remote sensing

MDPI

Article

Spatial Mapping of the Groundwater Potential of the Geum River Basin Using Ensemble Models Based on Remote Sensing Images

Jeong-Cheol Kim [1,2], Hyung-Sup Jung [2] and Saro Lee [3,4,*]

1 National Institute of Ecology (NIE), 1210 Geumgang-ro, Maseo-myeon, Seocheon-gun, Chungcheongnam-do 33657, Korea; jckim@nie.re.kr
2 Department of Geoinformatics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea; hsjung@uos.ac.kr
3 Geoscience Platform Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124, Gwahak-ro Yuseong-gu, Daejeon 34132, Korea
4 Korea University of Science and Technology, 217 Gajeong-ro Yuseong-gu, Daejeon 305-350, Korea
* Correspondence: leesaro@kigam.re.kr; Tel.: +82-42-868-3057

Received: 3 September 2019; Accepted: 26 September 2019; Published: 30 September 2019

Abstract: This study analyzed the Groundwater Productivity Potential (GPP) of Okcheon city, Korea, using three different models. Two of these three models are data mining models: Boosted Regression Tree (BRT) model and Random Forest (RF) model. The other model is the Logistic Regression (LR) model. The three models are based on the relationship between groundwater-productivity data (specific capacity (SPC) and transmissivity (T)) and the related hydro-geological factors from thematic maps, such as topography, lineament, geology, land cover, and etc. The thematic maps which are generated from the remote sensing images. Groundwater productivity data were collected from 86 wells locations. The resulting GPP maps were validated through area-under-the-curve (AUC) analysis using wells data that had not been used for training the model. When T was used in the BRT, RF, and LR models, the obtained GPP maps had 81.66%, 80.21%, and 85.04% accuracy, respectively, and when SPC was used, the maps had 81.53%, 78.57%, and 82.22% accuracy, respectively. The LR model, which is a statistical model, showed the highest verification accuracy, also the other two models showed high accuracies. These observations indicate that all three models can be useful for groundwater resource development.

Keywords: groundwater; remote sensing; GIS; random forest; Boosted Regression Tree; logistic regression

1. Introduction

According to The United Nations World Water Development Report (WWDR) 2018, more than 2 billion people in the World do not have access to safe drinking water and sanitation. If the current levels of water pollution and consumption are not reduced, nearly one-third of the world's population will suffer under severe water stress by approximately 2050 [1]. Climate change, increasing water scarcity, environmental degradation, population growth, and urbanization are already posing challenges for surface water supply systems [2]. Other means of meeting the demand for freshwater, such as using groundwater, will have to be determined. Presently around 20% of the total groundwater resources are being used globally [3].

Groundwater is a very efficient resource and can be used for agriculture, forestry, rearing of livestock, industrial purposes, and as a drinking water source for the community [4]. One of the most valuable benefits of groundwater is that it is less susceptible to environmental pollution than surface water [5]. Therefore, effort to find high quality groundwater is growing globally [6]. In South Korea, the rate of groundwater use has increased, and yet its supply does not meet the needs of the people [7]. Therefore, in South Korea, studies to evaluate the sustainability of groundwater in order to improve the use of groundwater and to evaluate the potential of groundwater in order to efficiently manage groundwater should be encouraged.

With scientific advancement in terms of GIS technology, various spatial modeling techniques have been developed and applied to evaluate the potential of groundwater productivity in recent years. GIS and remote sensing can be used to detail large areas in a more cost-effective manner [8–16]. In contemporary studies, Frequency Ratio (FR) [17–19], Random Forest (RF) [20–23], Logistic Regression (LR) [24–26], Boosted Regression Tree (BRT) [27,28], Support Vector Machine (SVM) [13,29–31], Artificial Neural Network (ANN) [32–36], Weights of Evidence (WoE) [37–39], Evidential Belief Function (EBF) [40–42], and various other ensemble models have been applied for Groundwater Productivity Potential (GPP) mapping.

Ensemble models such as RF and BRT were also used to study ecology, landslide, subsidence, flood vulnerability, and etc. [43]. Nsiah, et al. [44] evaluated the groundwater potential of Ghana's Nabogo basin using the weighted overlay technique. They achieved more accurate and reliable results by utilizing the commonly used specific capacity (SPC) values as wells as transmissivity (T) values. Park, Hamm, Jeon and Kim [24] performed GPP mapping using the LR and Multivariate Adaptive Regression Splines (MARS) models; these showed 84% and 87% verification accuracies, respectively. Lee, et al. [45] analysed the relationships between the groundwater pumping capacity and related factors using the FR and Boosted Classification Tree (BCT) models in Goyang-si in Gyeonggi-do province, South Korea. The results of the accuracy rates were 68.31% and 69.39%, respectively. In the previous studies, various ensemble models were used to predict GPP, and their accuracy showed a reliable level of results (approximately, >65%). However, for high accuracy of results, many studies will need to be carried out through the application of various topographical, geological and hydrological data (e.g., data obtained through remote sensing images) and various models that have not been utilized previously.

The purpose of the present study was to apply and analyze the LR (statistical model), RF & BRT (machine learning models) and determine their ability to perform accurate and effective GPP mapping. In addition, this study also intended to identify the important factors affecting GPP. Numerous preceding studies have used various models to analyze GPP, however, the LR, RF and BRT models have not yet been widely used. Therefore, we used them in correlation with hydrogeological factors related to groundwater productivity data to perform a more accurate GPP analysis and to verify and compare LR, RF and BRT models' accuracy and suitability. Also, various groundwater-related factors used in this study are derived from the thematic maps based on remote sensing data [46–49]. This study can be used as a reference to related future studies, such as the development of clean water resources, particularly groundwater [50].

2. Study Area and Spatial Data Set

2.1. Study Area

The research area in this study was Okcheon-gun in South Korea. The region is geographically located in the upstream area of the Geum River basin, which is the basin of one of South Korea's four major rivers. The Geum River flows from north to south in this area. Okcheon-gun lies between 36°10′N and 36°26′N latitudes and 127°29′E and 127°53′E longitudes. Its total area is 537.06 km^2, of which 347.04 km^2 is forest land, 55.82 km^2 is covered by fields, 45.63 km^2 is used as paddy fields, and other areas occupy 88.57 km^2 [6]. The annual precipitation in the area is 1297.4 mm, which is nearly equal to South Korea's annual average precipitation of 1277.4 mm (1978~2007) [33]. However, due to the influence of the East Asian Monsoon climate, rainfall is intense during summer and winter, while there is not enough water during spring and autumn. This area uses approximately

45,032,000 m^3 of groundwater per year. Of the total consumed groundwater, 67.2% is used for living, 32.1% for agriculture, and 0.5% for industrial purposes. As a result, the Okcheon-gun uses most of the groundwater as living and agricultural part, and the GPP map of Okcheon-gun is necessary for more efficient groundwater management [51].

Geologically, this area was developed from the Okcheon era and includes the unrecorded Okcheon supergroup. It also includes the Pyeongan supergroup, Paleozoic Choseon supergroup, the Triassic and Jurassic granitic rocks, the Cretaceous sedimentary, Quaternary alluvium, volcanic rocks, and intrusive igneous rocks (Figure 1). The Quaternary alluvium was found to be distributed along the tributaries in the Okcheon area. The alluvial layers in the plain are developed in the granite area, and the basin shape is narrow downstream in the plain. The Quaternary alluvium constitutes unconsolidated clastic sediments consisting of gravel, sand, silt, and clay. Relatively, silt and clay are more thickly deposited in the plain due to river flooding.

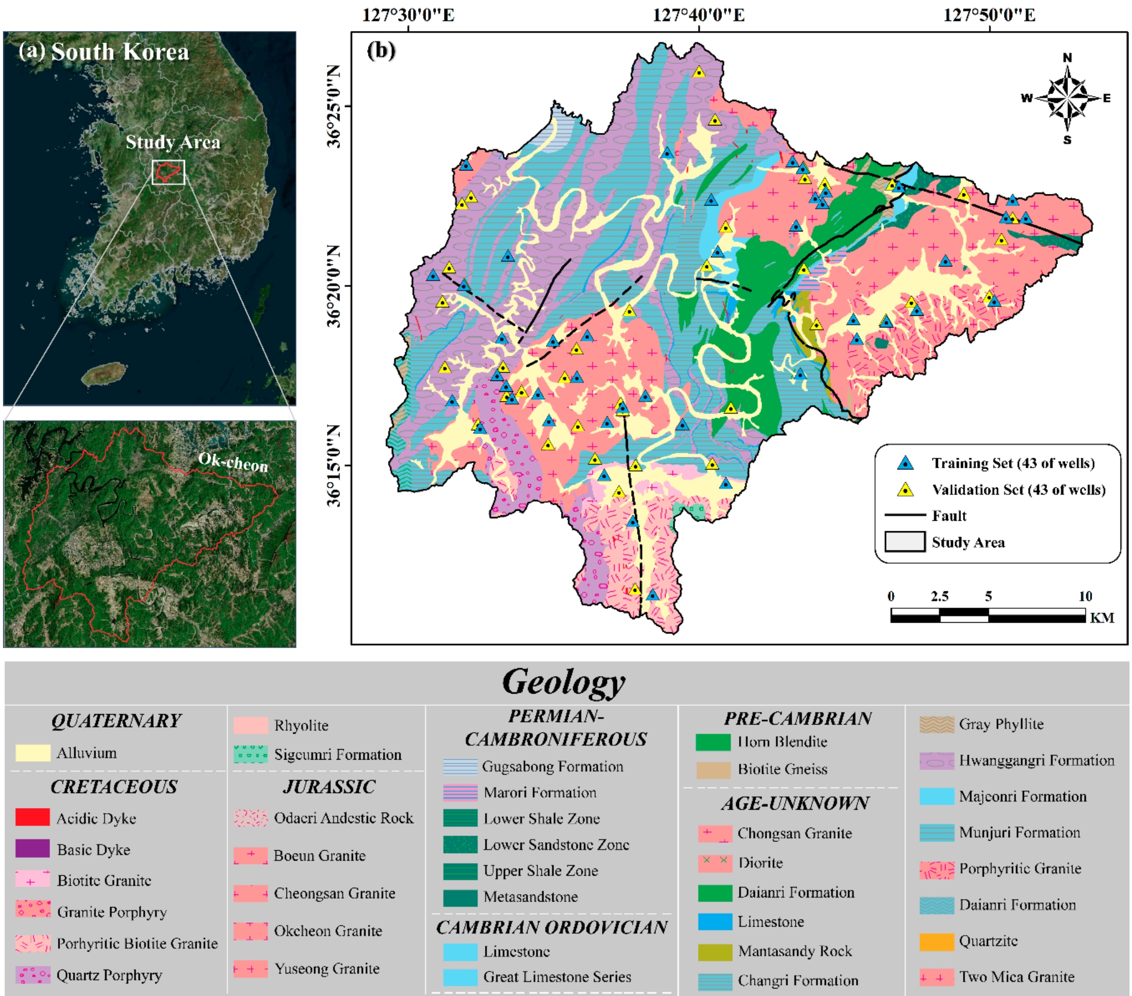

Figure 1. Study area (**a**) and geological map (**b**).

The representative geological structure of the Okcheon area shows the fault to the northwest of the Okcheon. It also shows the thrust fault in the Okcheon and Choseon supergroup, located over the upper Pyeongan Supergroup. The thrust fault developed in the northeast and south-northwest directions [7]. A strike-strip fault exists to the northwest of the Okcheon fault which occurs across Jurassic granite rocks, the Okcheon supergroup, the Paleozoic sequence, and the Triassic granite rocks. It stretches to tens of kilometers from the west side (Figure 1).

2.2. Spatial Data Set

This study used three models that are based on the relationship between groundwater productivity and geological factors (Table 1). To calculate GPP accurately, we used SPC and T as groundwater productivity values.

Table 1. Spatial data set related to groundwater of the study area.

Category	Factors	Data Type	Scale
Geological map [1]	Hydrogeology	Polygon	1:50,000
Land cover map [2]	Land cover	Polygon	1:5000
Soil map [3]	Soil texture	Polygon	1:25,000
Topographic map [4]	Slope gradient Hydraulic slope gradient Relative slope position Valley depth Topographic Wetness Index (TWI) Slope Length factor (LS-factor) Drainage basin Distance from lineament Line density Distance from fault Distance from channel network Depth of groundwater Terrain Ruggedness Index (TRI) Convergence index Plan curvature	GRID	1:5000

[1] The geology map offered by Ministry of Land, Transport and Maritime Affairs. [2] The land cover map offered by the Korea Ministry of Environment. [3] The soil map and land cover map offered by the National Institute of Agricultural Science and Technology. [4] Topographical maps offered by National Geographic Information Institute.

SPC is defined as the amount of water that can be produced by lowering a unit of the surface of water contained in wells through pumping. Its value is derived using the pumping test results of dividing the pumping rate by the drawdown. The pumping tests last for more than 24 h. The formula for calculating the SPC is as follows:

$$SPC = \frac{Q}{h_0 - h} \tag{1}$$

where SPC is the specific capacity of aquifer [L^2T^{-1}]; m^3/day/m), Q is the pumping rate ([L^3T^{-1}]; m^3/day), and $h_0 - h$ is the drawdown ([L]; m). T is defined as the flow rate under unit pressure. It is a function of the unit width of the entire aquifer. Therefore, T represents the ability to transfer the flow in aquifers at constant thickness. T is the measure of a material's capacity to transmit water according to Darcy's law. In other words, it indicates the volume of water flowing through a 0.3 m × 0.3 m cross-sectional area of an aquifer under a hydraulic gradient of 0.3 m/0.3 m in a given amount of time (usually 24 h).

$$K'(x,y) = \frac{1}{b}\int_0^b K(x,y,z)dz \tag{2}$$

$$T = Kb \tag{3}$$

where T is transmissivity (L^2T^{-1}), b is aquifer thickness (L), and K is hydraulic conductivity [6]. Generally, high values of T indicate wider unit widths of the aquifer and better drawdown. The mathematical calculations for the process of estimating T using SPC are explained in detail in [52,53]. All the T and SPC values in this study were extracted from the pumping test recorded in [7].

The Table 2 shows that the results of the pumping test Okcheon for about 120 min for each wells in Okcheon. The test were performed by Korea Institute of Geoscience and Mineral Resource, and all detailed experimental procedures and results are reported in the national groundwater survey report [7]. The groundwater productivity data used in this study were converted into the binary form, where 1 is displayed when there is more than a median value of groundwater productivity, and 0 is displayed otherwise. The split criterion was T (2.6 m^2/day) and the corresponding SPC (4.88 m^3/day/m), which is the median of the two values. In the present study, we applied T and SPC to the three models.

Table 2. The results of the pumping test of wells in the Okcheon-gun.

Type of Aquifers	SPC (m^3/day/m)				T (m^2/day)			
	Min	Max	Average	Median	Min	Max	Average	Median
Porous rock saturated aquifers	2.23	769.23	20.07	4.88	0.70	489.91	23.78	2.61
Alluvial aquifer	2.67	283.33	37.60		0.83	73.16	11.30	

18 various topographical factors were used for GPP analysis, including terrain and surface data derived from remote sensing images. (Figure 2). The selected factors were slope gradient, relative slope position, plan curvature, hydrogeology, hydraulic slope, distance from faults, distance from lineament, depth of groundwater, distance from channel network, lineament density, valley depth, Topographic Wetness Index (TWI), slope length (LS) factor, drainage basin, Terrain Ruggedness Index (TRI), convergence index, land-cover, and soil texture. The spatial database of these factors was reproduced using the ArcGIS software with SAGA-GIS.

The topographical data was obtained through digitizing using aerial photographs taken in 2006; additional corrections were performed and updated by other high-resolution satellite images. The satellite image used for correction was Pleiades 1A, spatial resolution of multi-spectral is 0.5 m, and the image was similar to that of the aerial photograph. Land cover maps were classified into 8 main categories using an unsupervised classification method from aerial photographs with a spatial resolution of 0.25 m taken in 2013. In addition, Kompsat-3 remote sensing image with spatial resolution of 0.7 m was used to evaluate the classification accuracy [45].

The digital elevation model (DEM) was generated from a topographic map with a resolution of 30 m using a 1:5000 digital topographic map from the National Geographic Information Institute (NGII). The slope gradient, plan curvature, relative slope position, valley depth, LS factor, convergence index, drainage basin, TWI, and TRI were calculated using the DEM [54]. Also, various thematic maps such as those depicting soil texture, land cover, and hydrogeology were resampling at 30 m resolution and used in this study [11].

Various parameters were used to analyze the GPP with more precision. The LS factor is the ratio of soil loss per unit catchment area to the slope length (L) and slope steepness (S). The formula proposed by Moore and Burch [55] for calculating the LS factor is as follows (Equation (4)):

$$LS = \left(\frac{A_s}{22.13}\right)^{0.6}\left(\frac{\sin\beta}{0.0896}\right)^{1.3} \quad (4)$$

where,

A_s is the catchment area
β represents the slope gradient measured in degrees

The convergence index represents the structure of the slope as a set of convergence and divergence sites. The index value for maximum convergence is +100. Conversely, the index value for maximum divergence is −100. If there is a flat, the index is 0 [56]. This means that the index value is closer to convergence (+100) for larger slope values.

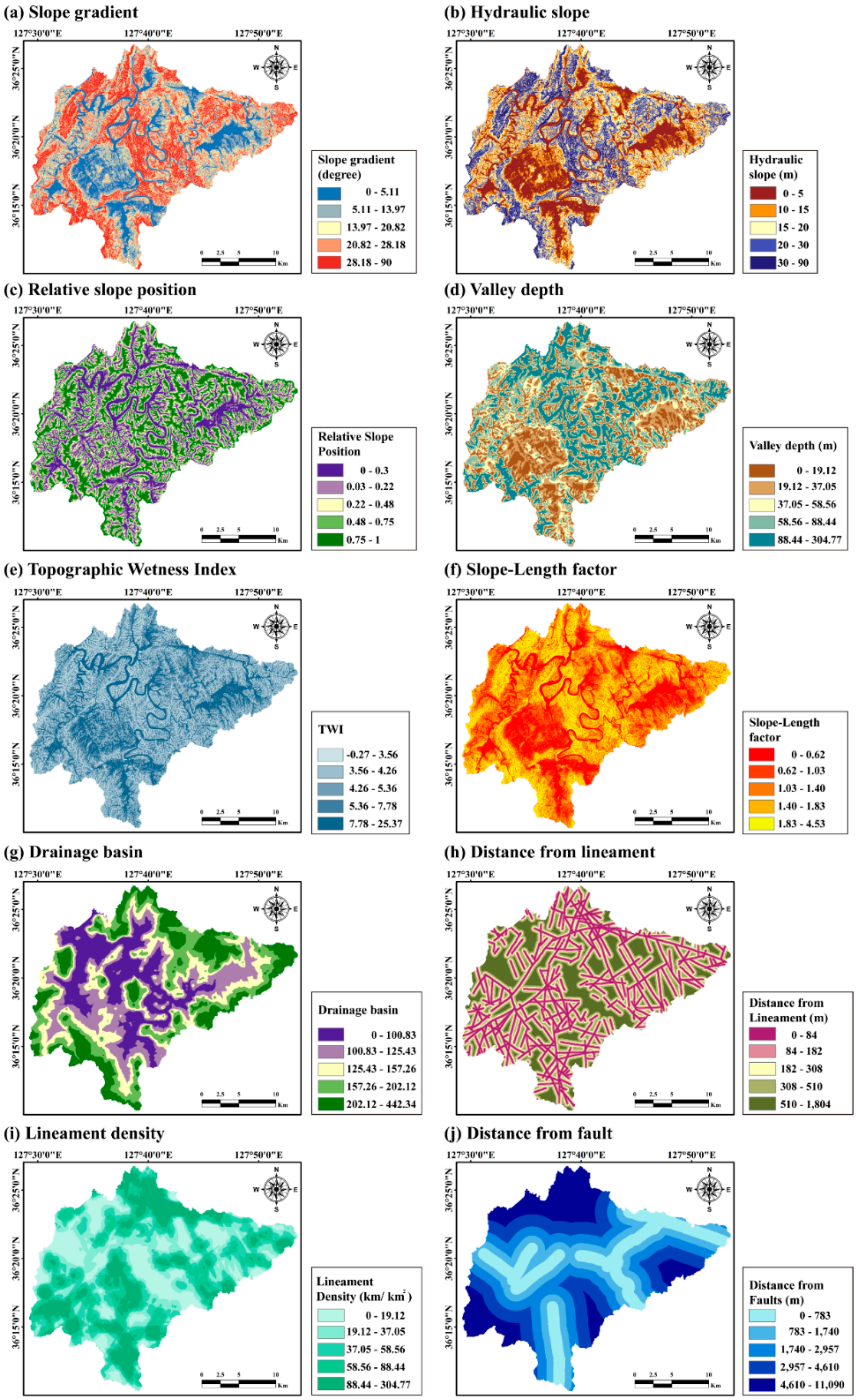

Figure 2. *Cont.*

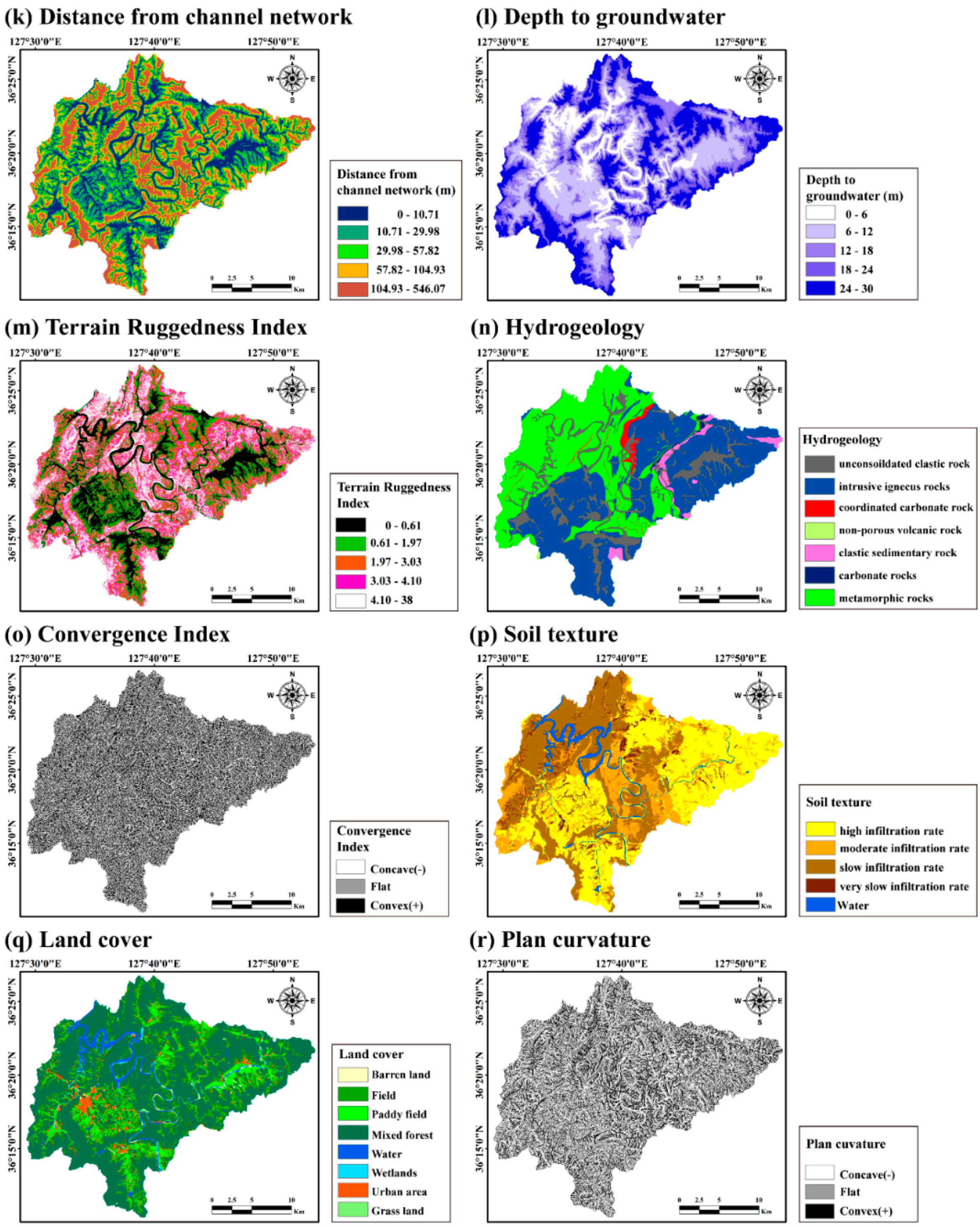

Figure 2. The spatial database constructed for Groundwater Productivity Potential (GPP). (**a**) slope gradient, (**b**) hydraulic slope, (**c**) relative slope position, (**d**) valley depth, (**e**) Topographic Wetness Index (TWI), (**f**) Slope length (LS) factor, (**g**) drainage basin, (**h**) distance from lineament, (**i**) lineament density, (**j**) distance from fault depth (continue). (**k**) distance from channel network, (**l**) depth of groundwater, (**m**) Terrain Ruggedness Index (TRI), (**n**) Hydrogeology, (**o**) convergence index, (**p**) soil texture, (**q**) land cover, (**r**) plan curvature.

The TRI is an index developed by Riley [57]. TRI represents the altitude difference between adjacent cells in a grid. The TRI index is calculated by determining the height differences between the center cell and the eight cells surrounding it. This difference corresponds to the average altitude change between any point on the grid and the surrounding area.

3. Methodology

The GPP mapping process is shown in Figure 3. A total of 84 groundwater wells were split: 50% were randomly demarcated as training data and the other 50% were retained as validation data. A total of 18 hydrogeology-related factors were combined into a spatial database. Then, the selected T and SPC data (T values $\geq$ 2.61 m^2/day and SPC values $\geq$ 4.88 m^3/day/m) were used to train the three models. Finally, the results of GPP maps was verified using Area-Under-the-Curve (AUC) analysis.

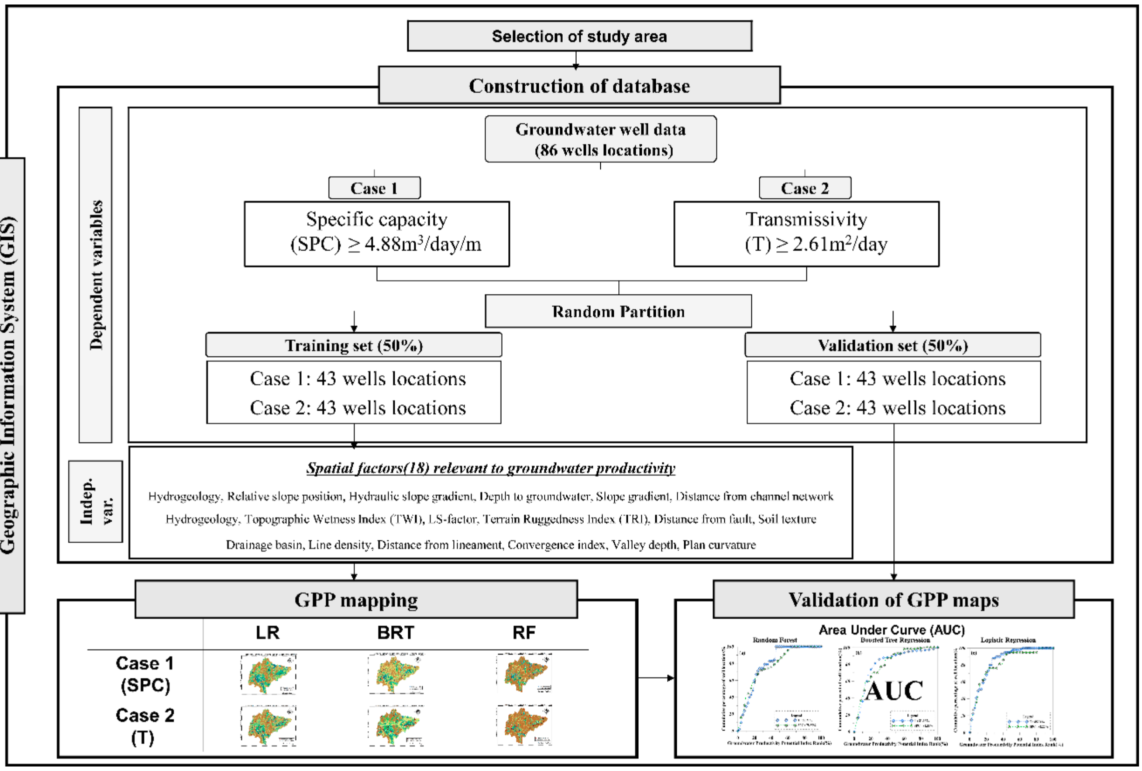

Figure 3. Flow chart of the study procedures.

The 18 factors were arranged in a grid format with 1016 rows by 1211 columns. There was a total of 601,320 cells in the grid. The T and SPC values corresponded to a total of 86 cells.

3.1. Random Forest(RF) Model

The random forest model is an ensemble classification technique that was developed as an extension of classification and regression trees (CART) to improve the prediction performance of the model [58]. The RF model constructs numerous decision trees to estimate the spatial relationship between groundwater and various topographic factors that consist of either categorical or continuous response variables. The RF model functions in two steps. First, it constructs the plurality of decision trees; this is the learning step. Second, a test to classify a loaded input value or predict its loading is performed. The advantages of the RF model include extremely high accuracy, simple and fast learning and testing algorithms, ability to handle thousands of input variables without deleting variables, good generalization performance through randomization, and multi-class algorithm characteristics.

Before running the RF model, we have defined two parameters. The first is the number of randomly sampled variables (m_{try}) to use in each tree building process and the number of trees (n_{tree}) to build in the forest to run. Both parameters should be optimized to minimize generalization errors.

Breiman [59], Liaw and Wiener [60] stated that even a single variable ($m_{try} = 1$) can performance to high accuracy, while Grömping [61] showed that more than two variables (i.e., $m_{try} = 2, 3, 4, \ldots, m$) should be used to increase the accuracy of the model.

The RF model consists of a combination of numerous trees generated by bootstrap samples using out-of-bag (OOB) errors. Two thirds of the samples are used for training, and the other 1/3 are used for verification. The OOB is an unbiased estimate of the generalization error. A detailed description of the mathematical formulation of RF model is found in Breiman [59].

The goal of the RF model is to analyze the relationship between independent and dependent variables in the model building stage to determine the weights for each factor. In this study, in order to analyze correlations between groundwater and related factors, groundwater productivity data (T or SPC) was used as a dependent variable, and 18 groundwater-related factors were used as independent variables. The parameters used in the RF model are as follows: (1) the number of randomly sampled variables in each spilt (m_{try}), (2) the number of trees to be grown (n_{tree}), and (3) the minimum size of the observations at the end node of the tree (node size). These parameters were set to 500, 10, and 5, respectively using the STATISTICA 10.1 software [62].

3.2. Logistic Regression(LR) Model

The LR model is useful for predicting whether groundwater exists in a particular location, based on the predictor variables. The primary reason for using the LR model is to explain the relationship between dependent variables and independent variables [63]. The advantage of the LR model is that variables need not have normal distribution, regardless of whether they are continuous, discrete, or a combination of both types [26]. In this study, dependent variables indicate the presence of groundwater using a binary variable. The following show the relationship between groundwater presence and the dependency of a variable:

$$P = \frac{e^Z}{1 + e^Z} \tag{5}$$

$$Z = a + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_mx_m \tag{6}$$

$$Z = log_e\left[\frac{p}{1-p}\right] = logit(p) \tag{7}$$

where a is the intercept of the LR model, $x_1, x_2, \cdots, x_m$ are regression coefficient of the logistic regression model, and Z is a linear combination function of the coefficient representing a linear relationship. The parameters $b_1, b_2, \cdots, b_m$ are the independent variable. The probability (p) represents to the estimated probability of potential groundwater. The value of Z is denoted in the binary form, where Z = 1 implies more than a specific amount of groundwater (T ≥ 2.61 m^2/day or SPC ≥ 4.88 m^3/day/m), and Z = 0 indicates either less than that specific amount or no groundwater. Function Z is represented as logit (p) is a likelihood ratio that the dependent variable Z is 1. The LR model coefficient is a value that represents the percentage (%) of the variance of the dependent that variable is explained by the independent variable, and has a value between 0.00 and 1.00. A value closer to 1.00 means closer to a perfect relationship, which is almost the same as the square of the multiple correlation coefficient in a linear regression.

3.3. Boosted Regression Tree (BRT) Model

BRT is one of the many ensemble models that combine two or more models to enhance the capability for prediction. This model can be used to effectively classify or perform regression analysis considering continuous and categorical data. The model constructs a binary tree that is divided into two samples. Each split node determines whether an observed value corresponds to a binary of 1 or 0. The residual and standard deviations of the node are calculated in the following step. It has also been used in research to detect natural resources such as groundwater and minerals. Basically, the model depends on the number of the regression trees produced. Thus, it is likely to be the same as the RF model. BRT adopts a machine learning technique to resolve regression problems and uses a predefined loss function to create each regression tree step by step. It measures the error in a step and fixes it in the subsequent steps. The BRT model does not need to transform original data or remove outlier data for training. It is also suitable for analyzing nonlinear complex relationships [64].

3.4. GPP Mapping Process

First, we applied the Frequency Ratio (FR) method to calculate the spatial relationships between groundwater presence and related topographical factors. In this study, we analyzed for correlation between the groundwater wells locations and 18 factors related to groundwater productivity using frequency ratio of each factor. The relationship analysis is the ratio of the area where groundwater productivity in the total study area. So, a value of 1 indicates an average. In other words, if the average value of FR is greater than 1, it has a higher correlation with groundwater. Also, the logistic regression coefficient correlates with the potential productivity of groundwater, and the higher the value, the higher the correlation. The details of FR calculations are described in more detail in [6]. To analyze the GPP, the results of the three models were compared based on the FR model results. The results of the RF and BRT models were calculated using the STATISTICA software and there were classified and re-summed over regression values at all nodes to calculate the importance of the predictor. The LR model was calculated using SPSS 21 statistical software. Groundwater productivity data (T and SPC) were randomly separated for each model. They were used as training (50%) and verification (50%) data.

Most of the maps used in GPP analysis were generated using the ArcGIS 10.5 software. A number of groundwater-related factor maps, including geological maps, were regenerated in the ASCII grid format at a 30 m resolution. The groundwater productivity values (T and SPC) were set as independent variables and used as training data. In the following step, all data were classified as categorical or continuous data. The continuous variables included the Terrain Ruggedness Index (TRI), slope length (LS) factor, hydraulic slope, depth of groundwater, lineament density, slope gradient, relative slope position, drainage basins, valley depth, distance from faults, Topographic Wetness Index (TWI), and distance from lineament. The categorical variables included the hydro-geology, soil texture, convergence index, plan curvature, and land cover.

To validate the algorithms, 86 T data points were divided into two different groups and randomly selected. Verification was performed using the previously segregated verification data. In the final verification process, a Receiver Operating Characteristics (ROC) curve was implemented. ROC is an index for the performance of models [13]. To quantitatively determine the accuracy of the models' verification, Area Under the Curve (AUC) of the ROC curve was recalculated for the total area and the correct prediction accuracies were obtained. Typically, the accuracy of the validation of the model is measured by the area under the ROC curve, and it can lie between 0.5 and 1. High AUC values indicate the superior performance of an algorithm.

4. Results

4.1. Correlation between GPP and the Variables

Generally, the productivity of groundwater is affected by various factors such as topography, hydrogeology, soil, forestation, and flow velocity [54]. To quantitatively analyze, we examined the relationship between the related factors using the FR and LR models.

Table 3 shows the coefficients of the factors in each class range; they were calculated with respect to groundwater T and SPC values. In general, the relationship between slope gradient and groundwater is inversely proportional. Thus, when the slope gradient is high, it is difficult for groundwater to accumulate in the aquifer. For slopes between 0° to 5°, the ratio was approximately 3.0, which indicated a high probability of GPP. The hydraulic slope gradient, relative slope position, valley depth, and slope-length (LS) were also found to be inversely proportional to GPP.

Table 3. Frequency ratio and logistic regression (LR) model's results between groundwater productivity data and related factors.

Factor	Class	No. of Pixels in Domain [a]	% of Domain	T ≥ 2.61 [b]			SPC ≥ 4.88 [c]			Logistic Regression Coefficient	
				No. of Data 1	% of Data 1	FR of Data 1	No. of Data 1	%of Data 1	FR of Data 1	T ≥ 2.61	SPC ≥ 4.88
Slope gradient (degree)	**0–5.11**	**113,319**	**18.85**	**25**	**58.14**	**3.09**	**24**	**55.81**	**2.96**		
	5.11–13.97	128,401	21.35	16	37.21	1.74	16	37.21	1.74		
	13.97–20.82	112,842	18.77	2	4.65	0.25	3	6.98	0.37	−0.58	−0.74
	20.82–28.18	118,770	19.75	0	0.00	0.00	0	0.00	0.00		
	28.18–90	127,988	21.28	0	0.00	0.00	0	0.00	0.00		
Hydraulic slope (degree)	**0–5**	**154,281**	**25.66**	**33**	**76.74**	**2.99**	**35**	**81.40**	**3.17**		
	5–10	113,239	18.83	8	18.60	0.99	7	16.28	0.86		
	10–20	178,303	29.65	2	4.65	0.16	1	2.33	0.08	−0.40	−0.84
	20–30	101,221	16.83	0	0.00	0.00	0	0.00	0.00		
	30–90	54,276	9.03	0	0.00	0.00	0	0.00	0.00		
Relative slope position	**0–0.0275**	**118,086**	**19.64**	**17**	**39.53**	**2.01**	**22**	**51.16**	**2.61**		
	0.0275–0.2235	122,639	20.40	19	44.19	2.17	14	32.56	1.60		
	0.2235–0.4784	121,248	20.16	2	4.65	0.23	2	4.65	0.23	0.32	−0.52
	0.4784–0.7529	120,084	19.97	1	2.33	0.12	3	6.98	0.35		
	0.7529–1	119,263	19.83	4	9.30	0.47	2	4.65	0.23		
Valley depth (m)	0–19.1231	118,571	19.72	8	18.60	0.94	6	13.95	0.47		
	19.1231–37.0510	126,094	20.97	17	39.53	1.89	12	27.91	0.79		
	37.0510–58.5645	122,756	20.41	5	11.63	0.57	11	25.58	1.25	−0.16	−0.28
	58.5645–88.4443	**119,947**	**19.95**	**10**	**23.26**	**1.17**	**8**	**18.60**	**1.42**		
	88.4443–304.7743	113,952	18.95	3	6.98	0.37	6	13.95	1.07		
TWI	−0.27–3.6	117,062	19.47	2	4.65	0.24	1	2.33	0.12		
	3.6–4.35	129,046	21.46	2	4.65	0.22	3	6.98	0.33		
	4.35–5.4	118,685	19.74	3	6.98	0.35	2	4.65	0.24	0.02	−0.11
	5.4–7.8	**117,832**	**19.59**	**17**	**39.53**	**2.02**	**19**	**44.19**	**2.26**		
	7.8–25.37	118,695	19.74	19	44.19	2.24	18	41.86	2.12		
LS factor	**0–1.0473**	**117,812**	**19.59**	**25**	**58.14**	**2.97**	**25**	**58.14**	**2.97**		
	1.0473–3.7223	119,314	19.84	14	32.56	1.64	13	30.23	1.52		
	3.7223–6.3280	123,649	20.56	3	6.98	0.34	3	6.98	0.34	−0.56	0.23
	6.3280–8.9336	123,837	20.609	1	2.33	0.11	2	4.65	0.23		
	8.9336–47.4598	116,708	19.41	0	0.00	0.00	0	0.00	0.00		
Lineament density (km/km^2)	0–0.6219	118,888	19.77	4	9.30	0.47	4	9.30	0.47		
	0.6219–1.0305	123,348	20.51	7	16.28	0.79	7	16.28	0.79		
	1.0305–1.4036	123,437	20.53	9	20.93	1.02	11	25.58	1.25	0.06	0.05
	1.4036–1.8300	**118,218**	**19.66**	**11**	**25.58**	**1.30**	**12**	**27.91**	**1.42**		
	1.8300–4.5306	117,429	19.53	12	27.91	1.43	9	20.93	1.07		
Distance from fault (m)	**0–783**	**116,641**	**19.40**	**10**	**23.26**	**1.20**	**11**	**25.58**	**1.32**		
	783–1740	122,116	20.31	14	32.56	1.60	10	23.26	1.15		
	1740–2957	122,194	20.32	8	18.60	0.92	8	18.60	0.92	−0.13	−0.27
	2957–4610	120,773	20.08	7	16.28	0.81	7	16.28	0.81		
	4610–11,090	119,596	19.89	4	9.30	0.47	7	16.28	0.82		
Distance from lineament (m)	**0–84**	**133,995**	**22.28**	**14**	**32.56**	**1.46**	**15**	**34.88**	**1.57**		
	84–182	119,978	19.95	8	18.60	0.93	9	20.93	1.05		
	182–308	119,286	19.84	8	18.60	0.94	6	13.95	0.70	0.12	0.05
	308–510	117,397	19.52	8	18.60	0.95	9	20.93	1.07		
	510–1804	110,664	18.40	5	11.63	0.63	4	9.30	0.51		
Distance from channel network (m)	**0–10.7073**	**126,750**	**21.08**	**25**	**58.14**	**2.76**	**28**	**65.12**	**3.09**		
	10.7073–29.9805	124,456	20.70	15	34.88	1.69	12	27.91	1.35		
	29.9805–57.8195	120,035	19.96	1	2.33	0.12	2	4.65	0.23	−0.68	−0.05
	57.8195–104.9317	115,499	19.21	1	2.33	0.12	1	2.33	0.12		
	104.9317–546.0730	114,580	19.05	1	2.33	0.12	0	0.00	0.00		
Depth of ground water (m)	0–6	77,398	12.87	8	18.60	1.45	8	18.60	1.45		
	6–12	**165,831**	**27.58**	**22**	**51.16**	**1.86**	**24**	**55.81**	**2.02**		
	12–18	118,083	19.64	9	20.93	1.07	9	20.93	1.07	−0.37	0.33
	18–24	87,655	14.58	2	4.65	0.32	1	2.33	0.16		
	24–30	152,353	25.34	2	4.65	0.18	1	2.33	0.09		
Drainage basin (km^2)	0–100.8281	120,219	19.99	6	13.95	0.70	7	16.28	0.81		
	100.8281–125.4287	**123,553**	**20.55**	**17**	**39.53**	**1.92**	**19**	**44.19**	**2.15**		
	125.4287–157.2648	121,267	20.17	7	16.28	0.81	7	16.28	0.81	0.60	0.32
	157.2648–202.1247	120,238	20.00	11	25.58	1.28	9	20.93	1.05		
	202.1247–442.3421	116,043	19.30	2	4.65	0.24	1	2.33	0.12		
Terrain Ruggedness Index (TRI)	**0–0.6067**	**114,532**	**19.05**	**22**	**51.16**	**2.69**	**24**	**55.81**	**2.93**		
	0.6067–1.9716	130,756	21.74	19	44.19	2.03	16	37.21	1.71		
	1.9716–3.0333	125,987	20.95	1	2.33	0.11	1	2.33	0.11	0.75	0.52
	3.0333–4.0950	115,656	19.23	1	2.33	0.12	2	4.65	0.24		
	4.0950–38.0000	114,389	19.02	0	0.00	0.00	0	0.00	0.00		

Table 3. *Cont.*

Factor	Class	No. of Pixels in Domain [a]	% of Domain	T ≥ 2.61 [b]			SPC ≥ 4.88 [c]			Logistic Regression Coefficient	
				No. of Data 1	% of Data 1	FR of Data 1	No. of Data 1	%of Data 1	FR of Data 1	T ≥ 2.61	SPC ≥ 4.88
Hydro geology	**Unconsolidated clastic rock**	**94,010**	**15.63**	**17**	**39.53**	**2.53**	**17**	**39.53**	**2.53**	−0.20	−1.44
	Intrusive igneous rocks	255,683	42.52	19	44.19	1.04	17	39.53	0.93	0.04	−0.99
	Dolomite rock	9173	1.53	1	2.33	1.52	0	0.00	0.00	0.75	−11.90
	Non-porous volcanic rock	431	0.07	0	0.00	0.00	0	0.00	0.00	−8.13	−8.27
	Clastic sedimentary rock	15,536	2.58	1	2.33	0.90	1	2.33	0.90	1.02	−0.26
	Carbonate rocks	3608	0.60	0	0.00	0.00	0	0.00	0.00	1.76	0.71
	Metamorphic rocks	222,879	37.06	5	11.63	0.31	8	18.60	0.50	0	0
Land cover	Barren land	5464	0.91	1	2.33	2.56	0	0.00	0.00	0.50	10.05
	Field	77,488	12.89	13	30.23	2.35	13	30.23	2.35	0.21	10.18
	Paddy field	**62,789**	**10.44**	**18**	**41.86**	**4.01**	**18**	**41.86**	**4.01**	0.37	1.08
	Mixed forest	395,630	65.79	4	9.30	0.14	5	11.63	0.18	−1.15	9.82
	Water	24,128	4.01	1	2.33	0.58	0	0.00	0.00	−0.53	0.36
	Wetlands	3963	0.66	0	0.00	0.00	0	0.00	0.00	−11.64	10.24
	Urban area	21,945	3.65	5	11.63	3.19	6	13.95	3.82	0.53	10.97
	Grass land	9913	1.65	1	2.33	1.41	1	2.33	1.41	0.00	0.00
Soil texture	High Infiltration rate	247,471	41.15	21	48.84	1.19	23	53.49	1.30	0.50	10.71
	Moderate infiltration rate	105,563	17.56	10	23.26	1.32	7	16.28	0.93	0.21	10.24
	Low Infiltration rate	199,492	33.18	7	16.28	0.49	8	18.60	0.56	0.37	10.03
	Very slow infiltration rate	**22,539**	**3.75**	**4**	**9.30**	**2.48**	**5**	**11.63**	**3.10**	−1.15	10.22
	Water	26,255	4.37	1	2.33	0.53	0	0.00	0.00	−0.53	0.00
Plan curvature	**Concave (−)**	**308,875**	**51.37**	**29**	**67.44**	**1.31**	**27**	**62.79**	**1.22**	0.35	−1.20
	0	1409	0.23	0	0.00	0.00	0	0.00	0.00	−9.79	−9.66
	Convex (+)	291,036	48.40	14	32.56	0.67	16	37.21	0.77	0.00	0.00
Convergence index	**Concave (−)**	**294,208**	**48.93**	**26**	**60.47**	**1.24**	**26**	**60.47**	**1.24**	0.43	0.91
	0	1434	0.24	0	0.00	0.00	0	0.00	0.00	−9.67	−7.87
	Convex (+)	305,678	50.83	17	39.53	0.78	17	39.53	0.78	0.00	0.00

[a] Total number of pixels is 601,320. [b,c] Total number of pixels of wells location is 43 (training set).

In the case of hydrogeological factors, the frequency ratio was higher for unconsolidated clastic sediment areas (Table 3). It was 0 for carbonate rocks, dolomite rock, and non-porous volcanic rocks. The areas with unconsolidated clastic sediments were shows that have a stronger GPP than areas with carbonated rocks because groundwater cannot easily penetrate between their particles that were so tiny and dense. With regards to land cover, groundwater potential values were higher for paddy fields and urban areas and lower in the mixed forest area. In fact, it is highly probable that there are many wells containing a large amount of groundwater in the mixed-forest covered area, but the frequency ratio was relatively low because 65% of the study area was covered with mixed-forest.

In the case of soil texture, the frequency ratio was higher for the D class (very slow infiltration rate) and lower for the C class (low infiltration rate). Sandy soil (D class) has an excellent effect on groundwater penetration because of its high permeability. Conversely, clay soil (A) has a low impact on groundwater accumulation because of its poor drainage capability and permeability.

In case of plan curvature, concave areas have a ratio of about 1.2 which is considerably higher than that of convex areas (approximately 0.70). Concave surfaces contain more water, particularly during periods of heavy rainfall. Therefore, areas with concave surfaces are more advantageous than areas with convex surfaces for storing groundwater. The GPP frequency ratio generally increases with increase in linear density. That is, the nearer the linear density is to 0, the lower the GPP generation is. When the value of linear density is larger, the GPP generation is also higher. With regards to the topological factors, such as distance from a fault, lineament, and channel network, the closer the area is to a river, the higher is the likelihood of the groundwater productivity. The longer the distance is, the lesser is the likelihood of groundwater generation. In other words, various linear structures and remote areas were leakier, while the nearby areas had better recharge and higher penetration.

The depth of groundwater was highest between 6 m and 12 m. The TWI index is defined as a function of the upstream contributing area per unit and the slope gradient. The results have shown that as the TWI value increases, the groundwater productivity ratio also increases. This is because high wetting index demonstrates better groundwater retention capability of an area.

The TRI index represents the altitude difference between adjacent cells in grid. That is, the higher the TRI value, the greater is the difference in altitude between adjacent areas. Therefore, low TRI value indicates high groundwater content because the GPP is higher at low altitude differences. A drainage basin represents a catchment area. When the TRI value is between 100 and 105, the GPP is the highest.

For mapping the GPP, the LR coefficients of the 18 factors were computed using the SPSS software (Table 3). The LR coefficient represents the probability of occurrence, and this value typically ranges between 0 and 1. If the value of multiple logistic coefficients is calculated to be less than 0, the GPP is low. This is so because GPP becomes smaller than 1 when converted to the corresponding log value.

Positive values were obtained for the relative slope position, TWI, lineament density, distance from lineament, drainage basin, and TRI when T productivity was used in the LR model. SPC productivity values were positive for LS factor, lineament density, distance from lineament, depth of groundwater, drainage basin, and TRI. Non-porous volcanic rocks had the lowest T and SPC values for the hydrogeology factor, thereby indicating the lowest impact on GPP. The urban areas had the highest values of land-cover factor. The flat item showed least influence on the GPP corresponding to plan curvature and convergence index variables.

Table 4 shows the importance of the values of each predictor variable in the BRT and RF models. The data in the table also explain the correlation between GPP and the related factors. The predictor importance ranges between 0 and 1. It indicates a factor near 1 that can be said to be closely related to the presence of groundwater. As shown in Table 4, the most important variable affecting groundwater productivity (both T and SPC values) when applying the BRT and RF models is land cover. Conversely, the least influential variable in the case of both models is plan curvature.

Table 4. Predictor of importance factor of the Boosted Regression Tree (BRT), Random Forest (RF) models.

Factor	Boosted Regression Trees		Random Forest	
	$T \geq 2.61$	**$SPC \geq 4.88$**	**$T \geq 2.61$**	**$SPC \geq 4.88$**
Land cover	1.000000	1.000000	1.000000	0.823951
Relative slope position	0.250930	0.690807	0.139181	0.698339
Hydraulic slope gradient	0.282250	0.509170	0.231625	0.723901
Depth of groundwater	0.176912	0.484824	0.121499	0.970446
Slope gradient	0.297003	0.480254	0.302406	0.677954
Distance from channel network	0.259162	0.457526	0.109057	0.845124
Hydrogeology	0.371401	0.455671	0.113639	0.843345
Topographic Wetness Index (TWI)	0.217793	0.349910	0.255783	0.642187
LS-factor	0.251877	0.323731	0.250476	0.850150
Terrain Ruggedness Index (TRI)	0.332975	0.311433	0.162339	0.723926
Distance from fault	0.170614	0.302349	0.523101	0.933228
Soil texture	0.181788	0.301434	0.556208	1.000000
Drainage basin	0.169170	0.253417	0.149756	0.841819
Line density	0.101006	0.214975	0.260951	0.612850
Distance from lineament	0.114429	0.201170	0.174736	0.403119
Convergence index	0.061917	0.135385	0.155390	0.347651
Valley depth	0.155604	0.133019	0.070290	0.532714
Plan curvature	0.052066	0.060235	0.134918	0.392855

4.2. GPP Mapping and Validation

The GPP map was generated using the predictor values determined by the three models. That is, the higher the value of probability for an area, the more likely it is to contain groundwater. The probabilities calculated by the three models was re-expressed in the form of the groundwater productivity potential index (Figure 4).

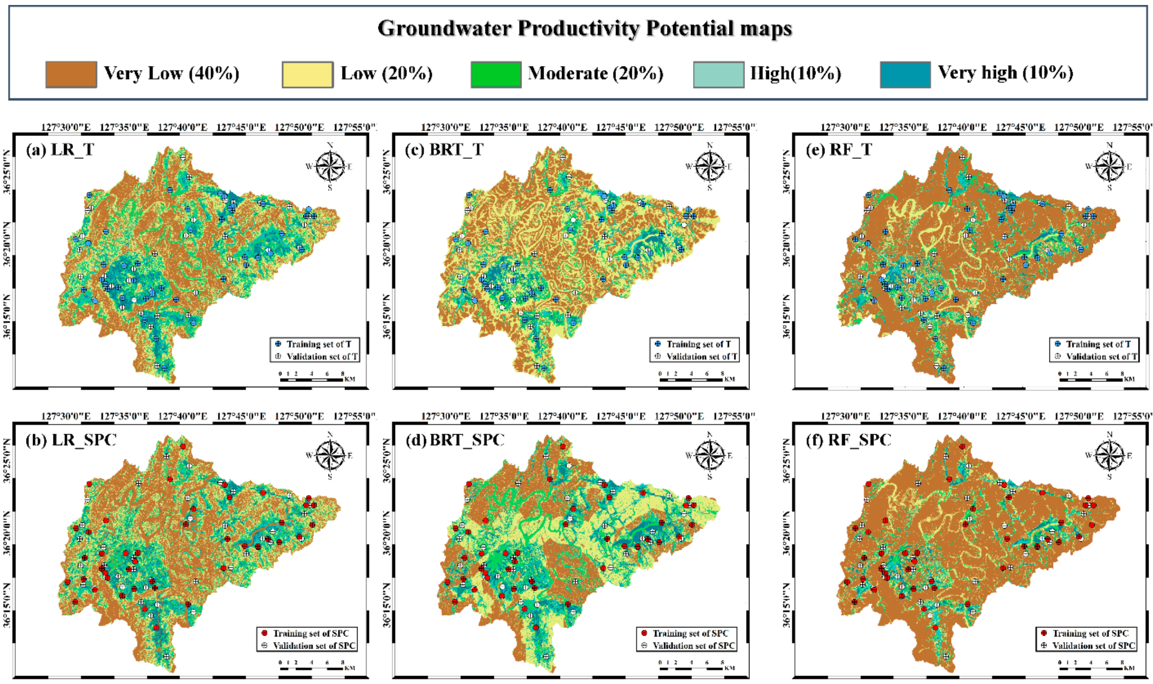

Figure 4. The results of GPP maps generated using LR, BRT and RF models. T/SPC values of (**a**,**b**) LR model, (**c**,**d**) BRT model and (**e**,**f**) RF model.

The next step was to validate the GPP maps created using the BRT, RF, and LR models. The prediction rate of the validations was determined by comparing the GPP maps created using the RF, LR, and BRT models with the remaining 50% of groundwater wells data not used in the training set. A GPP rank with more than 10% value could explain the presence of 30% of the groundwater wells and rank with more than 60% value could explain 90% of the groundwater wells identified by the three models.

AUC was used to comparatively analyze the results of the three models in order to quantitatively compare the results of each model. Upon validation of the GPP maps (Figure 5), the LR, RF, and BRT models produced AUC values of 0.8504, 0.8021, and 0.8166 with T values (i.e., the prediction accuracy was 85.04%, 80.21% and 81.66%), respectively, and 0.8222, 0.7857, and 0.8153 with SPC values (i.e., the prediction accuracy was 82.22%, 78.57% and 81.53%), respectively. All models indicated the presence of 90% of the potential groundwater wells in 60% of the study area.

This study applied the RF, BRT (data mining), and LR (statistics) models to estimate GPP. The RF and BRT data mining models showed good accuracy while spatially predicting GPP. Their accuracy amounted to approximately 80% and more. The LR statistical model showed the highest verification accuracy, reaching beyond 85%.

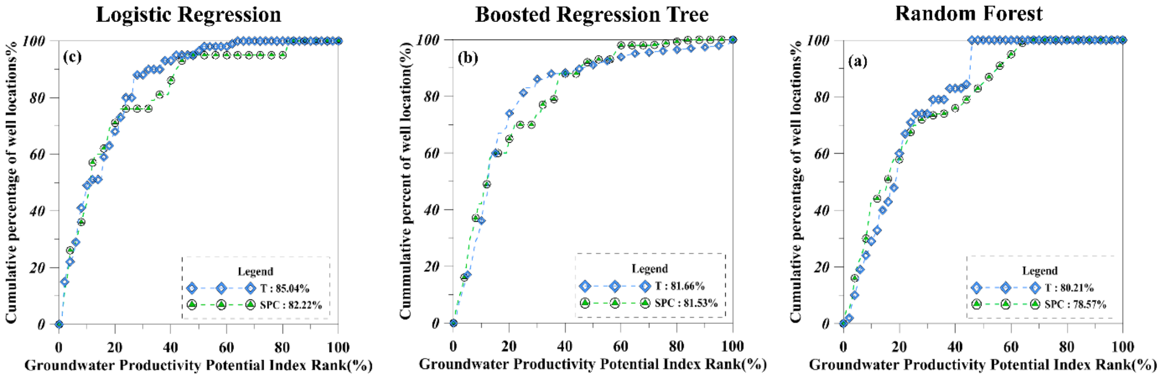

Figure 5. ROC curves for the GPP maps with T/SPC values produced by (**a**) LR model, (**b**) BRT model and (**c**) RF model.

5. Conclusions and Discussion

The region of Okcheon in South Korea needs a stable water management system that can provide high-quality drinking water and water for agricultural use; this system should be based on sources other than surface water. To provide a sufficient amount of water, it is very important to predict the locations of uncontaminated usable groundwater with accuracy. Therefore, this study estimated the groundwater of the un-surveyed area by analyzing the relationship between the well locations and the surrounding environment including the topographic factors using three models. For the models applied in this study, half of well location data (43 well locations) were used as training data and the other half were used as validation data. Total 18 topography, soil, and land cover variables were used as independent variables. Finally, the estimated potential groundwater maps were provided by using three models of LR, RF and BRT.

From the results of the LR, RF and BRT models, the following relationships between wells data and the examined factors could be established. GPP is higher in gentle slopes, hydraulic slopes, lower relative slope positions, and shorter slope lengths because rainfall running off from the upper regions accumulates in the lower regions. This in turn positively influences the aquifer. In addition, the TRI is an index representing the altitude difference between two adjacent areas in open terrain. The GPP is higher where the altitude difference is not significant. On the other hand, distance from fault, lineament, and channel network showed a negative correlation with GPP. Groundwater in aquifers hydrologically flows from high to low gradients like surface water. As a result, most of the groundwater charged in areas of low altitude and some are eventually discharged back to the river, the lowest zone. In the end, most areas with large amounts of groundwater are close to the river, which are clearly reflected in this study. These results indicate that closeness to rivers increases the GPP of an area, as is known from the hydrogeological point of view.

In case of "distance from the fault" factor, Bense, et al. [61] mentioned that the deformation along faults in the shallow crust (<1 km) introduces permeability heterogeneity and anisotropy, which has an important impact on processes such as groundwater. While the results in this paper show that voids between defects have a positive effect on groundwater recharge, direct assessment of the impact between defects and groundwater recharge remains a difficult discussion. We considered that needs to be further discussed through various experiments.

In conclusion, the three proposed models were able to estimate the location of groundwater wells with an average GPP probability of over 80%. These results validate the usefulness of the three models for groundwater resource development. The final GPP map proposed in this paper used 86 limited wells data, so there is a limit to reflect the real world. However, if we collect more wells data for this region in the future and perform a GPP analysis, you can expect better results. Also, it is showed from the results that the accuracy is higher when GPP is predicted using T values rather than SPC values. Despite the limitations, these GPP mapping methods can be efficiently applied in the future for national groundwater development and utilization planning in Korea.

Author Contributions: S.L. and H.-S.J. conceived and designed the experiments; J.-C.K. performed the experiments; All the authors analyzed the results, J.-C.K. wrote the paper.

Funding: This research received no external funding.

Acknowledgments: This research was conducted as parts of the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) and the Science and Technology Internationalization Project (NRF-2016K1A3A1A09915721) hosted by the National Research Foundation of Korea (NRF) Grant, which is funded by the Ministry of Science and ICT.

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

References

1. The United Nations World Water Development Report 2018 (WWDR). In *Nature-Based Solutions for Water*; UNESCO: Paris, France, 2018.
2. World Health Organization (WHO). *Progress on Sanitation and Drinking Water: 2015 Update and mdg Assessment*; WHO: Geneva, Switzerland, 2015.
3. World Economic Forum (WEF). *The Global RISK Report 2016*; WEF: Geneva, Switzerland, 2016.
4. Vadiati, M.; Asghari-Moghaddam, A.; Nakhaei, M.; Adamowski, J.; Akbarzadeh, A. A fuzzy-logic based decision-making approach for identification of groundwater quality based on groundwater quality indices. *J. Environ. Manag.* **2016**, *184*, 255–270. [CrossRef] [PubMed]
5. Naghibi, S.A.; Moghaddam, D.D.; Kalantar, B.; Pradhan, B.; Kisi, O. A comparative assessment of gis-based data mining models and a novel ensemble model in groundwater well potential mapping. *J. Hydrol.* **2017**, *548*, 471–483. [CrossRef]
6. Kim, J.-C.; Jung, H.-S.; Lee, S. Groundwater productivity potential mapping using frequency ratio and evidential belief function and artificial neural network models: Focus on topographic factors. *J. Hydroinf.* **2018**, *20*, 1436–1451. [CrossRef]
7. Ministry of Land, Transport and Maritime Affairs (MLTM). *National Groundwater Monitoring Network in Korea Annual Report 2016*; MLTM: Seoul, Korea, 2016.
8. Kim, J.-C.; Yoon, J.-D.; Park, J.-S.; Choi, J.-Y.; Yoon, J.-H. Utilizing the revised universal soil loss equation (rusle) technique comparative analysis of soil erosion risk in the Geumhogang Riparian Area. *Korean J. Remote Sens.* **2018**, *34*, 179–190.
9. Kim, D.; Jung, H.-S.; Kim, J.-C. Comparison of snow cover fraction functions to estimate snow depth of South Korea from modis imagery. *Korean J. Remote Sens.* **2017**, *33*, 401–410.
10. Kim, J.-C.; Jung, H.-S. Application of landsat tm/etm+ images to snow variations detection by volcanic activities at southern volcanic zone, Chile. *Korean J. Remote Sens.* **2017**, *33*, 287–299.
11. Park, S.-H.; Jung, H.-S.; Choi, J.; Jeon, S. A quantitative method to evaluate the performance of topographic correction models used to improve land cover identification. *Adv. Space Res.* **2017**, *60*, 1488–1503. [CrossRef]
12. Baek, W.-K.; Jung, H.-S.; Jo, M.-J.; Lee, W.-J.; Zhang, L. Ground subsidence observation of solid waste landfill park using multi-temporal radar interferometry. *Int. J. Urban Sci.* **2018**, 1–16. [CrossRef]
13. Lee, S.; Lee, M.-J.; Jung, H.-S. Data mining approaches for landslide susceptibility mapping in Umyeonsan, Seoul, South Korea. *Appl. Sci.* **2017**, *7*, 683. [CrossRef]
14. Kim, J.-C.; Kim, D.-H.; Park, S.-H.; Jung, H.-S.; Shin, H.-S. Application of landsat images to snow cover changes by volcanic activities at Mt. Villarica and Mt. Lliama, Chile. *Korean J. Remote Sens.* **2014**, *30*, 341–350. [CrossRef]
15. Lee, Y.-S.; Park, S.-H.; Jung, H.-S.; Baek, W.-K. Classification of natural and artificial forests from kompsat-3/3a/5 images using artificial neural network. *Korean J. Remote Sens.* **2018**, *34*, 1399–1414.
16. Baek, W.-K.; Jung, H.-S.; Chae, S.-H.; Lee, W.-J. Two-dimensional velocity measurements of uvêrsbreen glacier in Svalbard using terrasar-x offset tracking approach. *Korean J. Remote Sens.* **2018**, *34*, 495–506.
17. Al-Abadi, A.M. Modeling of groundwater productivity in northeastern Wasit Governorate, Iraq using frequency ratio and Shannon's entropy models. *Appl. Water Sci.* **2015**, *7*, 699–716. [CrossRef]

18. Jothibasu, A.; Anbazhagan, S. Spatial mapping of groundwater potential in ponnaiyar river basin using probabilistic-based frequency ratio model. *Model. Earth Syst. Environ.* **2017**, *3*, 33. [CrossRef]
19. Naghibi, S.A.; Pourghasemi, H.R. A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. *Water Resour. Manag.* **2015**, *29*, 5217–5236. [CrossRef]
20. 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]
21. Rahmati, O.; Pourghasemi, H.R.; Melesse, A.M. Application of gis-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *Catena* **2016**, *137*, 360–372. [CrossRef]
22. Sameen, M.I.; Pradhan, B.; Lee, S. Self-learning random forests model for mapping groundwater yield in data-scarce areas. *Nat. Resour. Res.* **2018**, *28*, 1–19. [CrossRef]
23. Zabihi, M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Behzadfar, M. Gis-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environ. Earth Sci.* **2016**, *75*, 1–19. [CrossRef]
24. Park, S.; Hamm, S.-Y.; Jeon, H.-T.; Kim, J. Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using r and gis. *Sustainability* **2017**, *9*, 1157. [CrossRef]
25. Zandi, J.; Ghazvinei, P.T.; Hashim, R.; Yusof, K.B.W.; Ariffin, J.; Motamedi, S. Mapping of regional potential groundwater springs using logistic regression statistical method. *Water Resour.* **2016**, *43*, 48–57. [CrossRef]
26. Lee, S.; Lee, M.-J. Susceptibility mapping of Umyeonsan using logistic regression (lr) model and post-validation through field investigation. *Korean J. Remote Sens.* **2017**, *33*, 1047–1060.
27. Lee, S.; Kim, J.-C.; Jung, H.-S.; Lee, M.J.; Lee, S. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul Metropolitan City, Korea. *Geomat. Nat. Haz. Risk.* **2017**, *8*, 1185–1203. [CrossRef]
28. Mousavi, S.M.; Golkarian, A.; Naghibi, S.A.; Kalantar, B.; Pradhan, B. Gis-based groundwater spring potential mapping using data mining boosted regression tree and probabilistic frequency ratio models in iran. *AIMS Geosci.* **2017**, *3*, 91–115.
29. Lee, S.; Hong, S.-M.; Jung, H.-S. Gis-based groundwater potential mapping using artificial neural network and support vector machine models: The case of Boryeong City in Korea. *Geocarto. Int.* **2018**, *33*, 847–861. [CrossRef]
30. Naghibi, S.A.; Ahmadi, K.; Daneshi, A. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resour. Manag.* **2017**. [CrossRef]
31. Oh, H.-J.; Kadavi, P.R.; Lee, C.-W.; Lee, S. Evaluation of landslide susceptibility mapping by evidential belief function, logistic regression and support vector machine models. *Geomat. Nat. Haz. Risk.* **2018**, *9*, 1053–1070. [CrossRef]
32. Alizadeh, M.; Alizadeh, E.; Asadollahpour Kotenaee, S.; Shahabi, H.; Beiranvand Pour, A.; Panahi, M.; Bin Ahmad, B.; Saro, L. Social vulnerability assessment using artificial neural network (ann) model for earthquake hazard in Tabriz City, Iran. *Sustainability* **2018**, *10*, 3376. [CrossRef]
33. Kim, D.; Jung, H.-S. Mapping oil spills from dual-polarized sar images using an artificial neural network: Application to oil spill in the Kerch Strait in November 2007. *Sensors* **2018**, *18*, 2237. [CrossRef]
34. Lee, S.; Lee, S.; Song, W.; Lee, M.-J. Habitat potential mapping of marten (martes flavigula) and leopard cat (prionailurus bengalensis) in South Korea using artificial neural network machine learning. *Appl. Sci.* **2017**, *7*, 912. [CrossRef]
35. Lee, S.; Song, K.-Y.; Kim, Y.; Park, I. Regional groundwater productivity potential mapping using a geographic information system (gis) based artificial neural network model. *Hydrogeol. J.* **2012**, *20*, 1511–1527. [CrossRef]
36. Sokeng, V.J.; Kouamé, F.; Ngatcha, B.N.; N'da, H.D.; You, L.A.; Rirabe, D. Delineating groundwater potential zones in western cameroon highlands using gis based artificial neural networks model and remote sensing data. *Int. J. Innovation Appl. Stud.* **2016**, *15*, 747–759.

37. Ghorbani Nejad, S.; Falah, F.; Daneshfar, M.; Haghizadeh, A.; Rahmati, O. Delineation of groundwater potential zones using remote sensing and gis-based data-driven models. *Geocarto. Int.* **2017**, *32*, 167–187. [CrossRef]
38. Lee, S.; Kim, Y.-S.; Oh, H.-J. Application of a weights-of-evidence method and gis to regional groundwater productivity potential mapping. *J. Environ. Manag.* **2012**, *96*, 91–105. [CrossRef] [PubMed]
39. Tahmassebipoor, N.; Rahmati, O.; Noormohamadi, F.; Lee, S. Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. *Arab. J. Geosci.* **2016**, *9*, 79. [CrossRef]
40. Mogaji, K.A.; San Lim, H. Application of dempster-shafer theory of evidence model to geoelectric and hydraulic parameters for groundwater potential zonation. *NRIAG J. Astron. Geophys.* **2018**, *7*, 134–148. [CrossRef]
41. Pourghasemi, H.R.; Beheshtirad, M. Assessment of a data-driven evidential belief function model and gis for groundwater potential mapping in the Koohrang Watershed, Iran. *Geocarto. Int.* **2015**, *30*, 662–685. [CrossRef]
42. Zeinivand, H.; Ghorbani Nejad, S. Application of gis-based data-driven models for groundwater potential mapping in Kuhdasht Region of Iran. *Geocarto. Int.* **2018**, *33*, 651–666. [CrossRef]
43. Kim, J.-C.; Lee, S.; Jung, H.-S.; Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto. Int.* **2018**, *33*, 1000–1015. [CrossRef]
44. Nsiah, E.; Appiah-Adjei, E.K.; Adjei, K.A. Hydrogeological delineation of groundwater potential zones in the nabogo basin, ghana. *J. Afr. Earth Sci.* **2018**, *143*, 1–9. [CrossRef]
45. Lee, S.; Hyun, Y.; Lee, M.-J. Groundwater potential mapping using data mining models of big data analysis in Goyang-Si, South Korea. *Sustainability* **2019**, *11*, 1678. [CrossRef]
46. Jo, M.J.; Jung, H.S.; Won, J.S. Detecting the source location of recent summit inflation via three-dimensional insar observation of Kilauea volcano. *Remote Sens.* **2015**, *7*, 14386–14402. [CrossRef]
47. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]
48. Watts, A.C.; Ambrosia, V.G.; Hinkley, E.A. Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use. *Remote Sens.* **2012**, *4*, 1671–1692. [CrossRef]
49. Atzberger, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens.* **2013**, *5*, 949–981. [CrossRef]
50. Tam, V.T.; Nga, T.T.V. Assessment of urbanization impact on groundwater resources in hanoi, vietnam. *J. Environ. Manag.* **2018**, *227*, 107–116. [CrossRef]
51. Ministry for Food, Agriculture, Forestry and Fisheries (MFAFF). *Rural Groundwater Survey Report (Okcehon Gun)*; MFAFF: Seoul, Korea, 2010.
52. Bradbury, K.R.; Rothschild, E.R. A computerized technique for estimating the hydraulic conductivity of aquifers from specific capacity data. *Groundwater* **1985**, *23*, 240–246. [CrossRef]
53. Razack, M.; Huntley, D. Assessing transmissivity from specific capacity in a large and heterogeneous alluvial aquifer. *Groundwater* **1991**, *29*, 856–861. [CrossRef]
54. Oh, K.-Y.; Jung, H.-S.; Lee, M.-J. Accuracy evaluation of dem generated from satellite images using automated geo-positioning approach. *Korean J. Remote Sens.* **2017**, *33*, 69–77. [CrossRef]
55. Moore, I.; Burch, G. Modelling erosion and deposition: Topographic effects. *Transact. ASAE* **1986**, *29*, 1624–1630. [CrossRef]
56. Claps, P.; Fiorentino, M.; Oliveto, G. Informational entropy of fractal river networks. *J. Hydrol.* **1996**, *187*, 145–156. [CrossRef]
57. Riley, S.J. Index that quantifies topographic heterogeneity. *Intermountain J. Sci.* **1999**, *5*, 23–27.
58. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning*; Springer: Berlin, Germany, 2012; pp. 157–175.
59. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
60. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.
61. Grömping, U. Variable importance assessment in regression: Linear regression versus random forest. *Am. Statist.* **2009**, *63*, 308–319. [CrossRef]

62. Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an afromontane landscape. *Ecol. Indic.* **2015**, *52*, 394–403. [CrossRef]
63. Lee, S.; Talib, J.A. Probabilistic landslide susceptibility and factor effect analysis. *Environ. Geol.* **2005**, *47*, 982–990. [CrossRef]
64. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* **2016**, *13*, 839–856. [CrossRef]