






Article

Soft Computing Ensemble Models Based on Logistic Regression for Groundwater Potential Mapping

Phong Tung Nguyen ^{1,*}, Duong Hai Ha ², Mohammadtaghi Avand ³ , Abolfazl Jaafari ⁴,
Huu Duy Nguyen ⁵, Nadhir Al-Ansari ^{6,*} , Tran Van Phong ⁷ , Rohit Sharma ⁸,
Raghvendra Kumar ⁹, Hiep Van Le ¹⁰, Lanh Si Ho ^{11,*}, Indra Prakash ¹²  and
Binh Thai Pham ^{10,*} 

¹ Vietnam Academy for Water Resources, Hanoi 100000, Vietnam

² Institute for Water and Environment, Hanoi 100000, Vietnam; hahaiduongcwe@yahoo.com

³ Department of Watershed Management Engineering, College of Natural Resources, TarbiatModares University, Tehran, P.O. Box 14115-111, Iran; mt.avand70@gmail.com

⁴ Research Institute of Forests and Rangelands, Agricultural Research, Education, and Extension Organization (AREEO), P.O. Box 64414-356, Tehran, Iran; jaafari@rifr-ac.ir

⁵ Faculty of Geography, VNU University of Science, Vietnam National University, 334 Nguyen Trai, Hanoi 100000, Vietnam; huuduy151189@gmail.com

⁶ Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden

⁷ Institute of Geological Sciences, Vietnam Academy of Sciences and Technology, 84 Chua Lang Street, Dong da, Hanoi 100000, Vietnam; tphong1617@gmail.com

⁸ Department of Electronics & Communication Engineering, SRM Institute of Science and Technology, Ghaziabad 201204, India; rohitapece@gmail.com

⁹ Department of Computer Science and Engineering, GIET University, Gunupur 765022, India; raghvendraagrawal7@gmail.com

¹⁰ University of Transport Technology, Hanoi 100000, Vietnam; hieplv@utt.edu.vn

¹¹ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

¹² Department of Science & Technology, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar 382002, India; indra52prakash@gmail.com

* Correspondence: phongicd@gmail.com (P.T.N.); nadhir.alansari@ltu.se (N.A.-A.); hosilanh@duytan.edu.vn (L.S.H.); binhpt@utt.edu.vn (B.T.P.)

Received: 6 March 2020; Accepted: 1 April 2020; Published: 3 April 2020



Abstract: Groundwater potential maps are one of the most important tools for the management of groundwater storage resources. In this study, we proposed four ensemble soft computing models based on logistic regression (LR) combined with the dagging (DLR), bagging (BLR), random subspace (RSSLR), and cascade generalization (CGLR) ensemble techniques for groundwater potential mapping in Dak Lak Province, Vietnam. A suite of well yield data and twelve geo-environmental factors (aspect, elevation, slope, curvature, Sediment Transport Index, Topographic Wetness Index, flow direction, rainfall, river density, soil, land use, and geology) were used for generating the training and validation datasets required for the building and validation of the models. Based on the area under the receiver operating characteristic curve (AUC) and several other validation methods (negative predictive value, positive predictive value, root mean square error, accuracy, sensitivity, specificity, and Kappa), it was revealed that all four ensemble learning techniques were successful in enhancing the validation performance of the base LR model. The ensemble DLR model (AUC = 0.77) was the most successful model in identifying the groundwater potential zones in the study area, followed by the RSSLR (AUC = 0.744), BLR (AUC = 0.735), CGLR (AUC = 0.715), and single LR model (AUC = 0.71), respectively. The models developed in this study and the resulting potential maps can assist decision-makers in the development of effective adaptive groundwater management plans.

Keywords: machine learning; ensemble modeling; dagging; bagging; random subspace; cascade generalization

1. Introduction

Groundwater is one of the most valuable resources of water supply for agricultural, urban, and industrial activities in many parts of the world [1]. Groundwater contributes to the economic development and biodiversity of an area [2]. The effective management of groundwater is critical for balancing different demands on water. Groundwater management is becoming progressively challenging due to rapid population growth and occurrences of intermittent and prolonged drought periods [3–6] that have mounted increasing pressure on groundwater resources worldwide [7]. Therefore, further research on the exploration and estimation of groundwater potential and resources is necessary for proper water management of an area. Among different management strategies, potential groundwater mapping is an effective approach that can assist managers to adopt more efficient management plans [8–12].

Various methods have been suggested and used for producing groundwater potential maps. They range from the traditional surface and sub-surface potential assessment methods to the recent predictive models derived by machine learning methods. The surface and sub-surface methods include esoteric, geomorphologic, geological, soil and micro-biological, surface geophysical test drilling of boreholes, and geophysical logging techniques [13]. Although these methods are highly efficient for obtaining an accurate estimate of groundwater potential, the lack of advanced technologies and sufficient hydrogeological data preclude their use in many countries.

Machine learning methods and GIS-based data processing techniques have provided opportunities for developing powerful tools for the assessment of groundwater potential in a time- and cost-effective manner. Examples of these methods include a classification and regression tree (CART) [14], logistic model tree (LMT) [11], multivariate adaptive regression spline (MARS) [15], random forest (RF) [16], boosted regression tree (BRT) [17], maximum entropy (ME) [18], support vector machine (SVM) [19,20], artificial neural network (ANN) [20], multivariate adaptive regression [21], and maximum entropy [22].

Recent advancement in groundwater potential mapping is the combined use of different machine learning methods towards developing hybrid ensemble machine learning models for obtaining the most accurate results. Chen et al. [23] proposed the integrated application of J48 decision trees with the random subspace (RSS), rotation forest (RF), AdaBoost, bagging, and dagging to identify the groundwater potential zones in Wuqi County, China. Naghibi et al. [14] improved performances of the BRT, CART, and RF classifiers using a rotation forest ensemble technique for modeling groundwater potential in Meshgin Shahr, Iran. Miraki et al. [24] developed an ensemble model based on the RF classifier and the RSS ensemble technique to map the groundwater potential in Kurdistan, Iran. Avand et al. [25] integrated the best first decision tree with the AdaBoost, multiboosting, and bagging ensembles for groundwater potential mapping in the Yasuj-Dena area, Iran. Al-Fugara et al. [26] developed a hybrid model based on the SVM and genetic algorithm (GA) for groundwater potential mapping in Jerash and Ajloun, Jordan. In another study, Naghibi et al. [19] used GA for optimizing the structure of the SVM and RF models for groundwater potential mapping in Iran. Khosravi et al. [27] used several metaheuristic algorithms for optimizing a neuro-fuzzy model for groundwater potential mapping in Lorestan, Iran. Banadkooki et al. [28] used the whale optimization algorithm for optimizing the ANN for groundwater potential mapping. All these studies demonstrated the enhanced predictive performance of the hybrid ensemble models compared to the single models. In fact, the premise of the application of a hybrid ensemble is that groundwater potential mapping requires big data of various geo-environmental variables [11,12,23,24] that largely make the single modeling approaches inefficient in many regions [29].

In this study, we developed four ensemble models for groundwater potential mapping. Each model consisted of a logistic regression combined with an ensemble learning technique, namely dagging, bagging, cascade generalization (CG), and RSS. The development and application of these models were underpinned by real-world data from Dak Lak Province, Vietnam, where groundwater is the main source of water supply for domestic use and agriculture. The reliability and accuracy of these models were evaluated by comparing their area under the receiver operating characteristic (ROC) curves (AUC) and the index values derived from several other validation methods.

2. Methods Used

2.1. Logistic Regression (LR)

LR is the most widely used empirical model in different fields of science, in particular for environmental studies [30–33]. In LR, the probability of occurrence of a phenomenon is estimated within the range of 0 to 1, and it is not necessary to assume the normality of the predictor variables. Binomial LR analysis is used when the dependent variable is at the binomial nominal level, and it is used to predict the presence or absence of an attribute based on a set of independent variables. In linear double regression, one variable is used to predict another variable (such as temperature–altitude prediction), while in multiple LR, the relationship between several independent variables is measured with one dependent variable. LR is a special type of multiple regression in which the dependent variable is discrete. In fact, the LR model describes the relationship between a two-dimensional response variable (the presence or absence of a variable) and a set of response variables. The response variable may be continuous or discrete and does not require a frequent distribution.

2.2. Bagging

The name bagging is derived from bootstrap aggregating, which is one of the first, most intuitive, and simple ensemble-based algorithms with excellent performance [34–37]. The diversity of classifiers is received through bootstrap copies of the training dataset. Different subsets (with placement) are selected from the whole training set. Each subset is then utilized to train a different classifier. The individual classifiers are thereafter combined with the majority vote in their decisions. The class selected by the largest number of classifiers is the final ensemble decision. Since training datasets may overlap, additional measures may be used to increase diversity, such as using a subset of training data to train each classifier, or using relatively weak classifiers such as decision stumps. The batch method is only effective in nonlinear models of instability where small changes in training data cause large changes in their classification and accuracy. Reducing variance in the results of unstable learners reduces the final error [38].

2.3. Dagging

Witten and Ting proposed dagging for the first time in 1977. Dagging uses some disjoint samples which replace bootstrap samples to derive basic variables. Dagging does not use the bootstrap samples for the extraction of basic classification [39]. Dagging uses certain samples for the basic classification extraction. For grouping the classifications, majority voting is used in the dagging technique.

2.4. Cascade Generalization (CG)

One of the effective algorithms for classifier fusion is the CG algorithm, which works on a meta-level [40]. In this algorithm, predictions of basic level classifiers are used to increase the dimension of the input space. To do this, the output of each base-level classifier is added as a new attribute for each of the training examples. Thus, base-level and meta classifiers use the main input features for training, while meta-level classifiers can also use additional features (predictors of base-level classifiers).

2.5. Random Subspace (RSS)

RSS was proposed in 1988 for improving the accuracy of the weak classifications and the individual classification performance. RSS is one of the most popular methods of random sampling [41,42], in which the original character varies randomly. RSS creates multiple subspaces with small dimensions and then uses a majority vote for grouping the characteristic series of each sub-classification formation [42,43]. RSS has been used in several fields such as economics [44] and medicine [45], but rarely in problems of determining groundwater potential.

2.6. Validation Methods

2.6.1. Receiver Operating Characteristic (ROC) Curve

In the ROC curve, the proportion of pixels correctly predicted by the presence or absence of groundwater potential on the horizontal axis (true positive or 1-specificity) and the proportion of pixels that are incorrectly predicted is plotted on the vertical axis (false positive or sensitivity) [46–51]. The area under this curve is called the AUC, and the model with the highest AUC has the highest relative performance [52–61]. The AUC values equal to 0.5 indicate random prediction for a model [62–66]. The AUC values range from 0 to 1. Greater values of the AUC indicate a higher prediction efficiency of a model [67–72].

2.6.2. Statistical Indices

Numerous statistical indices have been used in different studies to evaluate the performance of predictive models [46,73,74]. In this study, we used positive (TP), true negative (TN), false positive (FP), false negative (FN), positive predictive value (PPV), negative predictive value (NPV), sensitivity (SST), specificity (SPF), accuracy (ACC), root mean square error (RMSE), and Kappa for the comparison and validation of the models. The equation for each one of these indices is as follows [75–82]:

$$PPV = \frac{TP}{TP + FP} \quad (1)$$

$$NPV = \frac{TN}{TN + FN} \quad (2)$$

$$SST = \frac{TP}{TP + FN} \quad (3)$$

$$SPF = \frac{TN}{TN + FP} \quad (4)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$K = \frac{P_O - P_{est}}{1 - P_{est}} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (V_{predicted} - V_{actual})^2} \quad (7)$$

where n is the total of the variable, $V_{predicted}$ and V_{actual} are like the prediction and actual value of variable i -th, P_a is the relative observed agreement among raters, and P_{est} is the hypothetical probability of chance agreement.

2.7. OneR Feature Selection Method

To determine the groundwater potential using machine learning, it is necessary to select the appropriate predictive variables as the model inputs. In this study, we used the OneR feature technique

for ranking the predictive variables and selecting those factors that contribute the most to groundwater potential. This process can efficiently decrease noise and the over-fitting problem of the modeling process [83,84], leading to an increased quality of the results [46].

3. Study Area and Data Used

3.1. Study Area

Dak Lak Province ($12^{\circ}9'45''$ to $13^{\circ}25'06''$ north latitude; $107^{\circ}28'57''$ to $108^{\circ}59'37''$ east longitude) is located in the central highlands of Vietnam (Figure 1). The province covers about 13,085 km², which accounts for 3.9% of the area of Vietnam. The topography of the study area is predominantly mountainous. The elevation in the area varies from 400 m to 2442 m (Chu Yang Sin Peak). A flat stretch of highland is in the middle part of the province, covering about 50% of the area. The climate of the province varies from hot under an elevation (El) of 400 m, humid at 400 m (El), and cold over 800 m (El). There are two main seasons: dry (November to April) and rainy (May to October) with 90% annual rainfall.

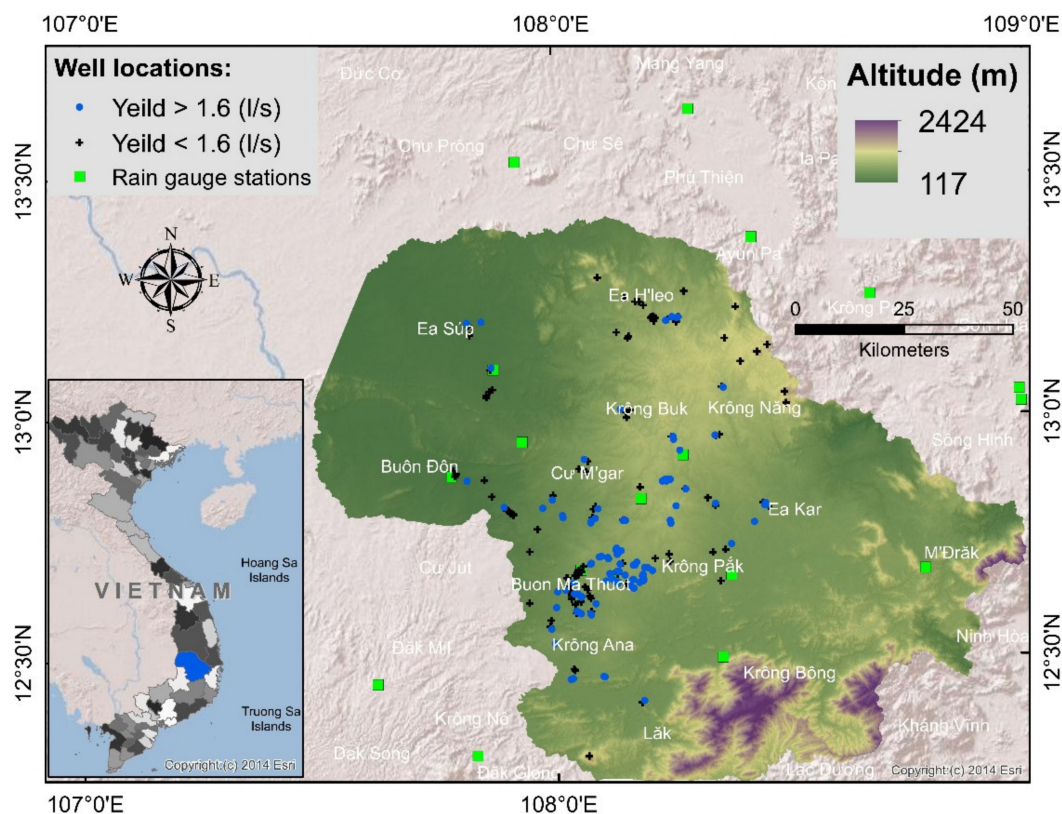


Figure 1. Location map of the study area (Dak Lak Province).

One-third of Dak Lak Province is covered by basalt rock and associated soil, which are favorable for developing rubber, coffee, and pepper plantations. About 660 million m³ of water is needed for 264,000 ha of coffee area, while availability of the surface water is only 250 million m³. During the dry months, the estimated amount of groundwater exploited for irrigation is about 500,000 m³/day. Due to climate change and El Nino-effect changes of the rainfall pattern in the central highland region, more dependency on the groundwater resources for the socio-economic development of the province is required. Therefore, the balance requirement needs to be met by groundwater through proper groundwater management. Thus, groundwater potential mapping is required, using advanced technology for the groundwater assessment and identifying an area suitable for the recharge.

3.2. Data Used

3.2.1. Well Yields

Groundwater potential assessment is generally done by the evaluation of well yields data [14,19,85,86]. In this study, the well yield data of 227 drilled wells was obtained from the national project of the Vietnam Academy for Water Resources (VAWR) and used for model development. The well data included: (1) water level at a non-operated and operated situation, (2) structure of the wells, (3) measurement of temperature and TDS (total dissolved solids), (4) yield, and (5) location. This dataset was randomly divided into separate sets such that 70% of the data was used for model building and the remaining data (30%) was used for model validation. The yield value of 1.6 l/s was selected as the threshold value for distinguishing between potential and non-potential areas of Dak Lak Province in terms of the groundwater potential.

3.2.2. Groundwater Influencing Factors

Selection of the groundwater influencing factors depends on the topography, geo-environment, meteorology, and anthropogenic activities of the area being investigated. In the present study, twelve influencing factors, namely soil type, geology, elevation, slope, Topographic Wetness Index (TWI), aspect, curvature, Sediment Transport Index (STI), river density, flow direction, land use, and rainfall were selected for groundwater potential mapping. A 30-m resolution Digital Elevation Model (DEM) collected from USGS (<https://earthexplorer.usgs.gov/>) was used to extract the maps of the slope, aspect, curvature, elevation, STI, TWI, flow direction, and river density factors.

The land use map was obtained from Dak Lak's Department of Natural Resources and Environment (DARD), Vietnam, at the scale of 1:50000. Geology and rainfall maps were extracted from the hydrogeological map (1:300,000 scale) collected from the central region of the Vietnam Division for Water Resources Planning and Investigation, Vietnam.

An aspect map, which indicates relationship with the ability to accumulate and retain water on the surface, was prepared and classified into nine classes (Figure 2a). Runoff will be greater on a convex surface and accumulation of water on a concave surface. The curvature in the study area ranged from -23.5 to 30.8 (Figure 2b). Runoff also depends on the meteorological conditions and elevation of the area. In the study area, the elevation ranged from 117 to 2424 m (Figure 2c). Flow direction is another important factor as it affects runoff and infiltration, and ranged in this area from 1 to 255 (Figure 2d). Slope has a direct relationship with the hydrological process, where surface water will accumulate, and thus more infiltration leads to groundwater recharge (Figure 2e).

A land use map is also a very important influencing factor for the assessment of groundwater potential that reflects the local conditions and anthropogenic activities in the area. The land use map was classified into different classes (Figure 2f). Concrete construction in an area increases runoff, and thus reduces recharge capacity of the ground. Similarly, deforestation increases runoff, and thus there is less infiltration in the soil.

TWI is important in assessing groundwater potential because this factor measures topography–moisture relationships (Figure 2h). River density is another important factor as it has an inverse relationship with infiltration and recharge. River density (7.565 km/km^2) in the study area was relatively high, which indicates more runoff and thus less probability of recharge (Figure 2i).

Rainfall is one of the most important factors for groundwater potential mapping as it directly affects groundwater recharge [10,12,22]. The yearly average rainfall of this area varies from 4.80 to 7.23 mm (Figure 2j). Soil and geology characteristics are very important for infiltration, recharge, and formation of aquifers [14,86]. The geology map of the study area was prepared into 31 classes (Figure 2k). Permeable soil helps in more infiltration [19]. In the study area, the soil map was prepared based on the soil types available in the study area (Figure 2l).

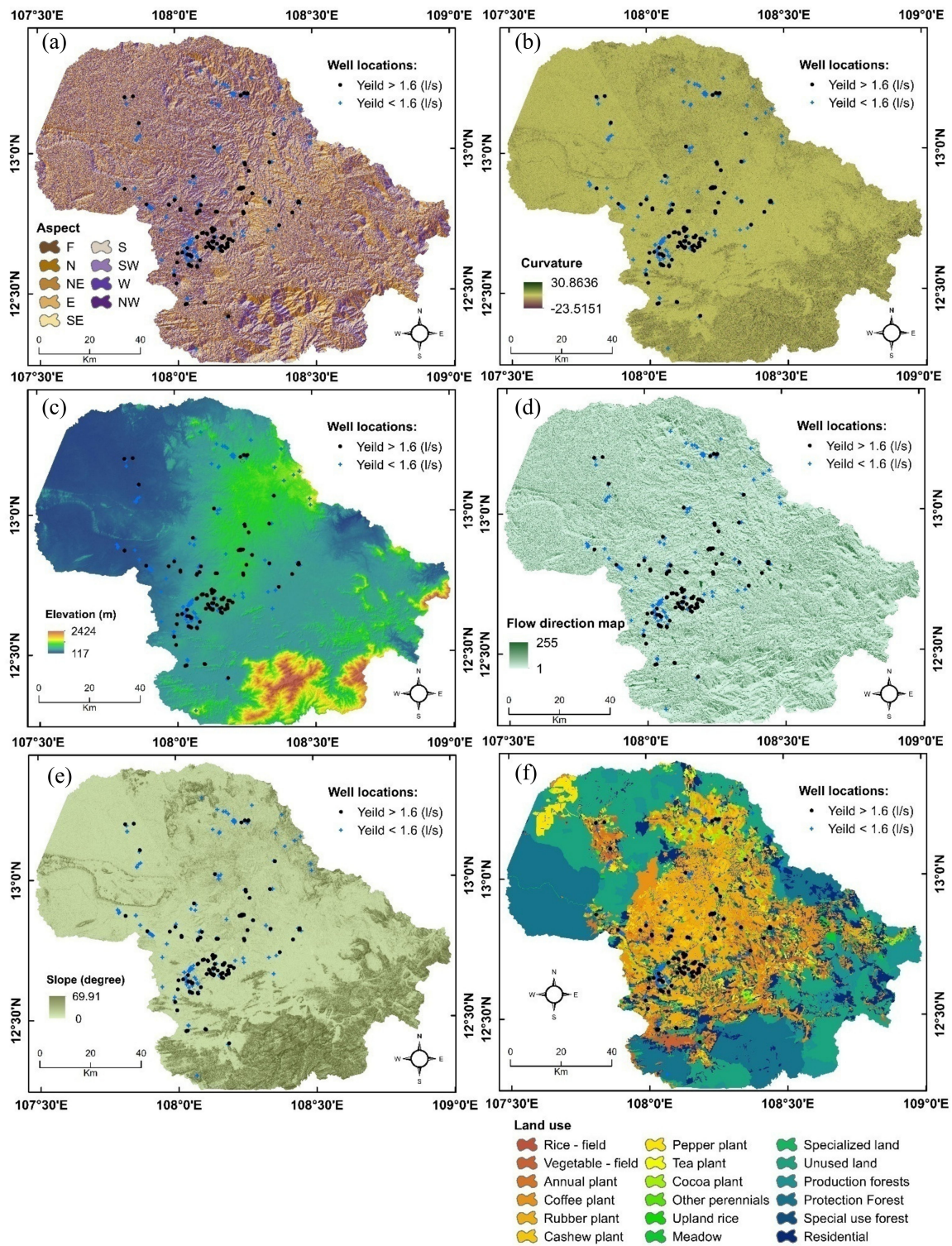


Figure 2. Cont.

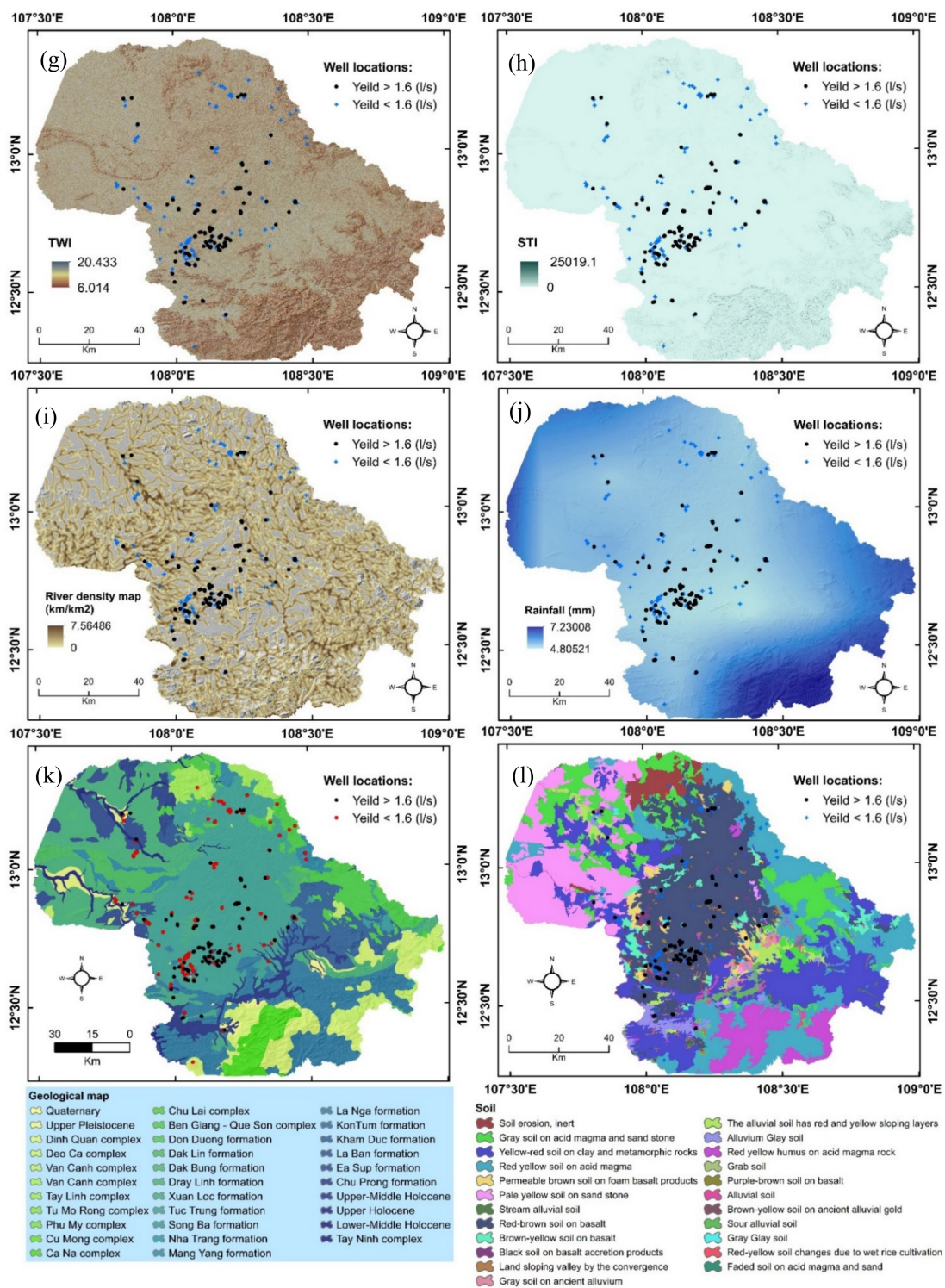


Figure 2. Maps of the influencing factors used in this study: (a) aspect, (b) curvature, (c) elevation, (d) flow direction, (e) slope, (f) land use, (g) Topographic Wetness Index (TWI), (h) Sediment Transport Index (STI), (i) river density, (j) rainfall, (k) geology, and (l) soil.

4. Modeling Methodology

The methodology of the development and validation of the groundwater potential models and their outcomes (potential maps) are shown in Figure 3, which involves the following steps: (1) Collection and preparation of data. The data was collected from various government sources, meteorological

departments, and satellite images. (2) Factor selection: The OneR feature selection method was used to validate and select the important factors for modeling use. (3) Construction of models: four ensemble models, namely DLR, BLR, CGLR, and RSSLR, were developed using the training dataset. More specifically, DLR is a combination of the LR model and the dagging ensemble technique, of which dagging was used to optimize the training datasets for the construction of the ensemble DLR model. BLR is a combination of the LR model and the bagging ensemble technique, of which bagging was used to optimize the training datasets for the construction of the ensemble BLR model. RSSLR is a combination of the LR model and the RSS ensemble technique, of which RSS was used to optimize the training datasets for the construction of the ensemble RSSLR model. CGLR is a combination of the LR model and the CG ensemble technique, of which CG was used to optimize the training datasets for the construction of the ensemble RSSLR model. (4) Validation of models: various statistical measures were used for the validation of groundwater potential models. These measures were AUC, NPV, SST, RMSE, ACC, SPF, PPV, and Kappa. Thereafter, the generation of groundwater potential maps was carried out using GIS software.

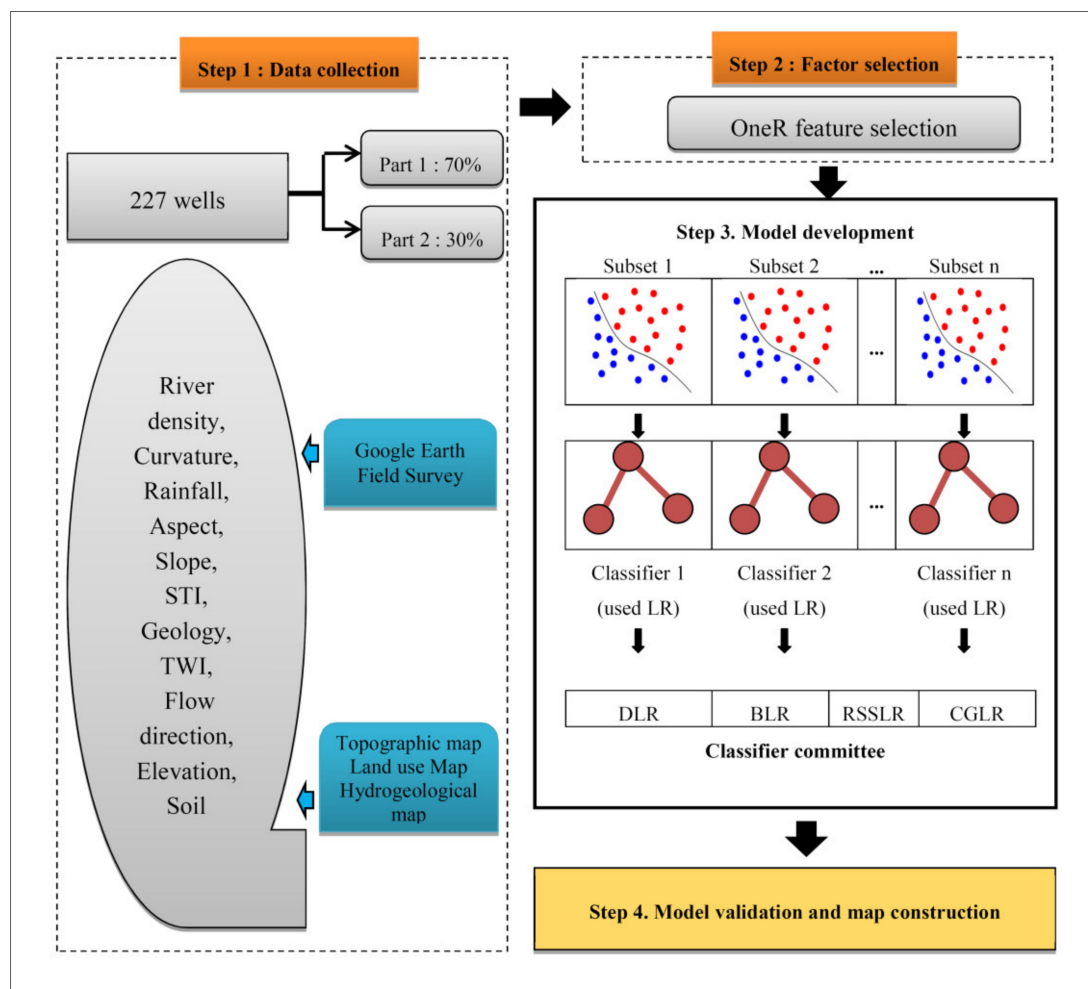


Figure 3. The flow chart of the methodology used in this study.

5. Results and Analysis

5.1. Factor Importance

The results of factor analysis using the OneR method measured the average merit (AM) for each influencing factor, which revealed that rainfall, land use, and elevation were the most important factors

for the assessment of groundwater potential in Dak Lak Province (Table 1). Apart from these three factors, the other factors that showed $AM \neq 0$ also had the efficiency of being used in the modeling process. Therefore, they were all selected for modeling of groundwater potential mapping in this study.

Table 1. Factor ranks extracted using the OneR feature selection method.

Rank	Factor	Average Merit (AM)
1	Rainfall	76.111
2	Land use	72.222
3	Elevation	68.333
4	STI	61.111
5	Soil	60
6	Slope	59.444
7	TWI	56.111
8	Curvature	53.889
9	Flow direction	53.889
10	River density	52.222
11	Geology	49.444
12	Aspect	44.444

5.2. Model Validation and Comparison

Validation of the models was done using various statistical indices (Table 2 and Figures 4 and 5). In terms of the training performance, the RSSLR model, which achieved the greatest values of NPV (82.28%), SST (84.95%), ACC (80%), and TN (65%) indices, ranked as the best model; followed by the BLR and CGLR models with NPV (79.75%), SST (83.33%), ACC (79.44%), ACC (79.44%), and TN (63%); followed by the LR model with NPV (75.95%), SST (80.381%), ACC (77.78%), and TN (60%); and the DLR model with NPV (69.62%), SST (76.47%), ACC (73.89%), and TN (55%), respectively. In terms of the TP and PPV indices, the BLR, CGLR, and LR models were better with 80% and 79.21%; followed by the RSSLR model with 79% and 78.22%; and the RLR model with 78% and 77.23%. For the FP index, the DLR model was better with 23%; followed by the RSSLR model with 22%; and the BLR, CGLR, and RL models with 21%. The DLR model was better in terms of the FN index (24%), followed by the LR model with 19%, the BLR and CGLR models with 16%, and the RSSLR model with 14%, respectively. For the SPF index, the BLR and CGLR models were better with 75%, followed by the RSSLR, LR, and DLR models with 74.71%, 74.07%, and 70.51%, respectively. In terms of the Kappa index, the RSSLR model performance with 0.5984 was the best, followed by the BLR, CGLR, LR, and DLR models with 0.586, 0.5855, 0.5501, and 0.469, respectively (Table 2).

In terms of the validation performance, RSSLR was better with TP = 32%, followed by DLR, BLR, and CGLR and LR with 31%, 30%, and 29%, respectively. For the TN index, DLR had the best performance with 25%, followed by the BLR (23%), CGLR and LR (22%), and RSSLR (21%) models. In terms of the FP index, the CGLR and LR models were better with 14%, followed by the BLR, DLR, and RSSLR models with 13%, 12%, and 11%, respectively. For the FN index, the RSSLR model was more precise with 12%, followed by the CGLR and LR (11%), BLR (10%), and DLR (8%) models. In terms of the PPV index, RSSLR was better with 74.42%, followed by the DLR, BLR, and CGLR and LR models with 72.09%, 69.77%, and 67.44% respectively. For the NPV index, DLR was better with 75.76%, followed by BLR (69.70%), CGLR (66.67%), LR (66.67%), and RSSLR (63.64%). For the SST index, DLR was better with 79.49%, followed by the BLR (75%), RSSLR (72.73%), and CGLR and LR (72.50%) models. In terms of the SPF index, DLR was better with 67.57%, followed by RSSLR with 65.63%, BLR with 63.89%, and CGLR and LR with 61.11%. In terms of the ACC and Kappa indices, DLR was more effective with 73.68% and 0.472, followed by BLR with 69.74% and 0.391, LR with 67.11% and 0.3375, and CGLR with 67.11% and 0.338 (Table 2).

Table 2. Model performance in the training and validation phases.

No	Index	Phase									
		Training					Validation				
		BLR	CGLR	DLR	RSSLR	LR	BLR	CGLR	DLR	RSSLR	LR
1	TP	80	80	78	79	80	30	29	31	32	29
2	TN	63	63	55	65	60	23	22	25	21	22
3	FP	21	21	23	22	21	13	14	12	11	14
4	FN	16	16	24	14	19	10	11	8	12	11
5	PPV (%)	79.21	79.21	77.23	78.22	79.21	69.77	67.44	72.09	74.42	67.44
6	NPV (%)	79.75	79.75	69.62	82.28	75.95	69.70	66.67	75.76	63.64	66.67
7	SST (%)	83.33	83.33	76.47	84.95	80.81	75.00	72.50	79.49	72.73	72.50
8	SPF (%)	75.00	75.00	70.51	74.71	74.07	63.89	61.11	67.57	65.63	61.11
9	ACC (%)	79.44	79.44	73.89	80.00	77.78	69.74	67.11	73.68	69.74	67.11
10	Kappa	0.586	0.5855	0.469	0.5984	0.5501	0.391	0.338	0.472	0.3819	0.3375

RMSE was also used to assess the accuracy of the models. Figure 4 shows the RMSE value; the BLR model was better for training data with RMSE = 0.3681, then LR (RMSE = 0.3687), CGLR (RMSE = 0.3692), RSSLR (RMSE = 0.3962), and DLR (RMSE = 0.3954), respectively. For the validation data, DLR was better with RMSE = 0.4442, followed by RSSLR (RMSE = 0.4479), BLR (RMSE = 0.4772), CGLR (RMSE = 0.4895), and LR (RMSE = 0.49), respectively.

In addition, the ROC method was used to compare the performance of the developed models (Figure 5). Using this method, the performances of the models were presented in a graph with 100-specificity on the x-axis and sensitivity on the y-axis that showed that the BLR model with AUC = 0.888 had a better training performance, followed by the CGLR (AUC = 0.885), LR (AUC = 0.884), RSSLR (AUC = 0.869), and DLR (AUC = 0.856) models, respectively. For the validation performance, the DLR model was better with AUC = 0.769, followed by the RSSLR (AUC = 0.744), BLR (AUC = 0.735), CGLR (AUC = 0.715), and LR (AUC = 0.710) models, respectively.

5.3. Groundwater Potential Maps

Groundwater potential maps were constructed using the training results of the models. The output values varying from 0 to 1 were classified into 5 classes (very low, low, moderate, high, very high) (Figure 6) using the natural breaks method, as it is the most suitable and popular technique for classification and construction of groundwater susceptibility maps [84]. Comparison of groundwater potential maps indicates that 51.56% of the study area is located in a very low groundwater potential zone and 11.62% in a very high zone using the BLR model. For the CGLR model, 55.53% of the study area is in a region with a very low groundwater potential and 11.95% is in a region with very high potential. In the case of the DLR model, 34.49% of the study area is located in a very low potential zone and 11.14% in a very high potential zone. For the RSSLR model, 26.02% of the study area lies in a very low potential zone and 11.11% in a very high zone. In the case of the LR model, 55.06% of the area is in a very low zone and 11.78% is in a very high zone. This suggests that 50% to 60% of the area of Dak Lak Province falls in the very low groundwater potential zone (Figure 7).

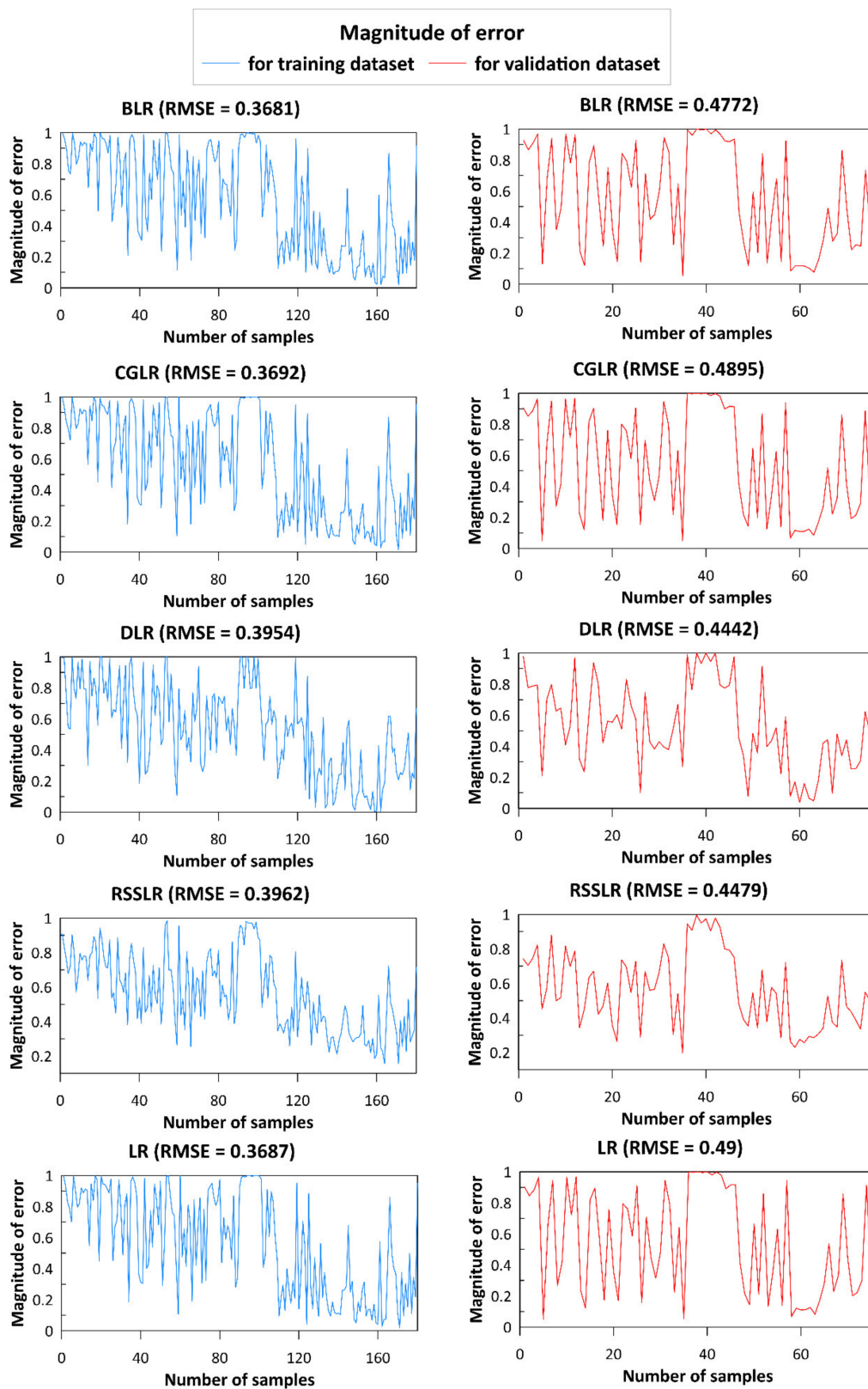


Figure 4. Magnitude of modeling error in the training and validation datasets.

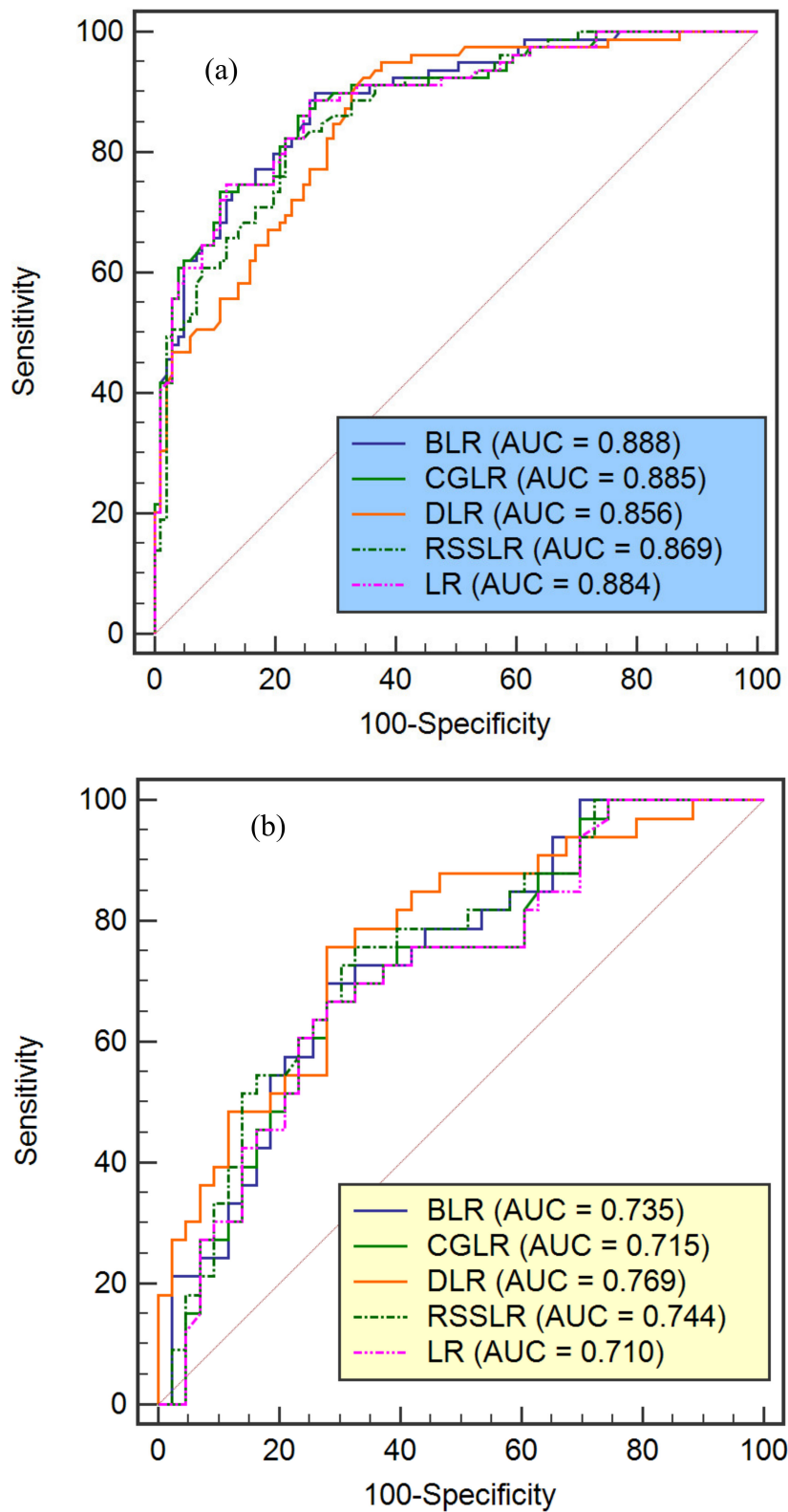


Figure 5. ROC curve of the models for the (a) training and (b) validation datasets.

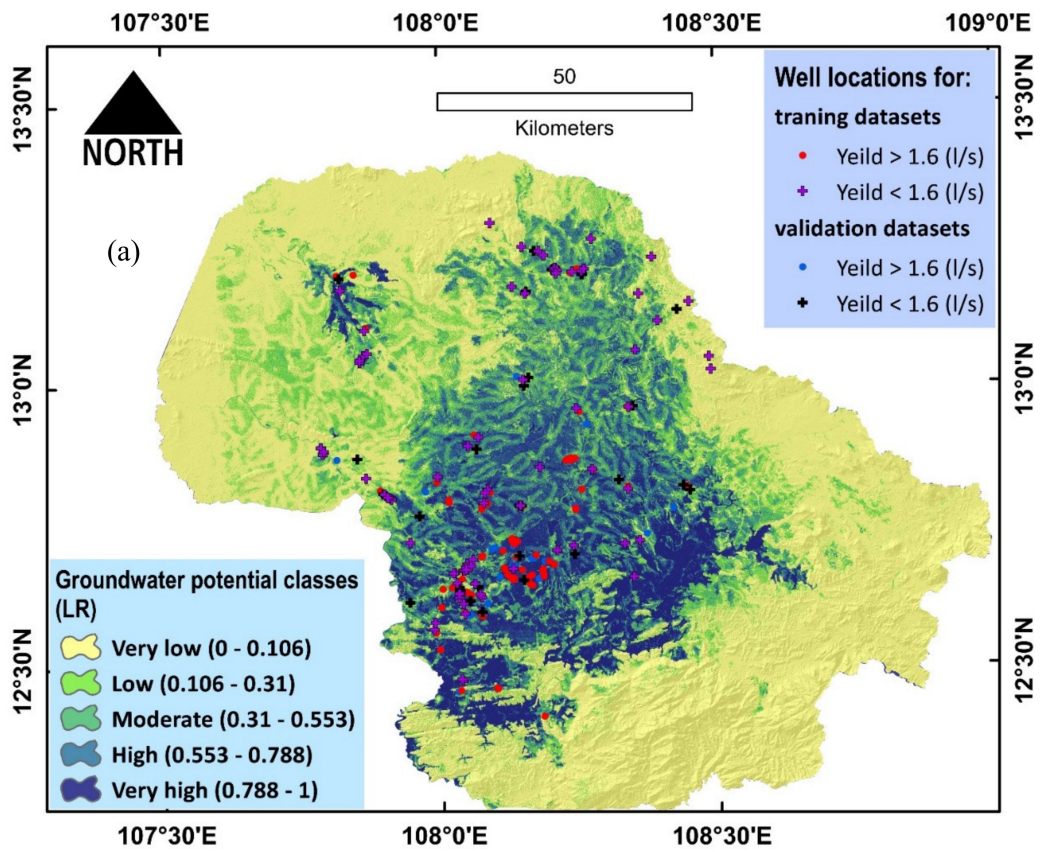


Figure 6. Cont.

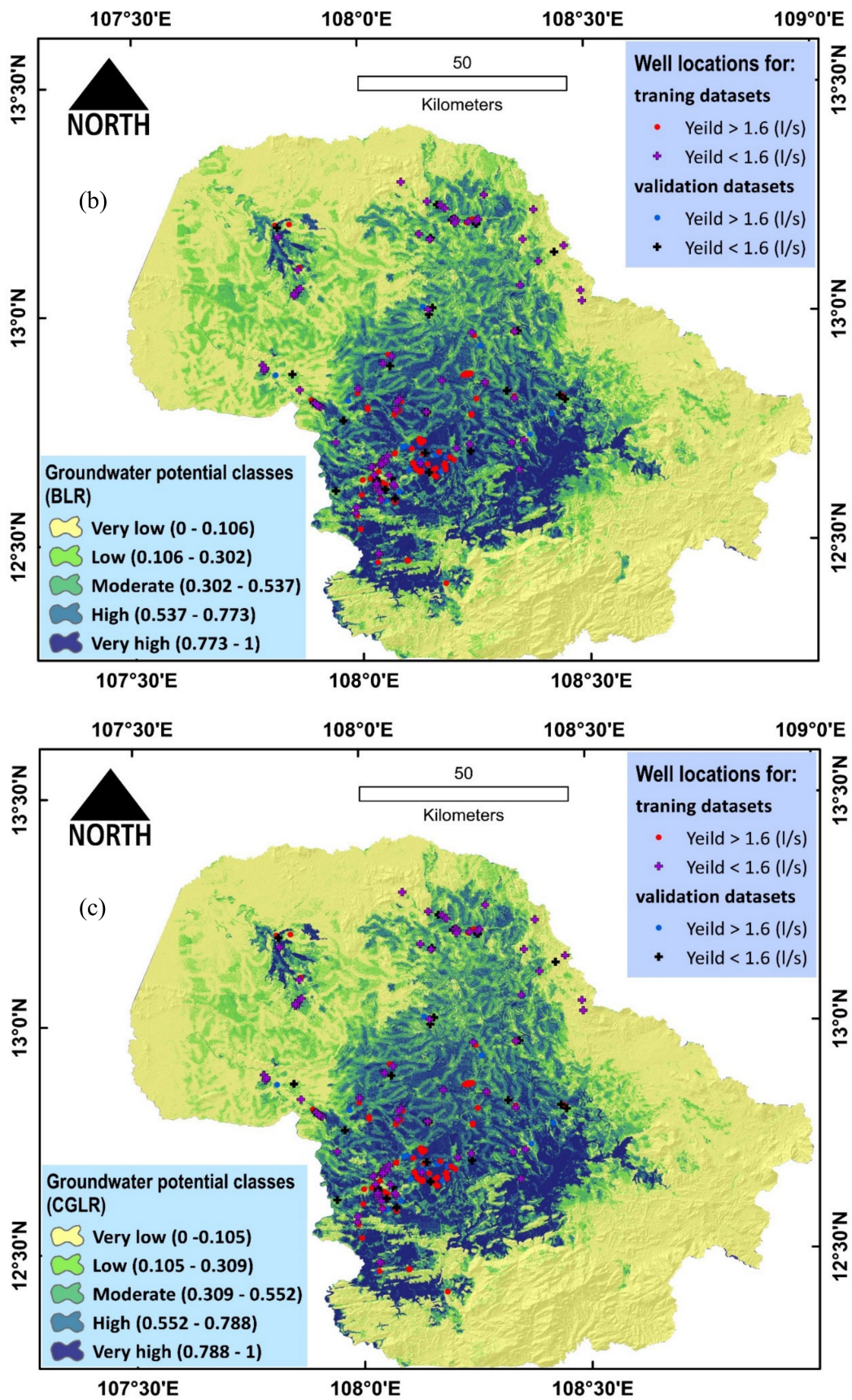


Figure 6. Groundwater potential maps: (a) LR, (b) BLR, (c) CGLR, (d) DLR, and (e) RSSLR models.

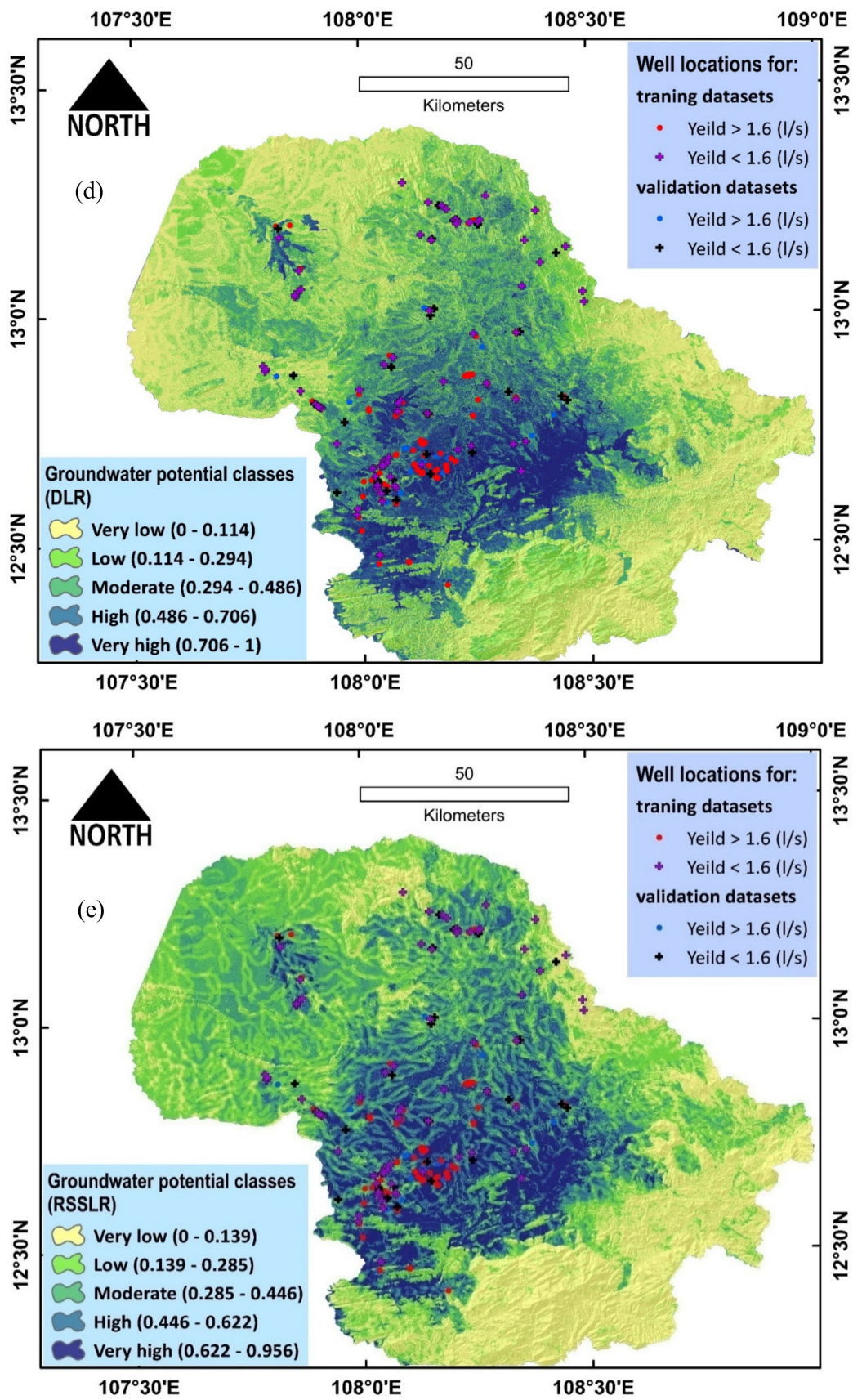


Figure 6. Groundwater potential maps: (a) LR, (b) BLR, (c) CGLR, (d) DLR, and (e) RSSLR models.

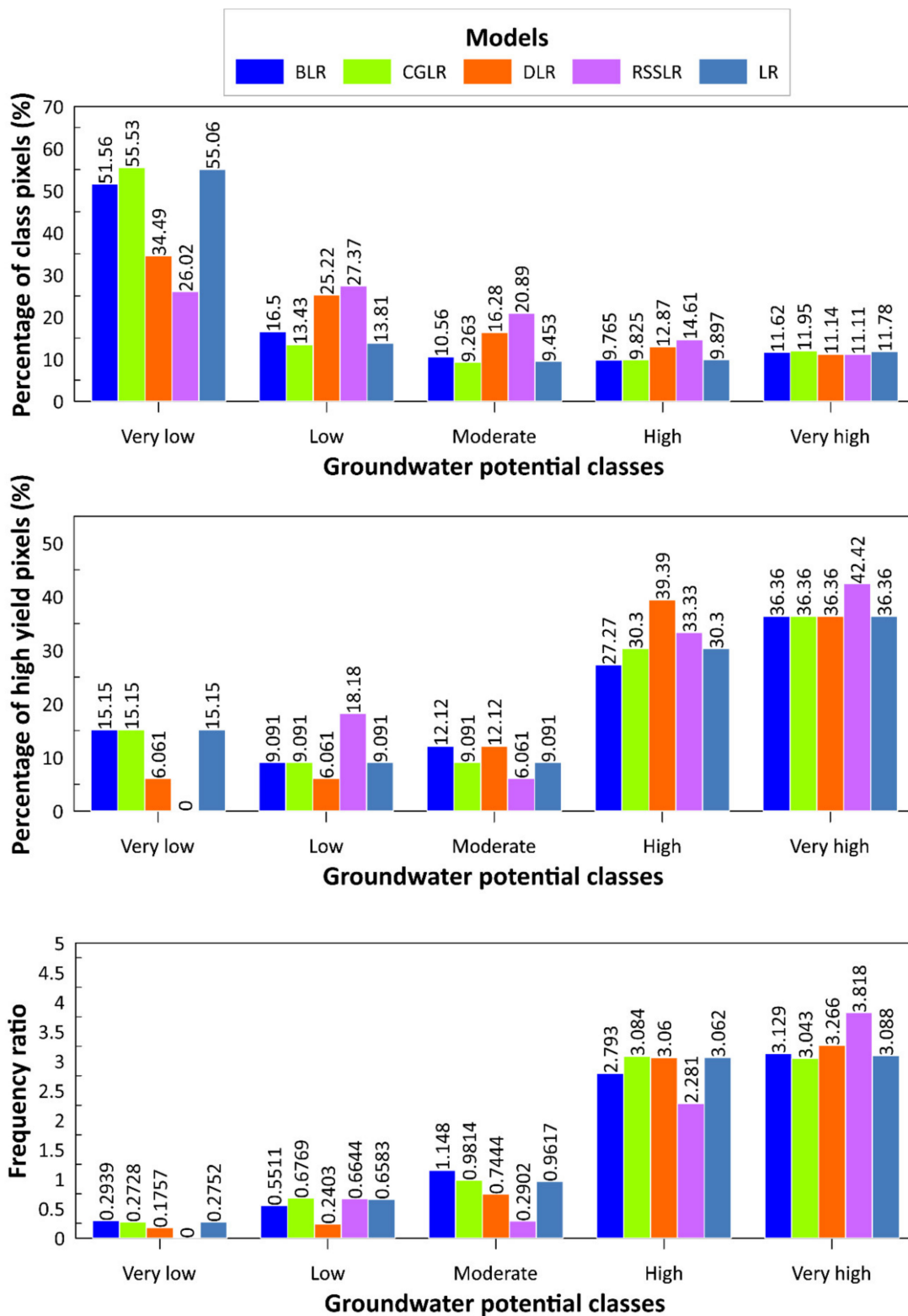


Figure 7. Analysis of the groundwater potential maps.

6. Discussion

Machine learning modeling of environmental problems has gained popularity because machine learning methods show promise when dealing with manifold geospatial data [69,87–89]. As such, machine learning modeling can effectively alleviate the difficulty associated with the identification

of groundwater potential zones over large-scale regions, which often suffer a lack of accurate and long-term geotechnical and hydrogeological data for the implementation of physically based and/or numerical models [11]. However, the utility of different machine learning methods should be broadly investigated via their applications in different regions with different geo-environmental settings to find the best model with the highest accuracy and lowest sensitivity to noisy input data [33,70,87,90,91].

The results of our study revealed that all four ensemble techniques used in this study proficiently optimized the performance of the base LR model and provided reliable estimations of groundwater potential based on different validation methods. More specifically, the results of the ROC–AUC method demonstrated that the single LR model and its four derived ensemble models with a mean AUC of 0.88 had an excellent goodness-of-fit with the training dataset. During the validation phase, performances of the models that indicate their capabilities to estimate groundwater potential [37] were decreased to a mean AUC of 0.73. Based on the relationship between AUC values and the predictive capability of the models that has been suggested in the literature [14,19,22,24,84], we can conclude that our models performed decently in estimating groundwater potential and developing distribution maps. Further, the results demonstrated the capability of ensemble learning techniques for improving LR performance. Among the four ensemble techniques used in this study, dagging was identified as the most effective technique, which was followed by RSS, bagging, and CG techniques, suggesting that dagging was the most capable model in reducing the variance, bias, and noise of the groundwater potential modeling.

Although the literature has mostly reported on the effectiveness of the ensemble techniques, these techniques showed different performances for different problems in different areas. For example, Nhu et al. [36] showed that reduced error pruning tree (REPT) performed better in combination with RSS than the bagging and AdaBoost techniques for gully erosion prediction, whereas Pham et al. [92] reported that the REPT model performed better with rotation forest and bagging than its combination with the RSS and multiboost for landslide prediction. Different results have also been reported for flood prediction based on the ensemble models [93,94]. From these studies, we can conclude that the machine learning and ensemble learning techniques are greatly case- and site-specific, and that their performances depend heavily on the local conditions that the training datasets are developed upon, indicating that the application of different methods in different regions should be continued to find the optimum method for each environmental setting [95].

7. Concluding Remarks

In Vietnam, there is a problem of water scarcity in Dak Lak Province due to enhanced requirements for agricultural development and the occurrence of frequent drought conditions in recent years. Erratic and decreased rainfall in the area has led to overexploitation of groundwater reservoirs to meet the supply of water for day to day requirements of the province for drinking, cultivation, and industrial uses. Therefore, there is a great need for the assessment of groundwater potential and identification of the suitable areas of recharge. To address this need, four ensemble models, namely BLR, DLR, CGLR, and RSSLR, were developed to produce groundwater potential maps for the province. All four models performed well ($AUC > 0.7$) for the assessment of groundwater potential and generation of potential zone maps. Among the studied models, the DLR model with $AUC = 0.769$ and $RMSE = 0.444$ was the best model in comparison to the other ensemble models and the single LR model. Therefore, a groundwater potential map generated using the DLR model can be used by decision-makers in the development of effective adaptive groundwater management plans for Dak Lak Province. The models proposed in the present study can be applied in other areas for better groundwater potential mapping considering local geo-environmental factors.

Author Contributions: Conceptualization, P.T.N., D.H.H., N.A.-A., and B.T.P.; methodology, B.T.P., P.T.N., T.V.P., A.J., and N.A.-A.; validation, B.T.P., P.T.N., N.A.-A., M.A., L.S.H., and I.P.; formal analysis, P.T.N., T.V.P., and D.H.H.; data curation, D.H.H., T.V.P., H.D.N., and H.V.L.; writing—original draft preparation, all authors; writing—review and editing, P.T.N., L.S.H., B.T.P., A.J., N.A.-A., and I.P.; visualization, T.V.P., L.S.H., R.K., R.S., H.V.L., and D.H.H.; supervision, P.T.N., B.T.P., N.A.-A., and I.P.; project administration, P.T.N., B.T.P., and N.A.-A.; funding acquisition, N.A.-A. and B.T.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.08-2019.03.

Acknowledgments: We thank the Vietnam Academy for Water Resources (VAWR) for providing the data for carrying out this research.

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

References

1. Khosravi, K.; Sartaj, M.; Tsai, F.T.-C.; Singh, V.P.; Kazakis, N.; Melesse, A.M.; Prakash, I.; Bui, D.T.; Pham, B.T. A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment. *Sci. Total Environ.* **2018**, *642*, 1032–1049. [[CrossRef](#)] [[PubMed](#)]
2. Nobre, R.; Rotunno Filho, O.; Mansur, W.; Nobre, M.; Cosenza, C. Groundwater vulnerability and risk mapping using GIS, modeling and a fuzzy logic tool. *J. Contam. Hydrol.* **2007**, *94*, 277–292. [[CrossRef](#)] [[PubMed](#)]
3. Mafi-Gholami, D.; Zenner, E.K.; Jaafari, A. Mangrove regional feedback to sea level rise and drought intensity at the end of the 21st century. *Ecol. Indic.* **2020**, *110*, 105972. [[CrossRef](#)]
4. Mafi-Gholami, D.; Zenner, E.K.; Jaafari, A.; Ward, R.D. Modeling multi-decadal mangrove leaf area index in response to drought along the semi-arid southern coasts of Iran. *Sci. Total Environ.* **2019**, *656*, 1326–1336. [[CrossRef](#)]
5. Mafi-Gholami, D.; Zenner, E.K.; Jaafari, A.; Bakhtiari, H.R.; Tien Bui, D. Multi-hazards vulnerability assessment of southern coasts of Iran. *J. Environ. Manag.* **2019**, *252*, 109628. [[CrossRef](#)]
6. Mafi-Gholami, D.; Zenner, E.K.; Jaafari, A.; Bui, D.T. Spatially explicit predictions of changes in the extent of mangroves of Iran at the end of the 21st century. *Estuar. Coast. Shelf Sci.* **2020**, *237*, 106644. [[CrossRef](#)]
7. Choubin, B.; Rahmati, O.; Soleimani, F.; Alilou, H.; Moradi, E.; Alamdari, N. Regional groundwater potential analysis using classification and regression trees. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 485–498.
8. Rahmati, O.; Melesse, A.M. Application of Dempster–Shafer theory, spatial analysis and remote sensing for groundwater potentiality and nitrate pollution analysis in the semi-arid region of Khuzestan, Iran. *Sci. Total Environ.* **2016**, *568*, 1110–1123. [[CrossRef](#)]
9. Rahmati, O.; Moghaddam, D.D.; Moosavi, V.; Kalantari, Z.; Samadi, M.; Lee, S.; Bui, D.T. An automated Python language-based tool for creating absence samples in groundwater potential mapping. *Remote Sens.* **2019**, *11*. [[CrossRef](#)]
10. Moghaddam, D.D.; Rahmati, O.; Panahi, M.; Tiefenbacher, J.; Darabi, H.; Haghizadeh, A.; Haghghi, A.T.; Nalivan, O.A.; Tien Bui, D. The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. *Catena* **2020**, *187*, 104421. [[CrossRef](#)]
11. Rahmati, O.; Naghibi, S.A.; Shahabi, H.; Bui, D.T.; Pradhan, B.; Azareh, A.; Rafiei-Sardooi, E.; Samani, A.N.; Melesse, A.M. Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. *J. Hydrol.* **2018**, *565*, 248–261. [[CrossRef](#)]
12. Rahmati, O.; Choubin, B.; Fathabadi, A.; Coulon, F.; Soltani, E.; Shahabi, H.; Mollafar, E.; Tiefenbacher, J.; Cipullo, S.; Ahmad, B.B.; et al. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. *Sci. Total Environ.* **2019**, *688*, 855–866. [[CrossRef](#)] [[PubMed](#)]
13. Berhanu, K.G.; Hatiye, S.D. Identification of Groundwater Potential Zones Using Proxy Data: Case study of Megech Watershed, Ethiopia. *J. Hydrol. Reg. Stud.* **2020**, *28*, 100676. [[CrossRef](#)]
14. Naghibi, S.A.; Dolatkordestani, M.; Rezaei, A.; Amouzegari, P.; Heravi, M.T.; Kalantar, B.; Pradhan, B. Application of rotation forest with decision trees as base classifier and a novel ensemble model in spatial modeling of groundwater potential. *Environ. Monit. Assess.* **2019**, *191*. [[CrossRef](#)] [[PubMed](#)]

15. 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*. [[CrossRef](#)]
16. Naghibi, S.A.; Pourghasemi, H.R.; Dixon, B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* **2016**, *188*, 44. [[CrossRef](#)]
17. 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.
18. 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](#)]
19. 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**, *31*, 2761–2775. [[CrossRef](#)]
20. 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](#)]
21. 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](#)]
22. 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*. [[CrossRef](#)]
23. Chen, W.; Zhao, X.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Xue, W.; Wang, X.; Ahmad, B.B. Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. *J. Hydrol.* **2020**, *583*, 124602. [[CrossRef](#)]
24. Miraki, S.; Zanganeh, S.H.; Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Pham, B.T. Mapping groundwater potential using a novel hybrid intelligence approach. *Water Resour. Manag.* **2019**, *33*, 281–302. [[CrossRef](#)]
25. Avand, M.; Janizadeh, S.; Tien Bui, D.; Pham, V.H.; Ngo, P.T.T.; Nhu, V.-H. A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *Int. J. Digit. Earth* **2020**, 1–22. [[CrossRef](#)]
26. Al-Fugara, A.K.; Ahmadlou, M.; Al-Shabeeb, A.R.; AlAyyash, S.; Al-Amoush, H.; Al-Adamat, R. Spatial mapping of groundwater springs potentiality using grid search-based and genetic algorithm-based support vector regression. *Geocarto Int.* **2020**, 1–20. [[CrossRef](#)]
27. Khosravi, K.; Panahi, M.; Tien Bui, D. Spatial prediction of groundwater spring potential mapping based on an adaptive neuro-fuzzy inference system and metaheuristic optimization. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 4771–4792. [[CrossRef](#)]
28. Banadkooki, F.B.; Ehteram, M.; Ahmed, A.N.; Teo, F.Y.; Fai, C.M.; Afan, H.A.; Sapitang, M.; El-Shafie, A. Enhancement of Groundwater-Level Prediction Using an Integrated Machine Learning Model Optimized by Whale Algorithm. *Nat. Resour. Res.* **2020**, 1–20. [[CrossRef](#)]
29. Arabameri, A.; Lee, S.; Tiefenbacher, J.P.; Ngo, P.T.T. Novel Ensemble of MCDM-Artificial Intelligence Techniques for Groundwater-Potential Mapping in Arid and Semi-Arid Regions (Iran). *Remote Sens.* **2020**, *12*, 490. [[CrossRef](#)]
30. Guns, M.; Vanacker, V.; Glade, T. Logistic regression applied to natural hazards: Rare event logistic regression with replications. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 1937–1947. [[CrossRef](#)]
31. Bayat, M.; Ghorbanpour, M.; Zare, R.; Jaafari, A.; Thai Pham, B. Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. *Comput. Electron. Agric.* **2019**, 164. [[CrossRef](#)]
32. Hong, H.; Jaafari, A.; Zenner, E.K. Predicting spatial patterns of wildfire susceptibility in the Huichang County, China: An integrated model to analysis of landscape indicators. *Ecol. Indic.* **2019**, *101*, 878–891. [[CrossRef](#)]
33. Jaafari, A.; Mafi-Gholami, D.; Pham, B.T.; Tien Bui, D. Wildfire probability mapping: Bivariate vs. multivariate statistics. *Remote Sens.* **2019**, *11*, 618. [[CrossRef](#)]
34. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]

35. Quinlan, J.R. Bagging, Boosting, and C4.5. 1996. Available online: <http://www.cs.ecu.edu/~jdingq/CSCI6905/readings/BaggingBoosting.pdf> (accessed on 3 April 2020).
36. Nhu, V.-H.; Janizadeh, S.; Avand, M.; Chen, W.; Farzin, M.; Omidvar, E.; Shirzadi, A.; Shahabi, H.; Clague, J.J.; Jaafari, A.; et al. GIS-Based Gully Erosion Susceptibility Mapping: A Comparison of Computational Ensemble Data Mining Models. *Appl. Sci.* **2020**, *10*, 2039. [[CrossRef](#)]
37. Pham, B.T.; Jaafari, A.; Prakash, I.; Singh, S.K.; Quoc, N.K.; Bui, D.T. Hybrid computational intelligence models for groundwater potential mapping. *Catena* **2019**, *182*. [[CrossRef](#)]
38. Bühlmann, P. Bagging, boosting and ensemble methods. In *Handbook of Computational Statistics*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 985–1022.
39. Ting, K.M.; Witten, I.H. *Stacking Bagged and Dagged Models*; University of Waikato, Department of Computer Science: Hamilton, New Zealand, 1997.
40. Gama, J.; Brazdil, P. Cascade generalization. *Mach. Learn.* **2000**, *41*, 315–343. [[CrossRef](#)]
41. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.
42. Skurichina, M.; Duin, R.P. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Anal. Appl.* **2002**, *5*, 121–135. [[CrossRef](#)]
43. Kuncheva, L.I.; Rodríguez, J.J.; Plumpton, C.O.; Linden, D.E.; Johnston, S.J. Random subspace ensembles for fMRI classification. *IEEE Trans. Med Imaging* **2010**, *29*, 531–542. [[CrossRef](#)]
44. Wang, G.; Ma, J. Study of corporate credit risk prediction based on integrating boosting and random subspace. *Expert Syst. Appl.* **2011**, *38*, 13871–13878. [[CrossRef](#)]
45. Bertoni, A.; Folgeri, R.; Valentini, G. Bio-molecular cancer prediction with random subspace ensembles of support vector machines. *Neurocomputing* **2005**, *63*, 535–539. [[CrossRef](#)]
46. Jaafari, A.; Zenner, E.K.; Pham, B.T. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecol. Inform.* **2018**, *43*, 200–211. [[CrossRef](#)]
47. Pham, B.T.; Tien Bui, D.; Indra, P.; Dholakia, M. Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS-based statistical approach of frequency ratio method. *Int. J. Eng. Res. Technol.* **2015**, *4*, 338–344.
48. Pham, B.T.; Bui, D.T.; Pham, H.V.; Le, H.Q.; Prakash, I.; Dholakia, M. Landslide hazard assessment using random subspace fuzzy rules based classifier ensemble and probability analysis of rainfall data: A case study at Mu Cang Chai District, Yen Bai Province (Viet Nam). *J. Indian Soc. Remote Sens.* **2017**, *45*, 673–683. [[CrossRef](#)]
49. Pham, B.T.; Prakash, I.; Jaafari, A.; Bui, D.T. Spatial prediction of rainfall-induced landslides using aggregating one-dependence estimators classifier. *J. Indian Soc. Remote Sens.* **2018**, *46*, 1457–1470. [[CrossRef](#)]
50. Nguyen, P.T.; Ha, D.H.; Nguyen, H.D.; Van Phong, T.; Trinh, P.T.; Al-Ansari, N.; Le, H.V.; Pham, B.T.; Ho, L.S.; Prakash, I. Improvement of Credal Decision Trees Using Ensemble Frameworks for Groundwater Potential Modeling. *Sustainability* **2020**, *12*, 2622. [[CrossRef](#)]
51. Pham, B.T.; Prakash, I.; Khosravi, K.; Chapi, K.; Trinh, P.T.; Ngo, T.Q.; Hosseini, S.V.; Bui, D.T. A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto Int.* **2019**, *34*, 1385–1407. [[CrossRef](#)]
52. Pham, B.T.; Khosravi, K.; Prakash, I. Application and comparison of decision tree-based machine learning methods in landside susceptibility assessment at Pauri Garhwal Area, Uttarakhand, India. *Environ. Process.* **2017**, *4*, 711–730. [[CrossRef](#)]
53. Pham, B.T.; Prakash, I.; Dou, J.; Singh, S.K.; Trinh, P.T.; Tran, H.T.; Le, T.M.; Van Phong, T.; Khoi, D.K.; Shirzadi, A. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto Int.* **2019**, 1–25. [[CrossRef](#)]
54. Nguyen, V.V.; Pham, B.T.; Vu, B.T.; Prakash, I.; Jha, S.; Shahabi, H.; Shirzadi, A.; Ba, D.N.; Kumar, R.; Chatterjee, J.M. Hybrid machine learning approaches for landslide susceptibility modeling. *Forests* **2019**, *10*, 157. [[CrossRef](#)]
55. Bui, D.T.; Tsangaratos, P.; Ngo, P.-T.T.; Pham, T.D.; Pham, B.T. Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. *Sci. Total Environ.* **2019**, *668*, 1038–1054. [[CrossRef](#)] [[PubMed](#)]
56. 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](#)]

57. Thai Pham, B.; Tien Bui, D.; Prakash, I. Landslide susceptibility modelling using different advanced decision trees methods. *Civ. Eng. Environ. Syst.* **2018**, *35*, 139–157. [[CrossRef](#)]
58. Pham, B.T.; Prakash, I. Machine learning methods of kernel logistic regression and classification and regression trees for landslide susceptibility assessment at part of Himalayan area, India. *Indian J. Sci. Technol.* **2018**, *11*, 1–11. [[CrossRef](#)]
59. Nguyen, P.T.; Tuyen, T.T.; Shirzadi, A.; Pham, B.T.; Shahabi, H.; Omidvar, E.; Amini, A.; Entezami, H.; Prakash, I.; Phong, T.V. Development of a novel hybrid intelligence approach for landslide spatial prediction. *Appl. Sci.* **2019**, *9*, 2824. [[CrossRef](#)]
60. Thai Pham, B.; Shirzadi, A.; Shahabi, H.; Omidvar, E.; Singh, S.K.; Sahana, M.; Talebpour Asl, D.; Bin Ahmad, B.; Kim Quoc, N.; Lee, S. Landslide susceptibility assessment by novel hybrid machine learning algorithms. *Sustainability* **2019**, *11*, 4386. [[CrossRef](#)]
61. Pham, B.T.; Prakash, I. Evaluation and comparison of LogitBoost Ensemble, Fisher’s Linear Discriminant Analysis, logistic regression and support vector machines methods for landslide susceptibility mapping. *Geocarto Int.* **2019**, *34*, 316–333. [[CrossRef](#)]
62. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Han, Z.; Pham, B.T. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* **2019**, 1–18. [[CrossRef](#)]
63. Pham, B.T.; Bui, D.T.; Prakash, I. Application of classification and regression trees for spatial prediction of rainfall-induced shallow landslides in the Uttarakhand area (India) using GIS. In *Climate Change, Extreme Events and Disaster Risk Reduction*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 159–170.
64. Nguyen, V.-T.; Tran, T.H.; Ha, N.A.; Ngo, V.L.; Nadhir, A.-A.; Tran, V.P.; Duy Nguyen, H.; MA, M.; Amini, A.; Prakash, I. GIS Based Novel Hybrid Computational Intelligence Models for Mapping Landslide Susceptibility: A Case Study at Da Lat City, Vietnam. *Sustainability* **2019**, *11*, 7118. [[CrossRef](#)]
65. Pham, B.T.; Prakash, I.; Chen, W.; Ly, H.-B.; Ho, L.S.; Omidvar, E.; Tran, V.P.; Bui, D.T. A Novel Intelligence Approach of a Sequential Minimal Optimization-Based Support Vector Machine for Landslide Susceptibility Mapping. *Sustainability* **2019**, *11*, 6323. [[CrossRef](#)]
66. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Pham, B.T.; Bin Ahmad, B.; Tien Bui, D. A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment. *Geocarto Int.* **2019**, *34*, 1427–1457. [[CrossRef](#)]
67. Thanh, D.Q.; Nguyen, D.H.; Prakash, I.; Jaafari, A.; Nguyen, V.-T.; Van Phong, T.; Pham, B.T. GIS based frequency ratio method for landslide susceptibility mapping at Da Lat City, Lam Dong province, Vietnam. *Vietnam J. Earth Sci.* **2020**, *42*, 55–66. [[CrossRef](#)]
68. Dao, D.V.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayedi, H.; Phong, T.V.; Ly, H.-B.; Le, T.-T.; Trinh, P.T.; et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* **2020**, *188*, 104451. [[CrossRef](#)]
69. Moayedi, H.; Tien Bui, D.; Gör, M.; Pradhan, B.; Jaafari, A. The feasibility of three prediction techniques of the artificial neural network, adaptive neuro-fuzzy inference system, and hybrid particle swarm optimization for assessing the safety factor of cohesive slopes. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 391. [[CrossRef](#)]
70. Janizadeh, S.; Avand, M.; Jaafari, A.; Phong, T.V.; Bayat, M.; Ahmadisharaf, E.; Prakash, I.; Pham, B.T.; Lee, S. Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. *Sustainability* **2019**, *11*, 5426. [[CrossRef](#)]
71. Jaafari, A. LiDAR-supported prediction of slope failures using an integrated ensemble weights-of-evidence and analytical hierarchy process. *Environ. Earth Sci.* **2018**, *77*. [[CrossRef](#)]
72. Pham, B.T.; Phong, T.V.; Nguyen-Thoi, T.; Parial, K.K.; Singh, S.; Ly, H.-B.; Nguyen, K.T.; Ho, L.S.; Le, H.V.; Prakash, I. Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto Int.* **2020**, 1–23. [[CrossRef](#)]
73. Pham, B.T.; Jaafari, A.; Prakash, I.; Bui, D.T. A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 2865–2886. [[CrossRef](#)]
74. Ly, H.-B.; Le, L.M.; Phi, L.V.; Phan, V.-H.; Tran, V.Q.; Pham, B.T.; Le, T.-T.; Derrible, S. Development of an AI model to measure traffic air pollution from multisensor and weather data. *Sensors* **2019**, *19*, 4941. [[CrossRef](#)]

75. Phong, T.V.; Phan, T.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Chapi, K.; Ly, H.-B.; Ho, L.S.; Quoc, N.K.; Pham, B.T. Landslide susceptibility modeling using different artificial intelligence methods: A case study at Muong Lay district, Vietnam. *Geocarto Int.* **2019**, 1–24. [[CrossRef](#)]
76. Pham, B.T.; Phong, T.V.; Nguyen, H.D.; Qi, C.; Al-Ansari, N.; Amini, A.; Ho, L.S.; Tuyen, T.T.; Yen, H.P.H.; Ly, H.-B. A Comparative Study of Kernel Logistic Regression, Radial Basis Function Classifier, Multinomial Naïve Bayes, and Logistic Model Tree for Flash Flood Susceptibility Mapping. *Water* **2020**, *12*, 239. [[CrossRef](#)]
77. Pham, B.T.; Bui, D.T.; Prakash, I.; Nguyen, L.H.; Dholakia, M. A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS. *Environ. Earth Sci.* **2017**, *76*, 371. [[CrossRef](#)]
78. 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](#)]
79. Pham, B.T.; Nguyen, M.D.; Van Dao, D.; Prakash, I.; Ly, H.-B.; Le, T.-T.; Ho, L.S.; Nguyen, K.T.; Ngo, T.Q.; Hoang, V. Development of artificial intelligence models for the prediction of Compression Coefficient of soil: An application of Monte Carlo sensitivity analysis. *Sci. Total Environ.* **2019**, *679*, 172–184. [[CrossRef](#)]
80. Pham, B.T.; Nguyen-Thoi, T.; Ly, H.-B.; Nguyen, M.D.; Al-Ansari, N.; Tran, V.-Q.; Le, T.-T. Extreme Learning Machine Based Prediction of Soil Shear Strength: A Sensitivity Analysis Using Monte Carlo Simulations and Feature Backward Elimination. *Sustainability* **2020**, *12*, 2339. [[CrossRef](#)]
81. Pham, B.T.; Qi, C.; Ho, L.S.; Nguyen-Thoi, T.; Al-Ansari, N.; Nguyen, M.D.; Nguyen, H.D.; Ly, H.-B.; Le, H.V.; Prakash, I. A Novel Hybrid Soft Computing Model Using Random Forest and Particle Swarm Optimization for Estimation of Undrained Shear Strength of Soil. *Sustainability* **2020**, *12*, 2218. [[CrossRef](#)]
82. Qi, C.; Ly, H.-B.; Chen, Q.; Le, T.-T.; Le, V.M.; Pham, B.T. Flocculation-dewatering prediction of fine mineral tailings using a hybrid machine learning approach. *Chemosphere* **2020**, *244*, 125450. [[CrossRef](#)]
83. Vasu, N.N.; Lee, S.-R. A hybrid feature selection algorithm integrating an extreme learning machine for landslide susceptibility modeling of Mt. Woomyeon, South Korea. *Geomorphology* **2016**, *263*, 50–70. [[CrossRef](#)]
84. 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](#)]
85. Kordestani, M.D.; Naghibi, S.A.; Hashemi, H.; Ahmadi, K.; Kalantar, B.; Pradhan, B. Groundwater potential mapping using a novel data-mining ensemble model. *Hydrogeol. J.* **2019**, *27*, 211–224. [[CrossRef](#)]
86. Kalantar, B.; Al-Najjar, A.H.H.; Pradhan, B.; Saeidi, V.; Halin, A.A.; Ueda, N.; Naghibi, A.S. Optimized Conditioning Factors Using Machine Learning Techniques for Groundwater Potential Mapping. *Water* **2019**, *11*. [[CrossRef](#)]
87. Jaafari, A.; Razavi Termeh, S.V.; Bui, D.T. Genetic and firefly metaheuristic algorithms for an optimized neuro-fuzzy prediction modeling of wildfire probability. *J. Environ. Manag.* **2019**, *243*, 358–369. [[CrossRef](#)]
88. Bui, D.T.; Ngo, P.T.T.; Pham, T.D.; Jaafari, A.; Minh, N.Q.; Hoa, P.V.; Samui, P. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *Catena* **2019**, *179*, 184–196. [[CrossRef](#)]
89. Tien Bui, D.; Moayed, H.; Gör, M.; Jaafari, A.; Foong, L.K. Predicting slope stability failure through machine learning paradigms. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 395.
90. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S.; et al. Spatial Prediction of Landslide Susceptibility Using GIS-Based Data Mining Techniques of ANFIS with Whale Optimization Algorithm (WOA) and Grey Wolf Optimizer (GWO). *Appl. Sci.* **2019**, *9*, 3755. [[CrossRef](#)]
91. Jaafari, A.; Zenner, E.K.; Panahi, M.; Shahabi, H. Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agric. For. Meteorol.* **2019**, *266–267*, 198–207. [[CrossRef](#)]
92. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Tran, T.T.T.; Bui, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [[CrossRef](#)]
93. Pham, B.T.; Avand, M.; Janizadeh, S.; Phong, T.V.; Al-Ansari, N.; Ho, L.S.; Das, S.; Le, H.V.; Amini, A.; Bozchaloei, S.K. GIS Based Hybrid Computational Approaches for Flash Flood Susceptibility Assessment. *Water* **2020**, *12*, 683. [[CrossRef](#)]

94. Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Bui, D.T.; Pham, B.T.; Khosravi, K. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ. Model. Softw.* **2017**, *95*, 229–245. [[CrossRef](#)]
95. Nasiri, V.; Darvishsefat, A.A.; Rafiee, R.; Shirvany, A.; Hemat, M.A. Land use change modeling through an integrated multi-layer perceptron neural network and Markov chain analysis (case study: Arasbaran region, Iran). *J. For. Res.* **2019**, *30*, 943–957. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).