

Article

Improvement of Credal Decision Trees Using Ensemble Frameworks for Groundwater Potential Modeling

Phong Tung Nguyen ^{1,*}, Duong Hai Ha ², Huu Duy Nguyen ³, Tran Van Phong ⁴ , Phan Trong Trinh ⁴, Nadhir Al-Ansari ^{5,*} , Hiep Van Le ⁶, Binh Thai Pham ^{6,*} , Lan Si Ho ^{7,*} and Indra Prakash ⁸ 

¹ Vietnam Academy for Water Resources, Hanoi 100000, Vietnam

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

³ Faculty of Geography, VNU University of Science, Vietnam National University, 334 Nguyen Trai, Hanoi 100000, Vietnam; huuduy151189@gmail.com

⁴ Institute of Geological Sciences, Vietnam Academy of Sciences and Technology, 84 Chua Lang Street, Dong da, Hanoi 100000, Vietnam; tphong1617@gmail.com (T.V.P.); phantrongt@yahoo.com (P.T.T.)

⁵ Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden

⁶ University of Transport Technology, Hanoi 100000, Vietnam; hieplv@utt.edu.vn

⁷ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

⁸ Department of Science & Technology, Bhaskaracharya 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.); binhpt@utt.edu.vn (B.T.P.); hosilanh@duytan.edu.vn (L.S.H.)

Received: 10 February 2020; Accepted: 24 March 2020; Published: 26 March 2020



Abstract: Groundwater is one of the most important sources of fresh water all over the world, especially in those countries where rainfall is erratic, such as Vietnam. Nowadays, machine learning (ML) models are being used for the assessment of groundwater potential of the region. Credal decision trees (CDT) is one of the ML models which has been used in such studies. In the present study, the performance of the CDT has been improved using various ensemble frameworks such as Bagging, Dagging, Decorate, Multiboost, and Random SubSpace. Based on these methods, five hybrid models, namely BCDT, Dagging-CDT, Decorate-CDT, MBCDT, and RSSCDT, were developed and applied for groundwater potential mapping of DakLak province of Vietnam. Data of 227 groundwater wells of the study area were utilized for the construction and validation of the models. Twelve groundwater potential conditioning factors, namely rainfall, slope, elevation, river density, Sediment Transport Index (STI), curvature, flow direction, aspect, soil, land use, Topographic Wetness Index (TWI), and geology, were considered for the model studies. Various statistical measures, including area under receiver operating characteristic (AUC) curve, were applied to validate and compare the performance of the models. The results show that performance of the hybrid CDT ensemble models MBCDT ($AUC = 0.770$), BCDT ($AUC = 0.731$), Dagging-CDT ($AUC = 0.763$), Decorate-CDT ($AUC = 0.750$), and RSSCDT ($AUC = 0.766$) improved significantly in comparison to the single CDT ($AUC = 0.722$) model. Therefore, these developed hybrid models can be applied for better ground water potential mapping and groundwater resources management of the study area as well as other regions of the world.

Keywords: Groundwater potential mapping; Machine learning; Ensemble Frameworks; Vietnam

1. Introduction

Groundwater is a vital natural resource for drinking water supply, irrigation and industries in many countries [1–3]. About 2.5 billion people all over the world depend on groundwater resources for drinking and agriculture [4]. Most of the world's groundwater resources are being overexploited, and thus acute water shortage is expected by 2025 all around the world as the fresh water resources are limited [5–7]. Population growth creates higher demand for water for domestic use, in addition to industrial development and extension of irrigated areas [8,9]. This problem is more prevalent in the arid and semi-arid regions, which have faced numerous drought events in recent years due to erratic scanty rainfall [10,11]. Thus, the identification and mapping of groundwater potential zones is an important task to recharge the aquifer. In recent years, several researchers, namely Magesh, Chandrasekar and Soundranayagam [1], Oikonomidis, Kazakis, Voudouris et al. [12], Rahmati, Samani, Mahdavi, et al. [13], and Zabihi, Pourghasemi, Pourtaghi, et al. [14], have studied groundwater potential, considering geological, hydrological and climatic factors using statistical methods, remote sensing, and geographic information system (GIS) technology [15]. Traditionally, expert's opinion-based models or weighted models have been used for groundwater potential mapping. However, these approaches are considered subjective and uncertainty [16,17].

Nowadays, artificial intelligence (AI)-based machine learning (ML) models are being utilized for mapping of groundwater potential with the advancement of spatial data acquisition and analysis. ML models are based on computational algorithms to deal with complex problems with complex datasets [18]. Chen, Li, Tsangaratos, et al. [19] used ML models based on Random Forest (RF), Kernel Logistic Regression (KLR), and Alternating Decision Tree (ADT) for groundwater potential mapping in China. Naghibi, Pourghasemi and Dixon [20] applied and compared several ML models, namely Classification and Regression Tree (CART), Boosted Regression Tree (BRT) and RF for GIS-based groundwater potential mapping in Iran. Lee, Hong and Jung [21] used Artificial Neural Network (ANN) and Support Vector Machines (SVM) models to develop groundwater potential maps in Korea. Park, Hamm, Jeon, et al. [22] compared two ML-based models of Multivariate Adaptive Regression Splines and Logistic Regression (LR) for groundwater potential mapping in Korea. Ozdemir [23] applied LR for mapping of groundwater potential in Turkey. Other popular ML-based models used for groundwater potential mapping are Adaptive Network-based Fuzzy Inference System [24], Naïve Bayes [25], K-nearest neighbor and Quadratic Discriminate Analysis [26]. Although all these single ML models performed well in the studied regions, there is no model available that can be applied to all regions including hybrid models [25] for optimal groundwater potential mapping.

With the above objective, the present study was carried out to fill the gap of suitable and better models by improving the predictive capability of Credal Decision Trees (CDT), which is a popular machine learning method but quite sensitive with tree construction [27,28]. Different ensemble frameworks namely Bagging, Dagging, Decorate, Multiboost, and Random SubSpace were used to develop five hybrid models with base classifier CDT such as BCDT, Dagging-CDT, Decorate-CDT, MCDT, and RSSCDT. For the model studies, the DakLak province of Vietnam was selected, where groundwater resources are required to be properly exploited as rainfall in this area is erratic due to climate change's effects [29,30]. To validate the models, several statistical measures, including area under the receiver operating characteristics (ROC) curve (*AUC*), were applied on the datasets. GIS and Weka software were used for data preparation, analysis and modeling.

2. Methods Used

2.1. Credal Decision Trees (CDT)

CDT is a classifier which is based on uncertainty measures and imprecise probabilities. CDT was first proposed in 2003 by Abell  n and Moral to solve the classification problem using credal sets [27]. To avoid complicated decision tree production, a new concept was developed, which is stopping the classification process from increasing the total uncertainty due to ramification of the decision tree [31].

Therefore, a new advanced method is built to measure quantitatively the total uncertainty from credal set based on the theory of Dempster and Shafer, as presented in following equation:

$$TU(x) = IG(x) + GG(x) \quad (1)$$

where x is defined as a credal set on frame X , TU is the total uncertainty value, IG is defined as a general function of non-specificity on the corresponding set of credits and GG is defined as a general randomness function for a credal set [32].

2.2. Bagging

Bagging is an ensemble technique that combines many ML classifiers together to create more accurate predictors. The Bagging algorithm is constructed from the combination of Bootstrap and Aggregating to create a unique overall model [33,34]. Bagging is a sensitive algorithm. In the Bagging method, small changes in the dataset can cause significant changes in the final results [35]. In this algorithm, learning data to be used for each learner is obtained by bootstrap sampling, and the learned learner is used for prediction and the final ensemble [36]. Bagging produces better accuracy as it can perform more independent learning.

2.3. Dagging

Dagging was first proposed by Ting and Witten in 1977. It uses certain separate samples instead of Bootstrap samples to extract the basic classifications [37]. The name of Dagging is the original of Bagging. In the Dagging algorithm, the dataset is used to classify once, and it is also disjointed [38]. In this model, majority voting is used to group the classifications to improve the accuracy of basic classification prediction [39].

2.4. Decorate

Decorate algorithm was introduced by Melville and Mooney in 2003 [40] to improve training data by creating artificial data. These data are constructed using the training variables of means and standard deviation according to the Gaussian distribution. They are added to the training samples. The difference between Decorate and other ensembles (Bagging and Adaboost) is that Adaboost and Bagging use only given training variables to create the various classifications [41], whereas Decorate builds the basic classifications using artificial data, which allows us to no longer be constrained by the training samples given when managing a set.

2.5. MultiBoost

Multiboost was introduced by Webb in 2000 [42]. This technique is produced by combining Adaboost and Wagging techniques to reduce the problem of variance and over-fitting [43]. The use of training boxes with different weights in the Wagging model can reduce the high bias in Adaboost model [44]. Combination of Adaboost and Wagging is an advantage in the classification process of weak learning and transforming it into strong learning [42]. Multiboost is formed in three stages. The first is randomly selecting a subset from the original data which are used to form the models. The second is the weights, which show the changes in the model prediction. In the third, the new subsets are chosen from the weighted instance to produce the new models [45].

2.6. Random SubSpace

Random SubSpace is considered to be one of the most popular random sampling methods which Ho proposed in 1988 to improve predictive capability of the individual classifications and accuracy of weak classifications [43,46]. In this technique, the original characteristic vector with the strong dimension is randomly divided to construct the subspace with a small dimension and then several classifications are randomly grouped in subspace at the final decision [46]. The subset characteristic

series of each sub-classification formation to the final prediction results are grouped using a majority vote [47].

2.7. Correlation-based Feature Selection

Selection of the appropriate factors is a very important task for constructing input variables and testing the ML models [20,48]. It can help to assess each variable in predicting outcomes by removing unnecessary factors from the input data. Therefore, the quality of the data is improved by reducing over-fitting and the noise-related problems. This leads to an increase in the model's predictive capacity [49]. There are several methods to select variables, such as ORAE, Gain information, and correlation-based feature selection [50]. Among them, correlation-based feature selection was selected in the present study. This method evaluates the attributes of the target class. It can be used to measure the correlations between each input variable and the output variable on which importance of input variables is evaluated and ranked [51].

2.8. Validation Methods

The performance of the models is evaluated by validation methods [52,53]. In addition, comparison of the training data and validation data plays an important role in determining the fit of the model [54]. For the validation of models, various statistical indices were used namely Negative Predictive Value (NPV), Positive Predictive Value (PPV), Specificity (SPF), Kappa (k), Sensitivity (SST), Accuracy (ACC), Root Mean Squared Error (RMSE) and Receiver Operating Characteristic (ROC).

PPV and NPV present the percentage of pixels which are correctly predicted as “potential groundwater” and “non-potential groundwater” [55,56]. Meanwhile, SST and SPF express the pixels which are correctly classified as “potential groundwater” and “non-potential groundwater” [57]. ACC shows the proportionality of classification “true negative” and “true positive” for the test, which are the pixel rate which is correctly classified from “potential groundwater” and “non-potential groundwater” [58,59]. False Positive (FP) and True Positive (TP) are considered to be the probability of a pixel which is incorrectly and correctly classified from “groundwater potential”, respectively, while False Negative (FN) and True Negative (TN) show the probability of a pixel which is incorrectly and correctly classified as “non-potential groundwater” [60]. These statistical measures can be calculated by following equations:

$$PPV = \frac{TP}{TP + FP} \quad (2)$$

$$NPV = \frac{TN}{TN + FN} \quad (3)$$

$$SST = \frac{TP}{TP + FN} \quad (4)$$

$$SPF = \frac{TN}{FP + TN} \quad (5)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Kappa (k) index is considered to be one of the most popular statistical measures for evaluating the ML models. Kappa presents the percentage of the agreement between the evaluators. The Kappa is often considered as a random chord. It was used to classify N objects into C mutually exclusive sets. The value of kappa ranges between -1 and 1 . If kappa equals 1 , the model has perfect performance [61–63]. Kappa (k) can be calculated by following equation:

$$k = \frac{P_p - P_{exp}}{1 - P_{exp}} \quad (7)$$

where P_p is the accuracy and P_{exp} is the expected agreements.

RMSE is the statistical index to assess differences between the predictive value and the target value [64–66]. RMSE is a good metric for the comparison between the performances of the models, which is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{predicted} - X_{actual})^2} \quad (8)$$

where n is defined as the total of variables, $X_{predicted}$ and X_{actual} are the prediction and actual values of variable i -th.

ROC is a graph commonly used in the validation of binary classification models. This curve is created by expressing sensitivity and specificity [67,68]. Therefore, the ROC curve will show the relationship, the trade-off and the significance of choosing an appropriate model of sensitivity and false alarm rate. Area under the ROC curve, called AUC, is often utilized quantitatively to validate and compare predictive capability of the models, which is calculated as follows:

$$AUC = \sum TP + \sum \frac{TN}{P} + N \quad (9)$$

where P and N are defined as the total number of “potential-groundwater” and “non-potential groundwater” samples, respectively.

3. Study Area

The study area of DakLak province is located in between 107°28′57″ to 108°59′37″ East longitude; and 12°9′45″ to 13°25′06″ North Latitude in the central highlands of Vietnam (Figure 1) on the upper course of Serepok River and a part of Ba River, covering about 13085 km². The topography of the DakLak province ranges from flat highland to mountainous. The highest peak in this area is Chu Yang Sin (2442 m). Other mountain peaks are Chu H’mu (2051 m), Chu De (1793 m), and Chu Yang Pel (1600 m). The average height of the highland is 450 m.

In general, the climate of the DakLak province varies as per the variation of topography. The area below 300 m elevation is hot, that between elevation 400 and 800 m is hot and humid; and that above 800 m is cold. In this region, about 90% of the annual rainfall occurs during the rainy season (May to October) and is almost negligible during summer (November to April).

Groundwater resource in the DakLak province is widely used for all needs, especially for irrigation. According to the Vietnam Academy for Water Resources (2018), in the dry season, the total volume of water needed is 264,000 hectares. For coffee cultivation, the total water requirement is about 660 million m³ against the availability of surface water of 250 million m³. Therefore, the remaining water requirement is to be met by groundwater for coffee production as well as for other crop cultivation to avoid drought conditions. Currently, the amount of water exploited in the dry months in this province is estimated to be about 500,000 m³/day for irrigation, which is mainly concentrated in the Basalt Complex. About one third of the study area is covered by basalt rock and remaining by quaternary sediments, Pliocene formation and Proterozoic metamorphic rocks.

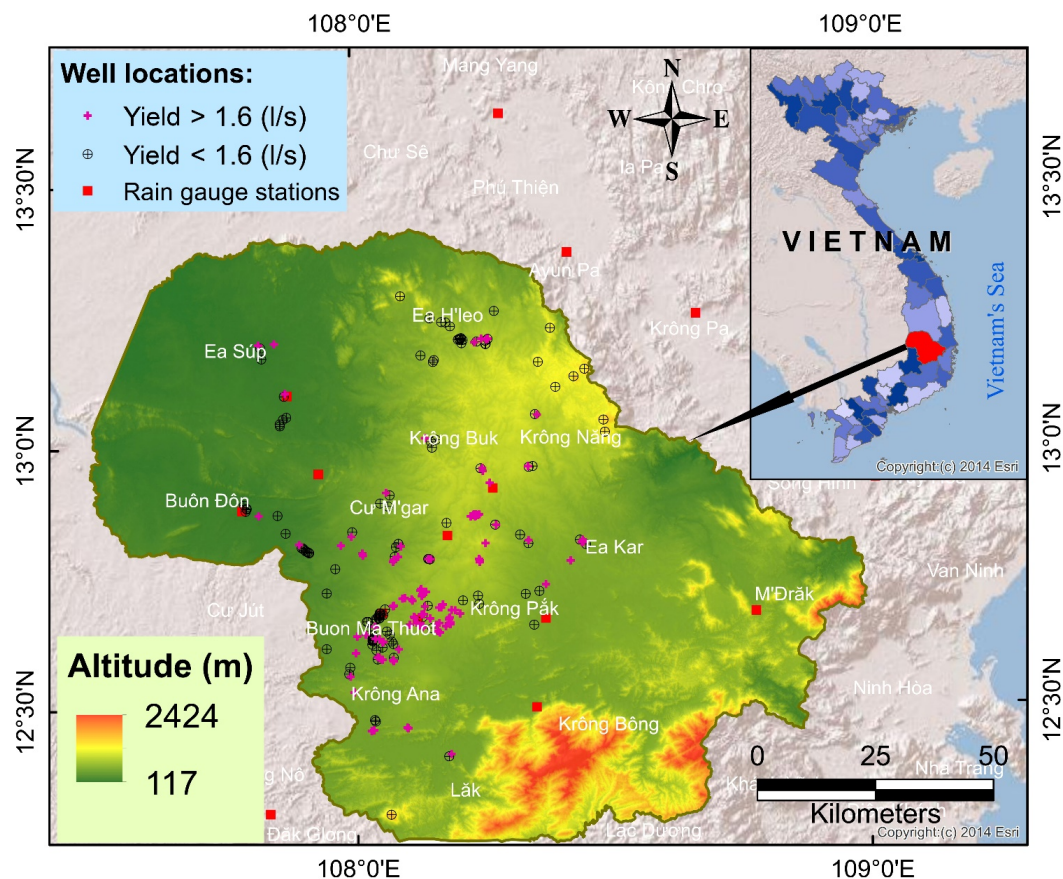


Figure 1. Location map of the study area showing well and rain gauge locations.

4. Data Used

4.1. Well Yields

Well yield data of 227 wells of the DakLak province obtained from the Vietnam Academy for Water Resources (VAWR) were used in the present study (VAWR 2018). The data were split into two parts: 70% of the data were used to train the model, and the remaining 30% of the data for the validation of the model. Based on the local conditions and requirements, 1.6 l/s yield of wells was used as a threshold value for the model study [69].

4.2. Groundwater Influencing Parameters

In the groundwater model study, the groundwater influencing parameters or conditioning factors based on topography, hydrology, geo-environmental conditions and anthropogenic activities play an important role in the model's predictive capacity [70]. In the present study, 12 groundwater affecting factors, namely aspect, curvature, elevation, slope, Sediment Transport Index (STI), flow direction, rainfall, river density, soil type, Topographic Wetness Index (TWI), land use, and geology (lithology), were selected for modeling. Topography and hydrology factors were extracted from the Aster Digital Elevation Model (DEM) of 30m resolution from the United States Geological Survey (USGS) website (<https://earthexplorer.usgs.gov/>) using GIS application and SAGA software [71]. Land use map (scale 1:50000) and soil map (scale 1:100000) was obtained from the Daklak Department of Natural Resources and Environment (DARD). Geology and rainfall maps were extracted from the hydrogeological map (1:300.000 scale) of South Central and Central Highland Vietnam conducted by the Central region of Vietnam Division for Water Resources Planning and Investigation (CEVIWRPI).

The aspect map shows the direction of the slope [72–74]. In this study, the aspect map is divided into nine classes (Figure 2a). The curvature map indicates the relationship with the ability to accumulate

and retain water on the surface. Normally, the concave slope accumulates more water [48,75,76]. In this region, curvature ranges from 23.5 to 30.8 (Figure 2b). Elevation is considered as one of the most important factors in the groundwater potential model as it has the inverse proportionality with the potential of underground water [77]. In the study region, elevation ranges from 117 to 2424 m (Figure 2c). Slope has a direct relationship with the hydrological process. On flat ground, the accumulation of surface water would be more and thus more infiltration is likely, which would help in groundwater recharge [76]. Slope in the DakLak province is grouped in different classes based on the natural break method between 0 and 69.9 degrees (Figure 2d).

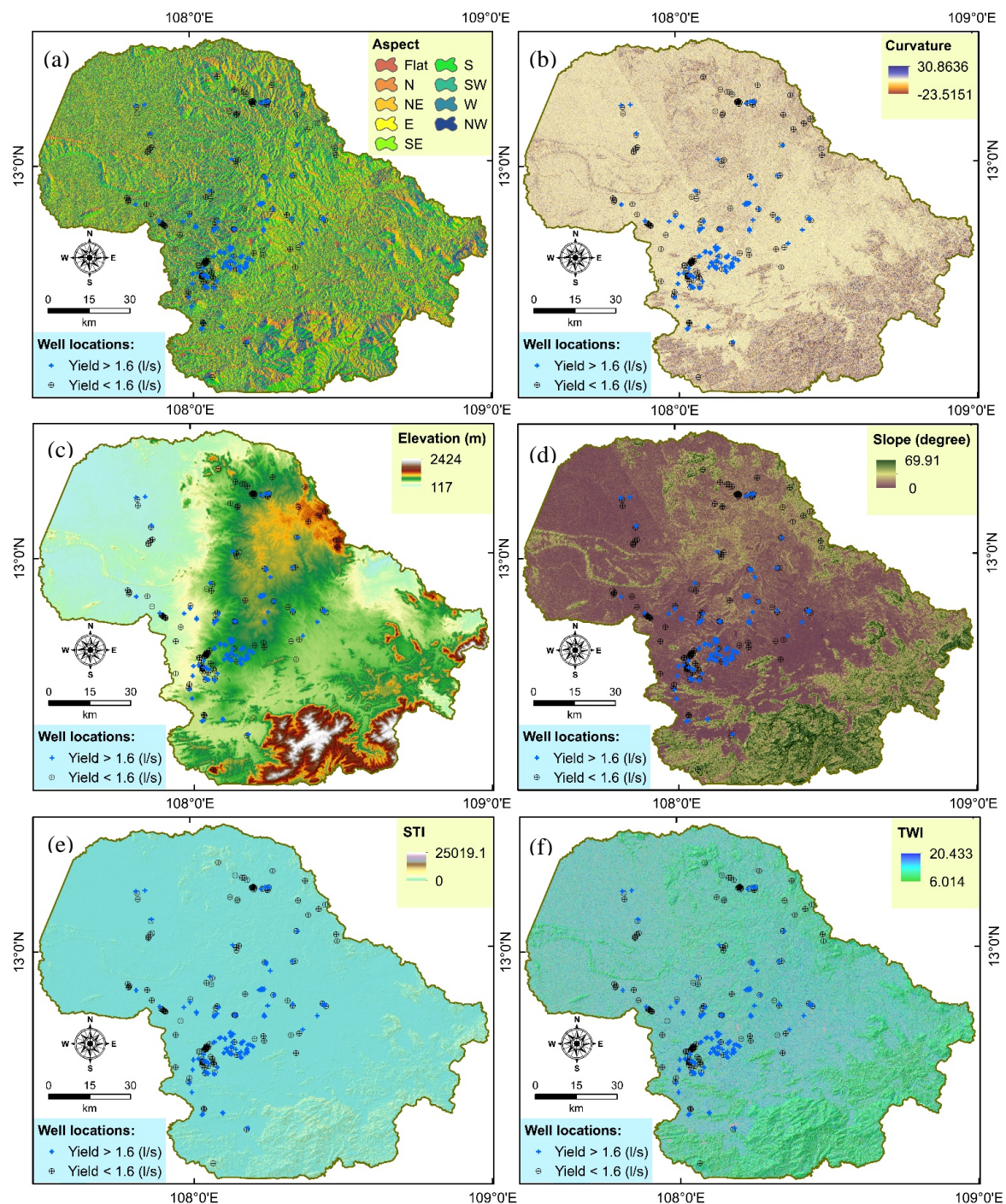


Figure 2. Cont.

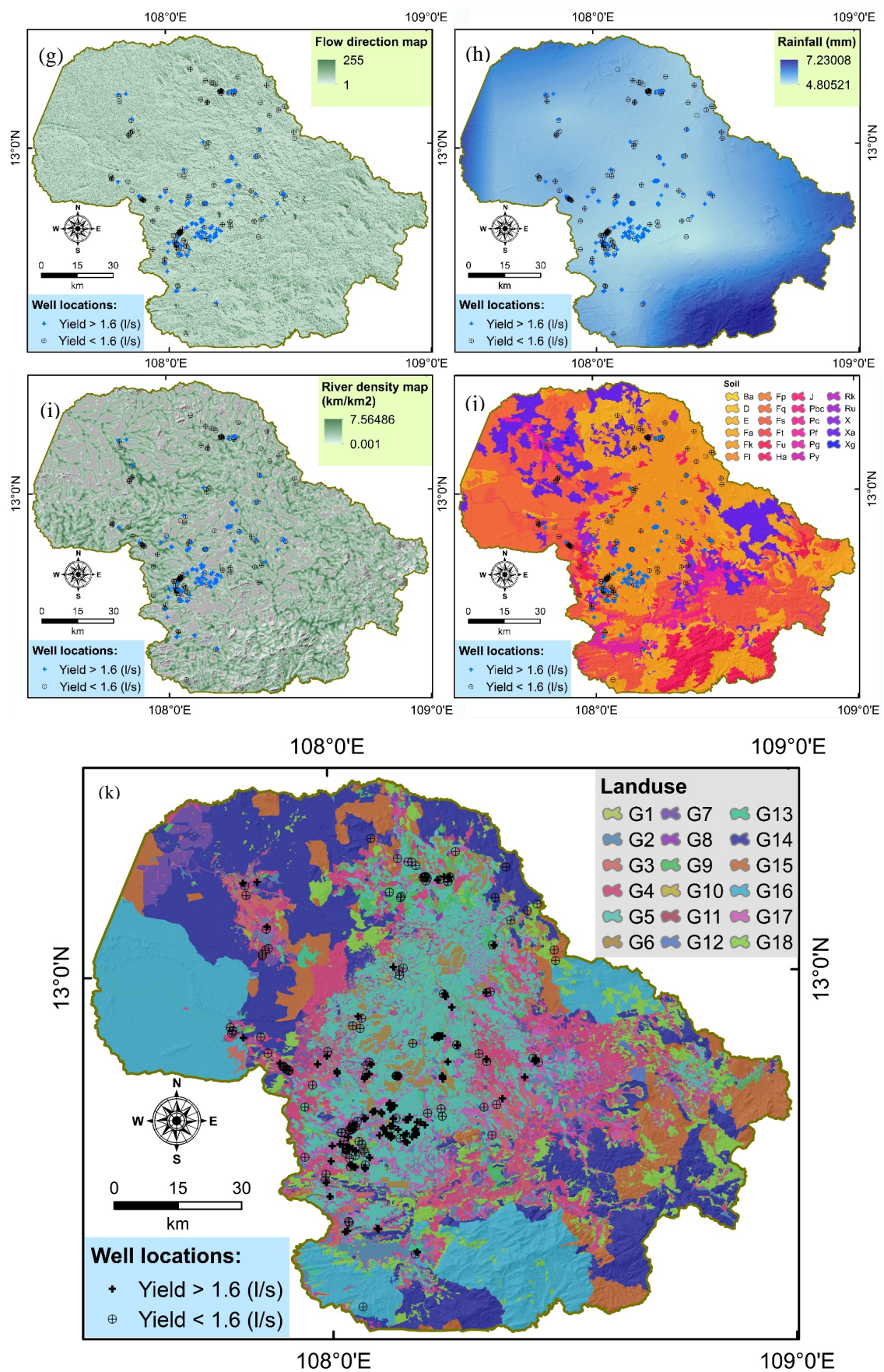


Figure 2. Cont.

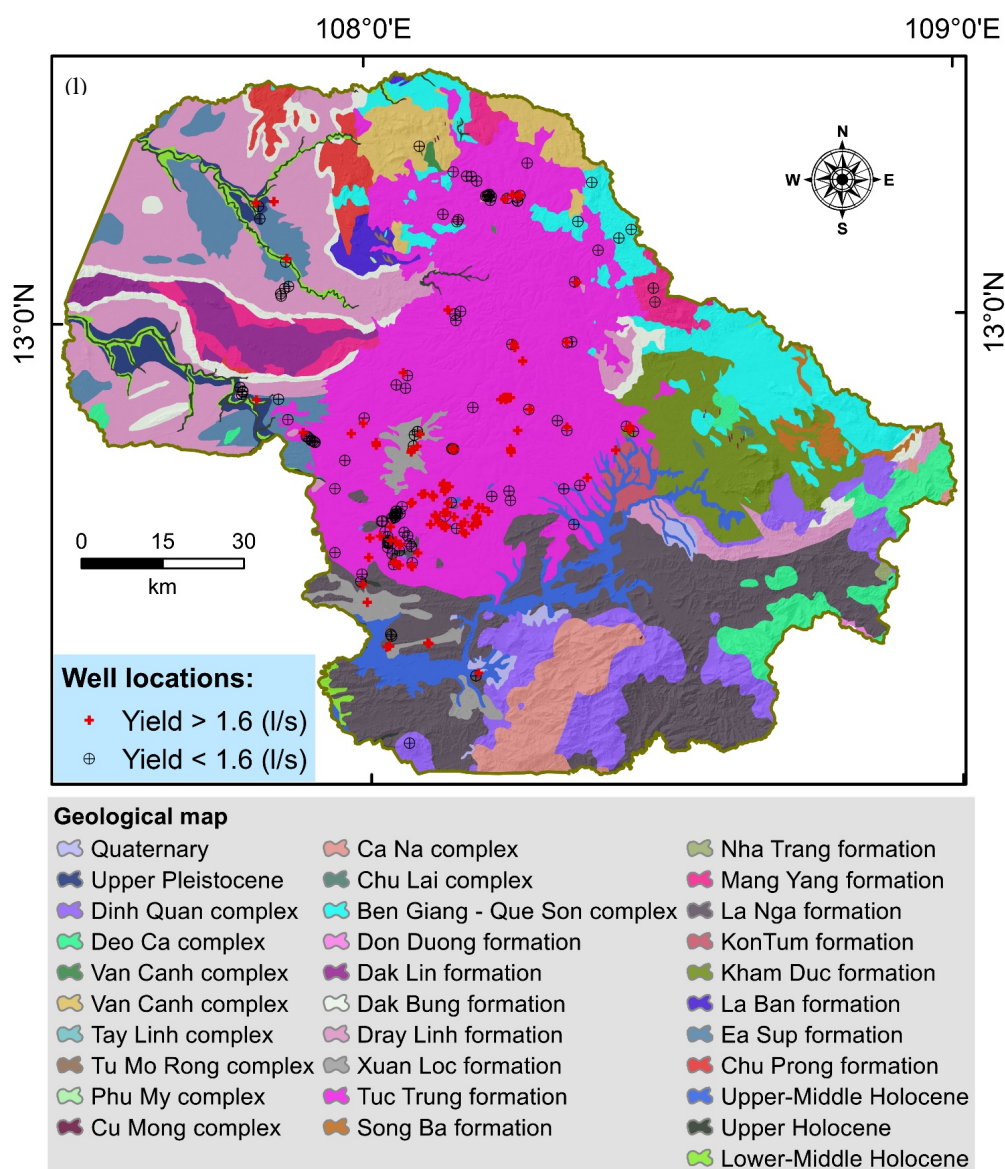


Figure 2. Groundwater conditioning factor maps used in this study: (a) aspect, (b) curvature, (c) elevation, (d) slope, (e) Sediment Transport Index (STI), (f) Topographic Wetness Index (TWI), (g) flow direction, (h) rainfall, (i) river density, (j) soil, (k) land use, and (l) geology.

STI helps in assessing erosion and deposition [78,79]. In this region, it varies from 0 to 25,019 (Figure 2e). TWI reflects the relationship between topography and the condition of the groundwater occurrence [80]. In this area, the value of TWI ranges from 6.04 to 20.433 (Figure 2f). Flow direction indicates the direction of runoff from higher to low region thus affecting infiltration [81,82]. In this area, the flow direction value ranges from 1 to 255 (Figure 2g). Rainfall is considered as an important factor for groundwater potential mapping because the chances of infiltration are greater in cases of high precipitation, thus leading to more recharge [83,84]. The average yearly rainfall value in the study area ranges from 4.80 to 7.23 mm (Figure 2h). River density is the inverse proportionality with infiltration [48,83–85]. The study area has a high river density (7.565 km/km^2) thus less probability of recharge (Figure 2i).

Soil is also an important factor in the modeling of groundwater potential. Permeability of the soil depends on its texture and structure which reflects the infiltration capacity of the soil [86–88]. The soil map of the study area is grouped into different classes based on local variations of soil properties (Figure 2j and Table 1). Land use depends on the topography, nature of the soil, hydrology, meteorology

and human (anthropogenic) requirement. Anthropogenic activities generally change the land use pattern, thus affecting groundwater potential locally [48,89]. In this study, the land use map was classified into various classes (G1 to G18) (Figure 2k and Table 2). Geology plays an important role in groundwater occurrence and thus in modeling of groundwater potential. Geological structure affects surface water infiltration (recharge) and groundwater movement. The porosity and permeability of rocks are important for assessing the characteristics of the ground surface and aquifer [90,91]. The geology map of the region was classified into different types of formation based on the characteristics of rocks (Figure 2l).

Table 1. Information of soil map.

No.	Code	Description	No.	Code	Description
1	Ba	Faded soil on acid magma and sand	13	J	Grab soil
2	D	Land sloping valley by the convergence	14	Pbc	Sour alluvial soil
3	E	Soil erosion, inert	15	Pc	Alluvial soil
4	Fa	Red yellow soil on acid magma	16	Pf	The alluvial soil has red and yellow sloping layers
5	Fk	Red-brown soil on basalt	17	Pg	Alluvium Clay soil
6	Fl	Red-yellow soil changes due to wet rice cultivation	18	Py	Stream alluvial soil
7	Fp	Brown-yellow soil on ancient alluvial gold	19	Rk	Black soil on basalt accretion products
8	Fq	Pale yellow soil on sand stone	20	Ru	Permeable brown soil on foam basalt products
9	Fs	Yellow-red soil on clay and metamorphic rocks	21	X	Gray soil on ancient alluvium
10	Ft	Purple-brown soil on basalt	22	Xa	Gray soil on acid magma and sand stone
11	Fu	Brown-yellow soil on basalt	23	Xg	Gray Clay soil
12	Ha	Red yellow humus on acid magma rock			

Table 2. Information of land use map.

No.	Code	Description	No.	Code	Description
1	G1	Rice - field	10	G10	Other perennials
2	G2	Vegetable - field	11	G11	Upland rice
3	G3	Annual plant	12	G12	Meadow
4	G4	Coffee plant	13	G13	Specialized land
5	G5	Rubber plant	14	G14	Unused land
6	G6	Cashew plant	15	G15	Production forests
7	G7	Pepper plant	16	G16	Protection Forest
8	G8	Tea plant	17	G17	Special use forest
9	G9	Cocoa plant	18	G18	Residential

5. Methodological Flow Chart

The methodology of the present groundwater potential model study is divided into four main stages: (1) GIS data collection and preparation, (2) correlation-based feature selection and generation of datasets, (3) hybrid model construction, and (4) performance assessment and final trained hybrid models (Figure 3). More specifically, (1) groundwater inventory map and conditioning factor maps were prepared and analyzed to develop groundwater potential map. As the original data of these maps were on different scales (units), they were normalized to values from 0 to 1 for the use as model input data [92]; (2) correlation-based feature selection was used to validate and select the suitable conditioning input factors for groundwater potential assessment, and then inventory data was split into two parts: the first part was used to build the model with 70% of the data (training data), and another 30% (testing data) were used to validate the model; (3) various hybrid ensemble framework-based models in the combination of single models, namely single CDT, BCDT, Dagging-CDT, Decorate-CDT, MBCDT, and RSSCDT, were constructed using training datasets. A list of the model parameters utilized for training the models is presented in Table 3; (4) groundwater potential models were validated using various statistical measures: *SST*, *SPF*, *ACC*, *K*, *PPV*, *NPV*, *RMSE* and *AUC*. After the validation of the models, groundwater potential maps were constructed using the studied models. These maps were classified into five classes: very high, high, moderate, low and very low based on the natural break classification method [93] in GIS application.

Table 3. List of the parameters used in different models.

No	Parameter	Models					
		CDT	BCDT	Dagging-CDT	Decorate-CDT	MBCDT	RSSCDT
1	KTH Root Attribute	1	-	-	-	-	-
2	S Value	1.0	-	-	-	-	-
3	Initial class value count	0	-	-	-	-	-
4	Maximum tree depth	-1	-	-	-	-	-
5	Minimum total weight of instances in a leaf	2.0	-	-	-	-	-
6	Minimum proportion of the variance	0.001	-	-	-	-	-
7	Number of Decimal Places	2	2	2	2	2	2
8	Number of Folds	3	-	9	-	-	-
9	Seed	1	1	1	1	1	1
10	Size of each bag	-	100	-	-	-	-
11	Batch Size	-	100	100	100	100	100
12	Number of Execution Slots	-	1	-	-	-	1
13	Number of Iterations	-	17	-	4	7	6
14	Artificial Size	-	-	-	1.0	-	-
15	Desired Size	-	-	-	15	-	-
16	Number of Subcommittees	-	-	-	-	3	-
17	Weight threshold	-	-	-	-	100	-
18	Size of each SubSpace	-	-	-	-	-	0.5

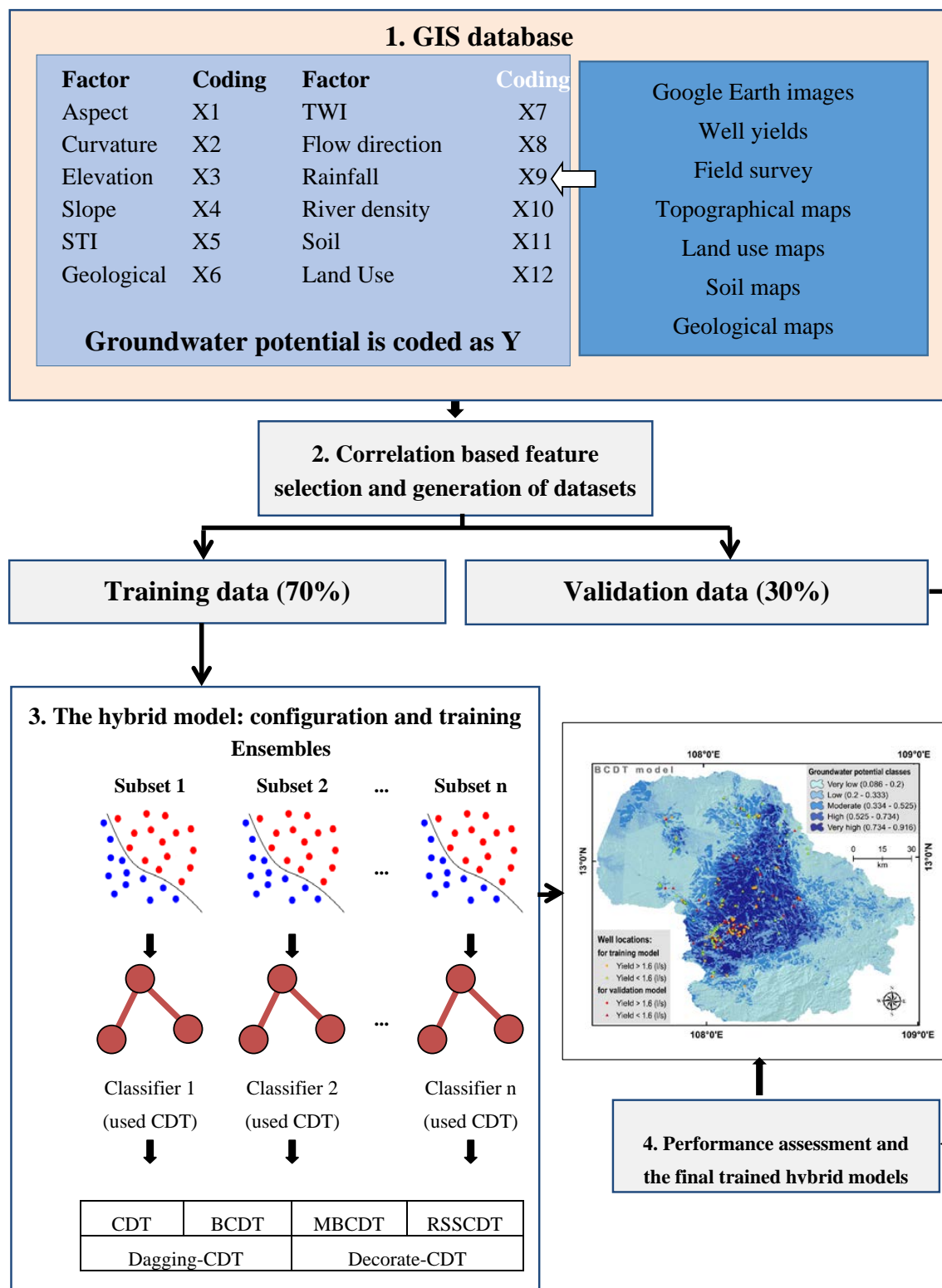


Figure 3. Methodological flow chart of this study.

6. Results and Analysis

6.1. Analysis of Feature Selection of Groundwater Potential Influencing Factors

Groundwater potential influencing factors are selected based on the field knowledge of the area, including geology, topography, geomorphology, meteorology, land use pattern and anthropogenic

activities [48,94,95]. At present, there is currently no known best method which can help in selecting the appropriate influencing factors for the groundwater potential assessment universally for all the areas [54,96,97]. However, to accomplish this task, at present correlation based feature selection method is considered to be one of the most popular methods due to its ability to take into account the impacts of each variable [49]. Therefore, in this study, this method was applied to 12 initially considered factors: land use, slope, elevation, river density, STI, curvature, TWI, flow direction, aspect, soil, geology, and rain fall. The results show that all these factors (variables) contributed to the groundwater potential model, but among these, land use and rainfall are the most important factors in the study area (Figure 4).

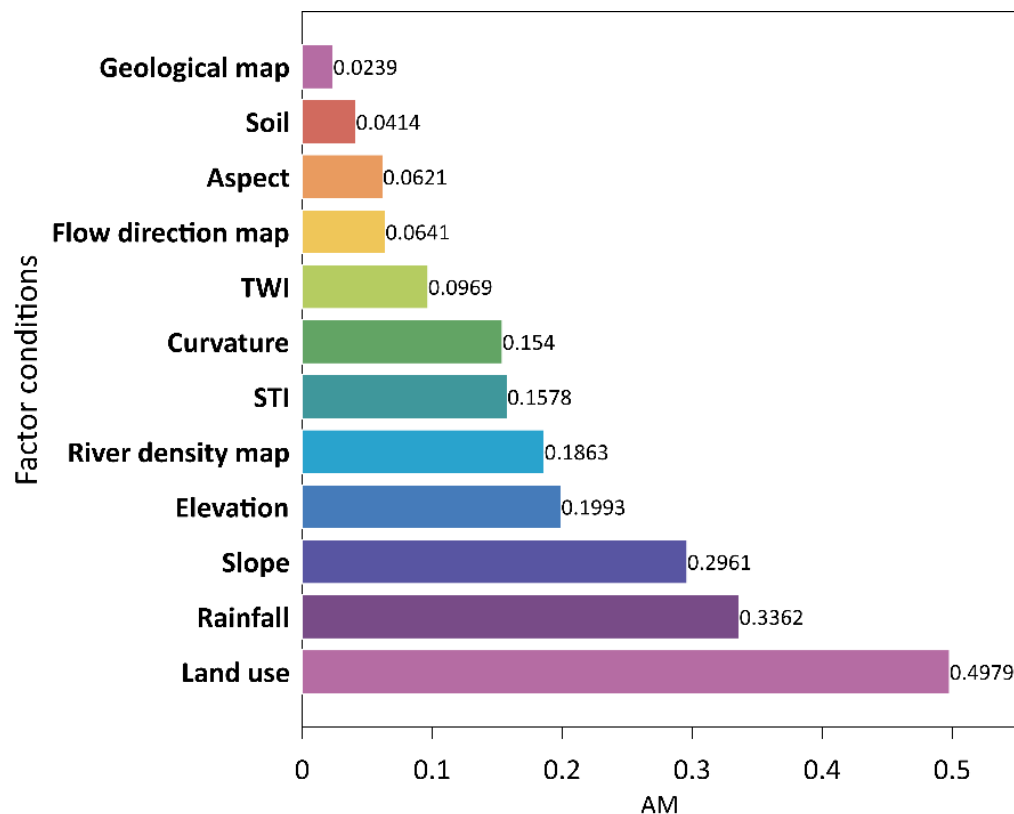


Figure 4. Importance of the factors using correlation-based feature selection.

6.2. Evaluation of Models Performance Using Statistical Methods

Groundwater potential models were constructed using training data and validated by testing data [98,99]. Weka software was used for the modeling. For training data, the results indicate that the MBCDT model is better in terms of *PPV* and *SPF*, whereas, in terms of *NPV* value, the Dagging-CDT model is better in comparison to other models. However, the RSSCDT model is more efficient than the other models for *SST* Kappa and *ACC* values (Figures 5 and 6). The results of the validation data suggest that the RSSCDT model is more efficient than other models in terms of *NPV*, *SST*, *ACC* and Kappa values (Figures 5 and 6).

Analysis of the model's performance was also done using *RMSE* values. The results indicate that the BCDT model is the best in terms of training data (Figure 7), whereas the RSSCDT model is more efficient for the validation data in comparison to other models (Figure 8).

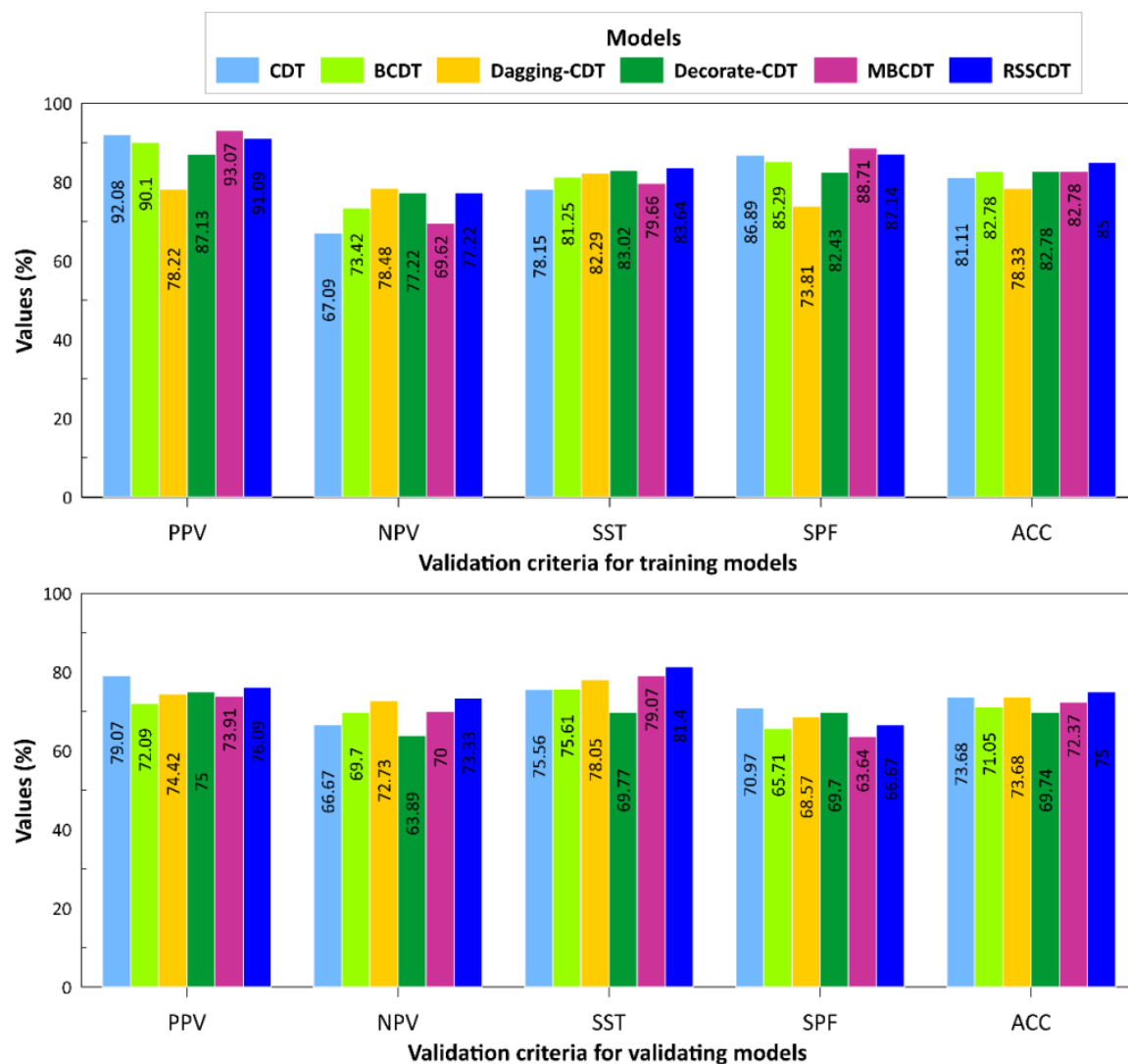


Figure 5. Performance of the models using Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity (SST), Specificity (SPF) and Accuracy (ACC).

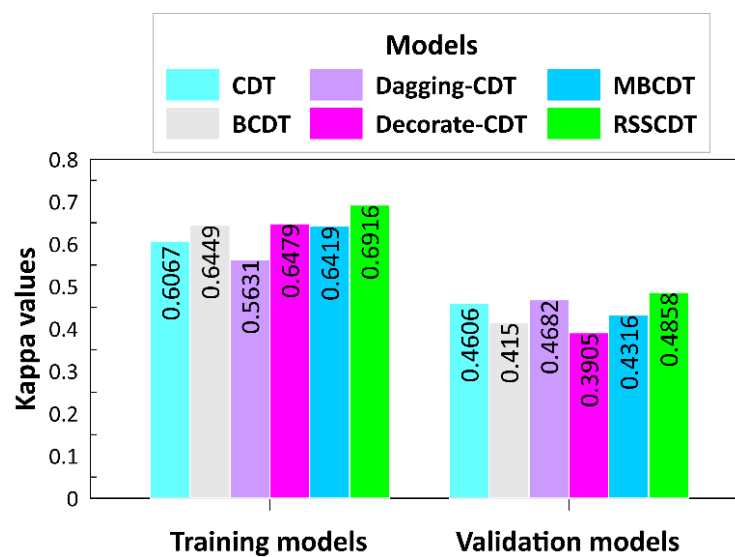


Figure 6. Performance of the models using Kappa criteria.

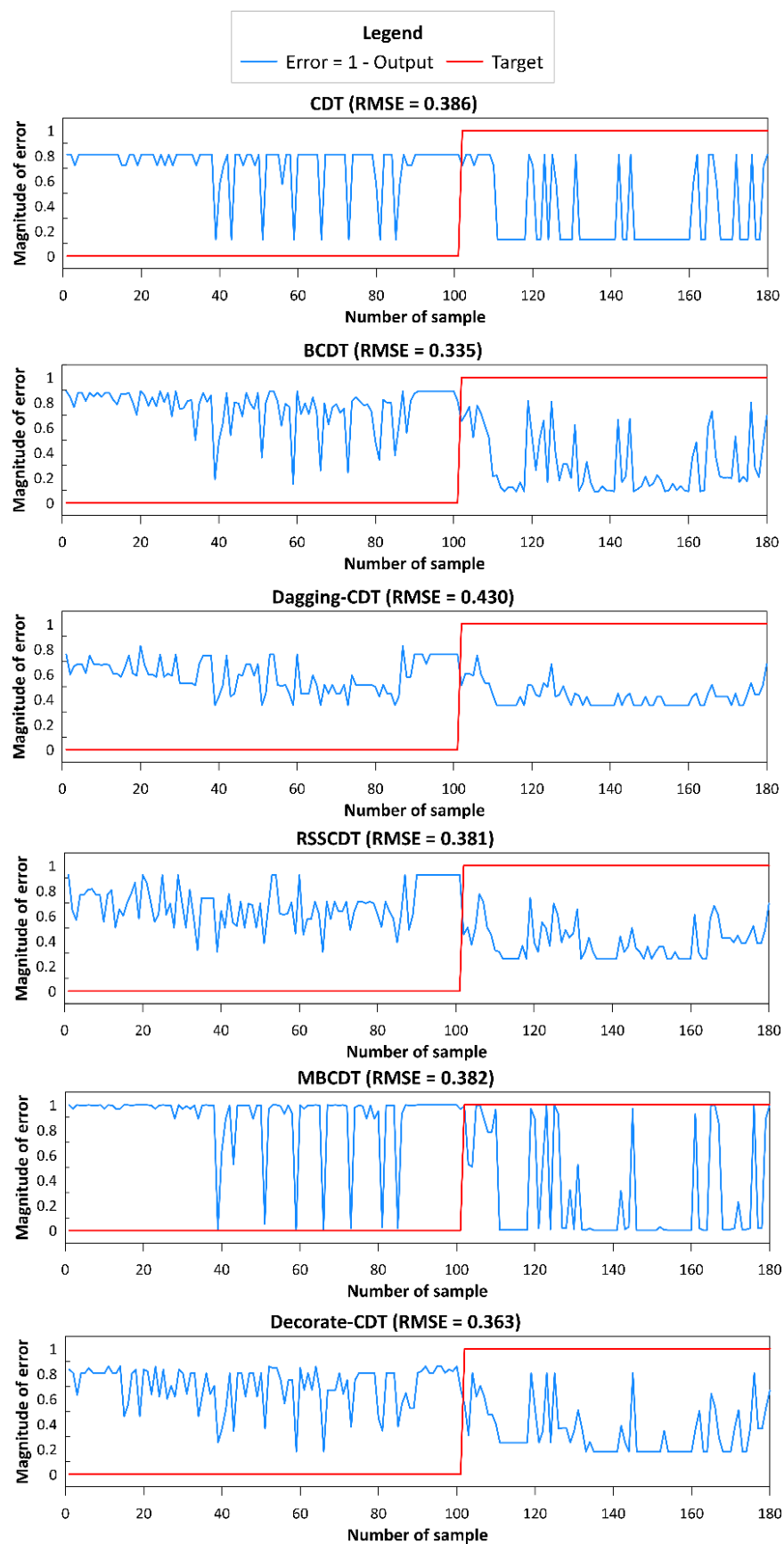


Figure 7. Root Mean Squared Error (RMSE) analysis of the models using training dataset.

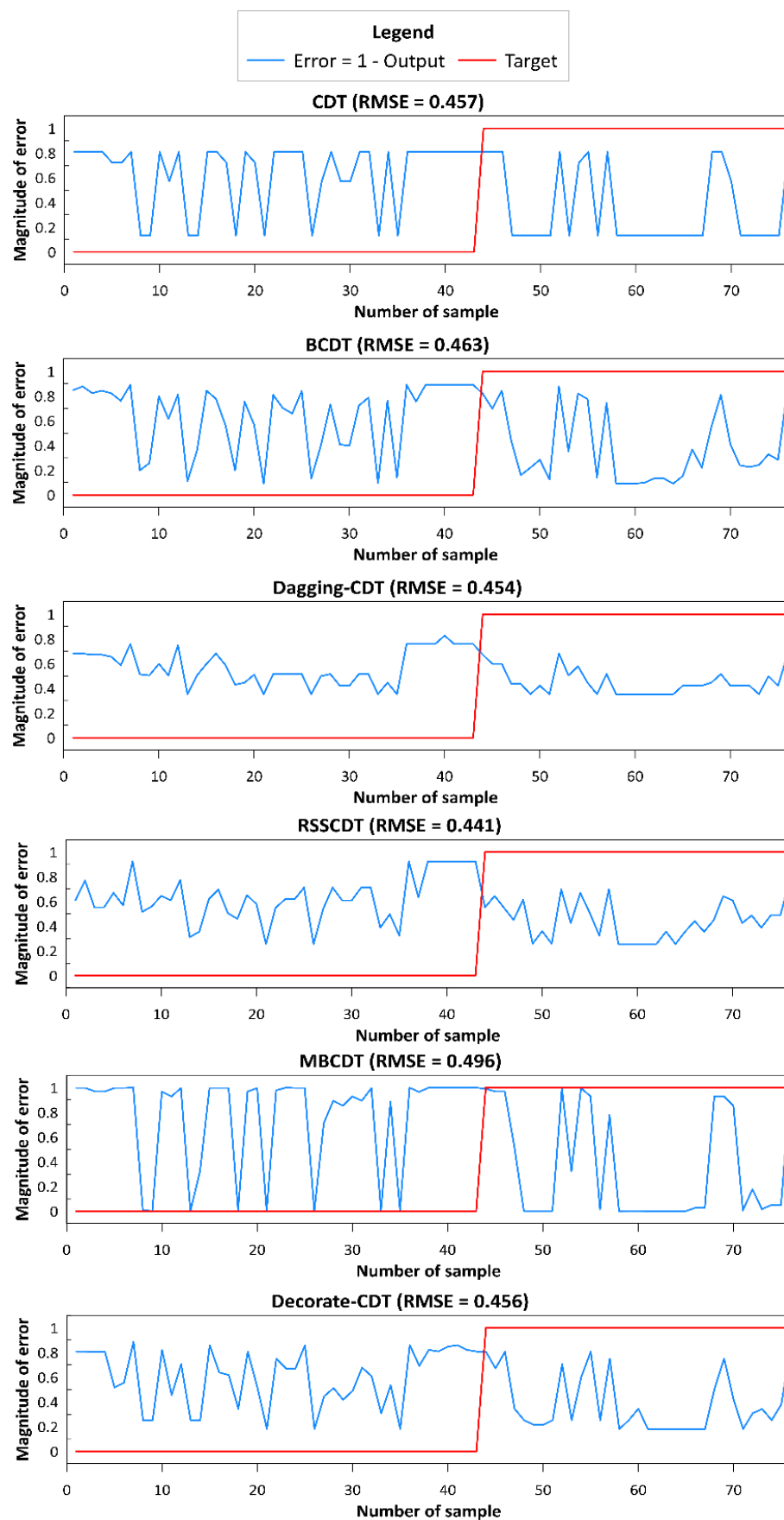


Figure 8. RMSE analysis of the models using testing data set.

Comparative analysis of models' performance using AUC values indicated that the BCDT model is better with AUC: 0.933, followed by the RSSCDT model (0.909), Decorate-CDT model (0.901), MBCDT (0.899), Dagging-CDT (0.856) and CDT (0.819), respectively, in terms of training data (Figure 9). In terms of validation data, the MBCDT model showed better predictive performance with AUC: 0.77, followed by RSSCDT (0.766), Dagging-CDT (0.763), Decorate-CDT (0.75), BCDT (0.731), and CDT (0.722), respectively. In general, the results of the model study show that all the models have AUC > 0.7, thus they are all efficient in building the groundwater potential maps.

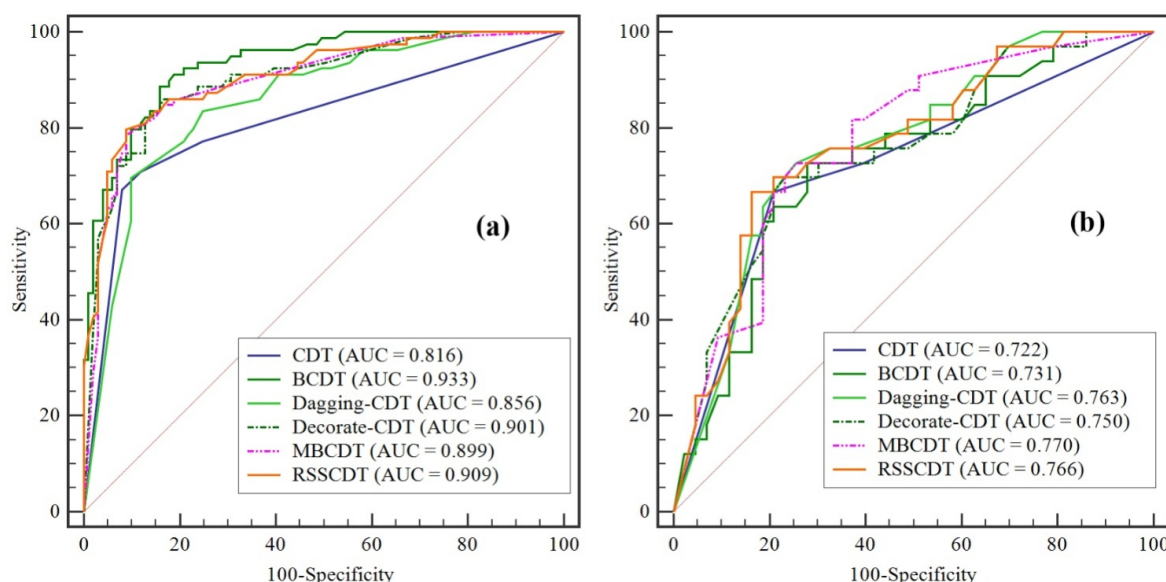


Figure 9. Evaluation of the models using the Receiver Operating Characteristic (ROC) curve: (a) training dataset; (b) validation dataset.

6.3. Evaluation and Validation of Groundwater Potential Maps

In the present study, groundwater potential maps were developed using six models: CDT, BCDT, Dagging-CDT, Decorate-CDT, MBCDT, and RSSCDT. These maps were constructed in five groups (very low, low, moderate, high and very high) of groundwater potential zones (Figure 10). Analysis of groundwater potential maps suggested that in case of CDT model; about 80% of the area is located in very low, 5% in low, and 15% in very high potential zones. For the BCDT model, 50% of the area is in very low, 20% in low, 10% in moderate, 7% in high and 13% in the very high potential zones. In the case of the Dagging-CDT model, about 35% is located in very low, 25% in low, 10% in moderate, 7% in high and 13% in the very high zones. For the Decorate-CDT model, 35% of the area is located in very low, 26% in low, 20% in moderate, 8% in high and 11% in the very high zone. In case of MBCDT model, 70% of the study area is located in very low, 7% in low zones, 3% in moderate, 2% in high and 13% in very high zones. Groundwater potential map in case of RSSCDT model showed that 10% of the area is in very low zones, 40% in low zones, 20% in moderate zones, 25% in the high zone and 15% in the very high zone (Figure 11). All the generated maps showed that high to very high groundwater potential areas are located in the central part of the study area. Thus, these groundwater potential maps can be used as scientific documents to assist decision-makers in land use planning and water resource management.

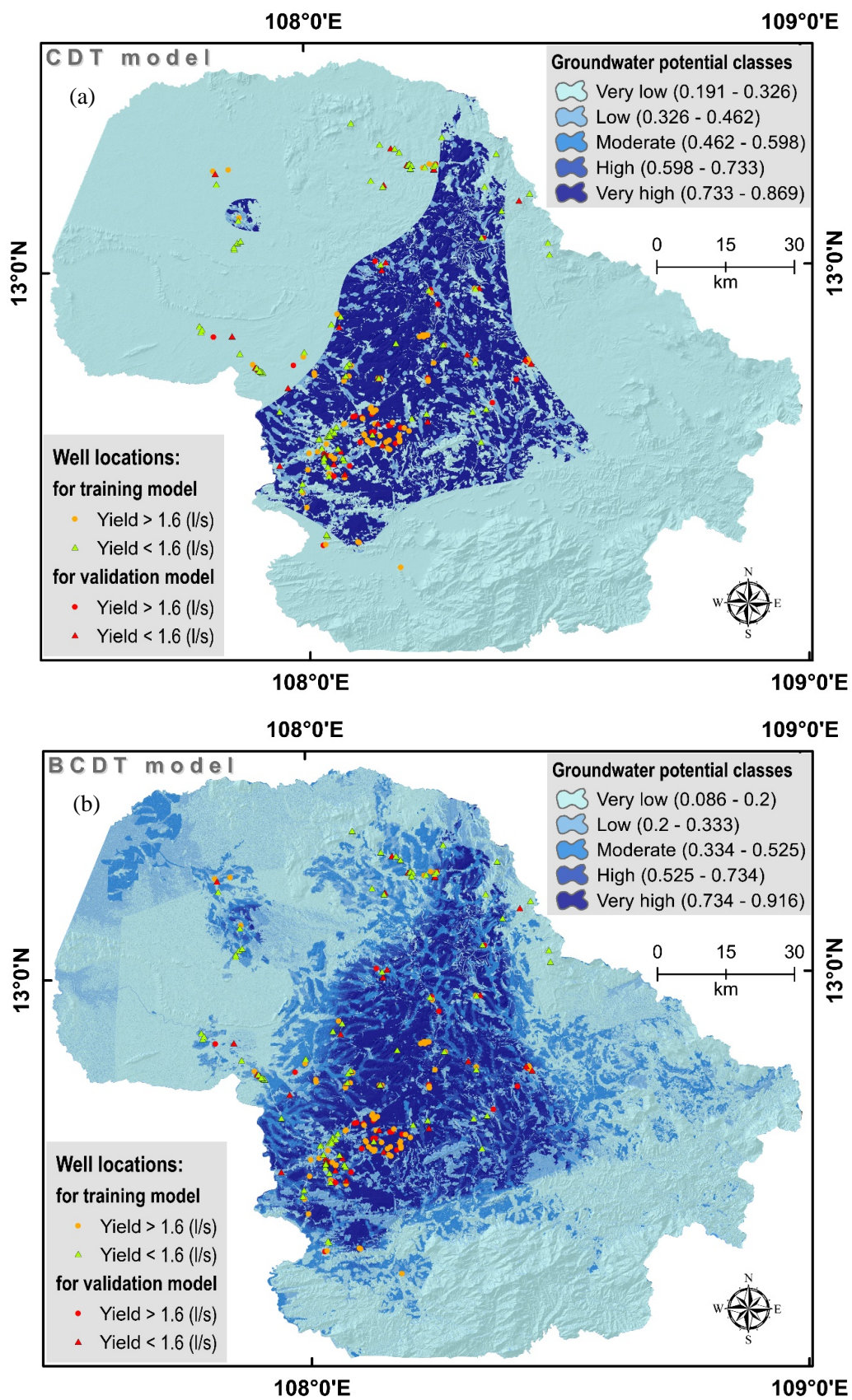


Figure 10. Cont.

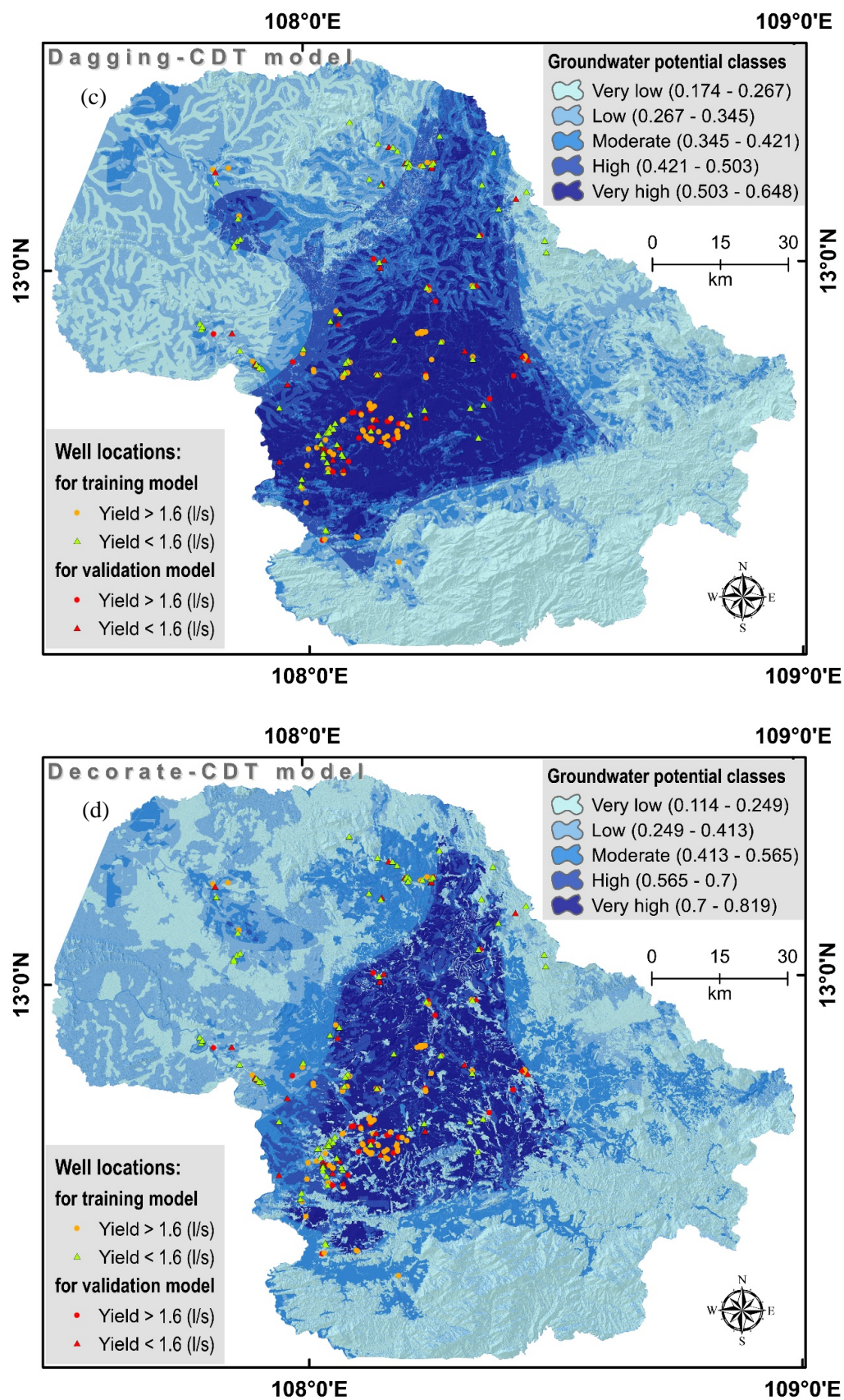


Figure 10. Cont.

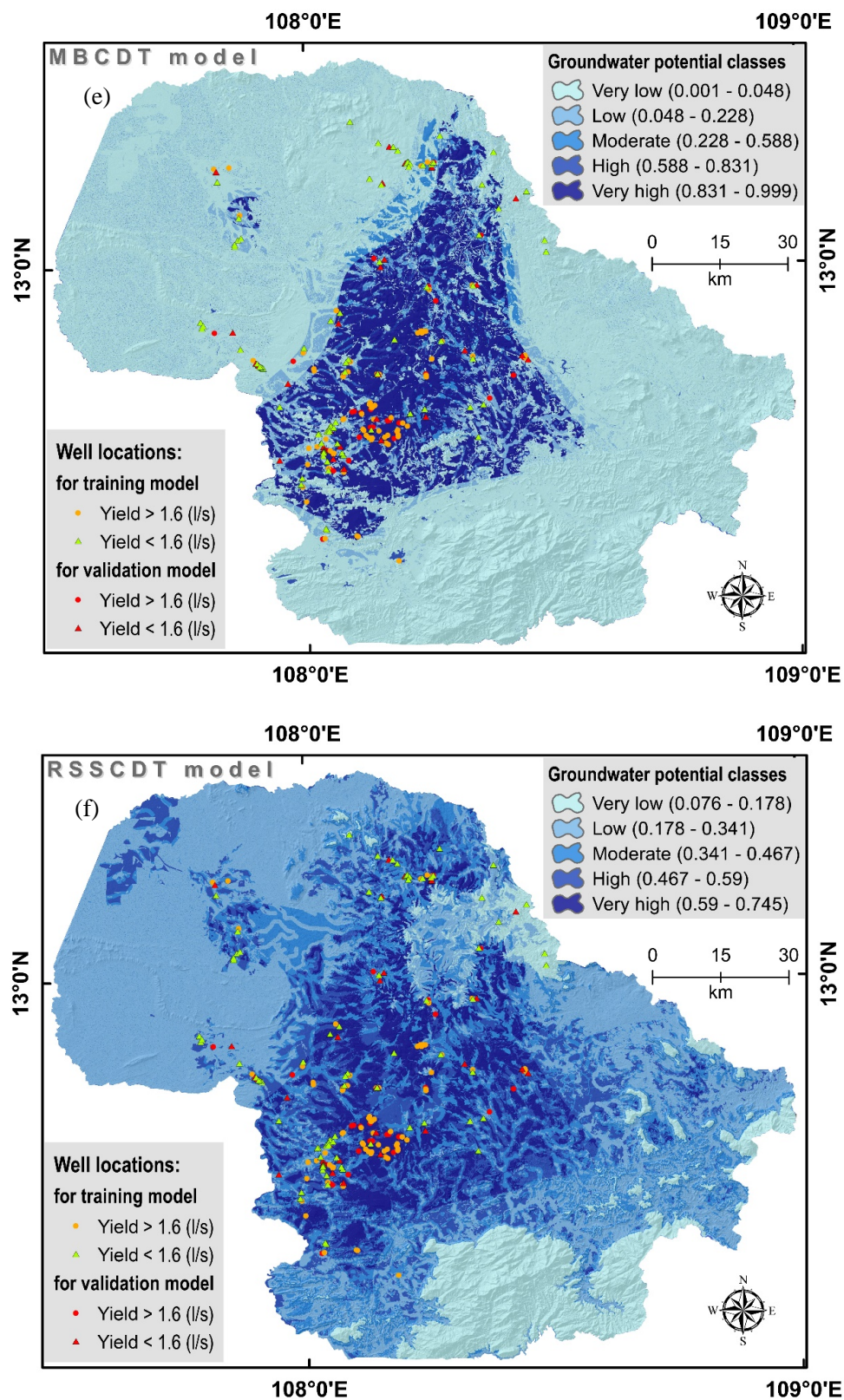


Figure 10. Groundwater potential maps using various models: (a) CDT, (b) BCDT, (c) Dagging-CDT, (d) Decorate-CDT, (e) MBCDT, and (f) RSSCDT.

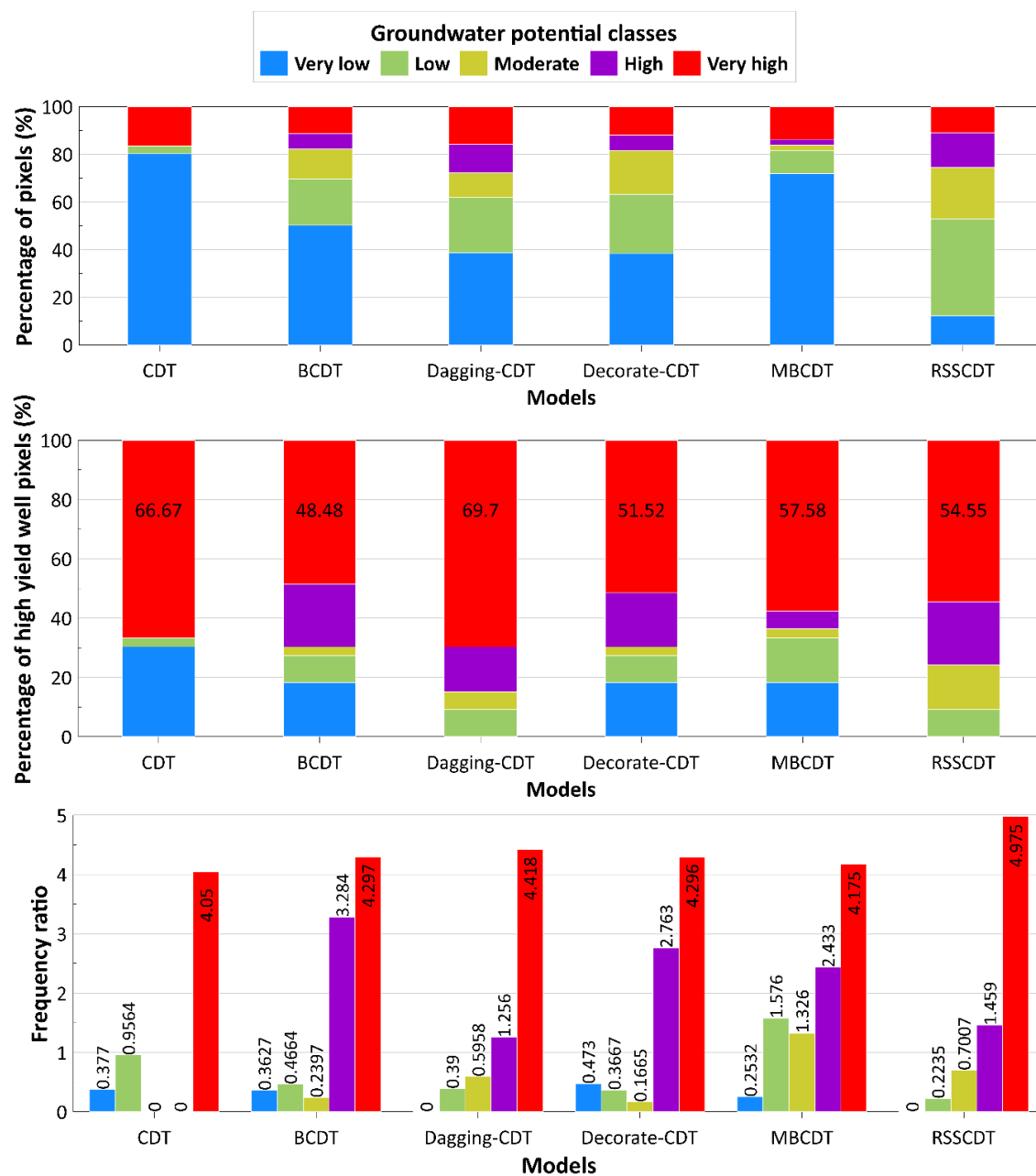


Figure 11. Validation of groundwater potential maps.

7. Discussion

Groundwater resources are an important source for potable water, which is also used for agriculture and industry [100–103]. The mapping of groundwater potential is an essential task to assess groundwater potential of the area for better groundwater resource management. Even though many studies have been carried out to map the groundwater potential in various regions of the world using different approaches [54,104], but more efforts are needed to improve the quality of these maps for predicting accurate groundwater potential zones [16]. Nowadays, advanced ML techniques are being used for this purpose [25,105]. In this study, different ensemble ML techniques, namely Bagging, Dagging, MultiBoost, Random SubSpace, and Decorate, were used to improve the performance of a single ML model, namely CDT, to develop various hybrid models (BCDT, Dagging-CDT, Decorate-CDT, MBCDT and RSSCDT) for the improvement of the performance of groundwater potential mapping in the DakLak province, Vietnam.

Based on the results of model validation, it can be stated that the proposed ensemble frameworks improved the performance of a single the CDT base classifier model for better groundwater potential mapping. This may be due to the fact that in CDT algorithm, the sub-dataset formed is different from a given problem domain which produces quite different trees [106,107]. This feature is very necessary to build the appropriate classification to increase the classification capacity of Random SubSpace, Bagging and Multiboost models [106–108]. Bagging is considered to be an important algorithm for improving the accuracy of individual classification prediction by creating different classifications together. In the present study, Bagging used the Radial Basis Function (RBF) kernel function to improve the stability capacity of CDT model. In addition, in Bagging algorithm, the Bootstrap sampling method was used to decrease the sensitivity of an individual classification for noise problem in training data [33]. In the Bagging model, the base classification generation errors are moved to the generation errors, which are calculated on the smaller training data and this model is useful for low classification [62,109]. The Dagging method has the advantage of reducing noise. Although Decorate is not known as the Bagging or Multiboost algorithm, it is the efficient algorithm as it enhances the original training data by creating artificial data and then producing various classifications on artificial samples. Therefore, this algorithm is presenting an advantage for small scale training datasets [41]. Literature survey indicated that Multiboost can reduce the average errors in terms of bias. In this method, the original training dataset is divided into several sub-datasets, which can be treated at the same time [42,43,57]. The findings of this study are also in line with the other studies [106,107].

In the present study, various validation criteria, namely *SST*, *SPF*, *ACC*, *K*, *PPV*, *NPV*, *RMSE* and *AUC*, were selected and used for validation and comparison of the models. It can be seen that the comparative performance of the models is different with different statistical criteria. For example, RSSCDT is better than other models in term of *NPV*, *SST*, *ACC* and Kappa (Figures 5 and 6), but MBCDT is better than other models in term of *AUC* (Figure 9). Thus, in this study, it can be stated that ensemble frameworks improved the performance of the single CDT base classifier but it is very difficult to assess which ensemble method is the best from the applied validation criteria.

8. Conclusions

In this study, various ensemble techniques, namely Bagging, Dagging, Decorate, MultiBoost, and Random SubSpace, were used to the improve performance of a single CDT base classifier for the generation of accurate groundwater potential maps. The performance of five developed hybrid models, namely BCDT, Dagging-CDT, Decorate-CDT, MBCDT, and RSSCDT, was evaluated and compared with the single CDT model.

Validation results show that although all the models are efficient in groundwater potential mapping in the study area ($AUC > 0.70$), the performance of the ensemble models MBCDT ($AUC = 0.770$), BCDT ($AUC = 0.731$), Dagging-CDT ($AUC = 0.763$), Decorate-CDT ($AUC = 0.750$), and RSSCDT ($AUC = 0.766$) improved significantly in comparison to single CDT model ($AUC = 0.722$). Thus, these developed hybrid models can be applied for better ground water resources management of the study area as well as other regions of the world.

Groundwater potential zones identified through mapping using developed hybrid (ensemble) models would help managers in prioritizing the area for future development of groundwater resources and their systematic exploitation, considering annual needs and recharge of the area by maintaining water balance. All the stakeholders, including government and non-government agencies and individuals, can use these maps for the sustainable development of the area. Based on these maps, local inhabitants can also be provided with technical help and monetary support in drought affected areas for the construction and maintenance of recharge structures at suitable locations.

The results of this study would be helpful not only in the proper management of the DakLak province of the Vietnam but also for the ground water potential mapping and assessment of other drought prone areas of the world.

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., and N.A.-A.; validation, B.T.P., P.T.N., N.A.-A., P.T.T., and I.P.; formal analysis, P.T.N., P.T.T., 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., N.A.-A., and I.P.; visualization, T.V.P., L.S.H., H.V.L., and D.H.H.; supervision, P.T.N., B.T.P., P.T.T., 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 to Vietnam Academy for Water Resources for providing the data to carry out this research.

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

References

1. Arabameri, A.; Rezaei, K.; Cerda, A.; Lombardo, L.; Rodrigo-Comino, J. GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *Sci. Environ.* **2019**, *658*, 160–177. [[CrossRef](#)] [[PubMed](#)]
2. Grönwall, J.; Oduro-Kwarteng, S. Groundwater as a strategic resource for improved resilience: A case study from peri-urban Accra. *Environ. Earth Sci.* **2017**, *77*, 6. [[CrossRef](#)]
3. Nkhuwa, D.; Abiye, T.; Oga, M.; Adelana, S.; Tindimugaya, C. Urban groundwater management and protection in Sub-Saharan Africa. In *IAH-Selected Papers on Hydrogeology*; Informa UK Limited: Colchester, UK, 2008; Volume 6152, pp. 1–7.
4. World Water Assessment Programme. *The United Nations World Water Development Report 4: Managing Water under Uncertainty and Risk*; UNESCO: Paris, France, 2012.
5. Aubriot, O. Baisse des nappes d’eau souterraine en Inde du Sud Forte demande sociale et absence de gestion de la ressource. *Geocarrefour* **2006**, *81*, 83–90. [[CrossRef](#)]
6. Amarasinghe, U.A.; Smakhtin, V. *Global Water Demand Projections: Past, Present and Future*; IWMI: Colombo, Sri Lanka, 2014.
7. Boretti, A.; Rosa, L. Reassessing the projections of the World Water Development Report. *NPJ Clean Water* **2019**, *2*, 15. [[CrossRef](#)]
8. Okello, C.; Tomasello, B.; Greggio, N.; Wambiji, N.; Antonellini, M. Impact of Population Growth and Climate Change on the Freshwater Resources of Lamu Island, Kenya. *Water* **2015**, *7*, 1264–1290. [[CrossRef](#)]
9. Carter, R.C.; Parker, A. Climate change, population trends and groundwater in Africa. *Hydrol. Sci. J.* **2009**, *54*, 676–689. [[CrossRef](#)]
10. Garcia-Franco, N.; Hobbey, E.; Hübner, R.; Wiesmeier, M. Chapter 23—Climate-Smart Soil Management in Semiarid Regions. In *Soil Management and Climate Change*; Muñoz, M.Á., Zornoza, R., Eds.; Academic Press: Waltham, MA, USA, 2018; pp. 349–368. [[CrossRef](#)]
11. Misra, A.K. Climate change and challenges of water and food security. *Int. J. Sustain. Built Environ.* **2014**, *3*, 153–165. [[CrossRef](#)]
12. Oikonomidis, D.; Dimogianni, S.; Kazakis, N.; Voudouris, K. A GIS/Remote Sensing-based methodology for groundwater potentiality assessment in Tirnavos area, Greece. *J. Hydrol.* **2015**, *525*, 197–208. [[CrossRef](#)]
13. Rahmati, O.; Samani, A.N.; Mahdavi, M.; Pourghasemi, H.R.; Zeinivand, H. Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arab. J. Geosci.* **2014**, *8*, 7059–7071. [[CrossRef](#)]
14. Zabihi, M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Behzadfar, M. GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environ. Earth Sci.* **2016**, *75*, 665. [[CrossRef](#)]
15. Al-Nahmi, F.; Saddiqi, O.; Rhinane, H.; Baiddar, L.; El Arabi, H.; Khanbari, K.; Hilali, A. Application of Remote Sensing in Geological Mapping, Case Study Al Maghrabah Area—Hajjah Region, Yemens. *ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2017**, *4*, 63. [[CrossRef](#)]
16. Khosravi, K.; Sartaj, M.; Tsai, F.T.-C.; Singh, V.P.; Kazakis, N.; Melesse, A.; Prakash, I.; Bui, D.T.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](#)]

17. Mehra, M.; Oinam, B.; Singh, C.K. Integrated Assessment of Groundwater for Agricultural Use in Mewat District of Haryana, India Using Geographical Information System (GIS). *J. Indian Soc. Remote. Sens.* **2016**, *44*, 747–758. [\[CrossRef\]](#)
18. Marsland, S. *Machine learning: An Algorithmic Perspective*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2014.
19. Chen, W.; Li, Y.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Xue, W.; Bian, H. Groundwater Spring Potential Mapping Using Artificial Intelligence Approach Based on Kernel Logistic Regression, Random Forest, and Alternating Decision Tree Models. *Appl. Sci.* **2020**, *10*, 425. [\[CrossRef\]](#)
20. 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.* **2015**, *188*, 44. [\[CrossRef\]](#)
21. 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.* **2017**, *33*, 847–861. [\[CrossRef\]](#)
22. 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. [\[CrossRef\]](#)
23. Ozdemir, A. GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *J. Hydrol.* **2011**, *411*, 290–308. [\[CrossRef\]](#)
24. Khosravi, K.; Panahi, M.; Khosravi, K.; Chen, W.; Rezaie, F.; Parvinnezhad, D. Spatial prediction of groundwater potentiality using ANFIS ensembled with teaching-learning-based and biogeography-based optimization. *J. Hydrol.* **2019**, *572*, 435–448.
25. 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.* **2018**, *33*, 281–302. [\[CrossRef\]](#)
26. Naghibi, S.A.; Pourghasemi, H.R.; Abbaspour, K. A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theor. Appl. Clim.* **2017**, *131*, 967–984. [\[CrossRef\]](#)
27. Abellán, J.; Moral, S. Building classification trees using the total uncertainty criterion. *Int. J. Intell. Syst.* **2003**, *18*, 1215–1225. [\[CrossRef\]](#)
28. Mantas, C.J.; Abellán, J. Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. *Expert Syst. Appl.* **2014**, *41*, 4625–4637. [\[CrossRef\]](#)
29. García-Tejero, I.F.; Durán Zuazo, V.; Muriel, J.; Rodriguez, C. *Water and Sustainable Agriculture*; Springer: Dordrecht, The Netherlands, 2011; Volume 1, pp. 1–94.
30. Kumar, M.D.; Sivamohan, M.V.K.; Bassi, N. *Water Management, Food Security and Sustainable Agriculture in Developing Economies*; Routledge: Abingdon, UK, 2013.
31. Abellán, J.; Masegosa, A.R. *Combining Decision Trees Based on Imprecise Probabilities and Uncertainty Measures*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4724, pp. 512–523.
32. Moral-García, S.; Mantas, C.J.; Castellano, J.G.; Benítez, M.D.; Abellán, J. Bagging of credal decision trees for imprecise classification. *Expert Syst. Appl.* **2020**, *141*, 112944. [\[CrossRef\]](#)
33. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
34. Quinlan, J.R. *Bagging, Boosting, and C4. 5*; University of Sydney: Sydney, Australia, 2006; Volume 1, pp. 725–730.
35. Dietterich, T.G. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Mach. Learn.* **2000**, *40*, 139–157. [\[CrossRef\]](#)
36. Hong, H.; Liu, J.; Zhu, A.-X. Landslide susceptibility evaluating using artificial intelligence method in the Youfang district (China). *Environ. Earth Sci.* **2019**, *78*, 488. [\[CrossRef\]](#)
37. Ting, K.; Witten, I. *Stacking Bagged and Dagged Models*; The University of Waikato: Hamilton, New Zealand, 1997.
38. Kotsianti, S.B.; Kanellopoulos, D. Combining Bagging, Boosting and Dagging for Classification Problems. *Comput. Vis.* **2007**, *4693*, 493–500.

39. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Tran, T.-T.-T.; Bui, D.T.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [\[CrossRef\]](#)
40. Melville, P.; Mooney, R.J. Creating diversity in ensembles using artificial data. *Inf. Fusion* **2005**, *6*, 99–111. [\[CrossRef\]](#)
41. Sun, B.; Chen, H.; Wang, J. An empirical margin explanation for the effectiveness of DECORATE ensemble learning algorithm. *Knowl. Based Syst.* **2015**, *78*, 1–12. [\[CrossRef\]](#)
42. Webb, G.I. MultiBoosting: A Technique for Combining Boosting and Wagging. *Mach. Learn.* **2000**, *40*, 159–196. [\[CrossRef\]](#)
43. Pham, B.T.; Bui, D.T.T.; Prakash, I.; Dholakia, M. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **2017**, *149*, 52–63. [\[CrossRef\]](#)
44. Kotti, M.; Benetos, E.; Kotropoulos, C.; Pitas, I. A neural network approach to audio-assisted movie dialogue detection. *Neurocomputing* **2007**, *71*, 157–166. [\[CrossRef\]](#)
45. Bui, D.T.T.; Ho, T.-C.; Pradhan, B.; Pham, B.T.; Nhu, V.-H.; Revhaug, I. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environ. Earth Sci.* **2016**, *75*, 1101.
46. Wang, X.; Tang, X. Random Sampling for Subspace Face Recognition. *Int. J. Comput. Vis.* **2006**, *70*, 91–104. [\[CrossRef\]](#)
47. Skurichina, M.; Duin, R.P.W. Bagging, Boosting and the Random Subspace Method for Linear Classifiers. *Pattern Anal. Appl.* **2002**, *5*, 121–135. [\[CrossRef\]](#)
48. Pham, B.T.; Jaafari, A.; Prakash, I.; Singh, S.K.; Quoc, N.K.; Bui, D.T.T. Hybrid computational intelligence models for groundwater potential mapping. *Catena* **2019**, *182*, 104101. [\[CrossRef\]](#)
49. Hoa, P.V.; Giang, N.V.; Binh, N.A.; Hai, L.V.H.; Pham, T.D.; Hasanlou, M.; Bui, D.T.T. Soil Salinity Mapping Using SAR Sentinel-1 Data and Advanced Machine Learning Algorithms: A Case Study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote. Sens.* **2019**, *11*, 128. [\[CrossRef\]](#)
50. Bui, Q.T.; Nguyen, Q.-H.; Nguyen, X.L.; Pham, V.D.; Nguyen, H.D.; Pham, V.-M. Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *J. Hydrol.* **2020**, *581*, 124379. [\[CrossRef\]](#)
51. Gnanambal, S.; Thangaraj, M.; Meenatchi, V.; Gayathri, V. Classification Algorithms with Attribute Selection: An evaluation study using WEKA. *Int. J. Adv. Netw. Appl.* **2018**, *9*, 3640–3644.
52. Kordestani, M.D.; Naghibi, S.A.; Hashemi, H.; Ahmadi, K.; Kalantar, B.; Pradhan, B. Groundwater potential mapping using a novel data-mining ensemble model. *Hydrogeol. J.* **2018**, *27*, 211–224. [\[CrossRef\]](#)
53. Kalantar, B.; Al-Najjar, H.A.H.; Pradhan, B.; Saeidi, V.; Halin, A.A.; Ueda, N.; Naghibi, S.A. Optimized Conditioning Factors Using Machine Learning Techniques for Groundwater Potential Mapping. *Water* **2019**, *11*, 1909. [\[CrossRef\]](#)
54. Termeh, S.V.R.; Khosravi, K.; Sartaj, M.; Keesstra, S.D.; Tsai, F.T.-C.; Dijkstra, R.; Pham, B.T. Optimization of an adaptive neuro-fuzzy inference system for groundwater potential mapping. *Hydrogeol. J.* **2019**, *27*, 2511–2534. [\[CrossRef\]](#)
55. Van Dao, D.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayedi, H.; Van Phong, T.; Ly, H.-B.; Le, T.-T.; Trinh, P.T. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* **2020**, *188*, 104451.
56. Lei, T.; Zhang, Y.; Lv, Z.; Li, S.; Liu, S.; Nandi, A.K. Landslide Inventory Mapping From Bitemporal Images Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote. Sens. Lett.* **2019**, *16*, 1–5. [\[CrossRef\]](#)
57. 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. Int. Assoc. Eng. Geol.* **2018**, *78*, 2865–2886. [\[CrossRef\]](#)
58. Pham, B.T.; Pradhan, B.; Bui, D.T.T.; Prakash, I.; Dholakia, M. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [\[CrossRef\]](#)
59. Janizadeh, S.; Avand, M.; Jaafari, A.; Van Phong, T.; 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. [\[CrossRef\]](#)

60. Nguyen, V.-T.; Tran, T.; Ha, N.; Ngo, V.L.; Al-Ansari, N.; Van Phong, T.; Nguyen, D.H.; Malek, M.; Amini, A.; Prakash, I.; et al. GIS Based Novel Hybrid Computational Intelligence Models for Mapping Landslide Susceptibility: A Case Study at Da Lat City, Vietnam. *Sustainability* **2019**, *11*, 7118. [\[CrossRef\]](#)
61. Pham, B.T.; Bui, D.T.T.; Dholakia, M.B.; Prakash, I.; Pham, H.V. A Comparative Study of Least Square Support Vector Machines and Multiclass Alternating Decision Trees for Spatial Prediction of Rainfall-Induced Landslides in a Tropical Cyclones Area. *Geotech. Geol. Eng.* **2016**, *34*, 1807–1824. [\[CrossRef\]](#)
62. Pham, B.T.; Shirzadi, A.; Shahabi, H.; Omidvar, E.; Singh, S.K.; Sahana, M.; Asl, D.T.; Bin Ahmad, B.; Quoc, N.K.; Lee, S.; et al. Landslide Susceptibility Assessment by Novel Hybrid Machine Learning Algorithms. *Sustainability* **2019**, *11*, 4386. [\[CrossRef\]](#)
63. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote. Sens.* **2019**, *11*, 196. [\[CrossRef\]](#)
64. Asteris, P.G.; Ashrafi, A.; Rezaie-Balf, M. Prediction of the compressive strength of self-compacting concrete using surrogate models. *Comput. Concr.* **2019**, *24*, 137–150.
65. Asteris, P.; Nozhati, S.; Nikoo, M.; Cavaleri, L.; Nikoo, M. Krill herd algorithm-based neural network in structural seismic reliability evaluation. *Mech. Adv. Mater. Struct.* **2018**, *26*, 1146–1153. [\[CrossRef\]](#)
66. Lemonis, M.E.; Asteris, P.G.; Zitouniatis, D.G.; Ntasis, G.D. Modeling of the lateral stiffness of masonry infilled steel moment-resisting frames. *Struct. Eng. Mech.* **2019**, *70*, 421–429.
67. Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Bui, D.T.T.; Pham, B.T.; Khosravi, K. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ. Model. Softw.* **2017**, *95*, 229–245. [\[CrossRef\]](#)
68. Pham, B.T.; Bui, D.T.T.; Prakash, I.; Dholakia, M.B. Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. *Nat. Hazards* **2016**, *83*, 97–127. [\[CrossRef\]](#)
69. Aladejana, O.; Anifowose, A.Y.B.; Fagbohun, B.J. Testing the ability of an empirical hydrological model to verify a knowledge-based groundwater potential zone mapping methodology. *Model. Earth Syst. Environ.* **2016**, *2*, 1–17. [\[CrossRef\]](#)
70. Andualem, T.G.; Demeke, G.G. Groundwater potential assessment using GIS and remote sensing: A case study of Guna tana landscape, upper blue Nile Basin, Ethiopia. *J. Hydrol. Reg. Stud.* **2019**, *24*, 100610. [\[CrossRef\]](#)
71. Schillaci, C.; Braun, A.; Kropáček, J. 2.4. 2. Terrain analysis and landform recognition. *Geomorphol. Tech.* **2015**, *2*, 1–18.
72. Solomon, S.; Quiel, F. Groundwater study using remote sensing and geographic information systems (GIS) in the central highlands of Eritrea. *Hydrogeol. J.* **2006**, *14*, 1029–1041. [\[CrossRef\]](#)
73. Moghaddam, D.D.; Rahmati, O.; Haghizadeh, A.; Kalantari, Z. A Modeling Comparison of Groundwater Potential Mapping in a Mountain Bedrock Aquifer: QUEST, GARP, and RF Models. *Water* **2020**, *12*, 679. [\[CrossRef\]](#)
74. Chen, W.; Li, H.; Hou, E.; Wang, S.; Wang, G.; Panahi, M.; Li, T.; Peng, T.; Guo, C.; Niu, C.; et al. GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci. Total. Environ.* **2018**, *634*, 853–867. [\[CrossRef\]](#)
75. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831. [\[CrossRef\]](#)
76. Ercanoglu, C.G.M.; Ercanoglu, M.; Gokceoglu, C. Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. *Environ. Earth Sci.* **2002**, *41*, 720–730.
77. Botzen, W.J.W.; Aerts, J.C.J.H.; Bergh, J.C.V.D. Individual preferences for reducing flood risk to near zero through elevation. *Mitig. Adapt. Strat. Glob. Chang.* **2012**, *18*, 229–244. [\[CrossRef\]](#)
78. Conforti, M.; Aucelli, P.P.C.; Robustelli, G.; Scarciglia, F. Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). *Nat. Hazards* **2010**, *56*, 881–898. [\[CrossRef\]](#)
79. Yilmaz, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). *Comput. Geosci.* **2009**, *35*, 1125–1138. [\[CrossRef\]](#)
80. Moore, I.D.; Grayson, R.B.; Ladson, A. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30. [\[CrossRef\]](#)

81. Rizeei, H.M.; Pradhan, B.; Saharkhiz, M.A. Surface runoff prediction regarding LULC and climate dynamics using coupled LTM, optimized ARIMA, and GIS-based SCS-CN models in tropical region. *Arab. J. Geosci.* **2018**, *11*, 53. [\[CrossRef\]](#)
82. Rizeei, H.M.; Pradhan, B.; Saharkhiz, M.A.; Lee, S. Groundwater aquifer potential modeling using an ensemble multi-adoptive boosting logistic regression technique. *J. Hydrol.* **2019**, *579*, 124172. [\[CrossRef\]](#)
83. Magesh, N.; Chandrasekar, N.; Soundranayagam, J.P. Delineation of groundwater potential zones in Theni district, Tamil Nadu, using remote sensing, GIS and MIF techniques. *Geosci. Front.* **2012**, *3*, 189–196. [\[CrossRef\]](#)
84. 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. [\[CrossRef\]](#)
85. Huy, T.D.; Thanh, T.N.; Van Lam, N.; Van Hoang, N. Inverse analysis for transmissivity and the Red river bed's leakage factor for Pleistocene aquifer in Sen Chieu, Hanoi by pumping test under the river water level fluctuation. *Vietnam J. Earth Sci.* **2017**, *40*, 26–38. [\[CrossRef\]](#)
86. 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**, *45*, 5–2775. [\[CrossRef\]](#)
87. Oanh, T.T.K.; Nguyen, V.L. High Arsenic Consntration in Groundwater related to Sedimentary Facies in the Mekong River Delta, Vietnam. *Vietnam J. Earth Sci.* **2016**, *38*, 178–187.
88. Thai, T.H.; Thao, N.P.; Dieu, B.T. Assessment and Simulation of Impacts of Climate Change on Erosion and Water Flow by Using the Soil and Water Assessment Tool and GIS: Case Study in Upper Cau River basin in Vietnam. *Vietnam J. Earth Sci.* **2017**, *39*, 376–392. [\[CrossRef\]](#)
89. Lerner, D.; Harris, B. The relationship between land use and groundwater resources and quality. *Land Use Policy* **2009**, *26*, S265–S273. [\[CrossRef\]](#)
90. Fashae, O.A.; Tijani, M.; Talabi, A.O.; Adedeji, O.I. Delineation of groundwater potential zones in the crystalline basement terrain of SW-Nigeria: An integrated GIS and remote sensing approach. *Appl. Water Sci.* **2013**, *4*, 19–38. [\[CrossRef\]](#)
91. Pinto, D.; Shrestha, S.; Babel, M.S.; Ninsawat, S. Delineation of groundwater potential zones in the Comoro watershed, Timor Leste using GIS, remote sensing and analytic hierarchy process (AHP) technique. *Appl. Water Sci.* **2015**, *7*, 503–519. [\[CrossRef\]](#)
92. Bui, Q.T.; Nguyen, Q.-H.; Pham, V.D.; Pham, M.H.; Tran, A.T.; Van Manh, P.; Hai, P.M.; Tuấn, T.A. Understanding spatial variations of malaria in Vietnam using remotely sensed data integrated into GIS and machine learning classifiers. *Geocarto Int.* **2018**, *34*, 1300–1314. [\[CrossRef\]](#)
93. 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. [\[CrossRef\]](#)
94. Dou, J.; Bui, D.T.T.; Yunus, A.P.; Jia, K.; Song, X.; Revhaug, I.; Xia, H.; Zhu, Z. Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan. *PLoS ONE* **2015**, *10*, e0133262. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Chen, W.; Pradhan, B.; Li, S.; Shahabi, H.; Rizeei, H.M.; Hou, E.; Wang, S. Novel Hybrid Integration Approach of Bagging-Based Fisher's Linear Discriminant Function for Groundwater Potential Analysis. *Nat. Resour. Res.* **2019**, *28*, 1239–1258. [\[CrossRef\]](#)
96. Pham, B.T.; Nguyen, M.D.; Bui, K.-T.T.; Prakash, I.; Chapi, K.; Bui, D.T.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. [\[CrossRef\]](#)
97. 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*, 248. [\[CrossRef\]](#)
98. Van Phong, T.; Phan, T.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Chapi, K.; Ly, H.-B.; Ho, L.; Quoc, N.K.; Pham, B.T.; et al. Landslide susceptibility modeling using different artificial intelligence methods: A case study at Muong Lay district, Vietnam. *Geocarto Int.* **2019**, 1–24. [\[CrossRef\]](#)

99. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Pham, B.T.; Bin Ahmad, B.; Bui, D.T.T. A novel hybrid approach of Bayesian Logistic Regression and its ensembles for landslide susceptibility assessment. *Geocarto Int.* **2018**, *34*, 1427–1457. [\[CrossRef\]](#)
100. Nas, B.; Berktaş, A. Groundwater quality mapping in urban groundwater using GIS. *Environ. Monit. Assess.* **2008**, *160*, 215–227. [\[CrossRef\]](#)
101. Mandel, S. *Groundwater Resources: Investigation and Development*; Elsevier: Rio de Janeiro, Brazil, 2012.
102. Nga, D.V.; Trang, P.T.K.; Duyen, V.T.; Mai, T.T.; Lan, V.T.M.; Viet, P.H.; Postma, D.; Jakobsen, R. Spatial variations of arsenic in groundwater from a transect in the Northwestern Hanoi. *Vietnam J. Earth Sci.* **2017**, *40*, 70–77. [\[CrossRef\]](#)
103. Van Hoang, N.; Hoa, P.T.L.; Tung, L.T. Study on the accuracy of the numerical modeling of the groundwater movement due to spatial and temporal discretization. *Vietnam J. Earth Sci.* **2015**, *36*, 424–431. [\[CrossRef\]](#)
104. Khosravi, K.; Barzegar, R.; Miraki, S.; Adamowski, J.; Daggupati, P.; Alizadeh, M.R.; Pham, B.T.; Alami, M.T. Stochastic Modeling of Groundwater Fluoride Contamination: Introducing Lazy Learners. *Ground Water* **2019**. [\[CrossRef\]](#) [\[PubMed\]](#)
105. Bui, D.T.T.; Shirzadi, A.; Chapi, K.; Shahabi, H.; Pradhan, B.; Pham, B.T.; Singh, V.P.; Chen, W.; Khosravi, K.; Bin Ahmad, B.; et al. A Hybrid Computational Intelligence Approach to Groundwater Spring Potential Mapping. *Water* **2019**, *11*, 2013.
106. Abellán, J.; Masegosa, A.R. An ensemble method using credal decision trees. *Eur. J. Oper. Res.* **2010**, *205*, 218–226. [\[CrossRef\]](#)
107. Abellán, J.; Masegosa, A.R. Bagging schemes on the presence of class noise in classification. *Expert Syst. Appl.* **2012**, *39*, 6827–6837. [\[CrossRef\]](#)
108. He, Q.; Xu, Z.; Li, S.; Li, R.; Zhang, S.; Wang, N.; Pham, B.T.; Chen, W. Novel Entropy and Rotation Forest-Based Credal Decision Tree Classifier for Landslide Susceptibility Modeling. *Entropy* **2019**, *21*, 106. [\[CrossRef\]](#)
109. Trawinski, K.; Cordon, O.; Quirin, A.; Sánchez, L. Multiobjective genetic classifier selection for random oracles fuzzy rule-based classifier ensembles: How beneficial is the additional diversity? *Knowl. Based Syst.* **2013**, *54*, 3–21. [\[CrossRef\]](#)



© 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/>).