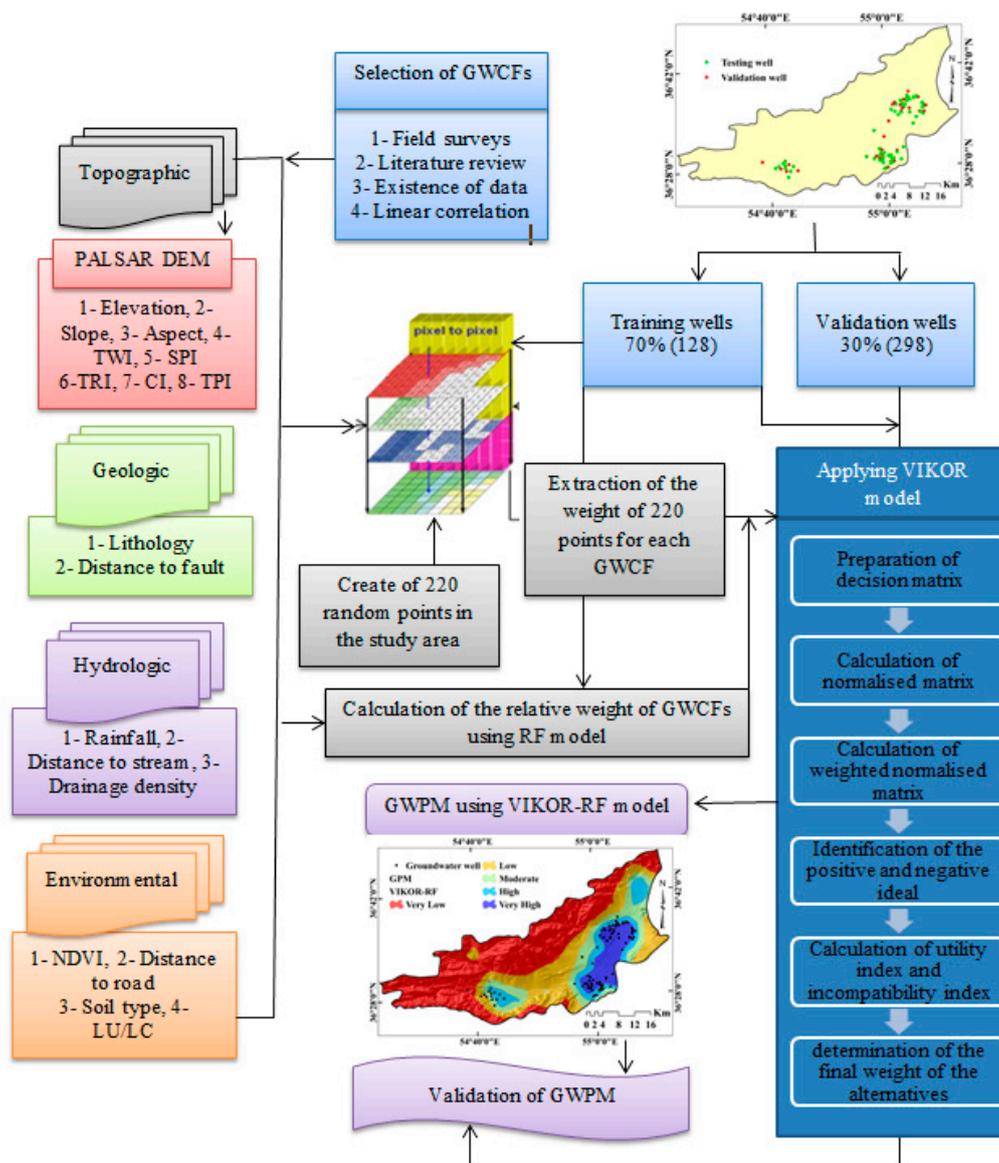


Figure 2. Flowchart of the research.

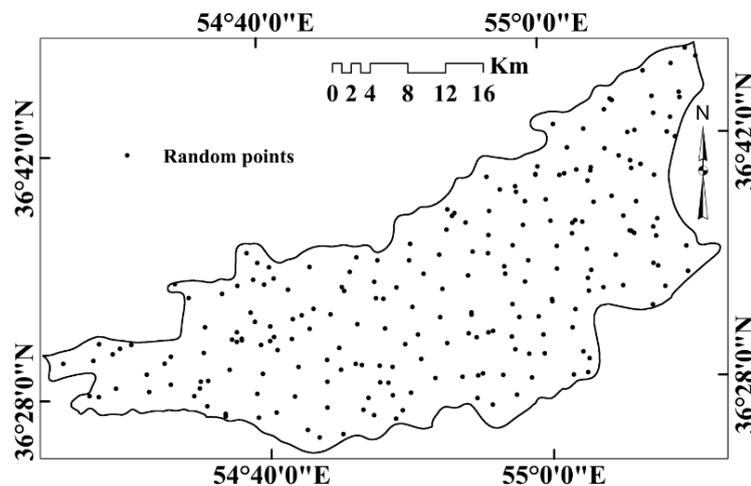


Figure 3. Location of random points in the study area.

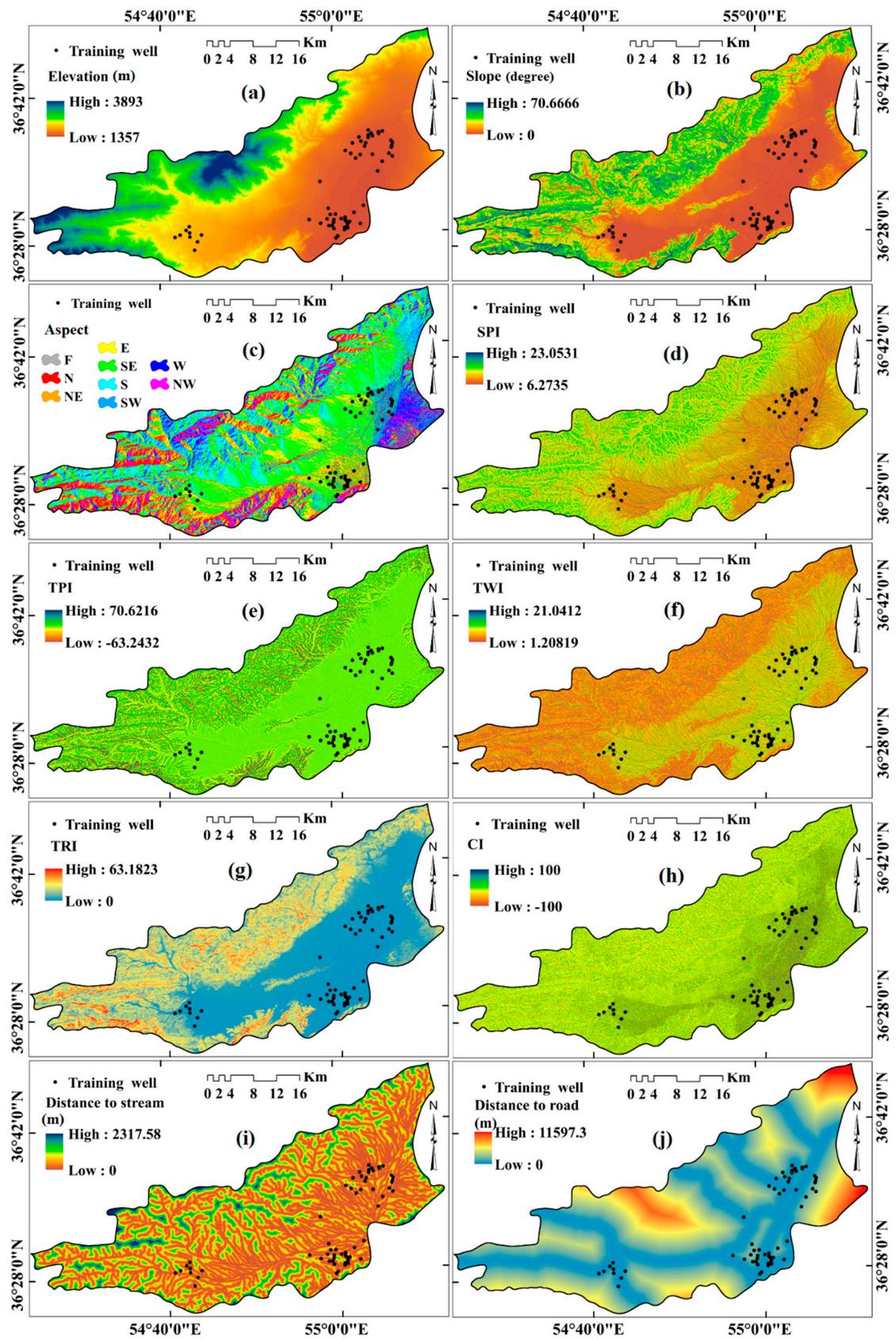
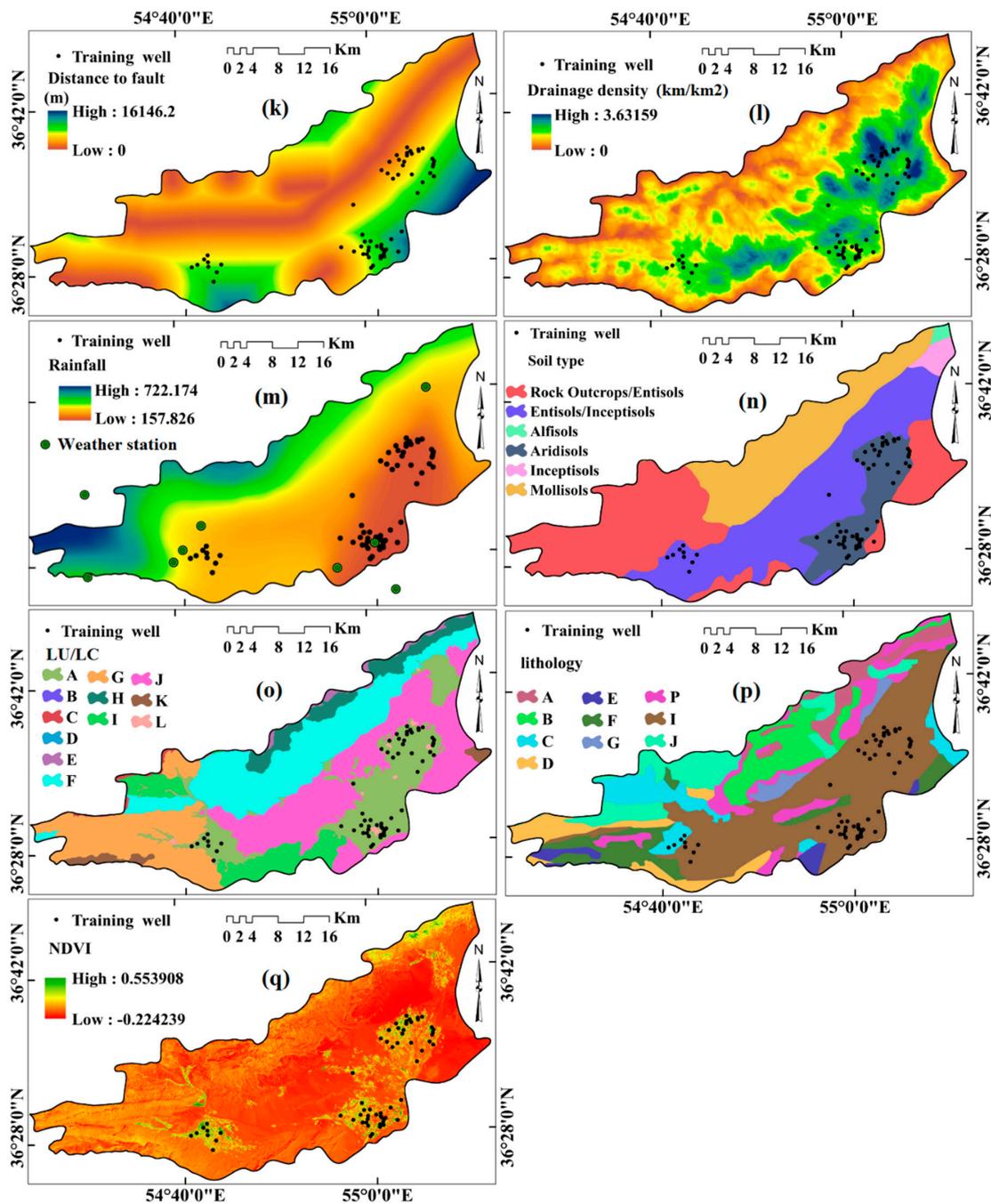


Figure 4. Cont.



**Figure 4.** Groundwater conditioning factors. (a) elevation, (b) slope, (c) aspect, (d) stream power index (SPI), (e) topography position index (TPI), (f) topography wetness index (TWI), (g) terrain ruggedness index (TRI), (h) convergence index (CI), (i) distance to stream, (j) distance to road, (k) distance to fault, (l) drainage density, (m) rainfall, (n) soil type, (o) land use/land cover (LULC), (p) lithology, (q) normalized difference vegetation index (NDVI). Classes of LULC: (A) Agriculture, (B) dense forest, (C) good range, (D) agri-dry farming, (E) dry farming, (F) low forest, (J) Woodland, (H) mod-forest, (I) mod-range, (J) poor-range, (K) rock, and (L) Urban. Classes of lithology: (A) limestone, (B) marl and shale, (C) tuffaceous shale, Andesitic and basaltic volcanics, (D) limestone, and shale, (E) marl, shale and detritic limestone, (F) Red conglomerate and sandstone, (G) siltstone and shale, (P) fluvial conglomerate, and sandstone, (I) braided channel and flood plain deposits, (J) dolomite, shale and sandstone.

Elevation, slope, and aspect were derived from ALOS PALSAR DEM in ArcGIS. Distances to streams, roads, and faults were computed as Euclidean distances pixel centroids to the nearest

polyline. The vector-based information was then rasterized to coincide with the resolution of the DEM. Road and fault-factor maps (scales of 1:50,000 and 1:100,000, respectively) were obtained from the National Geographic Organization of Iran ([www.ngo-org.ir](http://www.ngo-org.ir)) and the Geological Society of Iran (GSI) (<http://www.gsi.ir/>), respectively. Descriptions of lithological units in the study area were compiled. A factor map of streams was extracted from the DEM in ArcHydrov10.4 and ArcGISv10.4. Soil texture, lithology, and LULC were derived from three thematic maps: the soil map of the Semnan Agricultural and Natural Resources, Research and Education Center, the geological map of the Geological Society of Iran (GSI) (<http://www.gsi.ir/>), and the land use map created by the Iranian Soil Conservation and Watershed Management Research Institute (<https://www.scwmri.ac.ir>). All three of these maps were at the 1:100,000 scale.

Annual precipitation data from nine weather stations (at Mojen, Farahzad, Bastam, Abr, Karkhaneh, Semnan, Shahroud, Tarzeh, and Shah Kuh-e Bala) for the period from 1986 to 2016 were mapped and Kriging [1,24,51,54] was used to map annual rainfall patterns in ArcGIS 10.5 [69]. Land uses significantly influence the distribution of groundwater [1]. Greater vegetative cover tends to both indicate and promote the presence and extent of aquifers, as forests and denser vegetation reduce runoff rates and encourage infiltration more than do barren and sparsely vegetated lands. Equations (1)–(4) were used to calculate TWI, SPI, TRI, and TPI [77–79].

$$TWI = \ln(A_S / \tan \beta) \quad (1)$$

$$SPI = A_S \times \tan \beta \quad (2)$$

$$TRI = \sqrt{|x|(\max^2 - \min^2)} \quad (3)$$

$$TPI = \frac{E_{\text{pixel}}}{E_{\text{surrounding}}} \quad (4)$$

where  $A_S$  is the specific catchment area of the basin ( $\text{m}^2/\text{m}$ ),  $\beta$  is slope steepness in degrees,  $x$  indicate elevation of each neighbor cell to cell (0, 0) (m), and  $\max$  and  $\min$  show the largest and smallest elevation value among nine neighbor pixels, respectively.  $E_{\text{pixel}}$  depicts the elevation of the cell, and  $E_{\text{surrounding}}$  is the mean elevation of the neighboring pixels.

TPI compares the elevation of each pixel in the DEM with the average height of the pixels around it. This factor enables classification of landscapes into morphological classes. The positive and negative values indicate that a pixel is higher or lower than the pixels that surround it [80]. The plan curvature reflects the directional variations along a curve. The effect of plan curvature on slope erosion is that it causes convergence and divergence of water along and away from a flowing stream [81]. The profile curvature represents the amount of elevation variation along the flow path. When the concavity of the surface curvature is increasing, its values are negative, and, if decreasing, they are positive [82]. This index indicates the flow velocity, erosion (in negative amounts), sedimentation (in positive values), and geomorphology of the area [83]. TPI, plan curvature, profile curvature, and convergence index were extracted from the DEM in system for automated geoscientific analyses (SAGA-GIS).

NDVI is used to evaluate vegetation types and density using quantitative analysis of albedo or reflectance of visible and NIR portions of the electromagnetic spectrum. On a scale from  $-1$  to  $+1$ , high vegetation density usually has values between 0.05 and 1, bare soil areas are close to 0, and water bodies have values closer to  $-1$ . Vegetation depends upon moisture and promotes infiltration into soils and therefore is highly indicative of a likely presence of groundwater [1]. NDVI values were computed from a LANDSAT-8 image with 30 m spatial resolution acquired from the USGS (<https://earthexplorer.usgs.gov/>) (Table 1).

The GWCFs used in the GWPM were converted to raster layers at a 12.5 m resolution. The raster matrix was 2395 columns by 1460 rows.

**Table 1.** Minimum and maximum values and classes of groundwater conditioning factors (GWDFs).

Factors	Min	Max	Classes	Methods
Elevation (m)	1357	3893	1) <1659, 2) 1659–1990, 3) 1990–2342, 4) 2342–2724, 5) 2724–3163, 6) >3163	Natural break (Jenks)
Slope (°)	0	70.66	1) <5, 2) 5–10, 3) 10–15, 4) 15–20, 5) 20–30, 6) >30	Natural break (Jenks)
Aspect	–1	337.5	1) Flat (–1), 2) North (0–22.5), 3) Northeast (22.5–67.5), 4) East (67.5–112.5), 5) Southeast (112.5–157.5), 6) South (157.5–202.5), 7) Southwest (202.5–247.5), 8) West (247.5–292.5), 9) Northwest (292.5–337.5)	Directional units
SPI	6.27	23.05	1) <9.16, 2) 9.16–11.07, 3) 11.07–12.85, 4) 12.85–15.35, 5) >15.35	Natural break (Jenks)
TPI	–63.2	70.6	1) <–10.74, 2) –10.74––3.3, 3) –3.3–2.9, 4) 2.9–11.8, 5) >11.8	Natural break (Jenks)
TWI	1.20	21.04	1) <4.86, 2) 4.86–7.27, 3) 7.27–10.85, 4) >10.85	Natural break (Jenks)
TRI	0	63.1	1) <3.22, 2) 3.22–7.43, 3) 7.43–11.64, 4) 11.64–17.59, 5) >17.59	Natural break (Jenks)
CI (100/m)	–100	100	1) <–39.6, 2) –39.6––11.3, 3) –11.3–9.01, 4) 9.01–38.03, 5) >38.03	Natural break (Jenks)
Distance to stream (m)	0	2317.5	1) <100, 2) 100–200, 3) 200–300, 4) 300–400, 5) >400	Manual
Distance to road (m)	0	11,597	1) <500, 2) 500–1000, 3) 1000–1500, 4) 1500–2000, 5) >2000	Manual
Distance to fault(m)	0	16,146	1) <500, 2) 500–1000, 3) 1000–1500, 4) 1500–2000, 5) >2000	Manual
Drainage density (km/km <sup>2</sup> )	0.15	3.18	1) <0.85, 2) 0.85–1.49, 3) 1.49–2.19, 4) >2.19	Natural break (Jenks)
Rainfall (mm)	157	722	1) <248.5, 2) 248.5–350.3, 3) 350.3–461.02, 4) 461.02–582.7, 5) >582.7	Natural break (Jenks)
Soil type	-	-	1) Entisols, 2) Alfisols, 3) Aridisols, 4) Entisols/Inceptisols, 5) Inceptisols, 6) Mollisols	Soil types/Orders
LULC	-	-	1) Agriculture, 2) Dense forest, 3) Good range, 4) Agri-dry farming, 5) Dry farming, 6) Low forest, 7) Woodland, 8) Mod-forest, 9) Mod-range, 10) Poor-range, 11) Rock, 12) Urban	Supervised Classification
Lithology	-	-	1) A, 2) B, 3) C, 4) D, 5) E, 6) F, 7) G, 8) H, 9) I, 10) J.	Lithological Units
NDVI	–0.24	0.54	1) <–0.01, 2) –0.01–0.07, 3) 0.07–0.12, 4) 0.12–0.21, 5) 0.21–0.32, 6) >0.21	Natural break (Jenks)

## 2.4. Models

### 2.4.1. Frequency Ratio (FR)

FR is the probability of occurrence of a specific phenomenon. FR determines the relationship between GWCFs and well locations [61]. To calculate FR for a conditioning factor, Equation (5) is used [84]:

$$FR = \frac{A/B}{C/D} \quad (5)$$

where A is the number of wells in a class, B is the total number of wells, C is the number of pixels in that class, and D is the number of pixels [85].

### 2.4.2. VIKOR (Vlse Kriterijumsk Optimizacija Kompromisno Resenje)

The VIKOR method was first introduced by Opricovic [86]. VIKOR is derived from compromising programming. This model is an MCDM method that determines the best solution for a discrete decision that has disproportionate and conflicting criteria. This method focuses on ranking and selection from alternatives and it determines a compromise solution to guide decision-making. The compromise solution is a feasible solution that is the closest solution to the ideal solution. Compatibility is also based on concessions [87]. The following six steps are involved in the VIKOR model:

- (1) Prepare the decision matrix.
- (2) Calculate the normalized matrix, as shown in Equation (6):

$$r_{ij} = \frac{x_{ij}}{\sum_1^m x_{ij}} \quad (6)$$

where,  $x_{ij}$  is the  $i$ th alternative performance of the  $j$ th criterion, and  $m$  is the alternative numbers.

- (3) Calculate a weighted normalized matrix, as shown in Equation (7):

$$f_{ij} = r_{ij} \times W_j \quad (7)$$

where  $W_j$  is the weight of the criterion.

- (4) Identify the ideal positive (Equation (8)) and negative (Equation (9)) options:

$$f_i^+ = \max_{i_{f_{ij}}} \quad (8)$$

$$f_i^- = \min_{i_{f_{ij}}} \quad (9)$$

where  $f_i^+$  is the positive ideal solution for the  $i$ th criterion and  $f_i^-$  is the negative ideal solution for the  $i$ th criterion.

- (5) Calculate the utility index (Equation (10)) and incompatibility index (the distance from positive and negative ideal solution) (Equation (11)):

$$S_i = L_{1,i} = \frac{\sum_{i=1}^n W_j (f_i^+ - f_{ij})}{(f_i^+ - f_i^-)} \quad (10)$$

$$R_i = L_{\infty,i} = \max \left[ \sum_{i=1}^n W_j (f_i^+ - f_{ij}) / (f_i^+ - f_i^-) \right] \quad (11)$$

where  $S_i$  and  $R_i$  are the distance of the  $i$ th alternative to the ideal positive and negative solutions, respectively.

- (6) Calculate the VIKOR index and determine the final weight of the alternatives, as shown in Equation (12):

$$Q_i = V \times \frac{(S_i - S^*)}{(S^- - S^*)} + (1 - V) \times \frac{(R_i - R^*)}{(R^- - R^*)} \quad (12)$$

where  $V$  is a constant (0.5),  $S^*$  is  $\min S_i$ ,  $S^-$  is  $\max S_i$ ,  $R^*$  is  $\min R_i$ , and  $R^-$  is  $\max R_i$ . The greater the value of the VIKOR index in one alternative, the greater the importance of that alternative.

#### 2.4.3. Random Forest

RF is one of the most well-known machine-learning algorithms [88]. RF is nonparametric and based on decision trees. Many decision trees are grown in the classification of RF algorithms [89]. The RF algorithm:

- Considers the number of tests ( $N$ ) and the number of variables ( $M$ ),
- Enters  $R(m)$  variables to decide on each tree node ( $m$  should be less than  $M$ ),
- Selects test data for the tree by using the  $n$ -times placement of  $N$  samples, and the rest of the samples are used to estimate the tree-error,
- Selects  $M$  variables for each tree node, the basis for decision making in each node. The best groups are calculated on  $m$  variables in the test-sample, and
- Expands each tree completely without pruning.

The GWCFs are prioritized using the mean decrease accuracy and the mean decrease Gini. Using the mean decrease accuracy, as opposed to the mean decrease Gini, determines the priority of effective factors more effectively, especially when environmental factors are being compared [68].

#### 2.5. Validation of Results

AUC-ROC, FR, and SCAI criteria help to discern the quality of definite and probable identification, as well as the quality of prediction [65]. AUC-ROC represents a model's predictive accuracy by describing its ability to predict whether predefined events have occurred or will occur [90]. AUC-ROC curves indicate the sensitivity of the model to the percentage of cells or unstable units correctly predicted by the model against the percentage of predicted unstable cells within the total. These values reflect the model's ability to correctly distinguish between positive and negative observations in a validation sample. High sensitivity indicates that most predictions were correct (true positives), while specificity values of 1 indicate a high number of false positives. In the AUC-ROC, the false positive rate (1 specificity) is shown on the x axis (Equation (13)) and the true positive rate (sensitivity) on the y axis (Equation (14)):

$$x = 1 - \text{specificity} = 1 - \left[ \frac{\text{TN}}{(\text{TN} + \text{FP})} \right] \quad (13)$$

$$y = \text{sensitivity} = \left[ \frac{\text{TP}}{(\text{TP} + \text{FN})} \right] \quad (14)$$

where TN is true-negative, FP is false-positive, TP is true-positive, and FN is false-negative [85]. The quantitative–qualitative relationship classes between AUC-ROC and prediction accuracy, ranging from 0 to 1, are: excellent (0.9–1), very good (0.8–0.9), good (0.7–0.8), moderate (0.6–0.7), and weak (0.5–0.6) [80]. FR and SCAI criteria were used to evaluate the separations between classes. These criteria determine the classification accuracy of a model. The FR is the ratio of the percentage of wells in each class to the percentage area of that class [91], and SCAI is the ratio of the percentage area of each zoning class to the percentage of wells in each class [92].

### 3. Results

#### 3.1. Multi-Collinearity Analysis

The multicollinearity test shows that there is no collinearity among conditioning factors (Table 2). TOL among factors ranges from 0.15 to 0.84 and VIF ranges from 1.1 to 4.8. Therefore, all GWCFs were included in the GWPM process.

Table 2. Multi-collinearity test among groundwater conditioning factors.

Factors	Collinearity Statistics		Factors	Collinearity Statistics	
	Tolerance	VIF *		Tolerance	VIF
NDVI	0.6	1.4	Rainfall	0.2	4.3
Distance to fault	0.7	1.4	Distance to road	0.4	2.1
TRI	0.2	4.7	Slope	0.1	4.8
Slope aspect	0.8	1.1	Soil type	0.3	2.5
CI	0.7	1.3	SPI	0.5	1.7
Drainage density	0.3	2.7	Distance to stream	0.6	1.4
Elevation	0.3	3.2	TPI	0.3	2.7
Lithology	0.2	3.4	TWI	0.3	2.7
LU/LC	0.2	3.9			

\* VIF is variance inflation factor.

#### 3.2. Determining the Relative Importance of GWCFs using the RF Model

The RF model with an out-of-bag error rate = 8.57% was applied in R using the Caret package. This scenario implies that the accuracy of the model is equal to 91.43%, and exhibits the results of the confusion matrix. According to the results (Table 3), of the 80 non-well locations, 74 (92.5%) were predicted to be non-well and 6 (7.5%) were predicted to be locations of wells. Of the 80 well locations, 11 (13.75%) were predicted to be non-well locations and 69 (86.25%) were predicted to be wells. The determination of the relative importance of the GWCFs (using mean decrease accuracy of the RF model) (Figure 5) shows that LULC (18.5), lithology (13.9), and elevation (11.9) were the main factors that identified the locations of greater amounts of groundwater. These were followed by TPI (9.1), NDVI (7.9), TWI (7.7), distance to road (7.4), drainage density (6.3), slope aspect (5.7), soil type (5.3), distance to fault (5.03), TRI (4.9), slope degree (4.8), rainfall (4.7), CI (4.6), distance to stream (2.4), and SPI (1).

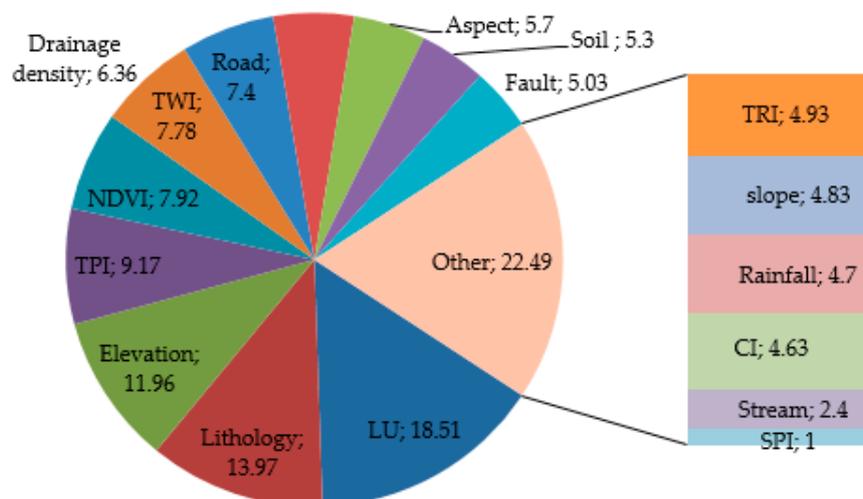


Figure 5. Relative importance of conditioning factors using random forest model.

**Table 3.** Confusion matrix from the RF model.

	No Well (0)	Well (1)
no well (0)	74	6
well (1)	11	69

### 3.3. Determining the Value of Each Pixel of GWCFs using the FR Model

The FR model revealed the spatial relationships between the GWCFs and wells (Table 4). Analysis of the values of each GWCF at each well location (Figure 6a–q) shows that locations with lower elevations, slope angles, CI, SPI, distance to road, distance to stream, rainfall, TRI, and NDVI had lower FR values (i.e., these factors have strong inverse relationships with groundwater potential). Locations with higher TWI and drainage density have higher values of FR (i.e., strong positive relationships to groundwater potential). In terms of LULC and lithology, agriculture and quaternary formations, not surprisingly, were strongly associated with well locations.

**Table 4.** Spatial relation between groundwater conditioning factors and wells using the frequency ratio model.

Factor	Class	No. of Pixels	%	No. of Wells	%	FR
Elevation (m)	<1659	479,814	32.48	53	85.48	2.63
	1659–1990	309,671	20.96	9	14.52	0.69
	1990–2342	239,854	16.24	0	0.00	0.00
	2342–2724	209,109	14.16	0	0.00	0.00
	2724–3163	146,912	9.95	0	0.00	0.00
	>3163	91,792	6.21	0	0.00	0.00
Slope (°)	< 5	588,363	39.83	62	100.00	2.51
	5–10	149,259	10.10	0	0.00	0.00
	10–15	124,159	8.41	0	0.00	0.00
	15–20	143,874	9.74	0	0.00	0.00
	20–30	280,859	19.01	0	0.00	0.00
	>30	190,637	12.91	0	0.00	0.00
Aspect	F	6262	0.42	2	3.23	7.61
	N	132,304	8.96	0	0.00	0.00
	NE	152,509	10.32	14	22.58	2.19
	E	245,198	16.60	11	17.74	1.07
	SE	341,071	23.09	17	27.42	1.19
	S	273,880	18.54	10	16.13	0.87
	SW	144,941	9.81	5	8.06	0.82
	W	89,718	6.07	2	3.23	0.53
SPI	NW	91,269	6.18	1	1.61	0.26
	<9.16	322,599	21.84	39	62.90	2.88
	9.16–11.07	437,271	29.60	15	24.19	0.82
	11.07–12.85	455,113	30.81	3	4.84	0.16
	12.85–15.35	203,521	13.78	4	6.45	0.47
TPI	>15.35	58,648	3.97	1	1.61	0.41
	<−10.74	69,120	4.68	0	0.00	0.00
	−10.74–−3.3	233,435	15.80	0	0.00	0.00
	−3.3–2.9	914,956	61.94	62	100.00	1.61
	2.9 0 11.8	201,144	13.62	0	0.00	0.00
TWI	>11.8	58,497	3.96	0	0.00	0.00
	<4.86	582,464	39.43	0	0.00	0.00
	4.86–7.27	585,130	39.61	38	61.29	1.55
	7.27–10.85	237,436	16.07	18	29.03	1.81
	>10.85	72,122	4.88	6	9.68	1.98

Table 4. Cont.

Factor	Class	No. of Pixels	%	No. of Wells	%	FR
TRI	<3.22	692,266	46.86	62	100.00	1.61
	3.22–7.43	286,708	19.41	0	0.00	0.00
	7.43–11.64	285,453	19.32	0	0.00	0.00
	11.64–17.59	174,864	11.84	0	0.00	0.00
	>17.59	37,860	2.56	0	0.00	0.00
CI (100/m)	<−39.6	91,462	6.22	19	30.65	4.92
	−39.6–−11.3	289,807	19.72	12	19.35	0.98
	−11.3–9.01	663,255	45.12	12	19.35	0.43
	9.01–38.03	341,553	23.24	9	14.52	0.62
	>38.03	83,751	5.70	10	16.13	2.83
Distance to stream (m)	<100	408,797	27.67	29	46.77	1.69
	100–200	298,966	20.24	16	25.81	1.28
	200–300	245,504	16.62	9	14.52	0.87
	300–400	157,383	10.65	6	9.68	0.91
	>400	366,502	24.81	2	3.23	0.13
Distance to road (m)	<500	224,545	15.20	21	33.87	2.23
	500–1000	193,768	13.12	15	24.19	1.84
	1000–1500	166,957	11.30	13	20.97	1.86
	1500–2000	147,431	9.98	9	14.52	1.45
	>2000	744,451	50.40	4	6.45	0.13
Distance to fault (m)	<500	102,353	6.93	0	0.00	0.00
	500–1000	103,021	6.97	0	0.00	0.00
	1000–1500	101,830	6.89	0	0.00	0.00
	1500–2000	101,938	6.90	0	0.00	0.00
	>2000	103,846	7.03	3	4.84	0.69
Drainage density (km/km <sup>2</sup> )	<0.85	398,195	26.96	0	0.00	0.00
	0.85–1.49	497,684	33.69	4	6.45	0.19
	1.49–2.19	368,973	24.98	34	54.84	2.20
	>2.19	212,300	14.37	24	38.71	2.69
Rainfall	<248.5	284,466	19.26	52	83.87	4.36
	248.5–350.3	535,319	36.24	10	16.13	0.45
	350.3–461.02	321,491	21.76	0	0.00	0.00
	461.02–582.7	208,018	14.08	0	0.00	0.00
	>582.7	127,857	8.66	0	0.00	0.00
Soil type	Entisols	486,282	32.92	2	3.23	0.10
	Alfisols	7086	0.48	0	0.00	0.00
	Aridisols	161,526	10.93	50	80.65	7.37
	Entisols/Inceptisols	479,698	32.47	10	16.13	0.50
	Inceptisols	24,777	1.68	0	0.00	0.00
	Mollisols	317,783	21.51	0	0.00	0.00
LU/LC	Agriculture	281,159	19.03	62	100.00	5.25
	Dense forest	155	0.01	0	0.00	0.00
	Good range	8660	0.59	0	0.00	0.00
	Agri-dry farming	243	0.02	0	0.00	0.00
	Dry farming	16,235	1.10	0	0.00	0.00
	Low forest	337,471	22.85	0	0.00	0.00
	Woodland	233,619	15.82	0	0.00	0.00
	Mod-forest	83,527	5.65	0	0.00	0.00
	Mod-range	105,592	7.15	0	0.00	0.00
	Poor-range	382,114	25.87	0	0.00	0.00
	Rock	23,154	1.57	0	0.00	0.00
Urban	5220	0.35	0	0.00	0.00	

Table 4. Cont.

Factor	Class	No. of Pixels	%	No. of Wells	%	FR
Lithology	A	76,475	5.18	0	0.00	0.00
	B	131,673	8.91	0	0.00	0.00
	C	114,748	7.77	5	8.06	1.04
	D	94,149	6.37	0	0.00	0.00
	E	33,722	2.28	0	0.00	0.00
	F	117,907	7.98	0	0.00	0.00
	G	31,564	2.14	0	0.00	0.00
	H	134,059	9.08	0	0.00	0.00
	I	594,531	40.25	57	91.94	2.28
	J	148,324	10.04	0	0.00	0.00
NDVI	<-0.01	18,508	1.26	18	29.03	23.12
	-0.01-0.07	44,145	2.99	15	24.19	8.08
	0.07-0.12	89,947	6.10	13	20.97	3.44
	0.12-0.21	362,491	24.59	10	16.13	0.66
	0.21-0.32	538,553	36.53	6	9.68	0.26
	>0.32	420,479	28.52	0	0.00	0.00

### 3.4. Application of VIKOR Model

Extracting the GWCF values at random locations, the constructed decision matrix, and the final weight of random locations using the VIKOR model indicate that groundwater potential is strongly indicated by weighting values. The weights of random locations range from 0 to 1. Higher values indicate greater groundwater potential.

### 3.5. Groundwater Potential Mapping Using the FR-RF-VIKOR Ensemble Model

Groundwater potential was mapped with the ensemble model (values varied from 0.01 to 0.92). Based on the natural break classification method, these values were divided into five classes: very low (0.01–0.14), low (0.14–0.28), moderate (0.28–0.44), high (0.44–0.65), and very high (0.65–0.92) groundwater potential (Figure 7). About 48.3% of the total area of 642.1 km<sup>2</sup> is classified (Figure 8) as having ‘very low’ groundwater potential, 22.8% (303.3 km<sup>2</sup>) has a ‘low’ groundwater potential zone, 11.8% of the area (157.8 km<sup>2</sup>) has ‘moderate’ potential, 8.5% (113.9 km<sup>2</sup>) has high groundwater potential, and 8.4% (112.05 km<sup>2</sup>) has very high groundwater potential.

### 3.6. Validation of Results

Validation of the ensemble model is based on the prediction rate curve (PRC) and the success rate curve (SRC) (Figure 9). The ensemble model, with PRC = 0.93 and SRC = 0.92, has excellent groundwater potential prediction accuracy. The FR and SCAI model results (Figure 10) show ascending and descending trends, respectively. In terms of statistical assessment of the modeling values, FR increased and SCAI decreased, indicating that the classification accuracy of the ensemble model was reasonable.

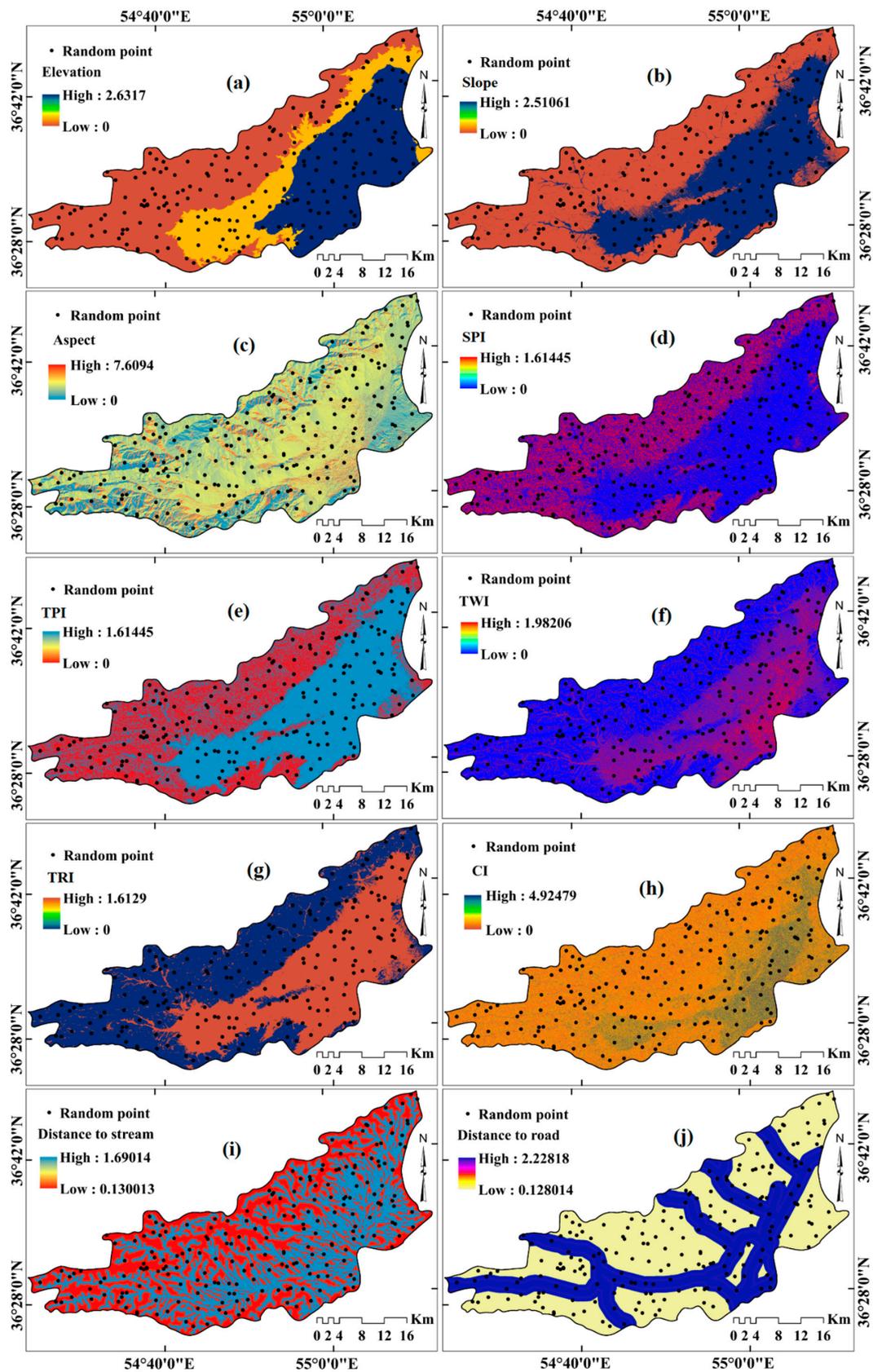
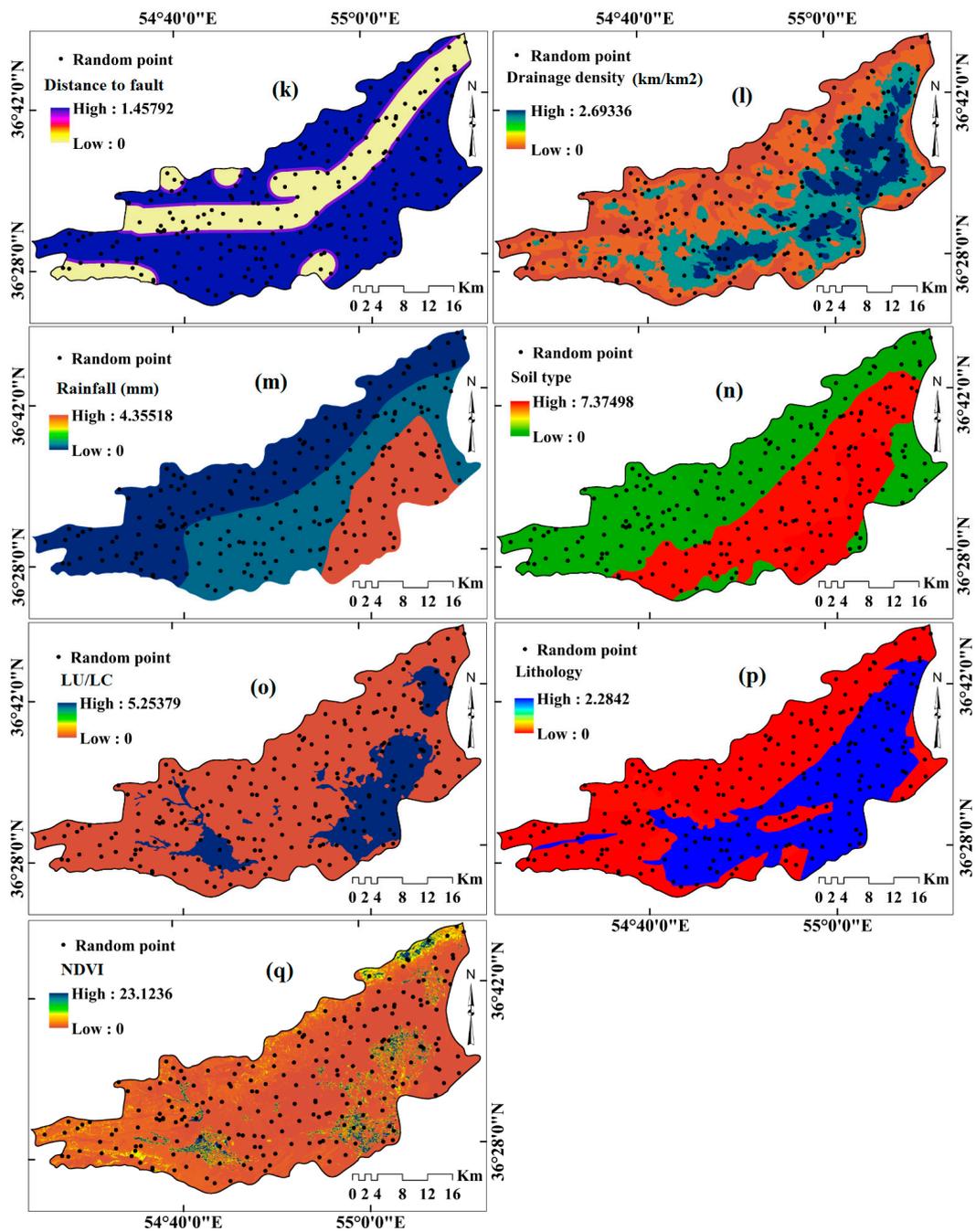


Figure 6. Cont.



**Figure 6.** Weighted GWCFs using the FR model for (a) elevation, (b) slope, (c) aspect, (d) stream power index (SPI), (e) topography position index (TPI), (f) topography wetness index (TWI), (g) terrain ruggedness index (TRI), (h) convergence index (CI), (i) distance to stream, (j) distance to road, (k) distance to fault, (l) drainage density, (m) rainfall, (n) soil type, (o) LU/LC, (p) lithology, (q) NDVI.

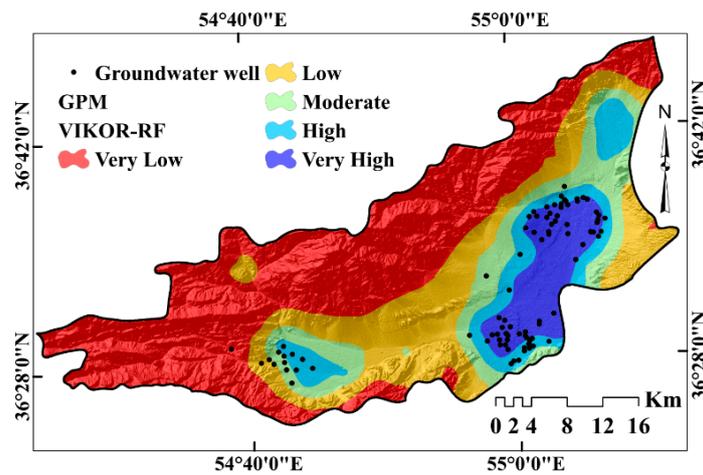


Figure 7. Groundwater potential map using the VIKOR-RF ensemble model.

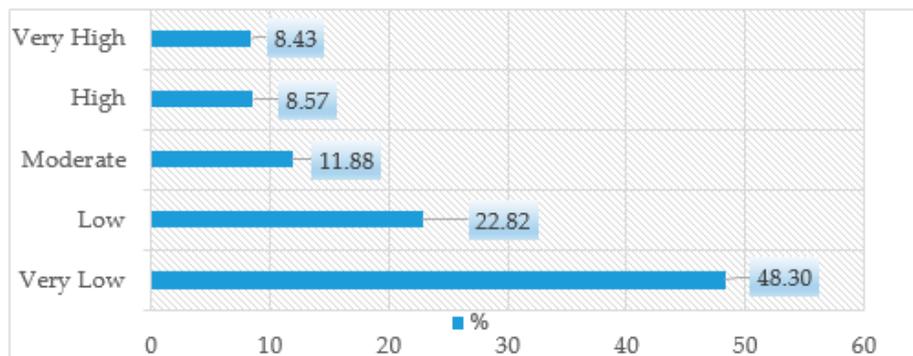


Figure 8. Percentage of potential classes.

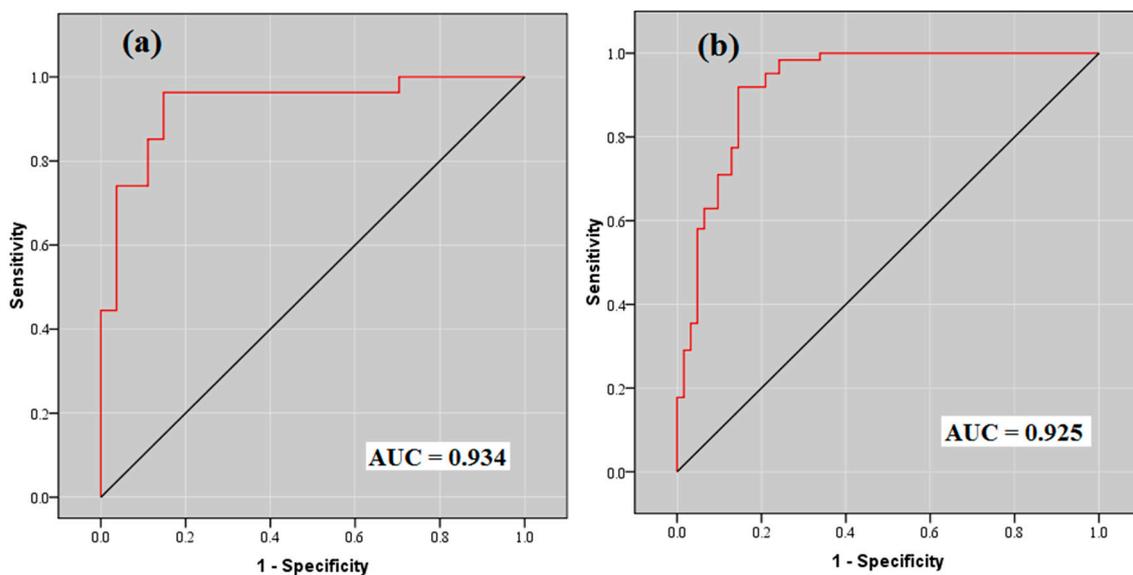


Figure 9. Area under curve (AUC). (a) Prediction rate curve (PRC), (b) Success rate curve (SRC).

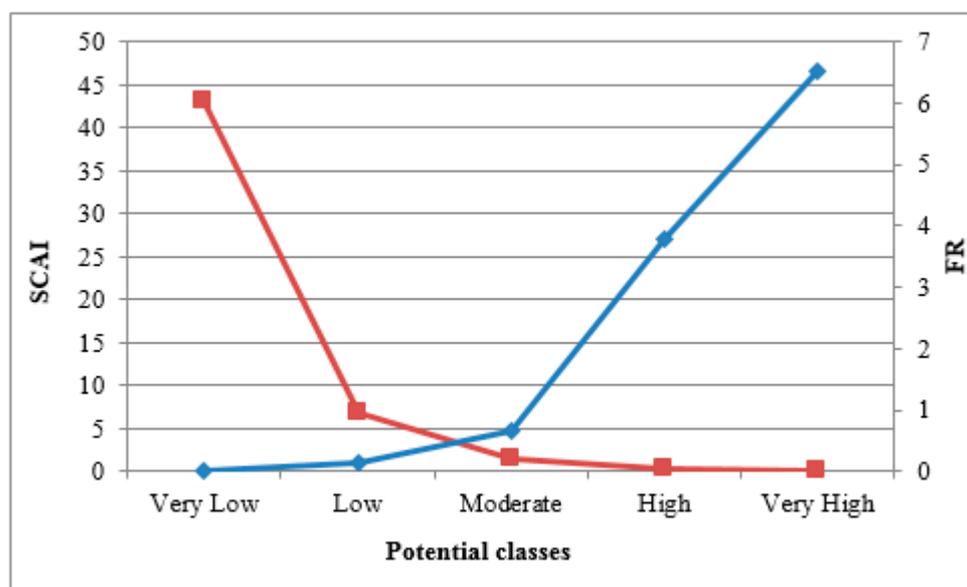


Figure 10. Trend of frequency ratio and seed cell area index indicators.

#### 4. Discussion

Groundwater resources are crucial in regions with limited surface water, particularly in arid and semi-arid climates [81]. When these regions have significant amounts of industrial and agricultural water demands, aquifers can be consumed at rates that exceed recharge rates and are at increasing risk of irreparable damage. Therefore, it is important to have accurate assessments of the spatial dimensions of groundwater in a watershed. Such assessments can improve water-management planning and use strategies [93]. Groundwater mapping can be improved and made more economical by identifying the geophysical and hydrological conditioning factors that are associated with sub-surface storage. Several methods have been developed to map groundwater potential, and each has advantages and disadvantages.

This research developed a novel state-of-the-art approach using an ensemble of statistical, machine-learning, and MCDM approaches in conjunction with RS and GIS techniques to map groundwater potential. MCDM models, statistical models, and machine-learning algorithms (like VIKOR, FR, and RF, respectively) can be used to support scientific decisions for problems that have a diversity of management criteria [94]. VIKOR can analyze both quantitative and qualitative data, uses simple mathematical formulas, produces stable results, is flexible, is highly efficient, is very accurate, it can factor uncertainty into analyses, is comprehensive in its analysis, produces visual representations of the data, enables comparison of alternatives, and can incorporate and analyze large amounts of data. FR was used for its simplicity of equations, easily interpretable results, and its high efficiency [95]. RS and GIS save time and are cost-effective ways to improve data gathering and analysis [96–99].

Machine-learning algorithms can model natural phenomena (e.g., groundwater potential and groundwater presence) with nonlinear relationships. They do not require removal of outliers, data transformations, or statistical assumptions. They can handle complex nonlinear relationships between conditioning factors and groundwater potentiality and can automatically analyze effective-factor (i.e., predictor) interactions. RF was the machine-learning model employed in this research because of its advantages. The RF method can accommodate different types of predictor variables as well as missing values. It can assess the relationships and interactions between effective factors. It can handle uncertainty, can apply several input factors without eliminating any, and can return a small set of categories to maintain high prediction accuracy [54,67]. Classification accuracy, however, is affected by the number, scale, types, and precision of inputs. Incorporating all suitable factors increases modeling accuracy. Moreover, compared to other models, RF can include more datasets [53]. It can overfit some datasets with

noisy classification and regression tasks. It can be biased toward attribute data with more categories, and it can also be biased toward attributes with more levels. Therefore, the variable-importance scores from RF are not reliable for such data. There are a few drawbacks to this algorithm, however. It does require significant computational memory, analysis of large datasets consumes a lot of time, there are temporal high costs for pruning, there are a high number of overlapping end nodes, and errors accumulate at each layer as trees grow.

The analysis of the spatial relationships between GWCFs and well locations using the FR model showed that elevation, slope, CI, SPI, distances to the nearest road, distances to the nearest stream, less rainfall, TRI, and NDVI are inversely correlated to groundwater potential. TWI and drainage density are directly correlated to well locations. These relationships are consistent with previous studies [29,33]. Elevation and relief play important roles in groundwater potential because weather and climatic conditions vary greatly between elevations and cause differential soil development and vegetation growth patterns. Areas at low elevations, with gentle slopes, and with low drainage density are usually associated with highly permeable material, higher vegetative cover, and less relief. These areas have higher groundwater potential. Conversely, higher elevations, steeper slopes, and higher drainage density usually have less soil, more impermeable subsurface material, sparse vegetation, and runoff-promoting relief. TWI, CI, TRI, and SPI topographic indices predict soil moisture, groundwater flow, and slope stability. These indices have been used to reflect spatial patterns of soil-moisture [77]. In terms of LULC and lithology, agricultural areas and regions with Quaternary formations have the strongest relationships to wells. Agriculture tends not to occur in places without access to water and wells are not often drilled in places that lack demand for water. The LULC categories covary with soil conditions and subsequently reflect groundwater occurrence [100]. Lithology influences groundwater because lithological variation is associated with varying porosity and hydraulic conductivity in rock formations and subsequent soils. Soil has an impact on recharge due to permeability factors and composition, and so it can influence the potential size of an aquifer [100].

The RF model indicated that LULC, lithology, and elevation were most strongly related to groundwater flow rates. This is consistent with other research [1,8,31,34,38,100]. Rahmati et al. [8] mapped groundwater potential in the Mehran Region of Iran by combining RF and maximum entropy models using data from 163 wells. They analyzed ten GWCFs and elevation, drainage density, lithology, and land use were found to be the greatest determinants of groundwater production. Arabameri et al. [1] combined RF, WoE, binary logistic regression, and the technique for order-preference by similarity to ideal solution (TOPSIS) for groundwater potential mapping on the Shahroud Plain in Iran. Analysis of the relationships between 122 wells and 15 GWCFs revealed that LULC, soil type, and slope were key correlates to groundwater. Naghibi et al. [38] used boosted regression tree, classification and regression tree, and RF machine-learning models for GWPM in Iran. Thirteen hydrological-geological-physiographical conditioning factors were assessed at 864 springs. Elevation and drainage density were the most important correlates to spring locations.

Comparing the goodness-of-fit and performance of the ensemble model using AUROC, FR, and SCAI methods revealed that the ensemble model had excellent prediction accuracy, and this was consistent with other research [30,43]. Combined or ensemble models are often more effective than are individual, stand-alone models. Kordestani et al. [30] showed that by combining EBF and BRT for GWPM, prediction accuracy (AUROC) climbed from 75% for EBF alone to 82% for EBF-BRT. Four data-mining models and FR were combined as an ensemble model [43] and showed better performance by the ensemble model because it reduced overfitting.

Overfitting occurs due to noise, limited training data, and classification complexity. Solutions are difficult because of the width of the array of potential causes. Three aspects of this study suggest that overfitting could have occurred: the area from which the non-well locations were selected was confined to the region that contained wells even though the study region extended well beyond the region covered by wells, the ratio of 70:30 training to validation datasets was used without an assessment of the accuracy of the sampling ratio, and despite the multicollinearity assessment, GWCFs still had noise.

A lack of confidence (uncertainty) might result from doubt about the completeness of knowledge. Data could be suspected of being unclear, inaccurate, unreliable, inconclusive, or potentially false [101–104]. Machine-learning techniques like RF can account for nonlinear relationships and can handle uncertainty in data. In an RF model, the prediction interval has two main components that determine its width: one represents the uncertainty of the predicted mean (or some other parameter)—this is the confidence interval—and the other represents the variability of individual observations around that mean. BRT uses uncertainty reduction to arrive at a final number of trees [105]. There are many sources of error in datasets used for modeling: measurement errors, sampling bias, limitations in field data collection, genetic variability, etc. These errors can affect model accuracy [106]. The strength of the novel ensemble model introduced here to model natural phenomena with nonlinear relationships is confirmed by the results of this study.

There are several things that can enhance GWPM accuracy and they should be the foci of future research. First, there should be a complete analysis of the implications of limiting or extending the area from which the selection of non-well locations is made to places where wells are located or to areas beyond the distribution of wells. Second, the conventional 70:30 ratio of the datasets used for training and validation should be revisited, as proportion may be causing issues. Third, the methods used to select features may be having unseen effects on the modeling process, and fourth, there should be a comparison of the VIKOR-RF-FR ensemble to other ensemble models to predict groundwater locations.

## 5. Conclusions

Many methods have been used to assess regional groundwater potential. This study created an MCDM statistical machine-learning ensemble model to spatially predict groundwater in Bastam Watershed, Iran. VIKOR, FR, and RF models were combined to map and assess the characteristics of groundwater-bearing landscapes. Like many other parts of the world, the Bastam Watershed suffers from a shortage of water. Seventeen conditioning factors (elevation, slope, aspect, SPI, TPI, TWI, TRI, CI, distance to stream, distance to fault, distance to road, drainage density, rainfall, soil type, LU/LC, lithology, and NDVI) were assessed at 89 well locations to determine each conditioning factors' statistical association with groundwater presence and productivity. RF modeling revealed that LULC, lithology, and elevation were the most important factors predicting groundwater potential and production. Validation confirmed that the ensemble model had high prediction accuracy. The VIKOR-RF-FR ensemble model is a systematic, objective, and scientific decision-making model that can be used to support groundwater resource development and management plans. A GWPM can help planners guide future groundwater exploration. The same analysis in other districts with similar topographic and geological conditions can save time and money in their quests for protecting and developing groundwater resources. A limitation of this research is that a fixed combination of GWCFs were considered for modeling. Different combinations of GWCFs should be explored in order to further improve groundwater potential modeling.

**Author Contributions:** Conceptualization, A.A.; Data curation, A.A.; Formal analysis, A.A.; Investigation, A.A. S.L., and P.T.T.N.; Methodology, A.A., S.L., J.P.T., and P.T.T.N.; Resources, A.A.; Software, A.A.; Supervision, A.A., S.L., and P.T.T.N.; Validation, A.A.; Writing—original draft, A.A.; Writing—review and editing, A.A., J.P.T., S.L., and P.T.T.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This research was conducted by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) funded by the Ministry of Science and ICT.

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

## References

1. Arabameri, A.; Rezaei, K.; Cerda, A.; Lombardo, L.; Rodrigo-Comino, J. GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *Sci. Total Environ.* **2019**, *658*, 160–177. [[CrossRef](#)]

2. Foster, S. Groundwater assessing vulnerability and promoting protection of a threatened resource. In Proceedings of the 8th Stockholm Water Symposium, Stockholm, Sweden, 10–13 August 1998.
3. UN. Water for people, water for life. In *The UN World Water Development Report (WWDR)*, UNESCO; Publishing and Berghahn Books: New York, NY, USA, 2003; p. 34.
4. ElNaqa, A.; AlShayeb, A. Groundwater protection and management strategy in Jordan. *Water Resour. Manag.* **2008**, *23*, 2379–2394. [[CrossRef](#)]
5. Schematization and Management Organ of Iran (SMOI). 2004. Available online: <http://www.ncc.org.ir/> (accessed on 12 August 2018).
6. Magesh, N.S.; Chandrasekar, N.; John, P. Delineation of groundwater potential zones in Theni district, Tamil Nadu, using remote sensing, GIS and MIF techniques. *Geosci. Front.* **2012**, *3*, 189–196. [[CrossRef](#)]
7. Lee, S.; Hyun, Y.; Lee, M.J. Groundwater Potential Mapping Using Data Mining Models of Big Data Analysis in Goyang-si, South Korea. *Sustainability* **2019**, *11*, 1678. [[CrossRef](#)]
8. Rahmati, O.; Pourghasemi, H.R.; Melesse, A.M.N. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran region, Iran. *Catena* **2016**, *137*, 360–372. [[CrossRef](#)]
9. Agarwal, R.; Garg, P.K. Remote Sensing and GIS Based Groundwater Potential & Recharge Zones Mapping Using Multi Criteria Decision Making Technique. *Water Resour. Manag.* **2016**, *30*, 243–260.
10. Mondal, S. Remote sensing and GIS based ground water potential mapping of kangshabati irrigation command area, west bengal. *Geogr. Nat. Disasters* **2012**, *1*, 1–8. [[CrossRef](#)]
11. Oikonomidis, D.; Dimogianni, S.; Kazakis, N.; Voudouris, K. A GIS/Remote Sensing-based methodology for groundwater potentiality assessment in Tirnavos area, Greece. *J. Hydrol.* **2015**, *525*, 197–208. [[CrossRef](#)]
12. Thomas, A.; Sharma, P.K.; Sharma, M.K.; Sood, A. Hydrogeomorphological mapping in assessing groundwater by using remote sensing data: a case study in Lehra Gage Block, Sangrur district, Punjab. *J. Indian Soc. Remote Sens.* **1999**, *27*, 31–42. [[CrossRef](#)]
13. Muralidhar, M.; Raju, K.R.K.; Raju, K.S.V.P.; Prasad, J.R. Remote sensing applications for the evaluation of water resources in rainfed area, Warangal district, Andhra Pradesh. *Indian Miner.* **2003**, *34*, 33–40.
14. Pinto, D.; Shrestha, S.; Babel, M.S.; Ninsawat, S. Delineation of groundwater potential zones in the Comoro watershed, Timor Leste using GIS, remote sensing and analytic hierarchy process (AHP) technique. *Appl. Water Sci.* **2017**, *7*, 503–519. [[CrossRef](#)]
15. Magaia, L.A.; Goto, T.N.; Masoud, A.A.; Koike, K. Identifying groundwater potential in crystalline basement rocks using remote sensing and electromagnetic sounding techniques in Central Western Mozambique. *Nat. Resour. Res.* **2018**, *27*, 275–298. [[CrossRef](#)]
16. Luís, A.M. Development of Regional Exploration Techniques for Groundwater Resources in Semiarid Areas Through Integration of Remote Sensing and Geophysical Survey. Ph.D. Thesis, Kyoto University, Kyoto, Japan, 2018.
17. Pradhan, B. Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. *Open Geosci.* **2009**, *1*, 120–129. [[CrossRef](#)]
18. Bera, K.; Bandyopadhyay, J. Ground water potential mapping in Dulung watershed using remote sensing & GIS techniques, West Bengal, India. *Int. J. Sci. Res. Publ.* **2012**, *2*, 1–7.
19. Al-Ruzouq, R.; Shanableh, A.; Merabtene, T.; Siddique, M.; Khalil, M.A.; Idris, A.; Almulla, E. Potential groundwater zone mapping based on geo-hydrological considerations and multi-criteria spatial analysis: North UAE. *CATENA* **2019**, *173*, 511–524. [[CrossRef](#)]
20. Shi, Z.H.; Cai, C.F.; Ding, S.W.; Wang, T.W.; Chow, T.L. Soil conservation planning at the small watershed level using RUSLE with GIS: A case study in the Three Gorge area of China. *Catena* **2004**, *55*, 33–48. [[CrossRef](#)]
21. Sharma, S.; Kumar Mahajan, A. Comparative evaluation of GIS-based landslide susceptibility mapping using statistical and heuristic approach for Dharamshala region of Kangra Valley, India. *Geoenviron. Disasters* **2018**, *5*, 4. [[CrossRef](#)]
22. Saro, L. Current and future status of GIS-based landslide susceptibility mapping: A literature review. *Korea J. Remote Sens.* **2019**, *35*, 179–193.
23. Saro, L.; Oh, H.J. Landslide Susceptibility Prediction using Evidential Belief Function, Weight of Evidence and Artificial Neural Network Models. *Korea J. Remote Sens.* **2019**, *35*, 299–316.
24. Arabameri, A.; Pourghasemi, H.R.; Yamani, M. Applying different scenarios for landslide spatial modeling using computational intelligence methods. *Environ. Earth Sci.* **2017**, *76*, 832. [[CrossRef](#)]

25. Singh, S.K.; Srivastava, K.; Gupta, M.; Thakur, K.; Mukherjee, S. Appraisal of land use/land cover of mangrove forest ecosystem using support vector machine. *Environ. Earth Sci.* **2014**, *71*, 2245–2255. [[CrossRef](#)]
26. Mahato, S.; Pal, S. Groundwater Potential Mapping in a Rural River Basin by Union (OR) and Intersection (AND) of Four Multi-criteria Decision-Making Models. *Nat. Resour. Res.* **2019**, *28*, 523–545. [[CrossRef](#)]
27. 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](#)] [[PubMed](#)]
28. Park, I.; Kim, Y.; Lee, S. Groundwater productivity potential mapping using evidential belief function. *Groundwater* **2014**, *52*, 201–207. [[CrossRef](#)]
29. Mogaji, K.A.; Omosuyi, G.O.; Adelusi, A.O.; Lim, H.S. 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](#)]
30. 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](#)]
31. Naghibi, S.A.; Pourghasemi, H.R.; Pourtaghie, Z.S.; Rezaei, A. Groundwater qanat potential mapping using frequency ratio and Shannon’s entropy models in the Moghan Watershed, Iran. *Earth Sci. Inf.* **2015**, *8*, 171–186. [[CrossRef](#)]
32. 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](#)]
33. Trabelsi, F.; Lee, S.; Khelifi, S.; Arfaoui, A. Frequency Ratio Model for Mapping Groundwater Potential Zones Using GIS and Remote Sensing; Medjerda Watershed Tunisia. In *Advances in Sustainable and Environmental Hydrology, Hydrogeology, Hydrochemistry and Water Resources*; Chaminé, H., Barbieri, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2019.
34. Razandi, Y.; Pourghasemi, H.R.; SamaniNeisani, N.; Rahmati, O. Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS. *Earth Sci. Inf.* **2015**, *8*, 867–883. [[CrossRef](#)]
35. Chen, W.; Li, H.; Hou, E.; Wang, S.; Wang, G.; Panahi, M. GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci. Total Environ.* **2018**, *1*, 853–867. [[CrossRef](#)]
36. Lee, S.; Lee, C.W.; Kim, J.C. Groundwater Productivity Potential Mapping Using Logistic Regression and Boosted Tree Models: The Case of Okcheon City in Korea. In *Advances in Remote Sensing and Geo Informatics Applications*; El-Askary, H., Lee, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2019.
37. Golkarian, A.; Rahmati, O. Use of a maximum entropy model to identify the key factors that influence groundwater availability on the Gonabad Plain, Iran. *Environ. Earth Sci.* **2018**, *77*, 369. [[CrossRef](#)]
38. 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](#)]
39. 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](#)]
40. Miraki, S.; Hedayati Zanganeh, S.; Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Thai Pham, B. Mapping Groundwater Potential Using a Novel Hybrid Intelligence Approach. *Water Resour. Manag.* **2019**, *33*, 281–302. [[CrossRef](#)]
41. Shahid, S.; Nath, S.K.; Kamal, A.S. GIS integration of remote sensing and topographic data using fuzzy logic for ground water assessment in Midnapur district, India. *Geocarto Int.* **2014**, *17*, 69–74. [[CrossRef](#)]
42. Chen, W.; Pradhan, B.; Li, S.; Shahabi, H.; Mojaddadi Rizeei, H.; 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](#)]
43. Naghibi, S.A.; Ahmadi, K.; Daneshi, A. Application of support vector machine, random forest, and genetic algorithm optimized randomforest models in groundwater potential mapping. *Water Resour. Manag.* **2017**, *31*, 2761–2775. [[CrossRef](#)]
44. Rahmati, O.; Naghibi, S.A.; Shahabi, H.; Tien Bui, D.; Pradhan, B.; Aareh, A. Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. *Hydrology* **2018**, *565*, 248–261. [[CrossRef](#)]

45. 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. Clim.* **2018**, *131*, 967–984. [[CrossRef](#)]
46. Naghibi, S.A.; Dashtpajardi, 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](#)]
47. Naghibi, S.A.; Moghaddam, D.D.; Kalantar, B.; Pradhan, B.; Kisi, O. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *J. Hydrol.* **2017**, *548*, 471–483. [[CrossRef](#)]
48. Razavi-Termeh, S.V.; Sadeghi-Niaraki, A.; Choi, S.-M. Groundwater Potential Mapping Using an Integrated Ensemble of Three Bivariate Statistical Models with Random Forest and Logistic Model Tree Models. *Water* **2019**, *11*, 1596. [[CrossRef](#)]
49. Alganci, U.; Besol, B.; Sertel, E. Accuracy Assessment of Different Digital Surface Models. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 114. [[CrossRef](#)]
50. Arabameri, A.; Pourghasemi, H.R.; Cerda, A. Erodibility prioritization of sub-watersheds using morphometric parameters analysis and its mapping: A comparison among TOPSIS, VIKOR, SAW, and CF multi-criteria decision making models. *Sci. Total Environ.* **2017**, *613*, 1385–1400.
51. Arabameri, A.; Pradhan, B.; Rezaei, K.; Yamani, M.; Pourghasemi, H.R.; Lombardo, L. Spatial modelling of gully erosion using Evidential Belief Function, Logistic Regression and a new ensemble EBF–LR algorithm. *Land Degrad. Dev.* **2018**, *29*, 4035–4049. [[CrossRef](#)]
52. Arabameri, A.; Rezaei, K.; Pourghasemi, H.R.; Lee, S.; Yamani, M. GIS-based gully erosion susceptibility mapping: A comparison among three data-driven models and AHP knowledge-based technique. *Environ. Earth Sci.* **2018**, *77*, 628. [[CrossRef](#)]
53. Arabameri, A.; Pradhan, B.; Pourghasemi, H.R.; Rezaei, K. Identification of erosion-prone areas using different multi-criteria decision-making techniques and GIS. *Geomat. Nat. Hazards Risk* **2018**, *9*, 1129–1155. [[CrossRef](#)]
54. Arabameri, A.; Rezaei, K.; Cerda, A.; Conoscenti, C.; Kalantari, Z. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *Sci. Total Environ.* **2019**, *660*, 443–458. [[CrossRef](#)]
55. Arabameri, A.; Pradhan, B.; Rezaei, K.; Conoscenti, C. Gully erosion susceptibility mapping using GIS-based multi-criteria decision analysis techniques. *CATENA* **2019**, *180*, 282–297. [[CrossRef](#)]
56. Arabameri, A. Application of the Analytic Hierarchy Process (AHP) for locating fire stations: Case Study Maku City. *Merit Res. J. ArtSoc. Sci. Humanit.* **2014**, *2*, 1–10.
57. Arabameri, A.; Ramesht, M.H. Site Selection of Landfill with emphasis on Hydrogeomorphological–environmental parameters Shahrood-Bastam watershed. *Sci. J. Manag. Syst.* **2017**, *16*, 55–80.
58. Yamani, M.; Arabameri, A. Comparison and evaluation of three methods of multi attribute decision making methods in choosing the best plant species for environmental management (Case study: Chah Jam Erg). *Nat. Environ. Chang.* **2015**, *1*, 49–62.
59. Arabameri, A. Zoning Mashhad Watershed for Artificial Recharge of Underground Aquifers Using Topsis Model and GIS Technique. *Glob. J. Hum. Soc. Sci. B Geogr. Geo Sci. Environ. Disaster Manag.* **2014**, *14*, 45–53.
60. Mousavi, S.M.; Golkarian, A.; Naghibi, S.A.; Kalantar, B.; Pradhan, B. GIS-based groundwater spring potential mapping using data mining boosted regression tree and probabilistic frequency ratio models in Iran. *Aims Geosci.* **2017**, *3*, 91–115.
61. Siahkamari, S.; Haghizadeh, A.; Zeinivand, H.; Tahmasebipour, N.; Rahmati, O. Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. *Geocarto Int.* **2018**, *33*, 927–941. [[CrossRef](#)]
62. Arabameri, A.; Pradhan, B.; Rezaei, K.; Sohrabi, M.; Kalantari, Z. GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms. *J. Mt. Sci.* **2019**, *16*, 595–618. [[CrossRef](#)]
63. Arabameri, A.; Pradhan, B.; Rezaei, K.; Saro, L.; Sohrabi, M. An Ensemble Model for Landslide Susceptibility Mapping in a Forested Area. *Geocarto Int.* **2019**, 1–26. [[CrossRef](#)]
64. Arabameri, A.; Pradhan, B.; Rezaei, K. Spatial prediction of gully erosion using ALOS PALSAR data and ensemble bivariate and data mining models. *Geosci. J.* **2019**, *24*, 669–686. [[CrossRef](#)]

65. Arabameri, A.; Pradhan, B.; Rezaei, K.; Lee, C.-W. Assessment of Landslide Susceptibility Using Statistical- and Artificial Intelligence-Based FR–RF Integrated Model and Multiresolution DEMs. *Remote Sens.* **2019**, *11*, 999. [[CrossRef](#)]
66. Arabameri, A.; Pradhan, B.; Pourghasemi, H.R.; Rezaei, K.; Kerle, N. Spatial Modelling of Gully Erosion Using GIS and R Programming: A Comparison among Three Data Mining Algorithms. *Appl. Sci.* **2018**, *8*, 1369. [[CrossRef](#)]
67. Arabameri, A.; Pourghasemi, H.R. Spatial Modeling of Gully Erosion Using Linear and Quadratic Discriminant Analyses in GIS and R. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, 1st ed.; Pourghasemi, H.R., Gokceoglu, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; p. 796.
68. Arabameri, A.; Pradhan, B.; Rezaei, K. Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS. *J. Environ. Manag.* **2019**, *232*, 928–942. [[CrossRef](#)]
69. IRIMO. Summary Reports of Iran’s Extreme Climatic Events. In *Ministry of Roads and Urban Development; Iran Meteorological Organization*: Tehran, Iran, 2012; Available online: <http://www.cri.ac.ir> (accessed on 12 August 2018).
70. GSI. Geology Survey of Iran. 1997. Available online: [http://www.gsi.ir/Main/Lang\\_en/index.html](http://www.gsi.ir/Main/Lang_en/index.html) (accessed on 12 August 2018).
71. IUSS Working Group WRB14. *World Reference Base for Soil Resources 2014, World Soil Resources Report*; FAO: Rome, Italy, 2014.
72. Noor, H. Analysis of Groundwater Resource Utilization and Their Current Condition in Iran. *Iran. J. Rainwater Catchment Syst.* **2018**, *5*, 29–38.
73. Ramesht, M.H.; Arabameri, A. Shahrood-Bastam Basin Zoning for the Purpose of Artificial Underground Aquifer Recharge by Using Linear Assignment Method and GIS Technique. *Geogr. Space* **2013**, *12*, 134–149.
74. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T. 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](#)]
75. Liang, C.-P.; Chen, J.-S.; Chien, Y.-C.; Chen, C.-F. Spatial analysis of the risk to human health from exposure to arsenic contaminated groundwater: A kriging approach. *Sci. Total Environ.* **2018**, *627*, 1048–1057. [[CrossRef](#)]
76. Haghizadeh, A.; Davoudi Moghadam, D.; Pourghasemi, H.R. GIS-based bivariate statistical techniques for groundwater potential analysis (an example of Iran). *J. Earth Syst. Sci.* **2017**, *126*, 109. [[CrossRef](#)]
77. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30. [[CrossRef](#)]
78. De Reu, J.; Bourgeois, J.; Bats, M.; Zwertvaegher, A.; Gelorini, V.; De Smedt, P.; Chu, W.; Antrop, M.; De Maeyer, P.; Finke, P.; et al. Application of the topographic position index to heterogeneous landscapes. *Geomorphology* **2013**, *186*, 39–49. [[CrossRef](#)]
79. Althuwaynee, O.F.; Pradhan, B.; Park, H.J.; Lee, J.H. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena* **2014**, *114*, 21–36. [[CrossRef](#)]
80. Jenness, J. *Surface Areas and Ratios from Elevation Grid*; Jenness Enterprises: Flagstaff, AZ, USA, 2012.
81. Oh, H.J.; Pradhan, B. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Comput. Geosci.* **2011**, *37*, 1264–1276. [[CrossRef](#)]
82. Hengl, T.; Gruber, S.; Shrestha, D.P. Digital terrain analysis in ILWIS. International Institute for Geo-Information Science and Earth Observation Enschede. *Int. Inst. Geoinf. Sci. Earth Obs. Enschede Neth.* **2003**, *62*, 1–56.
83. Yesilnacar, E.K. The Application of Computational Intelligence to Landslide Susceptibility Mapping in Turkey. Ph.D. Thesis, Department of Geomatics the University of Melbourne, Melbourne, Australia, 2005; p. 423.
84. Bonham-Carter, G.F. Geographic information systems for geoscientists: Modeling with GIS. In *Computer Methods in the Geosciences*; Pergamon: Bergama, Turkey, 1994.
85. Shafapour Tehrani, M.; Kumar, L. The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. *Environ. Earth Sci.* **2018**, *77*, 490. [[CrossRef](#)]
86. Opricovic, S. Multicriteria optimization of civil engineering systems. *Fac. Civ. Eng. Belgrade* **1998**, *2*, 5–21.
87. Opricovic, S.; Tzeng, G.H. Multicriteria planning of post-earthquake sustainable reconstruction. *Comput. Aided Civ. Infrastruct. Eng.* **2002**, *17*, 211–220. [[CrossRef](#)]

88. Breiman, L. Random forests. *J. Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
89. Hastie, T. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2001; p. 533.
90. Swets, J.A. Measuring the accuracy of diagnostic systems. *Science* **1988**, *240*, 1285–1293. [[CrossRef](#)]
91. Pradhan, B. Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. *Adv. Space Res.* **2010**, *45*, 1244–1256. [[CrossRef](#)]
92. Yilmaz, C.; Topal, T.; Suzen, M.L. GIS-based landslide susceptibility mapping using bivariate statistical analysis in Devrek (Zonguldak-Turkey). *Environ. Earth Sci.* **2012**, *65*, 2161–2178. [[CrossRef](#)]
93. Manap, M.A.; Nampak, H.; Pradhan, B.; Lee, S.; Sulaiman, W.N.A.; Ramli, M.F. Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. *Arab. J. Geosci.* **2012**, *7*, 711–724. [[CrossRef](#)]
94. Kiker, G.A.; Bridges, T.S.; Varghese, A.; Seager, T.P.; Linkov, I. Application of multi-criteria decision analysis in environmental decision making. *Integr. Environ. Assess. Manag.* **2005**, *1*, 95–108. [[CrossRef](#)] [[PubMed](#)]
95. Aditian, A.; Kubota, T.; Shinohara, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* **2018**, *318*, 101–111. [[CrossRef](#)]
96. Jha, M.K.; Chowdary, V.M.; Chowdhury, A. Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques. *Hydrogeol. J.* **2010**, *18*, 1713–1728. [[CrossRef](#)]
97. Arkoprovo, B.; Adarsa, J.; Prakash, S.S. Delineation of groundwater potential zones using satellite remote sensing and geographic information system techniques: A case study from Ganjam district, Orissa, India. *Res. J. Recent. Sci.* **2012**, *1*, 59–66.
98. Hammouri, N.; ElNaqa, A.; Barakat, M. An integrated approach to groundwater exploration using remote sensing and geographic information system. *J. Water Resour. Prot.* **2012**, *4*, 717–724. [[CrossRef](#)]
99. Davoodi, M.D.; Rezaei, M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Pradhan, B. Groundwater spring potential mapping using bivariate statistical model and GIS in the Taleghan watershed Iran. *Arab. J. Geosci.* **2015**, *8*, 913–929.
100. 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](#)]
101. Paris, G.; Robilliard, D.; Fonlupt, C. Exploring Overfitting in Genetic Programming. In Proceedings of the International Conference Evolution Artificielle, Marseilles, France, 17 October 2004; pp. 267–277.
102. 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](#)]
103. Arabameri, A.; Blaschke, T.; Pradhan, B.; Pourghasemi, H.R.; Tiefenbacher, J.P.; Bui, D.T. Evaluation of Recent Advanced Soft Computing Techniques for Gully Erosion Susceptibility Mapping: A Comparative Study. *Sensors* **2020**, *20*, 335. [[CrossRef](#)]
104. Arabameri, A.; Chen, W.; Blaschke, T.; Tiefenbacher, J.P.; Pradhan, B.; Tien Bui, D. Gully Head-Cut Distribution Modeling Using Machine Learning Methods—A Case Study of N.W. Iran. *Water* **2020**, *12*, 16. [[CrossRef](#)]
105. Pourtaghi, Z.S.; Pourghasemi, H.R. GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran. *Hydrogeol. J.* **2015**, *22*, 643–662. [[CrossRef](#)]
106. Sahoo, S.; Munusamy, S.B.; Dhar, A.; Kar, A.; Ram, P. Appraising the accuracy of multi-class frequency ratio and weights of evidence method for delineation of regional groundwater potential zones in canal command system. *Water Resour. Manag.* **2017**, *31*, 4399–4413. [[CrossRef](#)]

