

Article

# Application of Probabilistic and Machine Learning Models for Groundwater Potentiality Mapping in Damghan Sedimentary Plain, Iran

Alireza Arabameri <sup>1</sup>, Jagabandhu Roy <sup>2</sup>, Sunil Saha <sup>2</sup>, Thomas Blaschke <sup>3</sup>,  
Omid Ghorbanzadeh <sup>3</sup> and Dieu Tien Bui <sup>4,\*</sup>

<sup>1</sup> Department of Geomorphology, Tarbiat Modares University, Tehran 14117-13116, Iran; alireza.arab@ut.ac.ir

<sup>2</sup> Department of Geography, University of Gour Banga, Malda, West Bengal 732103, India; jagabandhuroy1991@gmail.com (J.R.); sunilgeo.88@gmail.com (S.S.)

<sup>3</sup> Department of Geoinformatics – Z\_GIS, University of Salzburg, 5020 Salzburg, Austria; thomas.blaschke@sbg.ac.at (T.B.); omid.ghorbanzadeh@stud.sbg.ac.at (O.G.)

<sup>4</sup> Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

\* Correspondence: buitiendieu@duytan.edu.vn

Received: 12 November 2019; Accepted: 10 December 2019; Published: 14 December 2019



**Abstract:** Groundwater is one of the most important natural resources, as it regulates the earth's hydrological system. The Damghan sedimentary plain area, located in the region of a semi-arid climate of Iran, has very critical conditions of groundwater due to massive pressure on it and is in need of robust models for identifying the groundwater potential zones (GWPZ). The main goal of the current research is to prepare a groundwater potentiality map (GWPM) considering the probabilistic, machine learning, data mining, and multi-criteria decision analysis (MCDA) approaches. For this purpose, 80 wells collected from the Iranian groundwater resource department and field investigation with global positioning system (GPS), have been selected randomly and considered as the groundwater inventory datasets. Out of 80 wells, 56 (70%) wells have been brought into play for modeling and 24 (30%) for validation purposes. Elevation, slope, aspect, convergence index (CI), rainfall, drainage density (Dd), distance to river, distance to fault, distance to road, lithology, soil type, land use/land cover (LU/LC), normalized difference vegetation index (NDVI), topographic wetness index (TWI), topographic position index (TPI), and stream power index (SPI) have been used for modeling purpose. The area under the receiver operating characteristic (AUROC), sensitivity (SE), specificity (SP), accuracy (AC), mean absolute error (MAE), and root mean square error (RMSE) are used for checking the goodness-of-fit and prediction accuracy of approaches to compare their performance. In addition, the influence of groundwater determining factors (GWDFs) on groundwater occurrence was evaluated by performing a sensitivity analysis model. The GWPMs, produced by technique for order preference by similarity to ideal solution (TOPSIS), random forest (RF), binary logistic regression (BLR), weight of evidence (WoE) and support vector machine (SVM) have been classified into four categories, i.e., low, medium, high and very high groundwater potentiality with the help of the natural break classification methods in the GIS environment. The very high groundwater potentiality class is covered 15.09% for TOPSIS, 15.46% for WoE, 25.26% for RF, 15.47% for BLR, and 18.74% for SVM of the entire plain area. Based on sensitivity analysis, distance from river, and drainage density represent significantly effects on the groundwater occurrence. validation results show that the BLR model with best prediction accuracy and goodness-of-fit outperforms the other five models. Although, all models have very good performance in modeling of groundwater potential. Results of seed cell area index model that used for checking accuracy classification of models show that all models have suitable performance. Therefore, these are promising models that can be applied for the GWPZs identification, which will help for some needful action of these areas.

**Keywords:** groundwater potential mapping (GWPM); probabilistic models; machine learning algorithms; sensitivity analysis; Damghan sedimentary plain

---

## 1. Introduction

Groundwater plays a crucial role in serving the heterogeneous need of human being such as drinking, agricultural, industrial, etc. [1]. In another way, groundwater availability and accessibility control sustainable development at a global, regional and local scale [2]. Large numbers of countries of the earth are facing the problem of water scarcity at the societal level [3]. In the arid and semi-arid regions, groundwater is the prime source of water and accounts for 80% groundwater resource [4]. Notably, in Iran, groundwater is more demanded source owing to its cleanness, tawdriness, constant chemical composition, constant temperature, lower pollution coefficient, and a high certainty [4,5]. Groundwater extensively affects economic development, biological diversity and community health [6]. Similar to Iran country, a major part of the largely arid and semi-arid physiographic regions suffers from the scarcity of water. Therefore, the groundwater is a main source of water to serve the different purpose and utilization of this region [7]. Iran has received average annual precipitation of 413 mm, and the evapo-transportation rate is 296 mm. Therefore, the 117 billion m<sup>3</sup> of water is stored as groundwater over the whole country. The global per capita annual renewal water is 7600 m<sup>3</sup> while the quantity of per capita global renewable water in Iran is 1900 m<sup>3</sup>. In this region, the average yearly water consumption is 3.4 billion m<sup>3</sup>, out of which about 65% is supplied from groundwater. In the present day, Iran is facing harsh water supply problems [8]. From these data, it is inevitable to implement water resource management policy for continuing the country's economic and societal development. However, this issue can be short out by taking some necessary steps and decisions such as watershed management, artificial recharge, and management of soil and water [8]. In present the decades, groundwater recharge level has fallen due to unnecessary use and unscientific management plans [9]. Hence, aquifer potential determination through groundwater potentiality analysis is a good strategy in this field [10,11]. Different methods and models, techniques and processes were induced and used for the groundwater potentiality map (GWPM) or identification of areas having good potentiality of groundwater recharge. A few decades back, conventional techniques were applied for GWPM. Day by day improvement in technology with regarding scientific approach, the measuring instrument and computerized data analysis able to recognize the groundwater level, flow and other aspects for GWPM. Comparatively, contemporary scientific methods are providing better outcomes than the conventional method. Recently, remote sensing (RS) and geographic information system (GIS) are playing important role in managing the groundwater resource without the computational requirements [12]. RS technique provides the spatial and non-spatial information—even over the inaccessible areas in a short duration [13]. Therefore, RS technique also a powerful, efficient, accurate tool for collecting, restoring, manipulating, analyzing the spatial data of the surface and sub-surface water research, e.g., groundwater recharge, potentiality, evaluation of water quality [2,14]. Specifically, satellite imagery can provide hydrological characteristics, i.e., drainage network, flow accumulation, drainage density, recharge, and other geomorphologic characteristics [15]. The modeling of groundwater potential zones (GWPZ) is not only dependent on the single factors but also dependent on the different geo-environmental factors such as elevation, slope, aspect, rainfall, geology, fault, rainfall, drainage density (Dd), land use/land cover (LU/LC), normalized difference vegetation index (NDVI), topographic wetness index (TWI), stream power index (SPI), soil permeability, topographic position index (TPI), convergence index (CI), infiltration rate, and soil texture. RS and GIS integration with modern groundwater mapping models such as probabilistic, knowledge-driven, machine learning, data mining could provide a powerful way to gain valuable decision-making information. The rapid development of probabilistic, machine learning, data mining, and ensemble models in recent decades is enhancing the basement to determine groundwater recharge opportunity, soil erosion susceptibility, gully erosion susceptibility, and other spatial modelings. Some

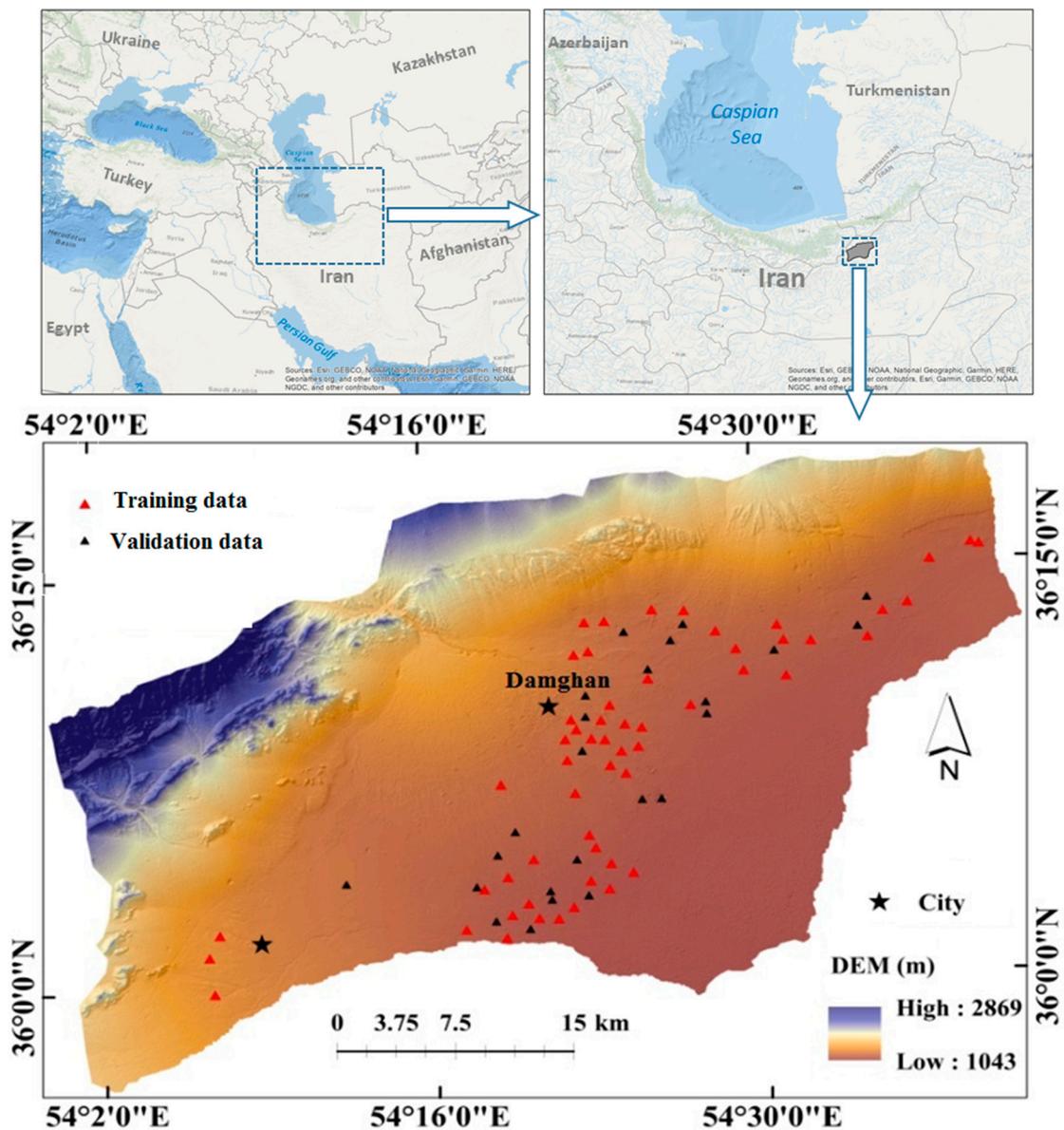
new methods which were used by the researcher for spatial hazards probability and groundwater potentiality modeling are: evidential belief function (EBF), weights of evidence (WoE), frequency ratio (FR), classification and regression tree (CART), boosted regression tree (BRT), decision tree (DT), artificial neural network (ANN), multivariate adaptive regression splines (MARS), binary logistic regression (BLR), Shannon's entropy (SE), analytic hierarchy process (AHP), maximum entropy (ME), random forest (RF), fuzzy logic (FL), support vector machine (SVM), multi-criteria decision analysis (MDCA), logistic model tree (LMT), quadratic discriminate analysis (QDA), K-nearest neighbor (KNN), and certainty factor (CF) [16–22].

In this work, we have used probabilistic, machine learning, data mining, and MDCA methods, namely WoE, BLR, SVM, RF, and a technique for order preference by similarity to ideal solution (TOPSIS). The outcomes of the same model vary depending on the physiographical situation in different regions. The suitable models help to demarcate the areas having groundwater potentiality. The models used in this research are accessible and efficiently capable of groundwater modeling and are used in various areas for environmental management [23–26]. Thus, the study aims to recognize the GWPZ using five models (RF, TOPSIS, WoE, SVM, and BLR,) in the Damghan sedimentary plain of Semnan province in Iran. The current study will help in determining the proper groundwater resource and to the decision-maker for managing the water resources.

## 2. Materials and Methods

### 2.1. Study Area

Damghan sedimentary plain, located within the Semnan province in Iran, covers an area of 1559 km<sup>2</sup>. Geographically, this plain region stretches from 35° 56' to 36° 18'N latitude and 54° 00'E to 54° 40'E longitude (Figure 1). The long-term average of precipitation and long-term evaporation are about 151.01 and 3000 mm, respectively [27]. The arid climate prevails in this plain because the annual evaporation is greater than annual precipitation [28]. The average temperature in the mountainous portion of the study area is 9.8 °C, and in the plain area, the mean temperature is 23.5 °C. In the south of Alborz zone, the upland area of the watershed is extended, and the plain's elevation ranges from 2860 m. a.s.l. in the northwest, to 1043 m a.s.l. in the southeast. Major portions of the study region are composed of Quaternary deposits [29]. The remaining parts of the plain are situated in the Alborz region and are covered by calcrete layers such as Cretaceous formations, as well as sandstone and Paleogene-related conglomerates. The low elevated area, composed of Quaternary deposits, has a high-water yield and recharge rate because of sediment nature and succession [30]. Nevertheless, the upland region in the Alborz zone is not suitable for recharge [31]. The mean depth of alluvial sediment varies from 150 m in north to 240 m in the south. In this area, the unconfined aquifer and bedrock consist of Neogene alluvium, such as marl and conglomerate, and the well logs set out the type of sediment.



**Figure 1.** Location of the study area in Iran and Semnan province and location of training and validation wells in the study area.

The sedimentary plain of Damghan is located in arid and semi-arid regions and facing the problem of water supply such as other arid regions. The main source of freshwater is sub-surface water storage and undergoes the problems of over pumping and lowering of groundwater. In the recent decade, due to excessive groundwater exploitation for irrigation and industrial purposes combining with the decreasing amount of rainfall, the water table is coming down rapidly. Therefore, immediate planning is needed to conserve the groundwater [30]. In this respect, the delineation of GWPZ is essential for proper planning and sustainable management of water.

## 2.2. Methodology

For assessing the groundwater potentiality (GWP), some spatial and non-spatial data have been gathered to prepare different datasets for modeling and validation of results. The data consists of two, i.e., primary and secondary, data. Primary data are pumping tests and yield measurements in the field. The secondary data are topographical map (scale 1:50,000), lithological map (scale

1:100,000), Sentinel 2A, Phased Array type L-band synthetic aperture radar (PALSAR) digital elevation model (DEM), rainfall of different metrological station of last 30 years, well location data from Water Resource Management, Iran, soil data from soil department of Iran. Thematic maps of all the data were extracted and analyzed by the RS and GIS. The present work methodologically consists of four phases (Figure 2) including; (1) preparation of groundwater inventory database thematic data layers of the groundwater conditioning factors including elevation, slope, aspect, CI, rainfall, lithology, soil type, LU/LC, Dd, distance to river, distance to fault, distance to road, NDVI, TPI, TWI, and SPI; (2) multicollinearity assessment of the effective groundwater determining factors (GWDFs); (3) application of models and preparation of GWPMs. The GWPMs were classified according to the four classification methods, namely quantile, natural breaks, equal interval, and geometrical interval, into four different groundwater susceptibility classes, including low, medium, high, and very high. By comparing the results of each classification method and the distribution of training and validation wells on the high and very high groundwater susceptibility classes, it was found that the natural break classification method gave the most accurate distribution. This agrees with the findings by Arabameri et al. [32], in that natural break method is a good classifier in susceptibility mapping; and (4), evaluation of the models performances using area under receiver operating characteristics (AUROC) curve, sensitivity (SE), specificity (SP), accuracy (AC), mean absolute error (MAE), root mean square error (RMSE) and seed cell area index (SCAI) methods.

### 2.3. Data Preparation

#### 2.3.1. Groundwater Inventory Map (GWIM)

The groundwater inventory database is of a key role in groundwater potentiality mapping. An inventory map is a target variable for any spatial modeling [32]. The well inventory database was prepared after extensive field visit with a hand GPS (global positioning system), and yield data were collected from the Department of Water Resources Management, Iran. Groundwater wells, with high yield of  $\geq 11 \text{ m}^3 \text{ h}^{-1}$  by pumping test analysis, have been considered for the GWPM. As a result, 80 wells have recognized in the study area. 56 wells (70%) of this dataset, were randomly selected to produce the GWPM models [32], whereas the remaining 24 (30%) wells were considered for validation of GWPMs [11]. The training and testing wells locations have been mentioned in Figure 1.

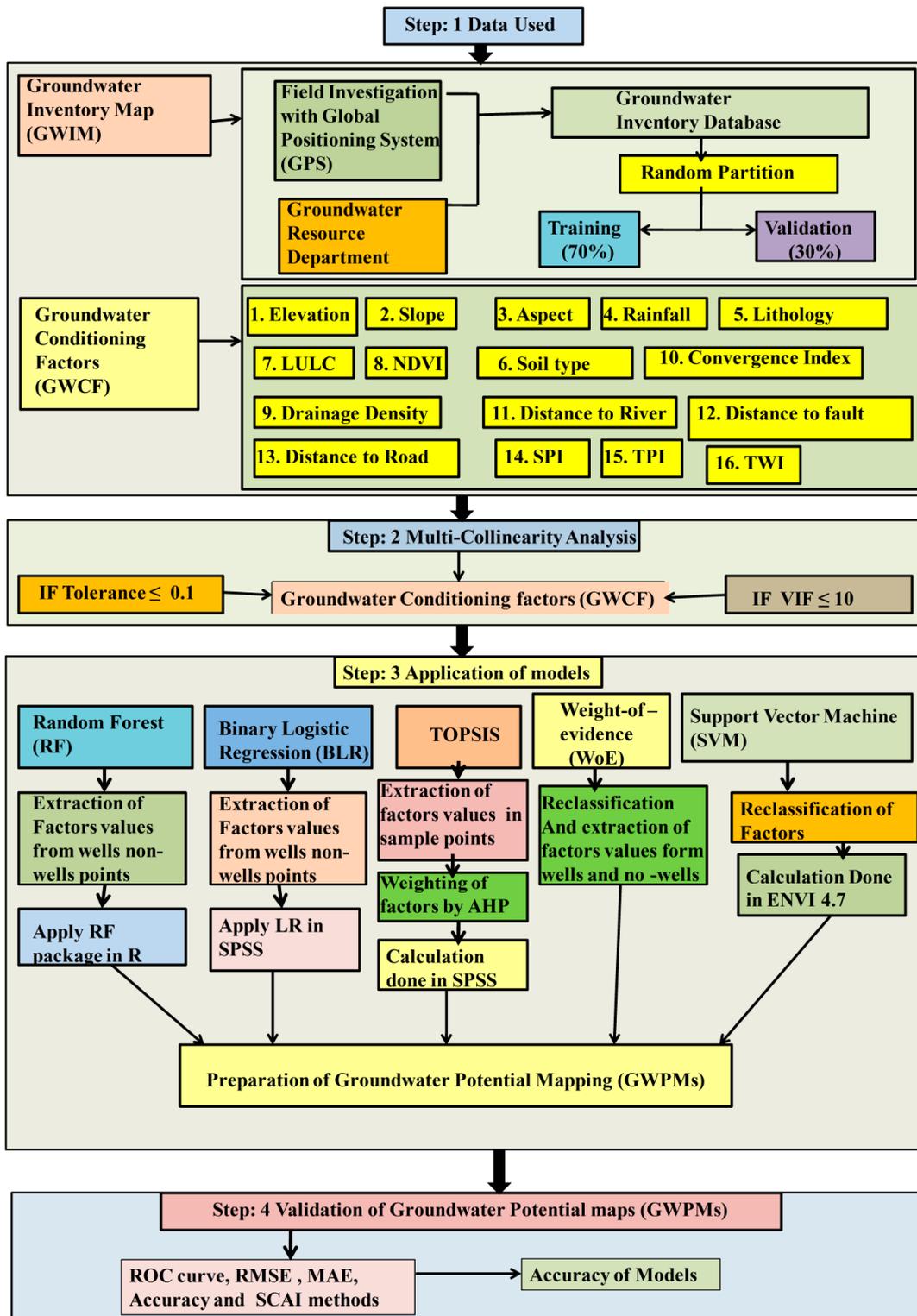


Figure 2. Methodological flowchart of the present work.

### 2.3.2. Groundwater Determining Factors (GWDFs)

The different geo-environmental components play a crucial role in determining the status of groundwater. GWPM represents the association between GWDFs and well locations [21,22]. For the GWPM, 16 GWDFs have been selected including elevation, slope, aspect, CI, rainfall, lithology, soil type, LU/LC, NDVI, Dd, distance to the river, distance to fault, distance to road, TWI, SPI and TPI

(Figure 3a–p). The PALSAR DEM (12.5 m resolution) downloaded from the Alaska Satellite Facility (ASF) Distributed Active Archive Center (DAAC). In this study, PALSAR DEM was used to extract the topographical, hydrological factors such as elevation, slope, aspect, CI, drainage, TWI, SPI, and TPI. The slope, aspect, and elevation are the major topographic components, used to determine the groundwater potentiality, erosion probability, etc. [21]. The DEM has been used as the elevation dataset (Figure 3a). The altitudinal fluctuation controls climatic conditions and helps to induce various vegetation types and soil development [33]. The slope data layer has been derived from PALSAR DEM by spatial analysis in the GIS environment (Figure 3b). In the same way, the aspect map has also been extracted from PALSAR DEM imagery using a spatial analysis tool (Figure 3d). CI is an important terrain factor that demonstrates the arrangement of relief as a set of channels and ridges. It is developed by Kiss [34]. The convergence index (CI) has been calculated using Equation (1).

$$CI = \left( \frac{1}{8} \sum_{i=1}^8 \theta_i \right) - 90^\circ, \quad (1)$$

where  $\theta$  indicates the average angle between the aspect of adjacent cells and the direction to the central cell. The CI value ranges from  $-100$  to  $+100$  (Figure 3c). The rainfall map was prepared by the kriging method considering the last 10-year annual rainfall of different stations (Figure 3e). The drainage was extracted from the topographical map and PALSAR DEM imagery. The Dd was computed based on Horton's morphometric formula (Equation (2)).

$$Dd = \frac{L_u}{A}, \quad (2)$$

where  $L_u$  means the total length of all orders streams,  $A$  is the area in square kilometer. Finally, the spatial data layer of the Dd has been built using the IDW interpolation method in the GIS environment (Figure 3f). The fault layer has been taken out from Landsat 7 imagery in ENVI software. The road network has been taken off from the topographical map and Google Earth imagery. The distance to river, fault, road data layers have been built using the Euclidian distance buffering tool and expressed in km (Figure 3g,h,i) [11]. The lithological information for the study area was gathered from the geological department of Iran [29]. The lithology map has been prepared by the digitization process (Figure 3j). Geologically, the region is composed of nine geological segments, namely A, B, C, D, E, F, G, H, and I (Figure 3j and Table 1). Soil data was collected from the soil department of Iranian and with the help of the digitized process, the thematic dataset of soil has been produced (Figure 3k). The LU/LC map has been produced from Sentinel 2A satellite image (12/08/2017) of 10 m, 20 m, and 60 m spatial resolution for each band using the supervised image classification method (Figure 3l). NDVI has been computed from satellite image (Figure 3m) using Equation (3):

$$NDVI = \frac{NIR - Red}{NIR + Red}, \quad (3)$$

where NIR is the near-infrared band or band 8 and red band or band 4. The TWI directly affects the topographic conditions, which control the hydrological process. TWI is the function of slope and the upstream area per unit width orthogonal to the direction of flow [35]. TWI plays a major role in the spatial heterogeneity of hydrological conditions such as soil moisture, underwater flow and slope steady-state [32]. The TWI has been introduced by Beven and Kirkby [36]. TWI is calculated from Equation (4):

$$TWI = \ln(A_S / \tan \beta), \quad (4)$$

where  $A_S$  represents the cumulative area of the catchment ( $m^2 m^{-1}$ ) and  $\beta$  is the slope gradient (degrees). The TWI value ranges from 1.11 to 21.54 (Figure 3n). The SPI is a calculation of water flow erosive power based on the assumption that discharge is commensurate with a given catchment

area [37]. One of the most important factors in controlling slope erosion processes is SPI. Regions with high stream power have high erosion potentiality [38]. From Equation (5), SPI has been calculated:

$$\text{SPI} = A_S \times \tan \beta, \quad (5)$$

where  $A_S$  is the upstream contributing area and  $\beta$  is slope gradient (in degrees). The spatial allocation of SPI ranges from 6.27 to 24.44 (Figure 3o) in the research area. TPI is defined as the difference between the middle point elevation ( $Z_0$ ) and the average elevation ( $\bar{Z}$ ) in a predetermined radius around it ( $R$ ) [39]:

$$\text{TPI} = Z_0 - \bar{Z}, \quad (6)$$

$$\bar{Z} = \frac{1}{n_R} \sum_{i \in R} Z_i. \quad (7)$$

The TPI has positive and negative value; a positive value demonstrates that the midpoint is located at a higher place than its average while a negative value indicates a lower place than the average. The TPI range depends not only on variations in altitude but also on landscape units ( $R$ ) [40]. Where large  $R$  values mainly depend on the main units of landscape, and small  $R$  values show up lower valleys such as small valleys and ridges. The TPI value ranges from 12.16 to 14.67 in this plain (Figure 3p). Spatial resolutions of the selected GWDFs are not the same. For preparing the groundwater potential maps of the study area the resolution of PALSAR DEM, i.e., 12.5 m\* 12.5 m has been selected as the base scale and all the GWCFs of which scale are greater or lesser than the PALSAR DEM have been resembled into a 12.5 m\* 12.5 m resolution. The data layers of elevation (Figure 3a), slope (Figure 3b), CI (Figure 3c), rainfall (Figure 3e), Dd (Figure 3f), distance to river (Figure 3g), distance to fault (Figure 3h), distance to road (Figure 3i), TWI (Figure 3n), SPI (Figure 3p), TPI (Figure 3o), and NDVI (Figure 3m) have been categorized into five sub-classes using the natural break classification method in GIS environment (Table 2). Aspect (Figure 3d), lithology (Figure 3j), soil type (Figure 3k), and LU/LC (Figure 3l) are the categorical factors. The categorical factors are also mentioned in Table 2.

**Table 1.** Description of lithology units in the study area.

Group	Unit	Description
A	COm	Dolomite platy and flaggy limestone containing trilobite; sandstone and shale (MILA FM).
	Cl	Dark red medium-grained arkosic to subarkosic sandstone and micaceous siltstone (LALUN FM).
B	Dckh	Yellowish, thin to thick-bedded, fossiliferous argillaceous limestone, dark grey limestone, greenish marl and shale, locally including gypsum
	Db	Grey and black, partly nodular limestone with intercalations of calcareous shale (BAHRAM FM).
C	Els	Sandstone, conglomerate, marl and sandy limestone.
	Ek	Well bedded green tuff and tuffaceous shale (KARAJ FM).
D	Jl	Light grey, thin-bedded to massive limestone (LAR FM).
	K2m,l	Marl, shale and detritic limestone.
E	K	Cretaceous rocks in general.
	Murmg	Gypsiferous marl.
F	Murc	Red conglomerate and sandstone.
	Plc	Polymictic conglomerate and sandstone.
G	PlQc	Fluvial conglomerate, Piedmont conglomerate and sandstone.
	P	Undifferentiated Permian rocks.
H	Pr	Dark grey medium-bedded to massive limestone (RUTEH LIMESTONE).
	Qft2	Low level piedmont fan and valley terrace deposits.
H	Qft1	High level piedmont fan and valley terrace deposits.
	Qcf	Clay flat.
I	Qal	Stream channel, braided channel and flood plain deposits.
	TRJs	Dark grey shale and sandstone (SHEMSHAK FM).

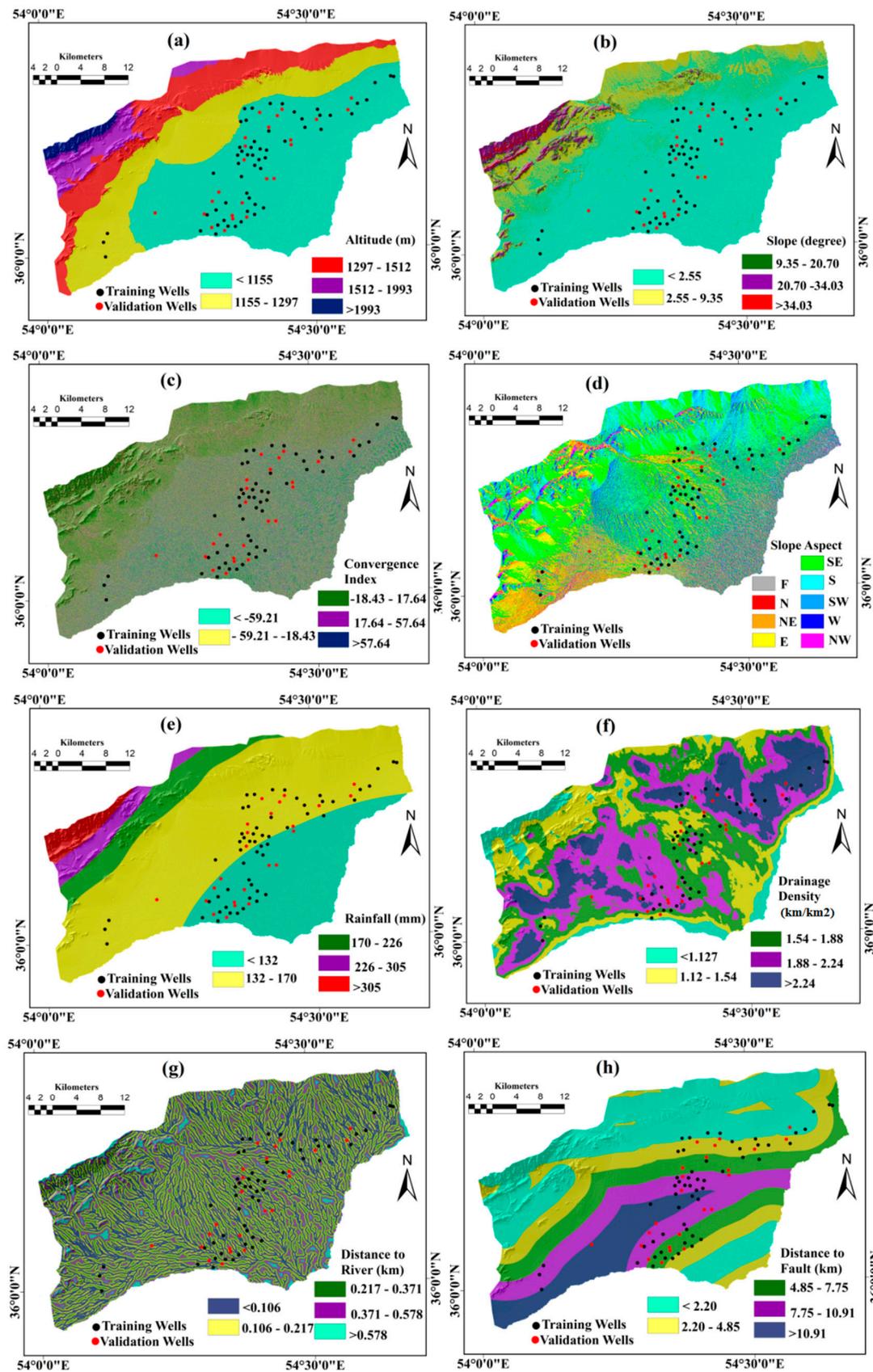
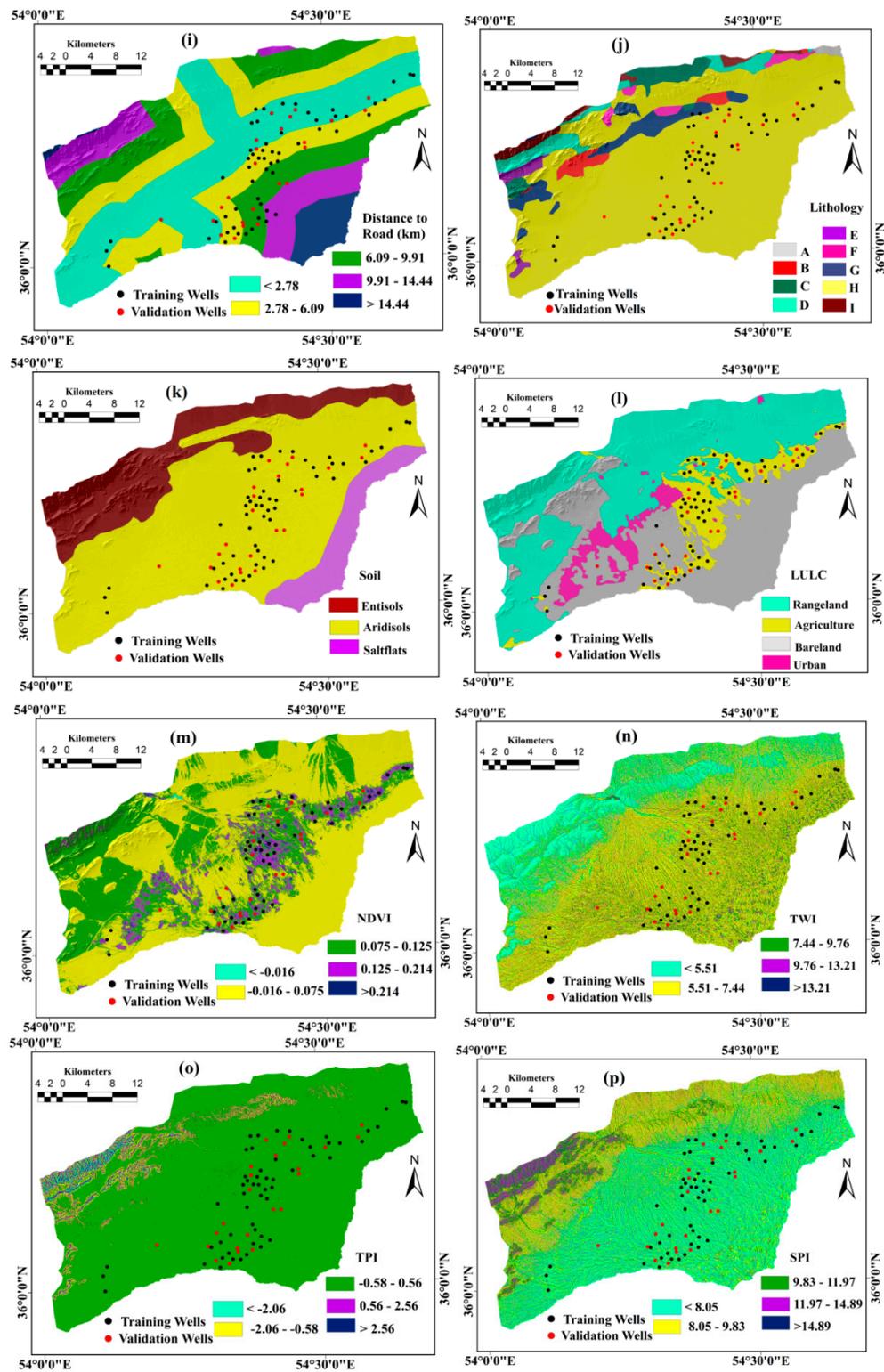


Figure 3. Cont.



**Figure 3.** Groundwater determining factors: (a) elevation, (b) slope, (c) aspect, (d) convergence index, (e) rainfall, (f) drainage density, (g) distance to river, (h) distance to fault, (i) distance to road, (j) lithology, (k) soil type, (l) land use/land cover (LULC), (m) normalized difference vegetation index (NDVI), (n) topographic wetness index (TWI), (o) topographic position index (TPI), (p) stream power index (SPI).

**Table 2.** Computation of statistics and classes of groundwater determining factors (GWDFs).

Factors	Min.	Max.	Classes	Methods
Elevation (m)	1043	2869	(1.) <1155, (2.) 1155–1297, (3.) 1297–1512, (4.) 1512–1993, (5.) >1993	Natural break (Jenks)
Slope (degree)	0	72.32	(1.) <2.55, (2.) 2.55–9.35, (3.) 9.35–20.70, (4.) 20.70–34.03, (5.) >34.03	Natural break (Jenks)
Aspect	-	-	(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
Convergence index	-100	100	(1.) <–59.21, (2.) –59.21–18.43, (3.) –18.43–17.64, (4.) 17.64–57.64, (5.) >57.64	Natural break (Jenks)
Rainfall (mm)	96	406	(1.) <132.95, (2.) 132.95–170.69, (3.) 170.69–226.68, (4.) 226.68–305.81, (5.) >305.81	Natural break (Jenks)
Lithology	-	-	(1.) A, (2.) B, (3.) C, (4.) D, (5.) E, (6.) F, (7.) G, (8.) H, (9.) I	Lithological Units
Soil type	-	-	(1.) Aridisols, (2.) Rock outcrops/entisols, (3.) Salt flats	Soil types/Orders
LULC	-	-	(1.) Bare land, (2.) Agriculture land, (3.) Rangeland, (4.) Urban	Supervised Classification
Drainage density (km/km <sup>2</sup> )	0.15	3.18	(1.) <1.12, (2.) 1.12–1.54, (3.) 1.54–1.88, (4.) 1.88–2.24, (5.) >2.24	Natural break (Jenks)
Distance to river (km)	0	1.35	(1.) <0.10, (2.) 0.10–0.21, (3.) 0.21–0.37, (4.) 0.37–0.57, (5.) >0.57	Natural break (Jenks)
Distance to fault (km)	0	16.08	(1.) <2.20, (2.) 2.20–4.85, (3.) 4.85–7.75, (4.) 7.75–10.91, (5.) >10.91	Natural break (Jenks)
Distance to road (km)	0	22.18	(1.) <2.78, (2.) 2.78–6.09, (3.) 6.09–9.91, (4.) 9.91–14.44, (5.) >14.44	Natural break (Jenks)
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	Natural break (Jenks)
TWI	1.11	21.54	(1.) <5.51, (2.) 5.51–7.44, (3.) 7.44–9.76, (4.) 9.76–13.21, (5.) >13.21	Natural break (Jenks)
TPI	–12.16	14.67	(1.) <–2.06, (2.) –2.06–0.58, (3.) –0.58–0.56, (4.) 0.56–2.56, (5.) >2.56	Natural break (Jenks)
SPI	6.27	24.44	(1.) <8.05, (2.) 8.05–9.83, (3.) 9.83–11.97, (4.) 11.97–14.89, (5.) >14.89	Natural break (Jenks)

## 2.4. Models

### 2.4.1. Weight of Evidence (WoE) Model

The WoE model is the main Bayesian probability system model in linear logic form and uses non-conditional and conditional probabilities [41]. WoE reveals the spatial association between dependent variable, i.e., well locations and independent variables, i.e., GWDFs. The weight of each class has been assigned by this method using the following equations (Equations (8)–(14)) [42]:

$$W_{WoE} = \frac{C}{S(C)}, \quad (8)$$

$$C = W_i^+ - W_i^-, \quad (9)$$

$$W_i^+ = \ln \frac{P(B/D)}{P(B/\bar{D})}, \quad (10)$$

$$W_i^- = \ln \frac{P(\bar{B}/D)}{P(\bar{B}/\bar{D})}, \quad (11)$$

$$S(C) = \sqrt{S^2(W^+) + S^2(W^-)}, \quad (12)$$

$$S^2(W^+) = \frac{1}{P(B/D)} + \frac{1}{P(B/\bar{D})}, \quad (13)$$

$$S^2(W^-) = \frac{1}{P(\bar{B}/D)} + \frac{1}{P(\bar{B}/D)}, \quad (14)$$

where  $P(B|D)$  is the conditional probability of  $B$  occurring given the presence of  $D$ ,  $B$  is the datasets of GWDFs related to the presence of groundwater well and  $\bar{B}$  indicates the groundwater is absent of the datasets of groundwater conditioning factors.  $D$  indicates the presence of well while  $\bar{D}$  stand for the absence of a well, and  $P$  is the probability. Whereas,  $W_i^+$  is a positive weight of GWDFs for groundwater occurrence. Conversely,  $W_i^-$  is a negative weight with respect to the absence of groundwater well (unfavorable factors). WoE computation has been started by the pixels counting process between groundwater well locations and GWDFs. The weighted GWDFs factors have been summed up in the raster calculation to generate the single layer of GWPM in the GIS environment using the following Equation (15).

$$\begin{aligned} \text{GWMP}_{\text{WoE}} = & (W_{\text{WoE}}\text{Elevation}) + (W_{\text{WoE}}\text{Slope}) + (W_{\text{WoE}}\text{Aspect}) + (W_{\text{WoE}}\text{Convergence Index}) \\ & + (W_{\text{WoE}}\text{Rainfall}) + (W_{\text{WoE}}\text{Drainage Density}) + (W_{\text{WoE}}\text{Distance to River}) + \\ & (W_{\text{WoE}}\text{Distance to Fault}) + (W_{\text{WoE}}\text{Distance to Road}) + (W_{\text{WoE}}\text{Lithology}) + \\ & (W_{\text{WoE}}\text{Soil Type}) + (W_{\text{WoE}}\text{LULC}) + (W_{\text{WoE}}\text{NDVI}) + (W_{\text{WoE}}\text{TWI}) + \\ & (W_{\text{WoE}}\text{TPI}) + (W_{\text{WoE}}\text{SPI}) \end{aligned} \quad (15)$$

#### 2.4.2. Random Forest (RF)

RF is the non-parametric multivariate model [43], which can be used for the analysis of regression and classification and variable selections. RF model creates thousands of trees, forming a ‘forest’ based on the decision rule. Each tree in the RF model depends on a sample of bootstrapped of data using a CART process with a random subset of variables selected at each node. The final decision of the class membership and model (output) has determined according to the majority priority of all decision trees [44]. The trees’ ensembles would have performed much better than a single tree. It is important to know that the program can be run by a large number of trees with taking large and too many computational requirements [45]. RF is a very reliable and flexible ensemble classifier, which depends upon the decision trees, that have so many attractive performances such a minimum costly, minimum tendencies for overfitting and also capability of the work with very high dimensional data [46]. The RF model is also a very fast machine learning solution, allowing a highly accurate classification with internal unbiased generalizability estimation during the process of forest construction [47]. The basic merits of RF model arise when the program proceeds including (i) no need of any assumptions regarding the data distribution, (ii) no overfitting problem, (iii) in case of single tree, a low correlation estimated, while the diversity of forest increases the usages of a number of factors, (iv) helps to estimate negative or error using ‘out-of-bag’ (OOB) data, (v) averages a large number of trees, resulting the low bias and low variance, subsequently, (f), resulting in the excellent prediction for performances [43,48]. Besides, the numerical and categorical data can be incorporated in the RF model. Using the OOB error-index, the variance and covariance between the grids cells can be estimated [48]. The predication values of this model are estimated by the huge amount of decision trees [43]. The presence and absence groundwater wells among the GWDFs can easily be estimated by RF model. In this algorithm, the mean decrease accuracy and Gini are estimated by the RF model to analysis the variable importance of the GWDFs [30]. This algorithm calculates untouched the proper count, the amount of correct classification applying the data out-of-bag as its test sample. In the out-of-bag instances, the values of the attributes are then randomly permuted. A new set of data will then be checked for proper classification. The average of this number is the raw importance score for the specific attribute over all trees in the forest. Therefore, factors importance in RF model is computed for variable  $Y_i$  by out of bag error (OOB) [49]. Factors importance of  $Y_i$  can be calculated using the Equation (16):

$$\text{VImp}(Y_j) = \frac{1}{n_{\text{tree}}} \sum_t \text{errOOB}_t^j - \text{errOOB}_t, \quad (16)$$

where  $n_{tree}$  stands for the number of trees,  $V_{Imp}(Y_i)$  denotes variable importance for variable  $Y_i$ ,  $err_{OOB_t}$  is an error when all the factors are included, and  $err_{OOB_t}^j$  denotes an error after the removal of the variable  $j$ . The Gini index was used to measure the variable significance based on the number of times that variable is picked by all trees [47,50]. In this study, the ‘randomForest’ package in R program has been installed and run the RF model for estimating the GWPZ [51]. Finally, the RF model-based GWPM has been produced in the GIS environment.

#### 2.4.3. Binary Logistic Regression (BLR)

BLR model is the most common statistical model which considers both dichotomous and continuous variables. However, the practical dependent variable must be in binary form, i.e., 1 and 0. Where, 0 represents the absence of groundwater well and 1 stands for the presence of the groundwater well [37,52]. For GWPM, it corresponds with the Bernoulli method, which determines the high groundwater potentiality over space depending on the Bernoulli probability [32]. The main target of the BLR analysis is to chalk out the correct and appropriate prediction of samples and probe the correlation between a dependent variable with a set of independent variables [32]. Among the different methods of the regressions, the BLR is fitting a logistic curve or function concerning data. As a result, BLR estimates the values that vary from 0 to 1, while 1 means the presence of groundwater well, conversely 0 means the chances of occurrence of groundwater well is nil. In this method, the target value is calculated using Equation (17):

$$Y = \text{Logit}(P) = \text{Ln}\left(\frac{P}{1-p}\right) = C_0 + C_1 \times X_1 + C_2 \times X_2 + \dots \dots \dots C_n \times X_n, \tag{17}$$

where Logit is the link function,  $P$  is the probability of occurrence of groundwater well ( $y$ ),  $p = 1 - p$  are the odds of groundwater occurrence (or probability of presence divided by the probability of absence) the,  $C_0$  is the model intercept and  $(C_1, \dots, C_n)$  are the regression coefficients for each GWCF ( $X_1, \dots, X_n$ ) [32]. In this contribution, the BLR model has been applied in R by using the ‘glm’ function based on the “stats” package [32]. In this study, the random point’s values have been extracted from each variable of GWCFs for presence and absence condition of the groundwater. Finally, GWPM by BLR model has been produced in GIS with regarding the prediction database.

#### 2.4.4. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

The TOPSIS method was introduced by Hwang and Yoon [53]. Presently, it is an important multi-criteria decision approach among the several MCDA processes utilized for water management practices [54]. This approach relies on the premise that the best alternative solution should have the shortest distance Euclidean from the positive ideal solution and the longest distance from the negative solution [32]. The ideal solution in GWPM is the best model to distinguish between groundwater well presences and absences. TOPSIS is a method that cannot understand categorical properties [55]. That is why the AHP has been used to assign the weight to each GWDF. The computation of the entire TOPSIS model was carried out step by step as follows:

Step 1: Preparation of a decision matrix with  $m$  criteria and  $n$  alternatives using Equation (18):

$$A_{ij} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}. \tag{18}$$

Step 2: Normalization of decision matrix using Equation (19):

$$r_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^m a_{ij}^2}}, \tag{19}$$

where  $i = 1, \dots, m; j = 1 \dots n$ .

Step 3: Determine the weight of criteria using the AHP model: AHP, first implemented by Saaty [56] is one of the most comprehensive MDCA approaches. This method assists decision-makers in receiving quantitative and qualitative parameters. The pairwise comparisons help for the judgment and computation of GWCFs [57]. After preparing paired comparisons, the resulting paired comparison matrix is normalized using Equation (20) and then the final weight ( $W_i$ ) of each parameter is obtained using Equation (21).

$$r_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^m a_{ij}}}, \quad (20)$$

$$W_i = \frac{\sum_{i=1}^n r_{ij}}{n}. \quad (21)$$

One of the advantages of this method is to show the inconsistency [58]. To assess the degree of weighting precision, an Inconsistency Index is used. The consistency test shows how much trust can be put in the priorities of a matrix. If this value is  $>0.1$ , this means that it is not consistent with the specified weights and should be checked. Inconsistency ratio is often being used to calculate the Inconsistency in judgments, Equations (22) and (24) have been used for its calculation:

$$IR = \frac{I.I}{I.I.R'}, \quad (22)$$

$$I.I = \frac{\lambda_{\max} - n}{n - 1}, \quad (23)$$

$$\lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \frac{\bar{a} \times W_{(ij)}}{W_{(ij)}}, \quad (24)$$

where IR refers to the inconsistency ratio, I.I. is an inconsistency index,  $n$  is the number of criteria,  $a$  is the geometric mean of matrix and  $W_{(ij)}$  is weight vector.

Step 4: Calculation of the weighted normalized decision matrix using Equation (25):

$$V_{ij} = r_{ij} \times W_j, \quad (25)$$

where  $W_j$  represents the weight of the criteria.

Step 5: Calculation of the positive and negative ideal solution using Equations (26) and (27), respectively:

$$A^+ = \left\{ \left( \max_{j \in J} V_{ij}, \left( \min_{j \in J'} V_{ij} \right) \right) ; i = 1, 2, \dots, m \right\} \\ = \left\{ V_1^+, V_2^+, \dots, V_j^+, \dots, V_n^+ \right\}, \quad (26)$$

$$A^- = \left\{ \left( \min_{j \in J} V_{ij}, \left( \max_{j \in J'} V_{ij} \right) \right) ; i = 1, 2, \dots, m \right\} \\ = \left\{ V_1^-, V_2^-, \dots, V_j^-, \dots, V_n^- \right\} \quad (27)$$

where,  $j$  and  $J'$  are related to increasing and decreasing criteria, respectively, where  $J$  is associated with the positive criteria and  $J'$  is associated with the negative criteria.

Step 6: Calculation of the distance from the positive and negative ideal solution using Equations (28) and (29), respectively:

$$d_{i+} = \sqrt{\sum_{j=1}^n (V_{ij} - V_j^+)^2}; i = 1, 2, \dots, m, \quad (28)$$

$$d_{i-} = \sqrt{\sum_{j=1}^n (V_{ij} - V_j^-)^2}; i = 1, 2, \dots, m. \quad (29)$$

Step 7: Calculation of the relative closeness to the ideal solution using Equation (30):

$$cl_{i+} = \frac{d_{i-}}{d_{i+} + d_{i-}}; 0 \leq cl_{i+} \leq 1; i = 1, 2, \dots, m, \quad (30)$$

where  $cl_{i+}$  is the closeness coefficient,  $d_{i+}$  is a positive ideal solution (PIS), and  $d_{i-}$  is negative ideal solution (NIS).

The value of  $cl_{i+}$  ranges between 0 and 1. The larger the  $cl_{i+}$  value indicates the better the performance of the alternatives. In this contribution, to perform the Mathematical calculation, 500 points have been randomly selected and derived the values of GWDFs for each point, then a table was made, consisting of 16 GWDFs columns and 500 rows. Subsequently, these values have entered into SPSS and done the process. Ultimately, the TOPSIS based GWPM has been prepared considering the point values using the IDW method in the GIS environment.

#### 2.4.5. Support Vector Machine (SVM)

SVM is the supervised learning system of machine learning associated with learning algorithms that analyze the data used for classification and regression analysis. It is developed by Bai et al. [59]. SVM helps in the transformations of nonlinear covariates into a higher dimensional feature space [60]. SVM is also a statistical learning theory associated with a training phase in which a training dataset of related input and target output values trains the model. The trained model will then be used to analyze a separate set of test data. SVM has two main underlying concepts for discriminating the problems, i.e., the optimum linear separating hyper-plane that separates patterns of data, and another is the kernel functions that convert the original nonlinear data pattern to a linearly separable format in a high-dimensional feature space [60]. A set of linear separable training vectors  $x_i$  ( $i = 1, 2, \dots, n$ ) consists of two classes, which are denoted as  $y_i = \pm 1$ . The SVM's goal is to find an  $n$ -dimensional hyperplane that differentiates the two groups by the total distance.

Mathematically, it can be minimized as:

$$\frac{1}{2} \|w\|^2, \quad (31)$$

Subject to the following constraints:

$$y_i = ((w \cdot x_i) + b) \geq 1, \quad (32)$$

where  $\|w\|$  is the norm of the normal hyper-plane,  $b$  is a scalar base, and  $(w \cdot x_i)$  denotes the scalar product operation. Introducing the Lagrangian multiplier, the cost function can be defined as:

$$L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \lambda_i (y_i ((w \cdot x_i) + b) - 1), \quad (33)$$

where  $\lambda_i$  is the Lagrangian multiplier. It is possible to achieve the solution by double minimizing Equation (32). The standard procedures for  $w$  and  $b$  and detailed discussions can be found in Vapnik [61], Tax and Duin [62] and Yao et al. [60]. For non-separable case, one can change the constraints by setting up slack variables  $\xi_i$ :

$$y_i ((w \cdot x_i) + b) \geq 1 - \xi_i. \quad (34)$$

Equation (32) will be modified as:

$$L = \frac{1}{2} \|w\|^2 - \frac{1}{vn} \sum_{i=1}^n \xi_i, \quad (35)$$

where  $v[0, 1]$  was introduced to account for misclassification [63]. Besides, a kernel function  $K(x_i, x_j)$  was introduced by Vapnik [61] to account for the nonlinear decision boundary [63]. The two-class SVM method was used in this study because it was reported that Yao et al. [60] produced a more accurate map of susceptibility from the two classes of SVM. That's why Radial Basis Function (RBF) was used for kernel in this study and the two-class SVM model was first trained and then used to construct a GWPM. In this method, 1 and 0 values indicate the positive and negative relationship of groundwater occurrence. To perform the GWP mapping using the SVM, we used the ENVI 4.3. The default RBF kernel, which works well in most cases, has been used. In addition, in many studies and cases (especially in nonlinear problems), RBF provides better prediction results compared to other kernels [64]. Finally, the GWPM by SVM has been produced in the GIS.

### 2.5. Validation of Models

In this study, to analyze the potentiality and performance of the selected models, we have used two thresholds dependent methods i.e., ROC curve and SCAI. The ROC curve and SCAI are the significant and accurate justification methods of different models [65]. For this purpose, 30% validation and 70% training datasets have been considered by the ROC curve and SCAI methods (11). The area under curve (AUC) of the ROC method range between 0.5 to 1. If the value is nearest to 1, it indicates excellent prediction accurateness of the models [66]. The accuracy value that is AUC of the ROC is mentioned in Table 3. The AUC values have been calculated using the Equation (36). In the case of SCAI method, if the sub-class values of models decrease from very low to very high sub-classes, it indicates that models are suitable and acceptable [67].

$$AUC = \frac{\sum TP + \sum TN}{P + N}. \quad (36)$$

**Table 3.** Area under curve (AUC) values and statements.

AUC Values	Accuracy Statements
0.5–0.6	Low
0.6–0.7	Moderate
0.7–0.8	High
0.8–0.9	Very high
0.9–1	Excellent

Source: Yesilnacar [66].

We also used five statistical techniques in this analysis to test models' performance, including SE, SP, AC, MAE, and RMSE. Based on four possible consequences i.e., true positive (TP), false positive (FP), true negative (TN) and false negative (FN), sensitivity (Equation (37)), specificity (Equation (38)) and accuracy (Equation (39)) have been measured: TP and FP are the counts of well pixel that are correctly identified as well pixel and non-well pixel, respectively. On the other hand, TN and FN are the numbers of well pixel which are correctly classified and incorrectly classified as non-well class. SE is the ratio of the number of well pixels properly classified to the total number of well pixels predicted. SP is the ratio between the number of well pixels wrongly classified and the total non-well pixels predicted. AC is the ratio between the number of properly classified well and non-well pixels. MAE (Equation (40)) and RMSE (Equation (42)) indices have been considered to assess the disparity between the observed and predicted data. The high values of Sensitivity, Specificity, and Accuracy and low value of MAE and RMSE value indicate the good capability of the models [68–72]. The following five formulas have been used for statistical measures.

$$SE = \frac{TP}{TP + FN} \quad (37)$$

$$SP = \frac{TN}{TN + FP}, \quad (38)$$

$$AC = \frac{TP + TN}{TP + TN + FP + FN}, \quad (39)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_{predicted} - X_{actual}|, \quad (40)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{predicted} - X_{actual})^2}, \quad (41)$$

where  $X_{predicted}$  and  $X_{actual}$  is the predicted and real values in the training dataset or testing dataset of the groundwater potentiality models and  $n$  is the total number of samples in the training data set or testing dataset.

### 2.6. Sensitivity Analysis (SA)

It is very difficult to completely remove the uncertainty in the preparation of data layers [73–75]. Refsgaard et al. [75] was used different techniques e.g., Monte Carlo analysis, error propagation equations, sensitivity analysis (SA), scenario analysis, etc. for measuring the uncertainty. Sensitive analysis has been used in various studies [75–77] for the measurement of the effect of variable variations on model outputs, allowing then a quantitative assessment of the relative importance of uncertainty sources. In the present study, map removal sensitivity analysis (MRSA) method has been used, which was developed by Lodwick et al. [78]. The MRSA method would help to evaluate the sensitivity of the groundwater potentiality maps by removing one or more parameters from the groundwater potentiality maps. This technique has been used by several researchers to address the significant role of the effective factors [79–81]. It helps to identify the quantitative contribution of each groundwater conditioning factor to the uncertainty of the model output [72,82]. The percentage of contribution (PC) of each groundwater conditioning factor has been estimated by the MRSA method to explore the relative importance on the model output using the following Equation (42) [83]:

$$PC = \frac{(AUC_{all} - AUC_i)}{AUC_i} \times 100, \quad (42)$$

where  $AUC_{all}$  and  $AUC_i$  indicate the AUC values obtained from modeling groundwater potential model using all GWDFs and the model when the  $i$ th GWDF has been excluded.

## 3. Results

### 3.1. Analyzing the Multi-Collinearity (MC) of Groundwater Determining Factors

The MC problem reduces some linear models' predictive accuracy [84]. Techniques were applied in this study to assess the MC problem between GWDFs, namely tolerance (TOL) and inflation factor variance (VIF) [85]. Tolerance values of <0.1 and VIF of <10 reveal no MC problem among the GWDFs [86]. Roy and Saha [87] and Arabameri et al. [32] were used the MC test for the landslide susceptibility and groundwater potentiality mapping. The selected 16 GWDFs have been tested by SPSS. No MC problem has been found among the GWDFs, as no one value of tolerance and VIF does exceed the threshold limit (Table 4). Therefore, the selected GWDFs are suitable for the prediction of groundwater potentiality. Here, maximum tolerance and VIF values are 0.91 and 5.91 (Table 4).

**Table 4.** Multi-collinearity test of groundwater conditioning factors.

Conditioning Factors	Collinearity Statistics	
	Tolerance	VIF
Elevation	0.281	4.275
Slope	0.256	3.908
Convergence Index	0.816	1.226
Rainfall	0.202	4.792
Drainage Density	0.542	1.846
Distance to River	0.855	1.170
Distance to Fault	0.527	1.897
Distance to Road	0.485	2.061
NDVI	0.704	1.420
TWI	0.201	4.911
TPI	0.891	1.122
SPI	0.202	4.713
Aspect	0.916	1.092
Lithology	0.580	1.724
LULC	0.612	1.634
Soil Type	0.492	2.032

### 3.2. Application of the Weight of Evidence (WoE)

Groundwater potentiality reclines on the positive and negative effects of the effective groundwater determining factors. The positive value of WoE indicates the chances of storage of groundwater and vice-versa. The zero value of WoE means the sub-class of factors has no role in determining the groundwater occurrence [88]. The results of WoE model have been put in Table 5. The low altitudinal zone is more potential for the accumulation of groundwater than abrupt slope and higher altitudinal areas due to the high infiltration rate and less surface runoff [89]. For elevation, 1043–1155 m altitude with a value of 4.88 is showing the strongest positive effects among these GWDFs in making the areas potential to groundwater. On the contrary, the others sub-layers such as 1155–1297 m (WoE = −3.05), 1297–1512 m (WoE = 0), 1512–1993 m (WoE = 0), and >1993 m (WoE = 0) altitudinal levels are representing the negative and less effect in the presence of the groundwater (Table 5). Among the five slope classes, the <2.55-degree class has the maximum value of WoE i.e., 2.97 which depicts the strong control on the occurrence of groundwater. On the other hand, the remaining four sub-classes of slope have no control over the groundwater at all (Table 5). The north-east aspect has the highest WoE value (WoE = 1.84), which indicates the strong positive effects on the storage of groundwater. CI is the parameters of topography that reflect the elevation as a collection of convergent (channel) and divergent (ridge) areas. The CI value ranges from +100 to −100. Among the five classes, the two sub-class of CI such as <−59.21 (WoE = 2.02) and >57.64 (WoE = 2.02) has the strongest positive relationship, while others sub-layers have a negative relationship with groundwater storage (Table 5). Rainfall is an important groundwater potentiality determining factor. Rainfall classes <132 mm (WoE = 2.47) and 132 mm–170 mm (WoE = 0.40) have the strongest positive relationship. Lithologically, the region is composed of nine geological regions, namely A, B, C, D, E, F, G, H, and I. Only the H geological region (quaternary sediments) with WoE of 2.64 has a strong positive effect on the groundwater formation. Among the GWDFs, the soil is of crucial part in the groundwater recharge. Pedologically, the region is composed of three soil orders i.e., aridisols, entisols or rock outcrops, and salt flats. Comparatively, aridisols have a great contribution (WoE = 3.50) in the storage of groundwater. The LU/LC are categories into four types namely rangeland, bare land, agriculture and urban. Only agriculture land with WoE = 8.80 has a strong positive relationship, indicating the high potential of groundwater comparatively the bare land, urban and rangeland areas. Among the other GWDFs, the sub-classes of 1.88–2.24 km<sup>2</sup>/km<sup>2</sup> (WoE = 1.62) of drainage density, <0.10 km (WoE = 0.74) of distance to river, 7.75–10.91 km (WoE = 3.25) of distance to fault, 2.78–6.09 km (WoE = 2.25) of distance to road, 0.12–0.21 (WoE = 6.08) of NDVI,

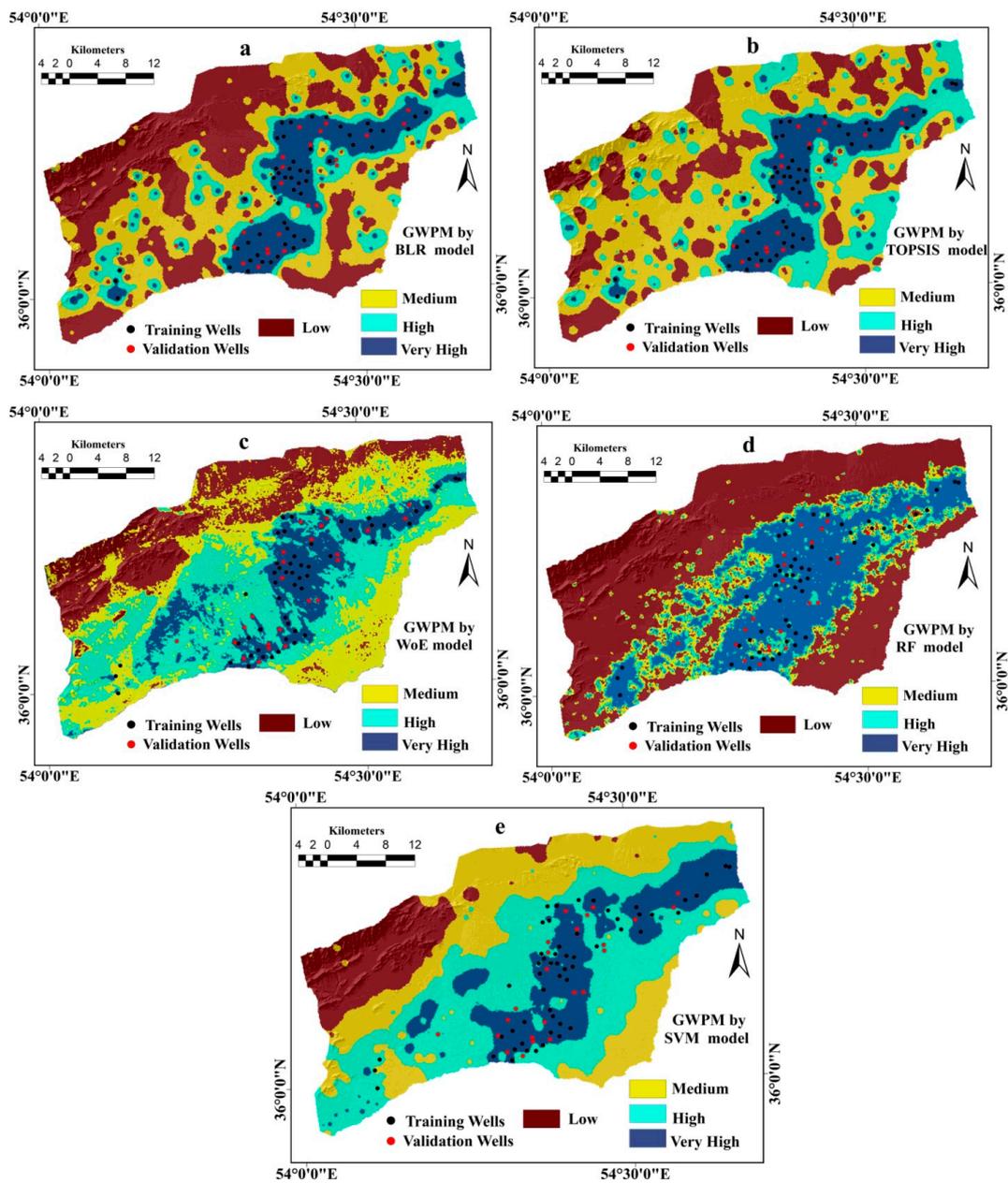
5.51–7.44 (WoE = 1.69) of TWI, −0.58–0.56 (WoE = 1.44) of TPI and <8.05 (WoE = 1.60) of SPI have the strong positive influence on the recharge of groundwater (Table 5). Subsequently, weights have been assigned to the sub-layers of GWDFs and converted as weighted WoE layers. All weighted GWDFs have been summed up and generated a single layer of GWPM (Figure 4c). The prepared GWPM has been classified into four categories i.e., low, medium, high and very high potential zones with the help of the natural break classification method (Figure 4c). The results of GWPM by WoE model are showing that only 297.33 km<sup>2</sup> (15.46%) area is of very high groundwater potentiality, followed by the 583.90 km<sup>2</sup> (30.36%) for high, 617.37 km<sup>2</sup> (32.1%) for medium and 424.85 km<sup>2</sup> (22.09%) for low GWPZs (Table 6 and Figure 5).

**Table 5.** The spatial relation between conditioning factors and well locations by weight of evidence model.

Elevation (m)	Pixels	% of Pixel	Well	% of Well	W+	W−	C	S2W+	S2W−	S⊙	C/S⊙
<1043–1155	855,560	49.36	53	94.64	0.65	−2.25	2.90	0.02	0.33	0.59	4.88
1155–1297	446,173	25.74	3	5.36	−1.57	0.24	−1.81	0.33	0.02	0.59	−3.05
1297–1512	303,221	17.50	0	0.00	0.00	0.19	0.00	0.00	0.02	0.00	0.00
1512–1993	101,149	5.84	0	0.00	0.00	0.06	0.00	0.00	0.02	0.00	0.00
>1993	27,036	1.56	0	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00
<b>Slope (degree)</b>											
<2.55	126,724,5	73.12	55	98.21	0.30	−2.71	3.01	0.02	1.00	1.01	2.98
2.55–9.35	335,864	19.38	1	1.79	−2.38	0.20	−2.58	1.00	0.02	1.01	−2.56
9.35–20.70	638,68	3.69	0	0.00	0.00	0.04	0.00	0.00	0.02	0.00	0.00
20.70–34.03	44,815	2.59	0	0.00	0.00	0.03	0.00	0.00	0.02	0.00	0.00
>34.03	21,347	1.23	0	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00
<b>Aspect</b>											
F	82,884	4.78	3	5.36	0.11	−0.01	0.12	0.33	0.02	0.59	0.20
N	89,279	5.15	3	5.36	0.04	0.00	0.04	0.33	0.02	0.59	0.07
NE	154,448	8.91	9	16.07	0.59	−0.08	0.67	0.11	0.02	0.36	1.85
E	296,877	17.13	10	17.86	0.04	−0.01	0.05	0.10	0.02	0.35	0.14
SE	431,538	24.90	8	14.29	−0.56	0.13	−0.69	0.13	0.02	0.38	−1.80
S	359,878	20.76	12	21.43	0.03	−0.01	0.04	0.08	0.02	0.33	0.12
SW	167,965	9.69	7	12.50	0.25	−0.03	0.29	0.14	0.02	0.40	0.71
W	853,21	4.92	3	5.36	0.08	0.00	0.09	0.33	0.02	0.59	0.15
NW	64,949	3.75	1	1.79	−0.74	0.02	−0.76	1.00	0.02	1.01	−0.75
<b>Convergence Index</b>											
<−59.21568627	145,566	8.40	9	16.07	0.65	−0.09	0.74	0.11	0.02	0.36	2.02
−59.21–−18.43	368,782	21.28	14	25.00	0.16	−0.05	0.21	0.07	0.02	0.31	0.68
−18.43–17.64	707,982	40.85	14	25.00	−0.49	0.24	−0.73	0.07	0.02	0.31	−2.36
17.64–57.64	364,974	21.06	10	17.86	−0.16	0.04	−0.20	0.10	0.02	0.35	−0.59
>57.64	145,835	8.41	9	16.07	0.65	−0.09	0.73	0.11	0.02	0.36	2.02
<b>Rainfall (mm)</b>											
<132	429,194	24.76	22	39.29	0.46	−0.21	0.68	0.05	0.03	0.27	2.47
132–170	100,66,02	58.08	34	60.71	0.04	−0.06	0.11	0.03	0.05	0.27	0.40
170–226	166,770	9.62	0	0.00	0.00	0.10	0.00	0.00	0.02	0.00	0.00
226–305	77,365	4.46	0	0.00	0.00	0.05	0.00	0.00	0.02	0.00	0.00
>305	53,208	3.07	0	0.00	0.00	0.03	0.00	0.00	0.02	0.00	0.00
<b>Lithology</b>											
A	7093	0.41	0	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
B	35,899	2.07	0	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00
C	57,180	3.30	0	0.00	0.00	0.03	0.00	0.00	0.02	0.00	0.00
D	72,339	4.17	0	0.00	0.00	0.04	0.00	0.00	0.02	0.00	0.00
E	23,837	1.38	0	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00
F	24,485	1.41	0	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00
G	86,958	5.02	2	3.57	−0.34	0.02	−0.36	0.50	0.02	0.72	−0.49
H	138,8009	80.09	54	96.43	0.19	−1.72	1.90	0.02	0.50	0.72	2.64
I	37,340	2.15	0	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00
Aridisols	118,6872	68.48	54	96.43	0.34	−2.18	2.52	0.02	0.50	0.72	3.50
Rock Outcrops/Entisols	392,588	22.65	0	0.00	0.00	0.26	0.00	0.00	0.02	0.00	0.00
Salt Flats	153,679	8.87	2	3.57	−0.91	0.06	−0.97	0.50	0.02	0.72	−1.34

Table 5. Cont.

Elevation (m)	Pixels	% of Pixel	Well	% of Well	W+	W−	C	S2W+	S2W−	SC	C/SC
<b>LULC</b>											
Bareland	654,072	37.74	4	7.14	−1.66	0.40	−2.06	0.25	0.02	0.52	−3.98
Agriculture	206,538	11.92	52	92.86	2.05	−2.51	4.57	0.02	0.25	0.52	8.80
Rangeland	777,361	44.85	0	0.00	0.00	0.60	0.00	0.00	0.02	0.00	0.00
Urban	95,167	5.49	0	0.00	0.00	0.06	0.00	0.00	0.02	0.00	0.00
<b>Drainage Density (km/square km)</b>											
<1.12	125,010	7.21	0	0.00	0.00	0.07	0.00	0.00	0.02	0.00	0.00
1.12–1.54	360,107	20.78	6	10.71	−0.66	0.12	−0.78	0.17	0.02	0.43	−1.81
1.54–1.88	480,396	27.72	19	33.93	0.20	−0.09	0.29	0.05	0.03	0.28	1.03
1.88–2.24	452,295	26.10	20	35.71	0.31	−0.14	0.45	0.05	0.03	0.28	1.62
>2.24	315,331	18.19	11	19.64	0.08	−0.02	0.09	0.09	0.02	0.34	0.28
<b>Distance to River (km)</b>											
<0.10	629,316	36.31	23	41.07	0.12	−0.08	0.20	0.04	0.03	0.27	0.74
0.10–0.21	519,863	30.00	19	33.93	0.12	−0.06	0.18	0.05	0.03	0.28	0.64
0.21–0.37	360,248	20.79	9	16.07	−0.26	0.06	−0.32	0.11	0.02	0.36	−0.87
0.37–0.57	170,585	9.84	4	7.14	−0.32	0.03	−0.35	0.25	0.02	0.52	−0.67
>0.57	53,127	3.07	1	1.79	−0.54	0.01	−0.55	1.00	0.02	1.01	−0.55
<b>Distance to Fault (km)</b>											
<2.20	634,007	36.58	7	12.50	−1.07	0.32	−1.40	0.14	0.02	0.40	−3.45
2.20–4.85	339,860	19.61	12	21.43	0.09	−0.02	0.11	0.08	0.02	0.33	0.34
4.85–7.75	295,667	17.06	14	25.00	0.38	−0.10	0.48	0.07	0.02	0.31	1.56
7.75–10.91	272,734	15.74	18	32.14	0.71	−0.22	0.93	0.06	0.03	0.29	3.25
>10.91	190,871	11.01	5	8.93	−0.21	0.02	−0.23	0.20	0.02	0.47	−0.50
<b>Distance to Road (km)</b>											
<2.78	584,777	33.74	22	39.29	0.15	−0.09	0.24	0.05	0.03	0.27	0.88
2.78–6.09	503,128	29.03	24	42.86	0.39	−0.22	0.61	0.04	0.03	0.27	2.25
6.09–9.91	341,304	19.69	8	14.29	−0.32	0.07	−0.39	0.13	0.02	0.38	−1.01
9.91–14.44	216,429	12.49	2	3.57	−1.25	0.10	−1.35	0.50	0.02	0.72	−1.87
>14.44	87,501	5.05	0	0.00	0.00	0.05	0.00	0.00	0.02	0.00	0.00
<b>NDVI</b>											
<−0.01	946	0.05	0	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
−0.01–0.07	995,879	57.46	8	14.29	−1.39	0.70	−2.09	0.13	0.02	0.38	−5.48
0.07–0.12	614,296	35.44	28	50.00	0.34	−0.26	0.60	0.04	0.04	0.27	2.24
0.12–0.21	95,540	5.51	15	26.79	1.58	−0.26	1.84	0.07	0.02	0.30	6.08
>0.21	26,478	1.53	5	8.93	1.77	−0.08	1.84	0.20	0.02	0.47	3.93
<b>TWI</b>											
<5.51	315,821	18.22	0	0.00	0.00	0.20	0.00	0.00	0.02	0.00	0.00
5.51–7.44	793,998	45.81	32	57.14	0.22	−0.23	0.46	0.03	0.04	0.27	1.69
7.44–9.76	391,225	22.57	14	25.00	0.10	−0.03	0.13	0.07	0.02	0.31	0.43
9.76–13.21	174,040	10.04	8	14.29	0.35	−0.05	0.40	0.13	0.02	0.38	1.05
>13.21	58,055	3.35	2	3.57	0.06	0.00	0.07	0.50	0.02	0.72	0.09
<−2.06	11,457	0.66	0	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00
−2.06–−0.58	55,078	3.18	0	0.00	0.00	0.03	0.00	0.00	0.02	0.00	0.00
−0.58–0.56	160,7635	92.76	55	98.21	0.06	−1.40	1.46	0.02	1.00	1.01	1.44
0.56–2.56	47,227	2.72	1	1.79	−0.42	0.01	−0.43	1.00	0.02	1.01	−0.43
>2.56	11,742	0.68	0	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00
<b>SPI</b>											
<8.05	538,726	31.08	23	41.07	0.28	−0.16	0.44	0.04	0.03	0.27	1.60
8.05–9.83	595,727	34.37	20	35.71	0.04	−0.02	0.06	0.05	0.03	0.28	0.21
9.83–11.97	329,339	19.00	6	10.71	−0.57	0.10	−0.67	0.17	0.02	0.43	−1.55
11.97–14.89	171,410	9.89	4	7.14	−0.33	0.03	−0.36	0.25	0.02	0.52	−0.69
>14.89	95,996	5.54	3	5.36	−0.03	0.00	−0.04	0.33	0.02	0.59	−0.06



**Figure 4.** Groundwater potentiality maps showing: (a), binary logistic regression (BLR); (b), technique for order preference by similarity to ideal solution (TOPSIS); (c), weight of evidence (WoE); (d), random forest (RF); (e), support vector machine (SVM).

**Table 6.** Areal distribution of groundwater potentiality maps.

Models	Potentiality Classes	Area in Square km	% of Area
TOPSIS	Low	446.1995	23.2
	Medium	787.3882	40.94
	High	399.4639	20.77
	Very high	290.222	15.09
WoE	Low	424.8511	22.09
	Medium	617.3708	32.1
	High	583.9059	30.36
	Very high	297.3381	15.46
RF	Low	1064.34	55.34
	Medium	198.6742	10.33
	High	174.4409	9.07
	Very high	485.8189	25.26
BLR	Low	744.3069	38.7
	Medium	594.4839	30.91
	High	286.9524	14.92
	Very high	297.5304	15.47
SVM	Low	248.6793	12.93
	Medium	569.0967	29.59
	High	744.8839	38.73
	Very high	360.4215	18.74

### 3.3. Application of Random Forest (RF) Model

The RF model has been used in the present study to identify GWPZ. For carrying out the RF model, the point values based on well and non-well locations have been derived from GWDFs. The out-of-bag (OOB) of RF model is 3.32% (Table 7). The results depict that elevation (385.72), rainfall (281.63), drainage density (152.98), distance to fault (258.36), NDVI (262.47), distance to the road (189.28) and LULC (137.42) factors have great contribution in the RF process (Figure 6). Conversely, factors such as lithology, soil type and convergence index have a tiny role in the RF process. Finally, GWPM by RF model has been prepared and classified into four classes such as low, medium, high and very GWPZs (Figure 4d). According to the GWPM, only 485.81 km<sup>2</sup> (25.26%) areas have very high groundwater potentiality. Other potentiality zones such as high, medium and low GWPZs are covered 9.07%, by 10.33% and 55.34% of the study area respectively (Table 6)

**Table 7.** Confusion matrix from random forest model (0 = non-well or negative, 1 = well or positive).

Observation	Predicted		Class Error	OOB (%)
	0	1		
0	8273	149	0.018	3.32
1	180	1319	0.120	

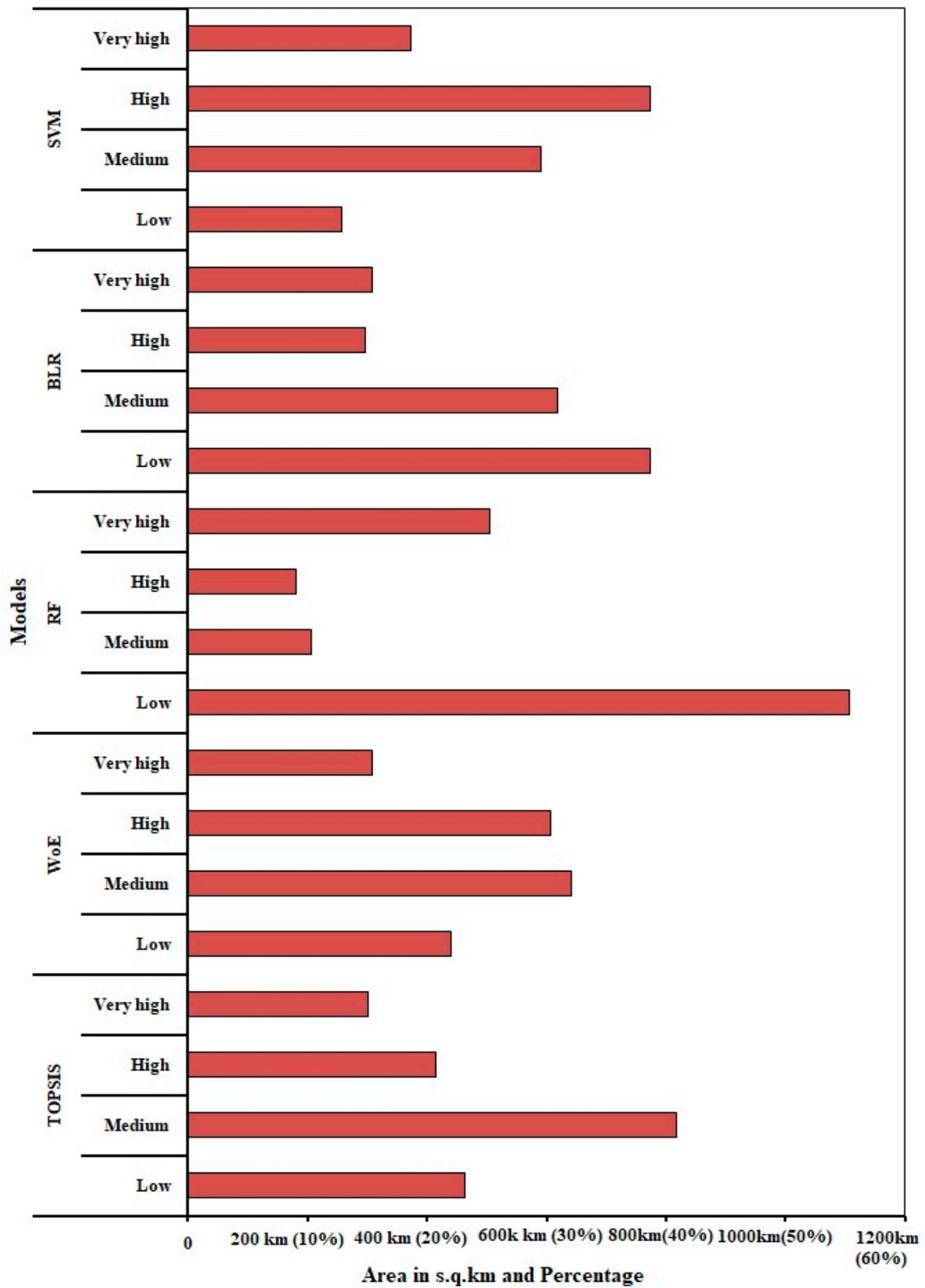
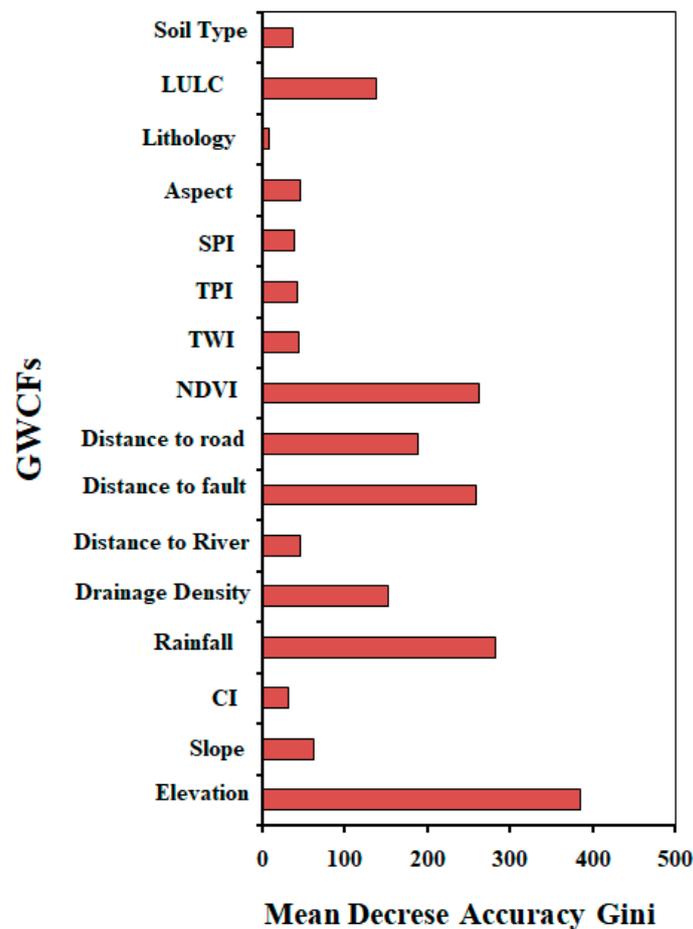


Figure 5. Areal distribution of groundwater potentiality maps.



**Figure 6.** Determining the weight of conditioning factors using random forest (RF).

#### 3.4. Application of Binary Logistic Regression (BLR)

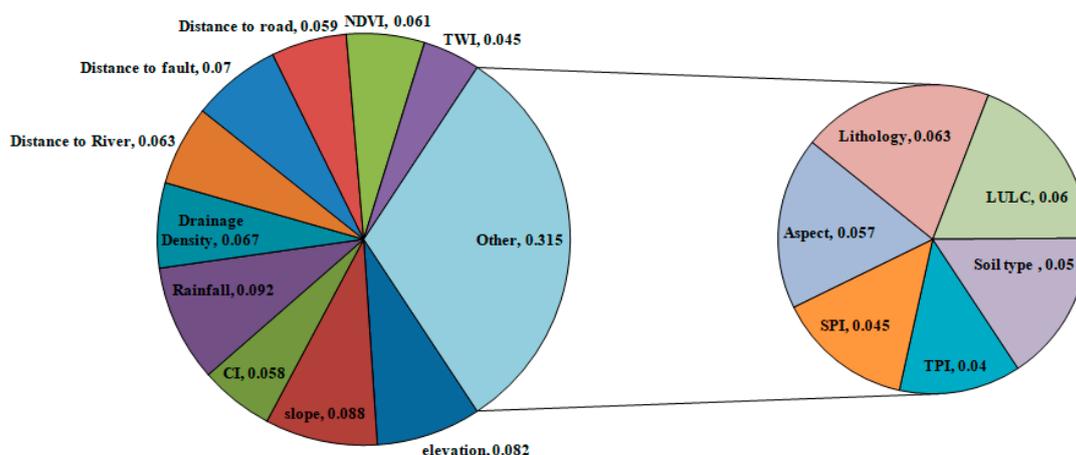
The BLR probabilistic model has been used for GWPZ estimation. The point base data for well and non-well locations have been extracted from each GWDF. Here, BLR is expressed with binary value, i.e., 0 and 1. 1 means that the presence of well and 0 means absence of well. The coefficients of regression values have been obtained by the BLR. The results of BLR (Table 8) show that slope (0.577), soil type (6.808), lithology (2.553) and LULC (2.2942) have the reciprocal and positive impacts on the occurrence of groundwater. Among the GWCFs, elevation (0.0237), convergence index (0.0029), rainfall (0.0131), distance to fault (0.0001), distance to road (0.0002), TWI (0.2892) and TPI (0.0426) have less importance for the formation of groundwater. Conversely, Dd, distance to the river, NDVI and SPI have negative impacts on the groundwater occurrence. Afterward, the weights have been assigned to GWCFs by BLR. Finally, GWPM by BLR has been built and categorized into four categories such as low, medium, high and very high GWPZs using the natural break method. According to this classification, 15.47% of the Damghan plain has very high groundwater potentiality, followed by 38.7%, 30.91%, and 14.92% for low, moderate and high GWPZs (Figure 4a and Table 6).

**Table 8.** Determining the weight of conditioning factors using logistic regression.

Parameters	Weight
Elevation	0.0237
Slope	0.5778
CI	0.0029
Rainfall	0.0131
Drainage Density	−1.739
Distance to River	−0.0008
Distance to Fault	0.0001
Distance to Road	0.0002
NDVI	−7.633
TWI	0.2892
TPI	0.0426
SPI	−0.1487
Aspect	−1.3488
Lithology	2.5531
LULC	2.2942
Soil types	6.8088

### 3.5. Application of TOPSIS

The TOPSIS an important MDCA approach used to delineate the GWPZs. In this research, 500 points were selected randomly and extracted point values from GWCFs. The AHP is an important knowledge-driven MDCA model, used to assign weights to GWCFs for performing the TOPSIS model. The weights of GWCFs are 0.082 (elevation), 0.088 (slope), 0.057 (aspect), 0.058 (convergence index), 0.092 (rainfall), 0.067 (drainage density), 0.063 (distance to river), 0.070 (distance to fault), 0.059 (distance to road), 0.063 (lithology), 0.050 (soil type), 0.060 (LULC), 0.061 (NDVI), 0.045 (TWI), 0.04(TPI) and 0.045 (SPI) (Figure 7). The weights of GWCFs by AHP and point base values of GWCFs have been computed using the Equations (17)–(26) and then, calculated the final weight. The GWPM by TOPSIS has been built considering the points values of GWCFs weights with the help of the inverse distance weighted (IDW) interpolation method (Figure 4b). The GWPM by TOPSIS has been classified into four classes such as low, medium, high and very high GWPZs with the help of natural break method. The results of GWMP by WoE model shows that only 290.22 km<sup>2</sup> (15.09%) area is very high groundwater potential, followed by the 399.46 km<sup>2</sup> (20.77%), 787.38 km<sup>2</sup> (40.94%) and 446.19 km<sup>2</sup> (23.2%) areas are high, medium and low groundwater potential (Table 6)



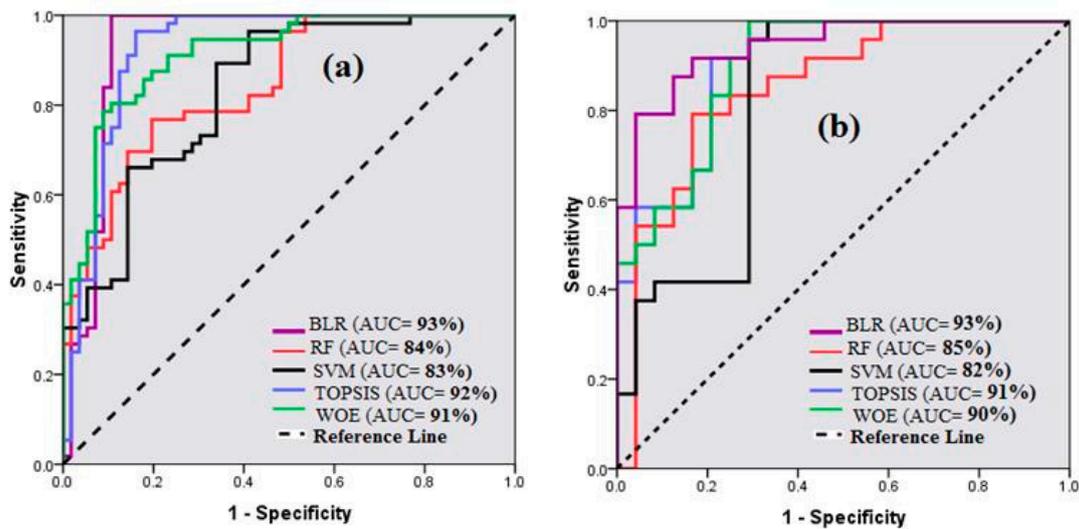
**Figure 7.** Determining the weight of conditioning factors using analytic hierarchy process (AHP) method.

### 3.6. Application of Support Vector Machine (SVM)

SVM is a vital machine learning data mining technique, used to recognize the potentiality of groundwater. All data layers have been reclassified into different classes using the SVM method. The SVM classification value ranges from 0 to 1, 0 indicating the absence of groundwater well in GWCFs and Conversely 1 value also indicates the presence and potentiality of groundwater formation. Among the GWDFs, the low altitude, low slopping, high rainfall, nearest distance to river, high Dd, nearest distance to road, far distance to fault, high vegetation index, H lithological units, agriculture land, arid soils, high TWI and low SPI have been considered as sub-layers of high groundwater potential, marked by the 1 values. Conversely, high altitude, slopping, low drainage density, far distance to a river, low fault distance, low vegetation index, rangeland, bare land, urban, entisols, salt flats, low TWI, and low rainfall have been identified as the 0 values because these conditions are not suitable for groundwater recharge. Thus, all GWDFs have been weighted by SVM and summed up in GIS to generate a single data layer of groundwater potential map (GWPM). The GWPM by SVM has been classified into four classes such as low, medium, high and very high GWPZs with the help of the natural break classification method (Figure 4e). The results of GWMP by WoE model show that only 360.42 km<sup>2</sup> (18.74%) area has very high groundwater potentiality and 248.67 km<sup>2</sup> (12.93%) area has low groundwater potentiality (Table 6).

### 3.7. Validations and Comparison of Models

Sometimes a single method of validation is not sufficient for judging the potentiality and performance of models because of the concentration of samples within a few places. Methods of AUROC, SE, SP, AC, MAE, RMSE, and SCAI were used to test the performance of models. Both training (goodness of fit) and validation (prediction accuracy) datasets have been used for judging the capability of models in producing the GWPMS of the study area. Considering the training dataset ROC curves (Figure 8) showing the AUC values of WoE, RF, TOPSIS, SVM, and BLR models are 0.914, 0.846, 0.924, 0.833, and 0.933, respectively (Figure 8). The SE values of the WoE, RF, TOPSIS, SVM, and BLR models are 0.807, 0.800, 0.833, 0.792, and 0.852, respectively. The SP values of the WoE, RF, TOPSIS, SVM, and BLR models are 0.818, 0.789, 0.810, 0.789, and 0.828, respectively (Table 9) The accuracy values of the WoE, RF, TOPSIS, SVM, and BLR models are 0.813, 0.795, 0.821, 0.791, and 0.839, respectively. The RMSE values of the WoE, RF, TOPSIS, SVM, and BLR models are 0.317, 0.367, 0.316, 0.377, and 0.314, respectively. MAE values of the WoE, RF, TOPSIS, SVM and BLR models are 0.221, 0.275, 0.219, 0.269 and 0.216, respectively. AUCROC, sensitivity, specificity, accuracy, MAE, and MRSE are depicting the consistency between the trained models and actual situation of groundwater. In the validation data context, the AUC values WoE, RF, TOPSIS, SVM, and BLR are 0.898, 0.816, 0.901, 0.851, and 0.943, respectively. The SE values of the WoE, RF, TOPSIS, SVM and BLR models are 0.800, 0.783, 0.870, 0.773 and 0.909, respectively. The SP values of the WoE, RF, TOPSIS, SVM and BLR models are 0.826, 0.760, 0.840, 0.760, and 0.846 respectively (Table 9). The accuracy value of the WoE, RF, TOPSIS, SVM, and BLR models are 0.813, 0.771, 0.854, 0.766 and 0.875, respectively. On the other hand the RMSE values of the WoE, RF, TOPSIS, SVM and BLR models are 0.332, 0.383, 0.321, 0.409, and 0.311 and MAE values are 0.235, 0.288, 0.233, 0.311, and 0.214, respectively (Table 9).



**Figure 8.** Validation of results using the area under the curve of the receiver operating characteristic (AUROC). (a) Training dataset (success rate curve) and (b) validation dataset (prediction rate curve).

**Table 9.** Analysis of performances using training dataset and validation dataset for the models.

Measures	Training Dataset					Validation Dataset				
	WoE	RF	TOPSIS	SVM	BLR	WoE	RF	TOPSIS	SVM	BLR
True positive	46	44	45	42	46	20	17	20	18	20
True negative	45	45	47	45	48	19	19	21	19	22
False positive	10	12	11	12	10	4	6	4	6	4
False negative	11	11	9	11	8	5	5	3	5	2
Sensitivity	0.807	0.800	0.833	0.792	0.852	0.800	0.773	0.870	0.783	0.909
Specificity	0.818	0.789	0.810	0.789	0.828	0.826	0.760	0.840	0.760	0.846
Accuracy	0.813	0.795	0.821	0.791	0.839	0.813	0.766	0.854	0.771	0.875
RMSE	0.317	0.367	0.316	0.377	0.314	0.332	0.383	0.321	0.409	0.311
MAE	0.221	0.275	0.219	0.269	0.216	0.235	0.288	0.233	0.311	0.214
AUC	0.914	0.846	0.924	0.833	0.933	0.898	0.81	0.901	0.851	0.943

All the statistical techniques and ROC curves used in this study for evaluating the performance of the models have judged all the models as good for mapping the groundwater potentiality in this plain. SCAI is another important validation method, used to validate the models. The SCAI values of sub-classes of all models have decreased from low potentiality to very high potentiality, indicating the appropriateness and suitability for the groundwater potentiality evaluation (Table 10). Above all, according to the threshold dependent, SCAI and statistical methods the BLR has the strongest predictability to evaluate the groundwater potentiality of the Damghan sedimentary plain, although other models have good capability in mapping the groundwater potentiality.

**Table 10.** Computation sheet of seed cell area index (SCAI) methods.

Models	Groundwater Potentiality Classes	% of Pixels	Training Datasets		Validation Datasets		Sum	SCAI
			No of Wells	% of Wells	No of Wells	% of Wells		
TOPSIS	Low	23.20	0	0.00	0	0.00	0.00	0.00
	Medium	40.94	0	0.00	0	0.00	0.00	0.00
	High	20.77	2	3.57	4	16.67	20.24	1.03
	Very high	15.09	54	96.43	20	83.33	179.76	0.08
WoE	Low	22.09	0	0.00	0	0.00	0.00	0.00
	Medium	32.10	2	3.57	1	4.17	4.17	7.70
	High	30.36	5	8.93	2	8.33	11.90	2.55
	Very high	15.46	49	87.50	21	87.50	96.43	0.16
RF	Low	55.34	1	1.79	0	0.00	1.79	30.99
	Medium	10.33	4	7.14	1	4.17	11.31	0.91
	High	9.07	8	14.29	3	12.50	26.79	0.34
	Very high	25.26	43	76.79	20	83.33	160.12	0.16
BLR	Low	38.70	0	0.00	1	4.17	4.17	9.29
	Medium	30.91	0	0.00	23	95.83	95.83	0.32
	High	14.92	2	3.57	0	0	3.57	4.18
	Very high	15.47	54	96.43	0	0	96.43	0.16
SVM	Low	12.93	0	0.00	0	0.00	0.00	0.00
	Medium	29.59	1	1.79	6	25.00	26.79	1.10
	High	38.73	14	25.00	18	75.00	100.00	0.39
	Very high	18.74	41	73.21			73.21	0.26

### 3.8. Sensitivity Analysis

To assess the influence of GWDFs on groundwater potentiality occurrence and to explore the effective factors with the strongest effect on the result of the groundwater potentiality prediction, a sensitivity analysis has been carried out (Table 11 and Figure 9). The results of sensitivity analysis showed in percentage contribution (PC) values of factors attained. The Pc values of the GWDFs are 7.5% (elevation), 11.35% (convergence index), 13.68% (drainage density), 11.81% (distance to road), 7.18% (distance to fault), 16.10% (distance to river), 6.19% (land use/land cover), 8.66% (lithology), 6.91% (NDVI), 9.67% (Rainfall), 7.86% (slope), 5.52% (soil), 10.10% (SPI), 12.58% (TPI), 9.51% (TWI) and 0.41% (aspect). The only slope aspect has very little contribution to the occurrence of groundwater potentiality. The results indicated that the groundwater potentiality maps of the study area are highly sensitive to elevation, lithology, drainage density, rainfall, distance to river, TPI, TWI, SPI, and distance to road. The sensitive analysis would help to reduce the variation in the model and to understand the significant geo-environmental factors that are vital for understanding the structure of model.

**Table 11.** Sensitivity result when each factor is excluded in the binary logistic regression model.

GWDFs	Decrease of AUC (in Percentage)
Elevation	7.5
CI	11.35
Drainage density	13.68
Distance from road	11.81
Distance from fault	7.18
Distance from river	16.10
LULC	6.19
Lithology	8.66
NDVI	6.91
Rainfall	9.67
Slope	7.86
Soil	5.52
SPI	10.70
TPI	12.58
TWI	9.51
Aspect	0.41

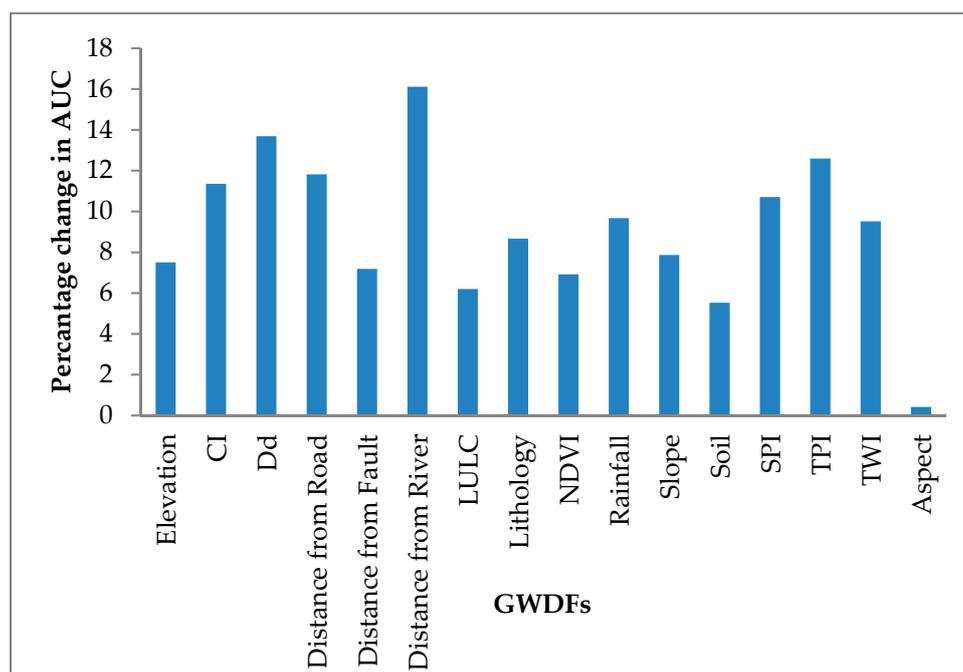


Figure 9. Sensitivity result when each factor is excluded in the binary logistic regression model.

#### 4. Discussion

In the recent decade, the demand for water has significantly increased because of the rapid growth of population, especially in arid and semi-arid areas. The large part of Damghan sedimentary plain covering the arid and semi-arid environments groundwater is the main source of water for living. In this region, groundwater planning and sustainable management are necessary. The hydrogeologist, engineers and decision need some basic tools for managing the groundwater. GWPM may meet the basic tool of groundwater management.

GWPM is the outcome of the lithology, tectonics, topography, vegetation, rainfall, and hydrology, which are available and accessible everywhere in the environment. In this research, a different type of data has been used as the input datasets. DEMs based study provides more accurate and significant results [90–92]. Different DEMs provide different results, e.g., ALOS DEM with 30 m spatial resolution provide suitable and excellent results, comparatively the ASTER and SRTM DEMs with 30 m resolution [93]. Here, the authors combined the geomorphology, geology and hydrology parameters to recognize the spatial groundwater potential. Spatial analysis is the core matter of the research for adopting the most performing approach and models for GWPMs, considering the argument topic [12–15]. Geo-environmental factors (i.e., elevation, slope, aspect, rainfall, lithology, land use/land cover, soil type, drainage density, distance to river, distance to fault, distance to road, NDVI, TWI, TPI, and SPI) were considered as the GWDFs that have been tested for the multi-collinearity problem by VIF and tolerance, and are the most effective for groundwater storage. The categorical variables such as aspect, lithology, soil type, LU/LC factors have been converted into the quantity continuous data through assigning the weight by the WoE and TOPSIS method. For the LR and RF, these GWDFs have been evaluated to prepare GWPMs taking the extracted values of GWDFs of the 500 points. The results of these models are more accurate than previous works [32]. In this work, we applied probabilistic (WoE, BLR), machine learning (SVM and RF) and multi-criteria decision approach (TOPSIS) models for building the GWPMs of Damghan sedimentary plain. These models have represented the excellent results as other works were done by Mohammady et al. [94] and Arabameri et al. [25,26]. All models, however, have very few variations in groundwater potential modeling accuracy. According to the AUROC, SE, SP, Accuracy, MAE and MRSE among the five models, the BLR models (for training dataset AUC = 0.933, SE = 0.852, SP = 0.828, AC = 0.839, MAE = 0.216 and RMSE = 0.314 and for

validation dataset  $AUC = 0.943$ ,  $SE = 0.909$ ,  $SP = 0.846$ ,  $AC = 0.875$ ,  $MAE = 0.214$  and  $RMSE = 0.311$ ) have better capability for mapping the groundwater potentiality than other models. Recognizing the significance of each variable for groundwater storage is very difficult. The soil, lithology, altitude, rainfall, LU/LC, NDVI, Dd, distance to fault factors are dominant factors among 16 GWDFs for the formation of groundwater. The SA is depicting the contribution in producing the uncertainty in the GWPM and the factor distance from the river has the highest contribution to the variation of output of model. Similar to the Shahroud plain, the Damghan sedimentary plain regions consists of the large bare land, rangeland, and urban land, interrupting the water infiltration into sub-surface layer, while agriculture land with aquifer locations are receiving the larger water into the sub-surface and also signify the hydrologic properties [95]. According to TOPSIS model, the rainfall, slope, elevation, LU/LC, soil type factors have been highly prioritized by the AHP model, suggesting the most potential for groundwater formation. Such findings are confirmed with the work of Arabameri et al. [32]. Among the 16 GWDFs, the elevation is the most important topographic component that influences the groundwater recharge. In fact, at a lower segment, the Damghan sedimentary plain is almost flat, where water stagnation and associated infiltration of water is maximized. On the contrary, high altitudes, associated with open and v-shaped slopes promote runoff due to local physiography. The methods applied for validating the GWPMs are showing outstanding accuracy, and ensemble models have better capabilities than the individual modes. Such ensemble models have been shown to be more reliable in this analysis than the other models used by the researchers for GWPM in various other locations [96,97]. The proper methods can have the ability to produce GWPMs, and that can be used for planning purposes. The used probabilistic, machine learning and ensemble models have excellent accuracy and may be used for groundwater management in this plain region.

## 5. Conclusions

Today, GWPM is an effective groundwater resource management method. Through the Over-extraction of groundwater in the low groundwater, the potential region can be limited by the GWPM. With the advancement in the technical field, different techniques for the spatial modeling of groundwater are introducing day by day. So, it very difficult to say what method would be best for spatial modeling. However, in the present research, five methods (BLR, TOPSIS, WoE, RF, and SVM) have been used for modeling the groundwater and the compared among them to answer the question of what model is relatively better for the Damghan sedimentary plain. The GWPM approaches are more appropriate to predict the potential of groundwater. The GWPMs have been produced with the help of RS and GIS techniques. RS and GIS both combinedly helped to perform the works such as identification of well, thematic data generation, classification, and final map generations. The RS and GIS-based study are cost and time saving, accurate, and provide meaningful results. The R studio is an important machine learning program that helps to perform different kinds of models such as logistic regression, random forest, naive Bayes tree, support vector machine, artificial neural network, and several other methods. R program based model performance is more easy, accurate, efficiency and perfect. The GWPMs, produced by the selected methods have been categorized into four categories i.e., low, medium, high and very high potential classes. The results of GWPMs show that the very high potentiality zones are covered with by an area of  $290.22 \text{ km}^2$  (15.09% by TOPSIS),  $297.34 \text{ km}^2$  (15.46% by WoE),  $485 \text{ km}^2$  (25.26% by RF),  $297.53 \text{ km}^2$  (15.47% by BLR) and  $360.42 \text{ km}^2$  (18.74% by SVM) out of  $1923.27 \text{ km}^2$  areas. The worthiness of GWPMs has been significantly validated by the ROC and SCAI methods and five statistical measures i.e., SE, SP, AC, MAE, and MRSE. According to the results of the ROC, SCAI methods and statistical measures, these models are excellent for the prediction of GWPZ. The very high or excellent GWPZs have been found in the low elevated and less sloppy area. The arid soils are covered by high potentiality of groundwater. In the case of the LU/LC and vegetation index, the agriculture land and high vegetation density areas have high potentiality of groundwater. Conversely, high altitude, sloppy land, urban area, rangeland, salt flats, entisols soil type have the low potentiality of groundwater formation. The GWPMs may be used as tool in

this study area for managing and developing the groundwater. The resulting maps can also assist decision-makers, planners, and engineers in choosing the ideal location, groundwater distribution for further groundwater exploration. Therefore, Damghan sedimentary plain region has high potentiality of groundwater storage which can be saved by sustainable use, obstructing groundwater pollution, increasing the people's awareness and suitable government policy regarding the amount and way water use.

**Author Contributions:** Methodology, A.A., J.R., and S.S.; formal analysis, A.A., J.R., and S.S.; investigation, A.A., J.R., and S.S.; writing—original draft preparation, A.A., J.R., and S.S.; writing—review and editing, A.A., J.R., S.S., T.B., O.G., and D.T.B.

**Funding:** This research was partly funded by the Austrian Science Fund (FWF) through the Doctoral College GIScience (DK W 1237-N23) at the University of Salzburg.

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

## References

1. Berhanu, B.; Seleshi, Y.; Melesse, A.M. *Surface Water and Groundwater Resources of Ethiopia: Potentials and Challenges of Water Resources Development*; Springer: Dordrecht, The Netherlands, 2014; pp. 97–117.
2. Zehtabian, G.; Khosravi, H.; Ghodsi, M. High demand in a land of water scarcity: Iran. In *Water and Sustainability in Arid Regions*, 1st ed.; Graciela, S.M., Courel, M.F., Eds.; Springer: Dordrecht, The Netherlands, 2001; pp. 75–86.
3. 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](#)]
4. National Geography Society. *National Geographic, Almanac of Geography*; National Geographic Books; National Geography Society: Washington, DC, USA, 2005.
5. Jha, M.K.; Kamii, Y.; Chikamori, K. Cost-effective approaches for sustainable groundwater management in alluvial aquifer systems. *Water Resour. Manag.* **2009**, *23*, 219. [[CrossRef](#)]
6. Gholizadeh, M.H.; Melesse, A.M.; Reddi, L. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors* **2016**, *16*, 1298. [[CrossRef](#)] [[PubMed](#)]
7. 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](#)]
8. Management and Planning Organization (MPO). *Water Resources State Report*; Management and Planning Organization (MPO): Tehran, Iran, 2004.
9. Nosrati, K.; Eeckhaut, M.V.D. Assessment of groundwater quality using multivariate statistical techniques in Hashtgerd Plain, Iran. *Environ. Earth Sci.* **2012**, *65*, 331–344. [[CrossRef](#)]
10. Rahmati, O.; Nazari Samani, A.; Mahdavi, M.; Pourghasemi, H.R.; Zeinivand, H. Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arab. J. Geosci.* **2014**, *8*, 7059–7071. [[CrossRef](#)]
11. Haghizadeh, A.; DavoudiMoghadam, 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](#)]
12. 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.
13. Kharazmi, R.; Tavili, A.; Rahdari, M.R.; Chaban, L.; Panidi, E.; Rodrigo-Comino, J. Monitoring and assessment of seasonal land cover changes using remote sensing: A 30-year (1987–2016) case study of Hamoun Wetland, Iran. *Environ. Monit. Assess.* **2018**, *190*, 356. [[CrossRef](#)]
14. He, B.; Wang, H.; Huang, L.; Liu, J.; Chen, Z. A new indicator of ecosystem water use efficiency based on surface soil moisture retrieved from remote sensing. *Ecol. Indic.* **2017**, *75*, 10–16. [[CrossRef](#)]
15. Thilagavathi, N.; Subramani, T.; Suresh, M.; Karunanidhi, D. Mapping of groundwater potential zones in Salem Chalk Hills, Tamil Nadu, India, using remote sensing and GIS techniques. *Environ. Monit. Assess.* **2015**, *187*, 1–17. [[CrossRef](#)] [[PubMed](#)]

16. 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.* **2018**, *27*, 211–224. [[CrossRef](#)]
17. 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](#)] [[PubMed](#)]
18. 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](#)] [[PubMed](#)]
19. 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](#)]
20. Saha, S. Groundwater potential mapping using analytical hierarchical process: A study on Md. Bazar Block of Birbhum District, West Bengal. *Spat. Inf. Res.* **2017**, *25*, 615–626. [[CrossRef](#)]
21. Rahmati, O.; Naghibi, S.A.; Shahabi, H.; Tien Bui, D.; Pradhan, B.; Aareh, A.; Rafiei-Sardooi, E.; Samani, A.N.; Melesse, A.M. Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. *Hydrology* **2018**, *565*, 248–261. [[CrossRef](#)]
22. 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. *Appl. Clim.* **2018**, *131*, 967–984. [[CrossRef](#)]
23. Arabameri, A.; Pourghasemi, H.R.; Cerda, A. Erodibility prioritization of subwatersheds using morphometric parameters analysis and its mapping: A comparison among TOPSIS, VIKOR, SAW, and CF multi-criteria decision making models. *Sci. Total Environ.* **2017**, *613*, 1385–1400.
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. Arabameri, A.; Pradhan, B.; Pourghasemi, H.R.; Rezaei, K.; Kerle, N. Spatial modeling of gully erosion using GIS and R programming: A comparison among three data mining algorithms. *Appl. Sci.* **2018**, *8*, 1369. [[CrossRef](#)]
26. 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](#)]
27. Islamic republic of Iran Meteorological Organization (IRIMO). 2012. Available online: <http://www.semnanmet.ir> (accessed on 12 August 2018).
28. Tang, Q.; Hu, H.; Oki, T. Groundwater recharge and discharge in a hyperarid alluvial plain (Akesu, Taklimakan Desert, China). *Hydrol. Processes* **2007**, *21*, 1345–1353. [[CrossRef](#)]
29. Geology Survey of Iran (GSI). 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).
30. Tehran Regional Water Cooperative (TRWC) Company. *Simulation Project for Optimum Excavation of Dasht-e-Damghan*; Principal Office of Water Resources: Washington, DC, USA, 2000; p. 46.
31. UNEP. *A Survey of Methods for Groundwater Recharge in Arid and Semi-Arid Regions*; UNEP/DEWA/RS: New York, NY, USA; Bilthoven, The Netherlands, 2002; pp. 5–10.
32. 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](#)] [[PubMed](#)]
33. Jothibas, A.; Anbazhagan, S. Modeling groundwater probability index in Ponnaiyar River basin of South India using analytic hierarchy process. *Model. Earth Syst. Environ.* **2016**, *2*, 109. [[CrossRef](#)]
34. Kiss, R. Determination of drainage network in digital elevation model. *Util. Limit. J. Hung. Geomath.* **2004**, *2*, 16–29.
35. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modeling: A review of hydrological, geomorphological and biological applications. *Hydrol. Processes* **1991**, *5*, 3–30. [[CrossRef](#)]
36. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [[CrossRef](#)]
37. Conforti, M.; Aucelli, P.C.; Robustelli, G.; Scarciglia, F. Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). *Nat. Hazards* **2011**, *56*, 881–898. [[CrossRef](#)]

38. Gómez-Gutiérrez, A.; Conoscenti, C.; Angileri, S.E.; Rotigliano, E.; Schnabel, S. Using topographical attributes to evaluate gully erosion proneness (susceptibility) in two mediterranean basins: Advantages and limitations. *Nat. Hazards* **2015**, *79*, 291–314. [[CrossRef](#)]
39. Gallant, J.C.; Wilson, J.P. Primary topographic attributes. In *Terrain Analysis: Principles and Applications*; Wilson, J.P., Gallant, J.C., Eds.; Wiley: New York, NY, USA, 2000; pp. 51–85.
40. Grohmann, C.H.; Riccomini, C. Comparison of roving-window and search-window techniques for characterising landscape morphometry. *Comput. Geosci.* **2009**, *35*, 2164–2169. [[CrossRef](#)]
41. Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Masuda, T.; Nishino, K. GIS based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ. Geol.* **2008**, *54*, 311–324. [[CrossRef](#)]
42. Armas, I. Weights of evidence method for landslide susceptibility mapping; Prahova Subcarpathians, Romania. *Nat. Hazards* **2012**, *60*, 937–950. [[CrossRef](#)]
43. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
44. Micheletti, N.; Foresti, L.; Robert, S.; Leuenberger, M.; Pedrazzini, A.; Jaboyedoff, M.; Kanevski, M. Machine learning feature selection methods for landslide susceptibility mapping. *Math. Geosci.* **2014**, *46*, 33–57. [[CrossRef](#)]
45. Strobl, C.; Boulesteix, A.L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinf.* **2008**, *9*, 307. [[CrossRef](#)]
46. Caruana, R.; Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; ACM: New York, NY, USA, 2006; pp. 161–168.
47. Reif, D.M.; Motsinger, A.A.; McKinney, B.A.; Crowe, J.E.; Moore, J.H. Feature Selection using a random forests classifier for the integrated analysis of multiple data type. In Proceedings of the 2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology, Toronto, ON, Canada, 28–29 September 2006.
48. Kuhnert, P.M.; Henderson, A.K.; Bartley, R.; Herr, A. Incorporating uncertainty in gully erosion calculations using the random forests modelling approach. *Environmetrics* **2010**, *21*, 493–509. [[CrossRef](#)]
49. Van Beijma, S.; Comber, A.; Lamb, A. Random forest classification of salt marsh vegetation habitats using quadpolarimetric airborne SAR, elevation and optical RS data. *Remote Sens. Environ.* **2014**, *149*, 118–129. [[CrossRef](#)]
50. Archer, K.J.; Kimes, R.V. Empirical characterization of random forest variable importance measures. *Comput. Stat. Data Anal.* **2008**, *52*, 2249–2260. [[CrossRef](#)]
51. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015; Available online: <http://www.Rproject.org> (accessed on 12 August 2018).
52. Lombardo, L.; Opitz, T.; Huser, R. Point process-based modeling of multiple debris flow landslides using INLA: An application to the 2009 Messina disaster. *Stoch. Environ. Res. Risk A* **2018**, *32*, 2179–2198. [[CrossRef](#)]
53. Hwang, C.L.; Yoon, K.P. *Multiple Attribute Decision Making: Methods and Applications*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 1981.
54. Zhang, Y.; Xu, Z. Efficiency evaluation of sustainable water management using the HF-TODIM method. *Int. Trans. Op. Res.* **2019**, *26*, 747–764. [[CrossRef](#)]
55. Vomm, V.B. TOPSIS with statistical distances: A new approach to MADM. *Decis. Sci. Lett.* **2017**, *6*, 49–66. [[CrossRef](#)]
56. Saaty, T.L. *The Analytic Hierarchy Process*; McGraw Hill: New York, NY, USA, 1980.
57. Saaty, T.L. *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*; RWS Publications: Pittsburgh, PA, USA, 2000.
58. Lootsma, F.A. *Multi-Criteria Decision Analysis via Ratio and Difference Judgement*, 1st ed.; Springer: New York, NY, USA, 2007.
59. Bai, S.B.; Wang, J.; Lu, G.N.; Kanevski, M.; Pozdnoukhov, A. GIS based landslide susceptibility mapping with comparisons of results from machine learning methods process versus logistic regression in Bailongjiang river basin, China. *Geophys. Res. Abstr. EGU* **2008**, *10*, A-06367.
60. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [[CrossRef](#)]

61. Vapnik, V. *Nature of Statistical Learning Theory*; Wiley: New York, NY, USA, 1995.
62. Tax, D.; Duin, E. Uniform object generation for optimizing one class classifiers. *J. Mach. Learn. Res.* **2002**, *2*, 155–173.
63. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining Inference and Prediction*; Springer: New York, NY, USA, 2001.
64. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [[CrossRef](#)]
65. Camilo, D.C.; Lombardo, L.; Mai, P.M.; Dou, J.; Huser, R. Handling high predictor dimensionality in slope-unit-based landslide susceptibility models through LASSO penalized Generalized Linear Model. *Environ. Model. Softw.* **2018**, *97*, 145–156. [[CrossRef](#)]
66. 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.
67. Süzen, M.L.; Doyuran, V. A comparison of the GIS based landslide susceptibility assessment methods: Multivariate versus bivariate. *Environ. Geol.* **2004**, *45*, 665–679. [[CrossRef](#)]
68. Dao, D.V.; Trinh, S.H.; Ly, H.-B.; Pham, B.T. Prediction of Compressive Strength of Geopolymer Concrete Using Entirely Steel Slag Aggregates: Novel Hybrid Artificial Intelligence Approaches. *Appl. Sci.* **2019**, *9*, 1113. [[CrossRef](#)]
69. Dao, D.V.; Ly, H.-B.; Trinh, S.H.; Le, T.-T.; Pham, B.T. Artificial Intelligence Approaches for Prediction of Compressive Strength of Geopolymer Concrete. *Materials* **2019**, *12*, 983. [[CrossRef](#)]
70. Ly, H.-B.; Monteiro, E.; Le, T.-T.; Le, V.M.; Dal, M.; Regnier, G. Prediction and Sensitivity Analysis of Bubble Dissolution Time in 3D Selective Laser Sintering Using Ensemble Decision Trees. *Materials* **2019**, *12*, 1544. [[CrossRef](#)] [[PubMed](#)]
71. Pham, B.T.; Nguyen, M.D.; Bui, K.-T.T.; Prakash, I.; Chapi, K.; Bui, D.T. A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil. *Catena* **2019**, *173*, 302–311. [[CrossRef](#)]
72. Pham, B.T. A novel classifier based on composite hyper-cubes on iterated random projections for assessment of landslide susceptibility. *J. Geol. Soc. India* **2018**, *91*, 355–362. [[CrossRef](#)]
73. Saltelli, A.; Chan, K.; Scott, E.M. *Sensitivity Analysis*; Wiley: New York, NY, USA, 2000.
74. Refsgaard, J.C.; Sluijs, J.P.V.D.; Højberg, A.L.; Vanrolleghem, P.A. Uncertainty in the environmental modelling process—A framework and guidance. *Water Resour. Manag.* **2007**, *22*, 1543–1556. [[CrossRef](#)]
75. Crosetto, M.; Tarantola, S. Uncertainty and sensitivity analysis: Tools for GIS-based model implementation. *Int. J. Geogr. Inf. Sci.* **2001**, *15*, 415–437. [[CrossRef](#)]
76. Ferretti, F.; Saltelli, A.; Tarantola, S. Trends in sensitivity analysis practice in the last decade. *Sci. Total Environ.* **2016**, *568*, 666–670. [[CrossRef](#)]
77. Chen, Y.; Yu, J.; Khan, S. Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. *Environ. Model. Softw.* **2010**, *25*, 1582–1591. [[CrossRef](#)]
78. Lodwick, W.A.; Monson, W.; Svoboda, L. Attribute error and sensitivity analysis of map operations in geographical information systems: Suitability analysis. *Int. J. Geogr. Inf. Syst.* **1990**, *4*, 413–428. [[CrossRef](#)]
79. 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](#)]
80. Fenta, A.A.; Kifle, A.; Gebreyohannes, T.; Hailu, G. Spatial analysis of groundwater potential using remote sensing and GIS-based multi-criteria evaluation in Raya Valley, northern Ethiopia. *Hydrogeol. J.* **2015**, *23*, 195–206. [[CrossRef](#)]
81. Tahmassebpour, 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*, 1–18. [[CrossRef](#)]
82. Convertino, M.; Muñoz-Carpena, R.; Chu-Agor, M.L.; Kiker, G.L.; Linkov, I. Untangling drivers of species distributions: Global sensitivity and uncertainty analyses of MAXENT. *Environ. Model. Softw.* **2014**, *51*, 296–309. [[CrossRef](#)]
83. Park, N.W. Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets. *Environ. Earth Sci.* **2015**, *73*, 937–949. [[CrossRef](#)]

84. Tien Bui, D.; Lofman, O.; Revhaug, I.; Dick, O. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat. Hazards* **2011**, *59*, 1413–1444.
85. Cama, M.; Lombardo, L.; Conoscenti, C.; Rotigliano, E. Improving transferability strategies for debris flow susceptibility assessment. Application to the Saponara and Itala catchments (Messina, Italy). *Geomorphology* **2017**, *288*, 52–65. [[CrossRef](#)]
86. 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](#)]
87. Roy, J.; Saha, S. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenvironmental Disasters* **2019**, *6*, 11. [[CrossRef](#)]
88. Regmi, N.R.; Giardino, J.R.; Vitek, J.D. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. *Geomorphology* **2010**, *115*, 172–187. [[CrossRef](#)]
89. Moghaddam, D.D.; Rezaei, M.; Pourghasemi, H.R.; Pourtaghie, Z.S.; Pradhan, B. Groundwater spring potential mapping using bivariate statistical model and GIS in the Taleghan Watershed, Iraq. *Arab. J. Geosci.* **2013**, *8*, 913–929. [[CrossRef](#)]
90. Pope, A.; Murray, T.; Luckman, A. DEM quality assessment for quantification of glacier surface change. *Ann. Glaciol.* **2014**, *46*, 189–194. [[CrossRef](#)]
91. Erasmi, S.; Rosenbauer, R.; Buchbach, R.; Busche, T.; Rutishauser, S. Evaluating the quality and accuracy of TanDEM-X digital elevation models at archaeological sites in the Cilician Plain, Turkey. *Remote Sens.* **2014**, *6*, 9475–9493. [[CrossRef](#)]
92. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [[CrossRef](#)]
93. Alganci, U.; Besol, B.; Sertel, E. Accuracy assessment of different digital surface models. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 114. [[CrossRef](#)]
94. Mohammady, M.; Pourghasemi, H.R.; Pradhan, B. Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster–Shafer, and weights-of-evidence models. *J. Asian Earth Sci.* **2012**, *61*, 221–236. [[CrossRef](#)]
95. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *J. Hydrol.* **2013**, *504*, 69–79. [[CrossRef](#)]
96. Hong, H.; Tsangaratos, P.; Ilia, L.; Chen, W.; Xu, C. Comparing the performance of a logistic regression and a random forest model in landslide susceptibility assessments. The Case of Wuyaun Area, China. In Proceedings of the Workshop World Landslide Forum, Ljubljana, Slovenia, 29 May–2 June 2017; pp. 1043–1050.
97. Hemasinghe, H.; Rangali, R.S.S.; Deshapriya, N.L.; Samarakoon, L. Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka). *Procedia Eng.* **2018**, *212*, 1046–1053. [[CrossRef](#)]

