

Article

Modeling of Groundwater Potential Using Cloud Computing Platform: A Case Study from Nineveh Plain, Northern Iraq

Ali ZA. Al-Ozeer ^{1,2}, Alaa M. Al-Abadi ² , Tariq Abed Hussain ³, Alan E. Fryar ⁴ , Biswajeet Pradhan ^{5,6,*} , Abdullah Alamri ⁷ and Khairul Nizam Abdul Maulud ^{6,8} 

¹ Department of Geology, College of Science, University of Mosul, Mosul 41001, Iraq; aalozeer@uomosul.edu.iq

² Department of Geology, College of Science, University of Basrah, Basrah 61004, Iraq; alaa.atiaa@uobasrah.edu.iq

³ Department of Civil Engineering, University of Technology, Baghdad 10001, Iraq; 40026@uotechnology.edu.iq

⁴ Department of Earth and Environmental Sciences, University of Kentucky, Lexington, KY 40506-0053, USA; alan.fryar@uky.edu

⁵ Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering & IT, University of Technology, Sydney, NSW 2007, Australia

⁶ Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Selangor, Malaysia; knam@ukm.edu.my

⁷ Department of Geology and Geophysics, College of Science, King Saud University, Riyadh 11362, Saudi Arabia; amsamri@ksu.edu.sa

⁸ Department of Civil Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Selangor, Malaysia

* Correspondence: Biswajeet.Pradhan@uts.edu.au



Citation: Al-Ozeer, A.Z.; Al-Abadi, A.M.; Hussain, T.A.; Fryar, A.E.; Pradhan, B.; Alamri, A.; Abdul Maulud, K.N. Modeling of Groundwater Potential Using Cloud Computing Platform: A Case Study from Nineveh Plain, Northern Iraq. *Water* **2021**, *13*, 3330. <https://doi.org/10.3390/w13233330>

Academic Editor: Roger Falconer

Received: 1 October 2021

Accepted: 16 November 2021

Published: 24 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Knowledge of the groundwater potential, especially in an arid region, can play a major role in planning the sustainable management of groundwater resources. In this study, nine machine learning (ML) algorithms—namely, Artificial Neural Network (ANN), Decision Jungle (DJ), Averaged Perceptron (AP), Bayes Point Machine (BPM), Decision Forest (DF), Locally-Deep Support Vector Machine (LD-SVM), Boosted Decision Tree (BDT), Logistic Regression (LG), and Support Vector Machine (SVM)—were run on the Microsoft Azure cloud computing platform to model the groundwater potential. We investigated the relationship between 512 operating boreholes with a specified specific capacity and 14 groundwater-influencing occurrence factors. The unconfined aquifer in the Nineveh plain, Mosul Governorate, northern Iraq, was used as a case study. The groundwater-influencing factors used included elevation, slope, curvature, topographic wetness index, stream power index, soil, land use/land cover (LULC), geology, drainage density, aquifer saturated thickness, aquifer hydraulic conductivity, aquifer specific yield, depth to groundwater, distance to faults, and fault density. Analysis of the contribution of these factors in groundwater potential using information gain ratio indicated that aquifer saturated thickness, rainfall, hydraulic conductivity, depth to groundwater, specific yield, and elevation were the most important factors (average merit > 0.1), followed by geology, fault density, drainage density, soil, LULC, and distance to faults (average merit < 0.1). The average merits for the remaining factors were zero, and thus, these factors were removed from the analysis. When the selected ML classifiers were used to estimate groundwater potential in the Azure cloud computing environment, the DJ and BDT models performed the best in terms of all statistical error measures used (accuracy, precision, recall, F-score, and area under the receiver operating characteristics curve), followed by DF and LD-SVM. The probability of groundwater potential from these algorithms was mapped and visualized into five groundwater potential zones: very low, low, moderate, high, and very high, which correspond to the northern (very low to low), southern (moderate), and middle (high to very high) portions of the study area. Using a cloud computing service provides an improved platform for quickly and cheaply running and testing different algorithms for predicting groundwater potential.

Keywords: cloud computing; groundwater potential mapping; GIS; machine learning; Iraq

1. Introduction

Groundwater is a vital resource for supplying drinking water for millions of people around the world, as well as for agriculture, industry, and preserving the natural environment. Relative to surface water, groundwater has a range of important benefits: it is less susceptible to seasonal and perennial variability, often more evenly dispersed over large areas, more insulated against pollution, and of acceptable quality for most common uses, such as drinking and irrigation [1]. Groundwater also acts as a key strategic reservoir during droughts [2]. Notwithstanding these benefits, groundwater is often under-monitored and under-controlled in comparison with more visible surface water. As a result, it is difficult to get a good picture of groundwater's temporal and spatial availability, which is the most important requirement for groundwater management. Hydrogeological mapping is a tool for comprehensive groundwater resource development and planning. Hydrogeological maps are necessary for collecting and analyzing aquifer and related geological data in order to establish a three-dimensional representation of the subsurface environment where groundwater exists [3]. Furthermore, mapping aids in determining the exposure of aquifers and their related ecosystems to pollution and overexploitation, as well as identifying areas for artificial recharge and communicating information to groundwater users [4]. Traditional groundwater mapping requires a lot of effort and money, especially in remote areas or developing countries. This necessitates the development of new methods to make groundwater exploration and assessment as effective as possible. One of these methods is the assessment of groundwater potential by relying on geographic information systems (GIS), which has also been associated with the use of remote sensing techniques and global positioning system technology, machine learning (ML), and soft computing techniques [5–7]. The term groundwater potential denotes the amount of groundwater available in an area, and it is a function of several hydrologic and hydrogeological factors [8]. This definition, from our point of view, is too simple, as the groundwater potential in an area is the outcome of a complex process that is not influenced by hydrological and hydrogeological elements. Many other factors such as geology and structural setting, geomorphology, and topography can influence groundwater accumulation. The most accurate definition should be as follows: groundwater potential denotes the amount of groundwater available in an area, and it is a function of several surface and subsurface factors.

Martínez-Santos and Renard [9] provide a more practical definition for this term as “an application of predictive mapping, a forecasting technique that consists in developing spatially distributed estimates for a target variable based on a series of indirect indicators (explanatory variables)”. The target outcome of groundwater potential maps is the feasibility of drilling successful boreholes in different parts of a given region. In general, a closer look at the literature shows two different meanings for groundwater potential. Some researchers have used groundwater potential analysis as an exploration tool by assuming that the study of surficial factors such as topography, geology, geomorphology, soil, land use/land cover (LULC), drainage characteristics, lineament density, and proximity to surface-water bodies offer an indirect exploration tool to find where the groundwater is more likely to occur [10–12]. Others have contended that groundwater potential maps show variation in groundwater storage across a given region and thus provide information on groundwater availability and productivity [5,13,14].

Worldwide, researchers have used three types of techniques to model groundwater potential: data-knowledge, data-driven, and hybrid models of knowledge and data-driven techniques [15]. In data-knowledge models, such as simple overlay technique, fuzzy logic, and multi-criteria decision making (MCDM), a specific number of groundwater-affecting occurrence factors are combined to generate the groundwater potential map [16–19]. In data-driven models, such as bivariate and multivariate statistical models, ML classifiers, and hybrids of these three models, or with knowledge-driven models, the relationship between the locations of wells with specified pumping capacity or specific capacity (S_c) and the groundwater influential occurrence factors is explored to map groundwater

potential [5,13,20–24]. The locations of the operating groundwater wells, in this case, are taken as the target (dependent) variable and the groundwater influential occurrence factors are taken as predictors (independent variable).

Cloud computing services enable the development of virtual servers, referred to as instances, inside a centralized computing facility operated by service providers such as Amazon, Microsoft, or Google, among others [25]. Cloud computing encompasses all the technologies provided as services over the Internet, as well as the hardware and software in the data centers that support those services [26]. The key benefit of cloud computing is that it allows users to access huge computational resources without having to invest in conventional servers or distributed computing infrastructure. The major benefit of the cloud is that any researcher with an internet connection, a username, a password, and a system may access the final output from any location in the world [27]. Anyone may perform the reengineering and test the results if necessary. The final assessed result can be made available to an application that can use it as a web service. Hayley [25] discussed how cloud computing can be effectively used for numerical modeling of groundwater flow and automatic calibration of groundwater simulation models that need vast resources to implement [28,29]. In this study, we use cloud computing in GIS-based groundwater potential mapping and examine the opportunity given by this remote, highly efficient platform instead of using stand-alone applications. Wang et al. [30] presented the only available study concerning the use of cloud computing to study groundwater potential mapping, employing a simple overlay technique supported by the analytical hierarchy process (AHP) method and employing only a few surface thematic layers such as topographic slope, aspect, water-density, land surface temperature, and normalized difference vegetation index (NDVI). The authors of the previous article used the groundwater potential maps as an exploration tool to indicate where the groundwater was likely to occur.

In this study, the Microsoft Azure cloud computing platform was used for modeling groundwater productivity in the Nineveh Plain, northern Iraq. Nine ML algorithms—namely, Artificial Neural Network (ANN), Decision Jungle (DJ), Averaged Perceptron (AP), Bayes Point Machine (BPM), Decision Forest (DF), Locally-Deep Support Vector Machine (LD-SVM), Boosted Decision Tree (BDT), Logistic Regression (LG), and Support Vector Machine (SVM)—were used for modeling the relationship between the locations of 520 boreholes with specific capacity data and 14 groundwater-influencing occurrence factors. The primary focus was to demarcate groundwater potential areas—e.g., areas with a sufficient yield for groundwater abstraction for a given region. The novelty of this work is that it is the first-time cloud computing has been used to reveal groundwater potential (storage variation) using both surface and subsurface groundwater influential occurrence factors, whereas in most groundwater potential studies, the subsurface factors (aquifer related factors) have been ignored.

2. Materials and Methods

2.1. The Study Area

The study area (Nineveh Plain) is located in northeastern Mosul governorate, northern Iraq, between $36^{\circ}47'27.47''$ N– $35^{\circ}59'3.57''$ N latitude and $42^{\circ}44'33.51''$ E– $43^{\circ}33'33.91''$ E longitude. It covers an area of 2540 km². It extends from Mosul Lake in the north to the Great Zab River in the south and is bordered by the Tigris River on the west (Figure 1). The elevation ranges from 1047 m above mean sea level (amsl) in the northeastern part of the study area, within the Alqoosh and Ain Sifni anticlines, to 190 m amsl in the Tigris River basin to the southwest. In general, the land surface is relatively flat in the middle and south of the area, with some hills and valleys in the north and northeast (Figure 1). The ages of exposed formations in the study area range from Middle–Upper Eocene to Quaternary. The youngest formations are exposed in the middle and south of the study area, while the oldest formations appear within the folded rocks in the north and northeast of the study (Figure 2). Table 1 gives a brief overview of the exposed formations in the study area (see [30] for more information). The study area contains many complex structures in the northern and

northeastern parts (Figure 2), including several major faults near the folds. The axes of the folds represent hydrologic divides. Geological formations that are exposed on the flanks of these folds represent recharge areas the Pila Spi Formation, which is jointed and fractured in the folded area. Six LULC classes are recognized in the study area: Urban, Agriculture, Grassland/Pasture, Rocky, Forest, and Water Bodies (Figure 3). Agriculture covers about 1769 km² (69.4%), while the other classes cover 771 km² (30.6%). The soil textures in the study area were mapped using 36 soil samples using the USDA web-based soil texture calculator (<https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey>) in 2 April 2020. Based on the assigned texture, the soils were classified into hydrological groups and mapped using the Thiessen polygon method (Figure 4).

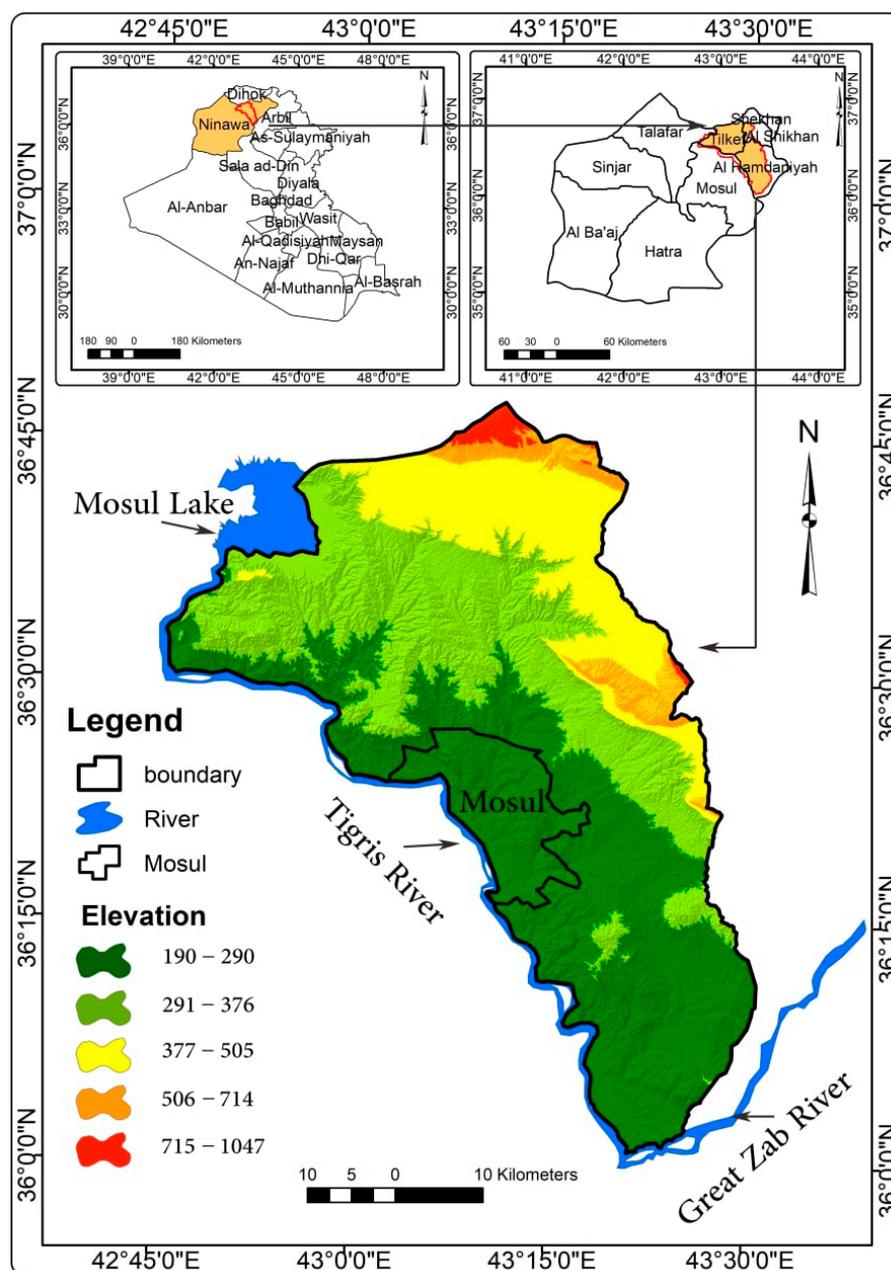


Figure 1. Location and topography of the study area.

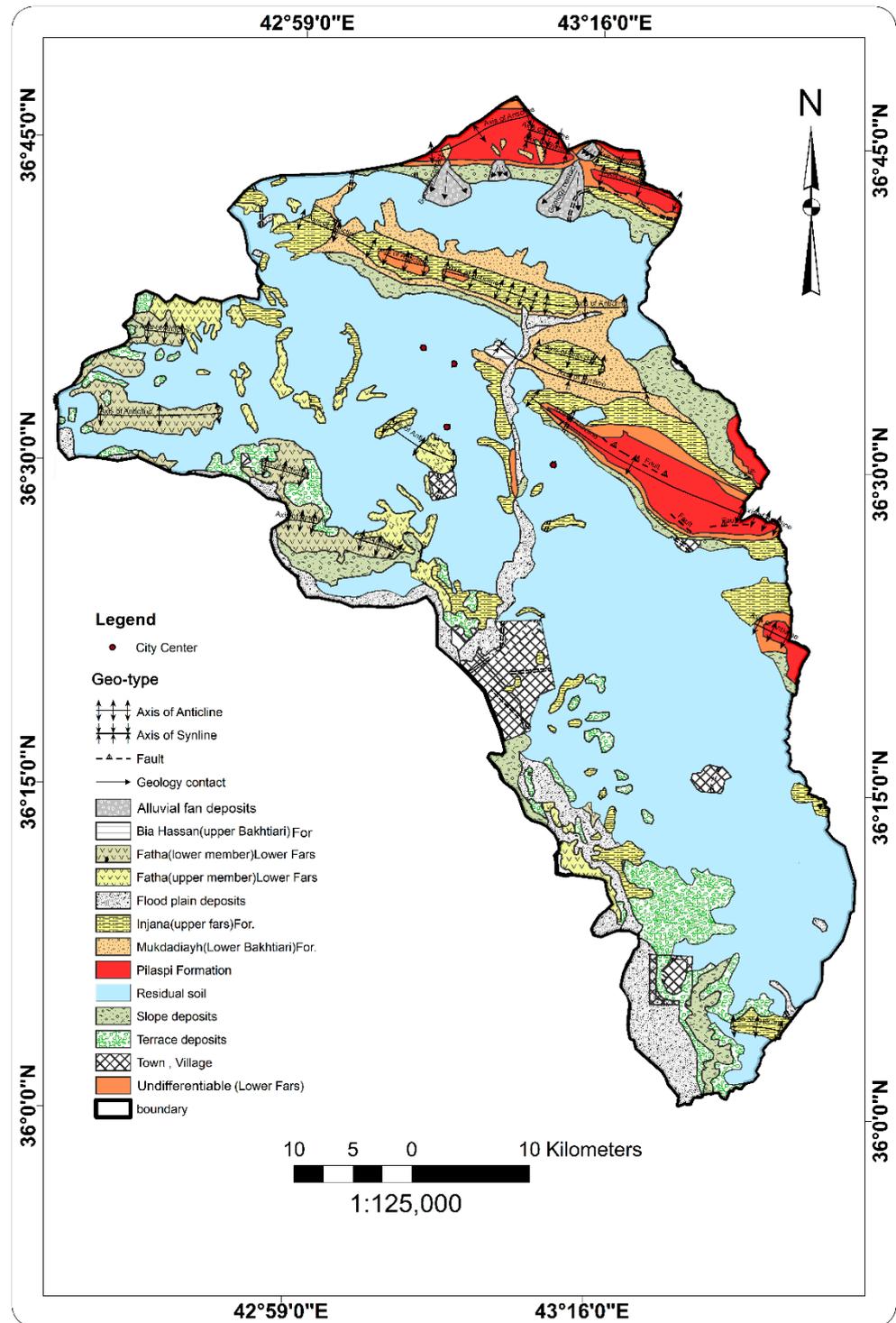


Figure 2. Geological map of the study area.

Table 1. A brief description of the exposed geological formations in the study area (After [31]).

Formation	Age	Environment	Description
Pila spi	Middle–Upper Eocene	Shallow marine	Limestone and dolomite limestone
Fatha	Middle Miocene	Shallow marine	Anhydrite, mudstone, and thin limestone
Injana	Upper Miocene	Sub-marine	Red or gray colored silty marl or clay stones and purple silt stones
Muqdadyia	Pliocene	Continental	Gravelly sandstone, sandstone, and red mudstone
Bai Hassan	Late Pliocene	Continental	Conglomerate with sandstone, silt stone, and claystone
Quaternary	Pleistocene–Holocene	Continental	Mixture of gravel, sand, silt, and clay

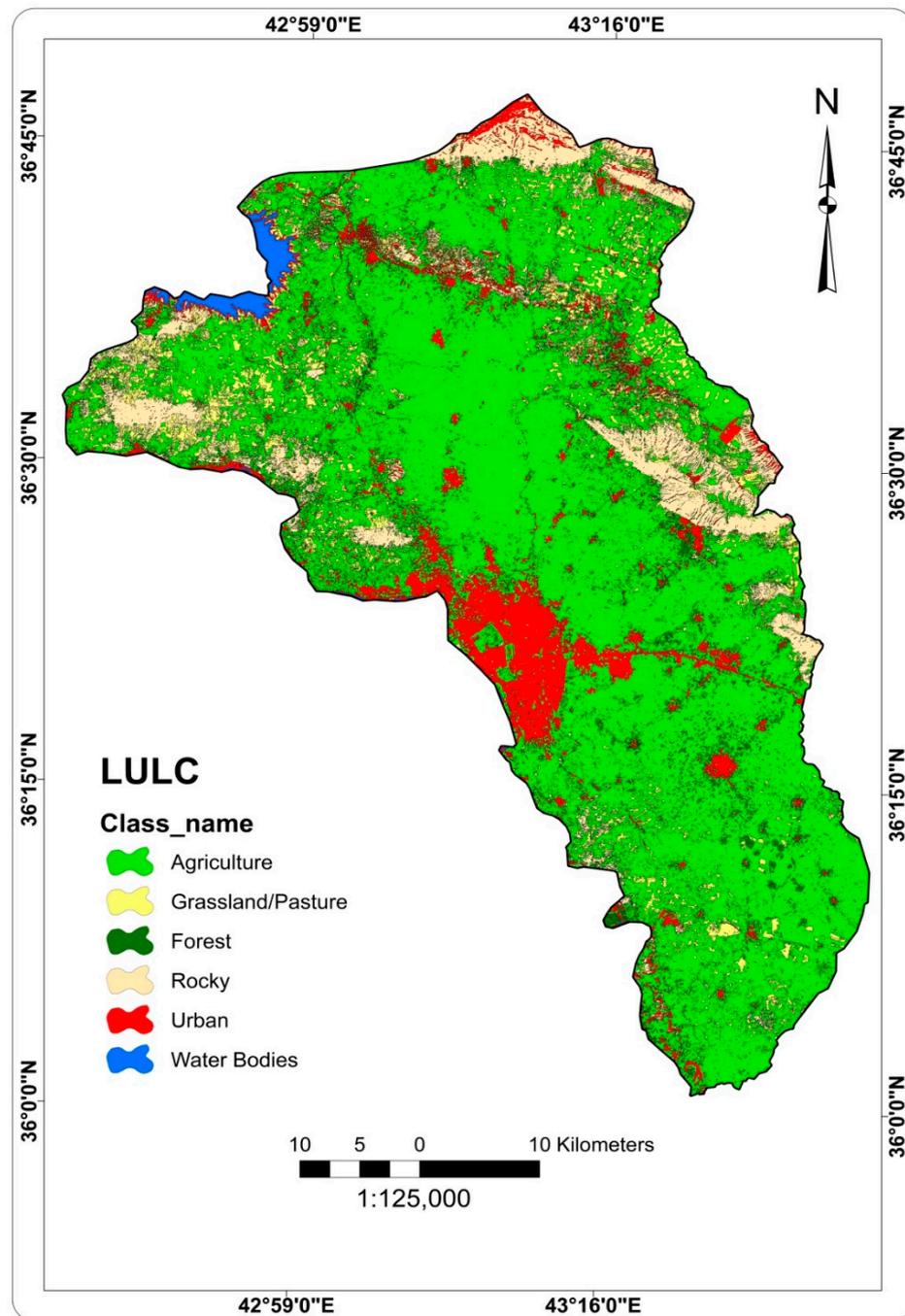


Figure 3. Major LULC in the study area.

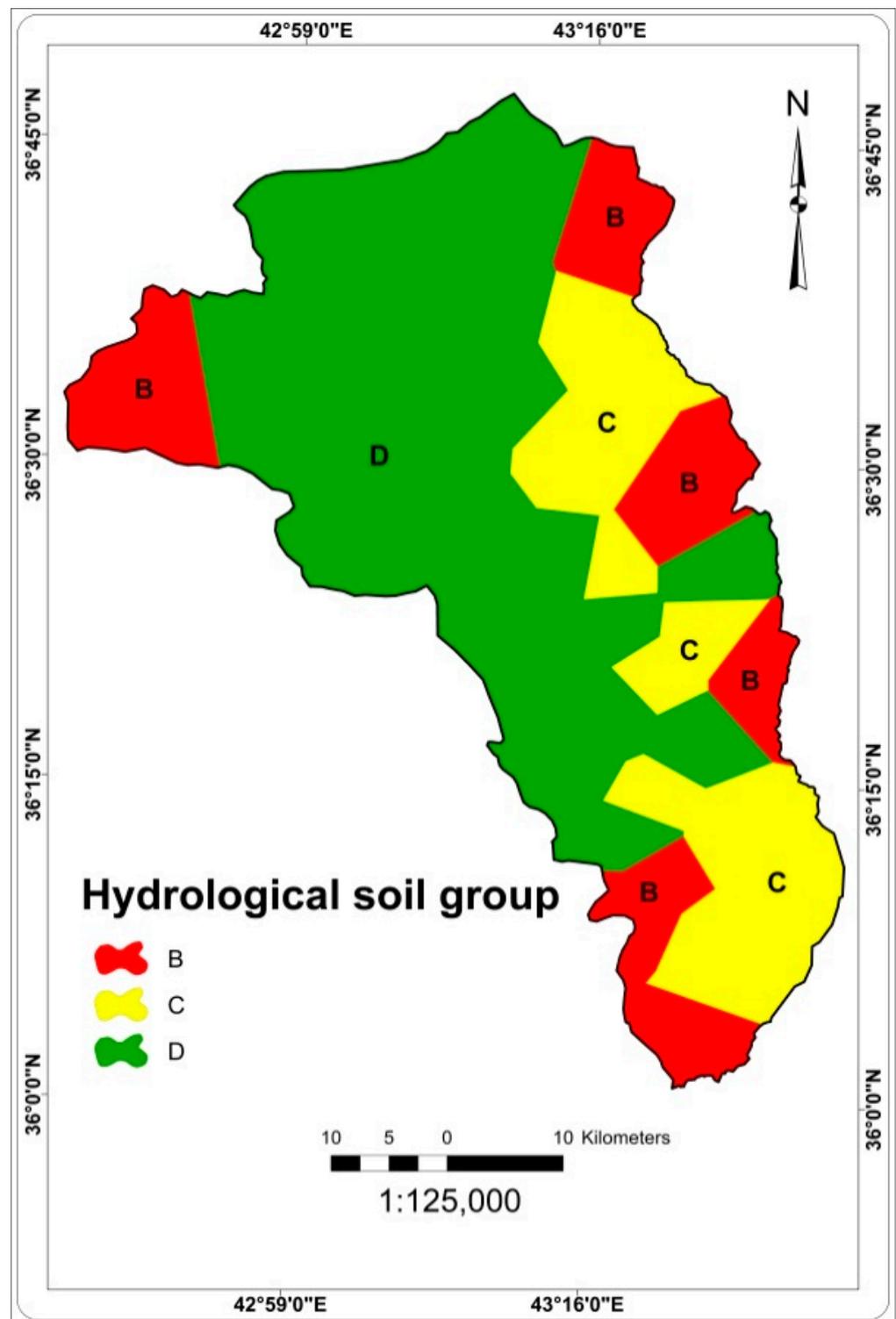


Figure 4. Hydrological soil groups.

The construction of the conceptual model for study area was based on the records data for boreholes obtained from the Groundwater Department in the city of Mosul. These data represent characteristics of a well, including construction data, well owner, well depth and geographic coordinates, lithology and stratigraphy data including intervals, material types, and well screens position. Additionally, information such as depth to groundwater, drawdown, pumping test data, and well discharge were also available. Thirty-four wells were selected to represent the study area. These wells had a semi-completed record, and

there was an even distribution of these wells throughout the area. The main water-bearing layers in the study area are located in the Quaternary and Tertiary formations (Mukdadiyah, Bai Hassan, and Injana Formations). Due to the lack of a regional confining unit in the study area and the existence of numerous joints, faults, and fractures that allow hydraulic connections between these formations, these water-bearing layers can be considered as a single unconfined aquifer system, Figure 5. The groundwater depth ranges from 11 to 66 m below land surface (bls) and averages 28.81 m bls. Figure 6a shows the spatial distribution of the groundwater depths in the study area based on the measurements made by the authors in 58 wells evenly distributed throughout the study area. The depth to groundwater increases from southeast to northeast. Overall, the shallow depths occur in the plain, and the greater depths appear in the folded areas, especially where the oldest formations are exposed. Groundwater levels (heads) were estimated by subtracting the ground elevation from the depth to groundwater available at 58 wells, and the result was interpolated by Empirical Bayesian Kriging technique to reveal the spatial distribution of groundwater level over the study area, Figure 7. The groundwater levels are high in the north and northeastern parts of the study area and decrease towards the Basin of Tigris River. The north and northeast areas of the study area represent the recharge zone while south and southwest areas represent discharge areas.

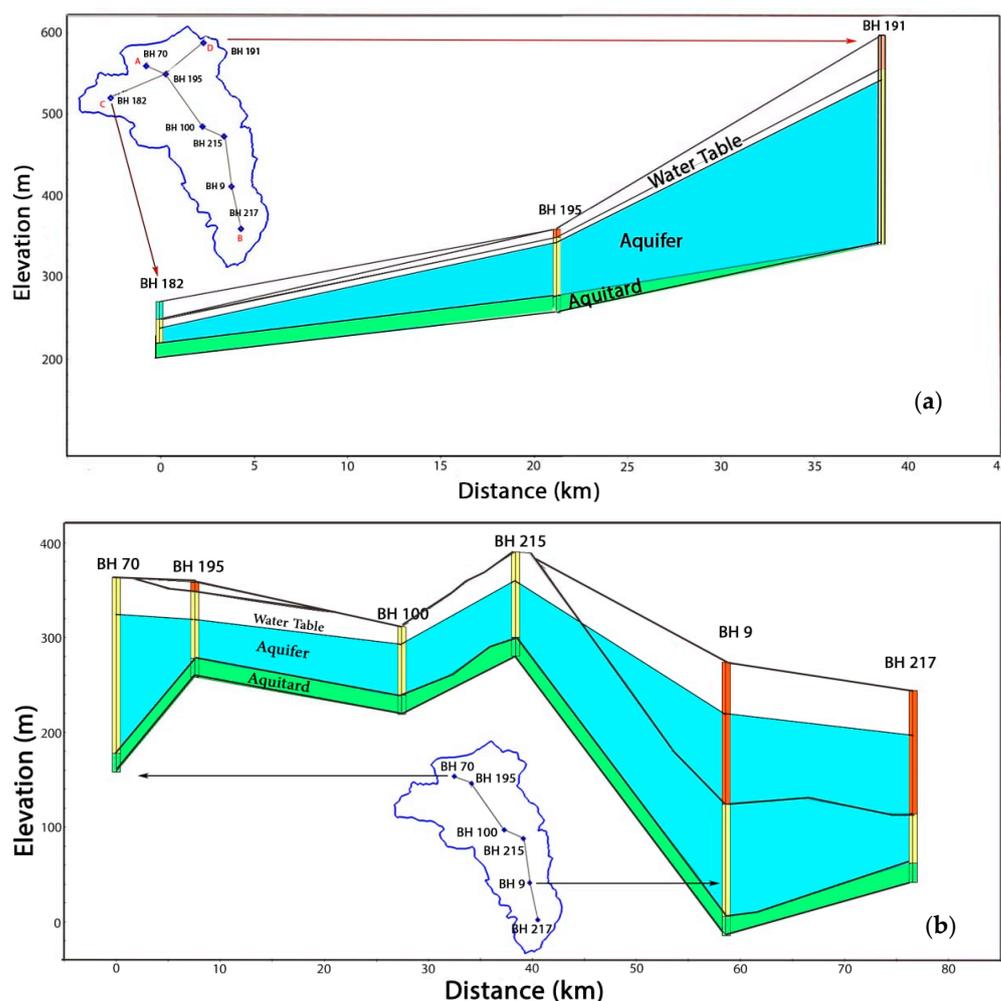


Figure 5. Selected cross sections in the study area (a) from NW–SE, and (b) from NE–SW.

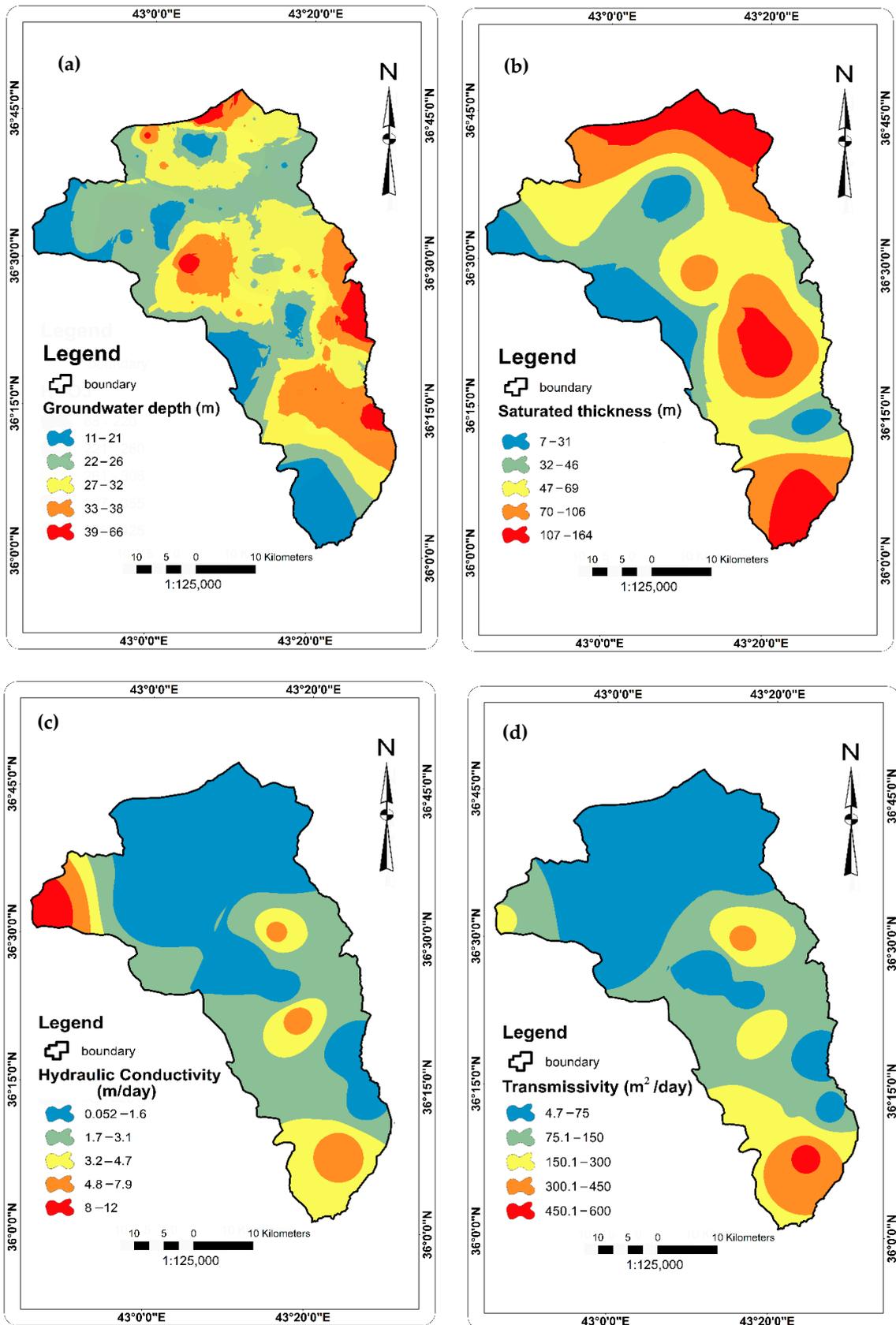


Figure 6. Cont.

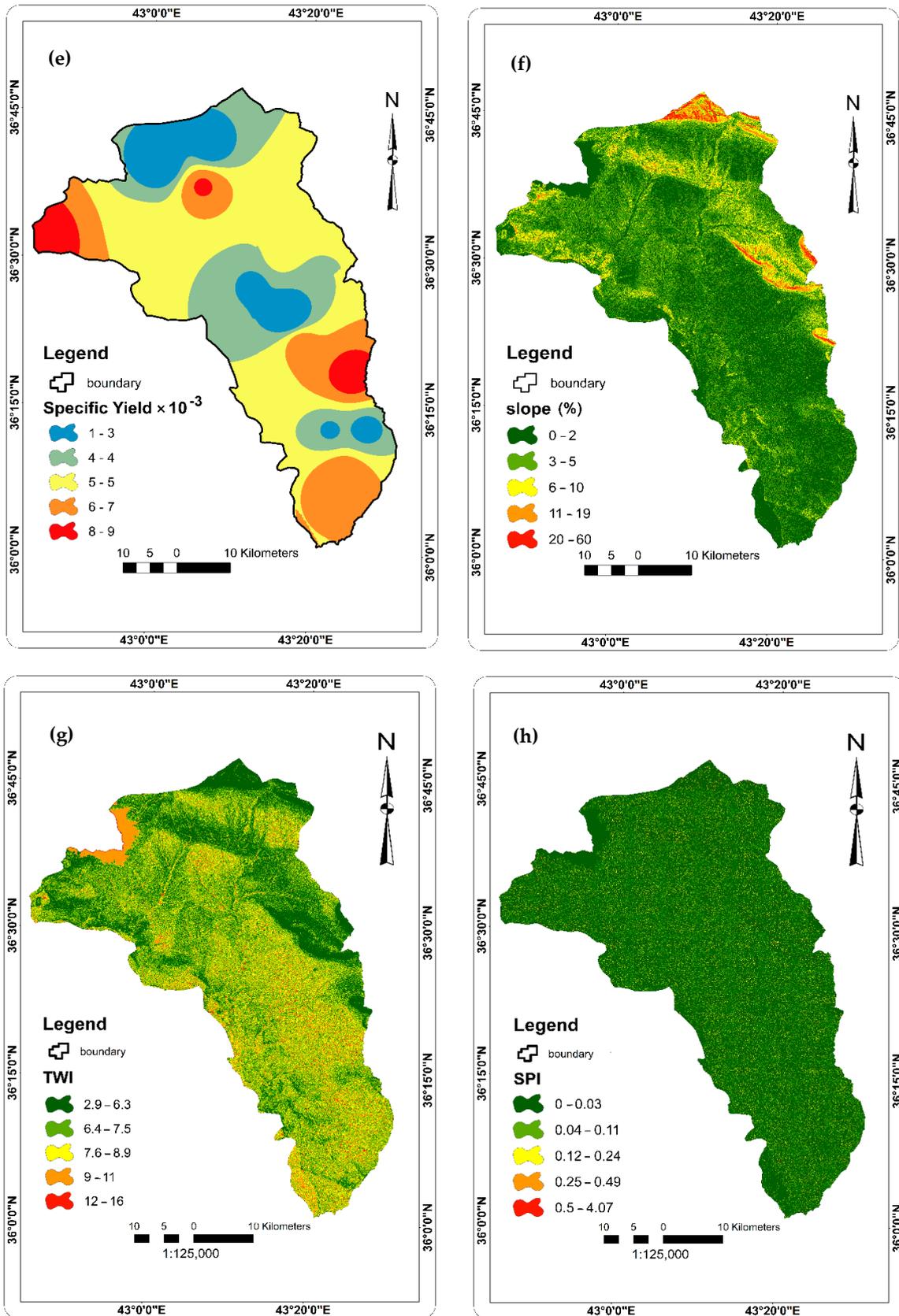


Figure 6. Cont.

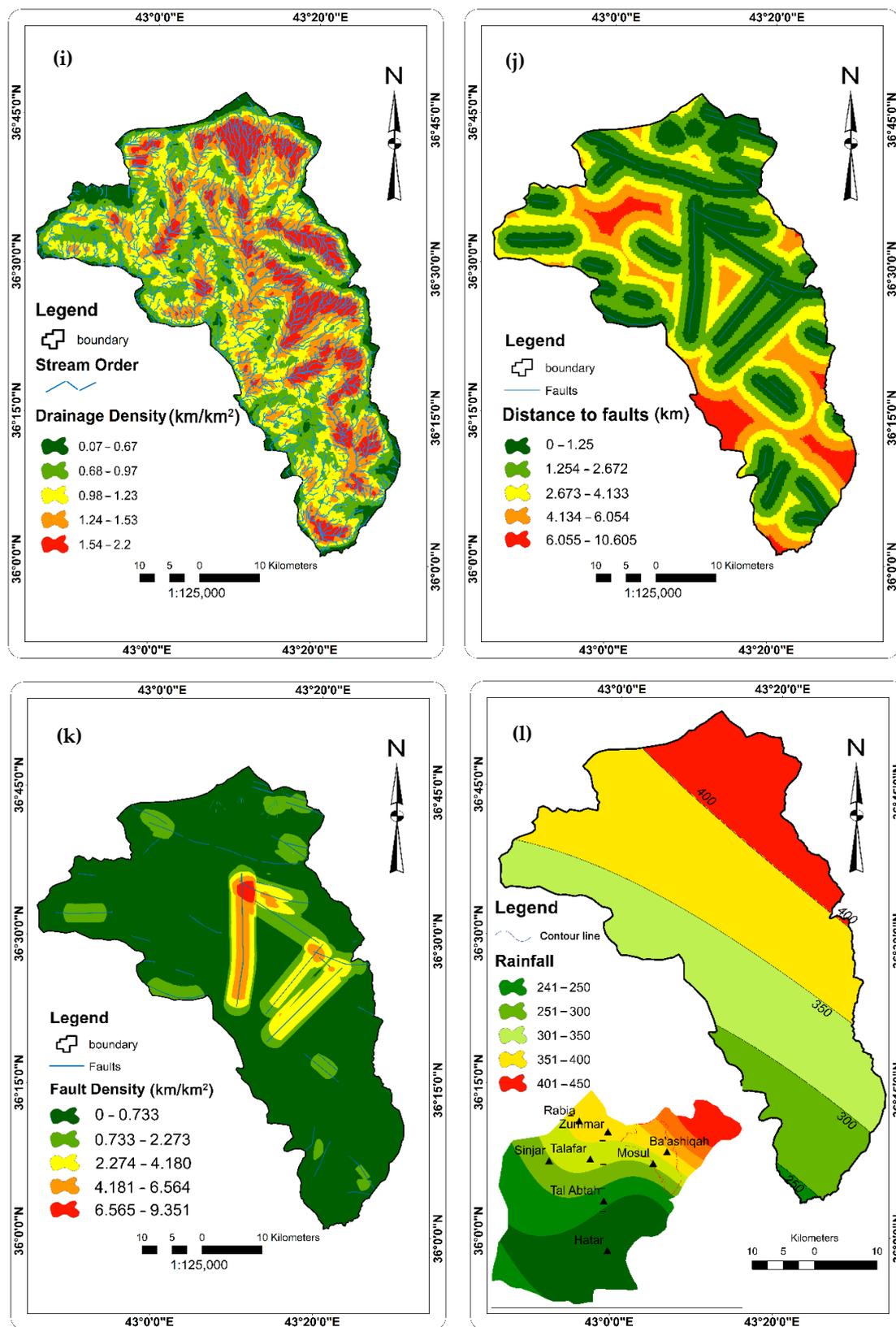


Figure 6. Aquifer-related factors: (a) depth to groundwater (m), (b) saturated thickness (m), (c) hydraulic conductivity (m/d), (d) transmissivity (m²/d), (e) specific yield (f), slope (%), (g) TWI, (h) SPI, (i) drainage density (km/km²), (j) distance to faults (km), (k) fault density (km/km²), (l) annual rainfall (mm).

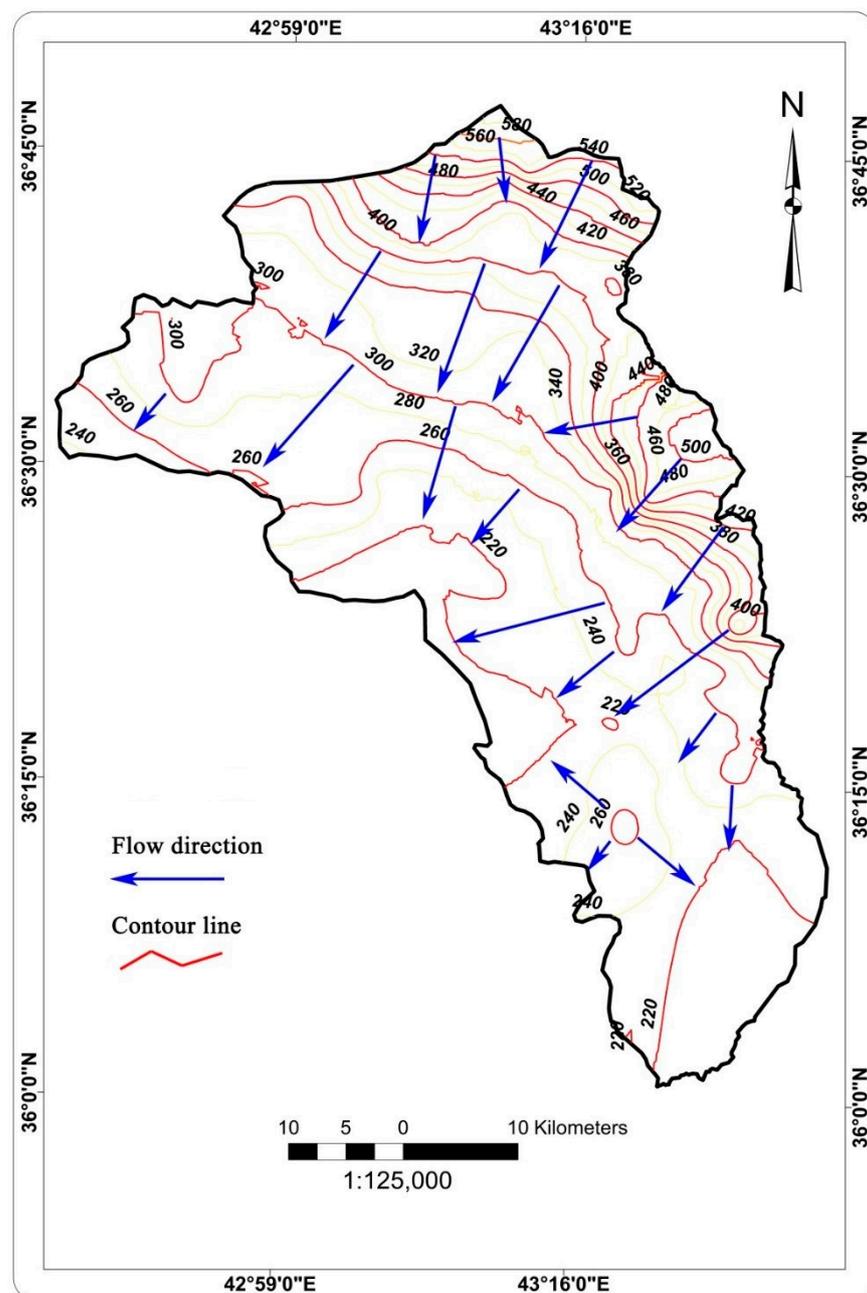


Figure 7. Groundwater levels (m) and the flow direction in the study area.

The spatial distribution of the saturated thickness of this aquifer is shown in Figure 6b. There is a clear decrease in the saturated thickness of the aquifer from the north and northeast to the west and southwest. The structural setting (the distribution of folds and faults) plays a major role in controlling the aquifer saturated thickness, in addition to the proximity to the recharge area.

The pumping test data of 14 wells in the study area was used to infer the hydraulic characteristics of the main aquifer, i.e., the hydraulic conductivity and specific yield. The estimated hydraulic conductivity was interpolated using inverse distance weighting to reveal its spatial distribution (Figure 6c). The hydraulic conductivity is low in the north and increases towards the Tigris River in the south and west, where there are high-permeability Quaternary and Tertiary sediments that mainly consist of pebbly sands. The transmissivity is higher in the middle of the study area and the south than in the northern and eastern parts (Figure 6d), which are mainly occupied by lower-permeability fractured rocks. The

specific yield of the aquifer ranges from 0.001 to 0.01 (average 0.004), with the relatively high values appearing in the northeast part of the study area and in small spots in the middle and southeast (Figure 6e).

2.2. Workflow Steps

The steps for conducting this work are summarized in Figure 8 and consisted of four stages. The inventory map of productive wells was prepared first from well capacity data. This map was used as a target variable in the analysis of groundwater potential. Second, 14 surface and subsurface factors influencing groundwater productivity—namely, hydraulic conductivity, saturated thickness, specific yield, depth to groundwater, rainfall, geology, soil, LULC, topography-related factors (slope, topographic wetness index (TWI), stream power index (SPI)), drainage density, fault density, and distance to faults—were collected from different sources such as field surveys, remote sensing data, and previous works. These factors were used as predictors in the groundwater potential analysis. Third, the relationship between the well inventory map and the influential groundwater occurrence factors was explored using nine ML algorithms through the Microsoft Azure cloud computing platform. Fourth, the performance of the used algorithms was compared using error statistics measures, and the best was selected to map the probability of groundwater potential.

2.2.1. Operating Wells Inventory Map

Water wells in the study area are mainly used for drinking, industrial, and agricultural (including livestock) purposes. The inventory map of the well locations was prepared using the data of the General Commission of Groundwater (GCG), Mosul branch, and an extensive field survey by the authors. We registered and mapped locations of 520 wells to use as the target variable to generate the map of groundwater potential. Values of S_c , which is the pumping rate divided by the drawdown [32], were considered. The S_c can be efficiently used to design the maximum yield of a well and to estimate the transmissivity of the aquifer tapped by the well. The pumping capacity of the well is not a good indicator for determining the productivity of the water-bearing layer because pumps may vary in capacity from one well to another. Therefore, the areas of low productivity may seem to be highly productive if the pumping capacity of the well (flow rate) is high and vice versa [33]. Values of S_c in the study area ranged from 0.006 to 9.994 L/s/m, with an average of 1.000 L/s/m. The S_c of the 520 wells was divided into high-potential (>2 L/s/m) and low potential (≤ 2 L/s/m) [5]. Wells with high and low productivity status were coded as yes and no, respectively, to use in the classification problem solved in this study. After assigning the codes for the wells, these wells were randomly partitioned into two groups: 70% (364 wells) were used for training the models, and the remaining 30% (156 wells) were allocated for testing (Figure 9).

2.2.2. Factors Affecting Groundwater Occurrence and Availability

The numbers and types of groundwater-influencing occurrence factors used in the analysis of groundwater potential vary depending on the complexity of the problem to be solved, data availability, and the method in which the groundwater potential map is used, i.e., as an exploration tool or to refer to the groundwater productivity and availability. These factors can be categorized into two main groups: surface and subsurface. The surface factors such as the topography-related factors, exposed lithological units, soil, LULC, and drainage density mainly control the groundwater renewal rate, while subsurface factors determine the aquifer capacity (the ability of the aquifer to store and transmit groundwater) and the capability of an aquifer to respond to external stresses (natural or induced). The subsurface factors, which include the transmissivity, storativity, aquifer type, aquifer architecture, groundwater depth, and presence of structural features such as faults, determine the strategic or fixed groundwater storage. In most cases, especially in arid and semi-arid regions, due to the precipitation scarcity, uneven distribution of precipitation,

and absence of perennial lakes and rivers that have direct contact with the aquifer system, the percentage of renewable storage is much lower than the fixed groundwater storage in stock. Therefore, many of the large aquifers in arid and semi-arid regions are depleted and enter a stage of mining due to the imbalance between aquifer inputs and outputs [34]. Mapping of groundwater availability or productivity using only surface factors pertains only to the renewable storage of the aquifer. To map the spatial distribution of groundwater productivity across the study area, we need to consider both the surface and subsurface factors.

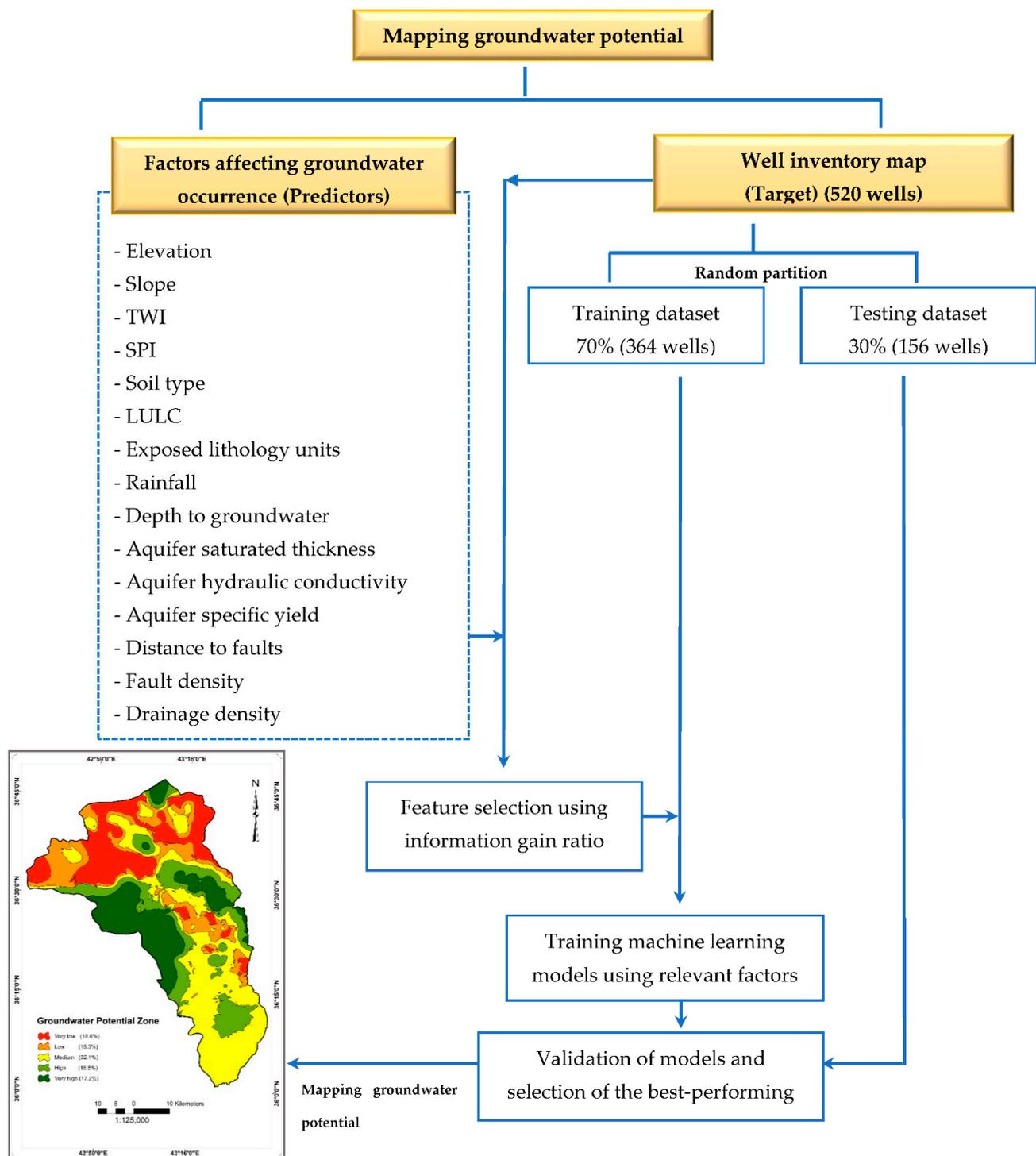


Figure 8. Steps to develop groundwater potential map in this study.

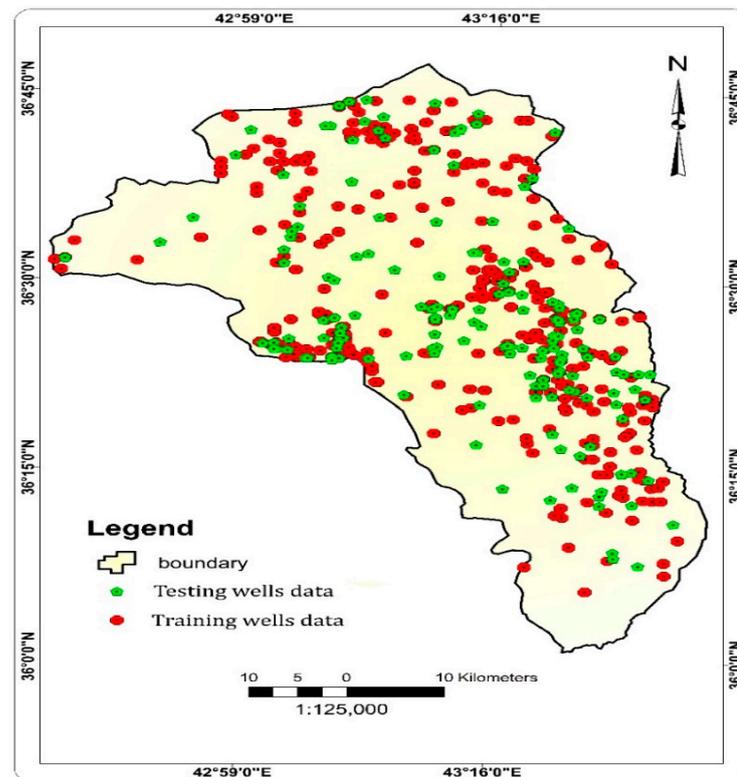


Figure 9. Groundwater wells inventory map.

The elevation, soil, LULC, geology, depth to groundwater, aquifer saturated thickness, hydraulic conductivity, and specific yield were generated using different sources such as the 30 m Shuttle Radar Topography Mission digital elevation model (DEM), archival data, and field test data (Figures 1–4 and Figure 6a–e), Section 2.1. To generate the slope, TWI, and SPI, the same DEM that was used to generate the elevation layer was used after pre-processing (fill sink) and re-projection. The importance of these factors for modeling groundwater potential is comprehensively discussed elsewhere [35–37]. The slope thematic layer was directly generated from the DEM (Figure 5f). Both TWI and SPI are used to describe the spatial soil moisture pattern. TWI was developed by [38] and is defined as

$$TWI = \ln \left[\frac{\alpha}{\tan \beta} \right] \quad (1)$$

where α is the local upslope area draining through a certain point per unit contour length, and $\tan \beta$ is the local slope in degrees. TWI reflects the tendency of water to accumulate at any point in a catchment (in terms of α) and the tendency of gravitational forces to move that water downslope (in terms of $\tan \beta$). The SPI measures the erosive power of the stream and is defined as [39]

$$SPI = A_s \tan \beta \quad (2)$$

where A_s is a specific catchment area. The TWI and SPI maps are shown in Figure 6g,h.

Drainage density (DD) is the total length of all channels (L) in the drainage basin divided by the area of the basin (A):

$$DD = \frac{L}{A} \quad (3)$$

DD represents both the drainage basin's ability to produce surface runoff and the erodibility of the surface material by indicating the extent of terrain dissection by channels. Areas with high DD are generally characterized by low infiltration rate, rapid runoff

generation, and moderately erodible surface materials. As a result, when the *DD* is high, the surface run-off is high, and the quantities of water infiltrated through the soil section, which contribute to groundwater recharge and storage, are reduced. When the *DD* is low, however, it is expected that water infiltration will be larger due to the availability of adequate circumstances such as topographic characteristics, soil, and land uses, and that this process will affect groundwater recharge. The *DD* in the middle and northern portions of the study area is relatively dense, indicating that these parts receive less recharge than other parts (Figure 6i).

The structural setting, represented by distance to faults and fault density, plays a role in controlling the fixed groundwater storage by controlling the architecture of the aquifer itself. In addition, faults allow water to enter the subsurface and enhance groundwater storage [40]. The fault features were firstly digitized from the tectonic map of Iraq with a scale of 1:1,000,000. The Euclidian distance command of ArcGIS 10.4 was used to create the distance to faults map (Figure 6j), and the line density command was used to create the fault density map (Figure 6k).

Rainfall is an important factor in delineating groundwater potential because it mainly affects the renewable recharge of the aquifer system. The average annual sum of rainfall recorded at eight stations inside and outside of the study area was used to map the interpolated surface of this factor (Figure 6l) [41]. Overall, the rainfall in the northern part of the study area is higher than in the south, indicating that recharge is higher in the north, taking into consideration the exposed lithology and type of soil.

2.3. Feature Selection

Feature selection is the process of selecting a subset of characteristics from a dataset based on their quality, importance, conventions, significance, and constraints [42]. Feature selection helps in (1) simplifying the predictive models so that they can be easily developed and interpreted; (2) shortening the time of model training, and (3) enhancing generalization by reducing model over-fitting. In this study, the information gain ratio was used to quantify which variables were important in the analysis of groundwater potentiality and which should have been discarded to obtain more accurate ML models. Information gain chooses features by assessing each variable's gain in the context of the target variable. The calculation is called mutual information between the two random variables. By dividing a dataset according to a specified value of a random variable, information gain evaluates the reduction in entropy or surprise. A higher information gain implies a lower entropy group or groups of samples, and hence a lower level of surprise [43]. More information about the mathematics of this technique can be found in [44].

2.4. Azure Cloud Platform and Machine Learning Classifiers Used

Azure is a cloud computing service for developing, testing, deploying, and managing applications and services through Microsoft-managed data centers [45]. It offers software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS), and it supports a wide range of programming languages, tools, and frameworks, including Microsoft-developed and third-party software.

For this research, nine ML classifiers were used to map the probability of groundwater potential in the study area. The first of these, ANN, is a computer system that mimics the way the human brain analyzes and processes data [46]. An ANN is made up of hundreds or thousands of artificial neurons called processing units that are connected via nodes. Based on an internal weighting mechanism, the input units receive diverse forms and structures of information, and the neural network strives to learn about the information provided to generate one output report [47]. Backpropagation is a set of learning principles used by ANNs to perfect their output results. Many hidden layers can be inserted between the input data and output data layers. Just one or a few hidden layers can perform most predictive tasks quickly.

AP is a special kind of neural network perceptron, which is a classification method that finds a separating hyperplane (a line that linearly divides a dataset) to make predictions. The perceptron is an online algorithm, which means it processes each instance in the training set individually [48]. Using a set of starting weights, the weighted sum of the features is calculated for each sample in the training set. The weights stay unchanged if this value has the same sign as the current example's label. The weights are updated in a generalization of this method by multiplying the feature vector by the learning rate and by the gradient of some loss function [49]. In the AP, a weight vector is produced for each iteration or pass over the training data. After that, the final prediction is determined by averaging the weighted sum of each weight vector and examining the result's sign.

SVM is an advanced supervised machine learning approach to classification and regression problems. It aims to determine the optimum hyperplane for dividing a dataset into distinct classes [50]. The data points near the hyperplane, called support vectors, are important parts of the SVM since deleting them changes the location of the hyperplane substantially. The SVM method selects a hyperplane with the largest possible margin between it and any point in the training dataset, allowing the new data to be correctly categorized. When defining a clear hyperplane is problematic, the two-dimensional data are transformed into three dimensions using kernelling, and the hyperplane becomes a plane [51]. The dataset is continuously mapped into higher and higher dimensions until an optimum segregation hyperplane can be constructed.

LD-SVM is a kind of SVM in which the kernel function for mapping data points to feature space is specially intended to minimize training time while keeping maximal classification accuracy. Because this approach is based on supervised learning, it requires a tagged dataset with a label column.

The BPM is a Bayesian method for linear classification. By selecting one "average" classifier, the Bayes Point efficiently approximates the theoretically optimal Bayesian average of linear classifiers (in terms of generalization performance) [52]. The BPM is not prone to overfitting the training data because it is a Bayesian classification machine.

DF is a fast supervised ensemble model, and it is a good choice for predicting a target with a maximum of two outcomes. The DF algorithm is a classification-focused ensemble learning approach. Ensemble techniques are founded on the basic idea that by developing several related models and integrating them in some way, one may achieve better results and a more generalized model than by depending on a single model. In general, ensemble models outperform single decision trees in terms of coverage and accuracy.

DJ is a recent extension to DF. A DJ consists of an ensemble of decision-directed acyclic graphs (DAGs). A DAG enables numerous pathways from the root to each leaf, unlike traditional decision trees that only allow one path to each node. Node splitting and merging are driven by the minimization of the same objective function during training, in this case, the weighted sum of entropies at the leaves. Results from a variety of datasets demonstrate that DJs need significantly less memory than DFs and various other baselines, while significantly enhancing generalization [53].

A BDT is an ensemble learning approach in which the first tree's errors are corrected by the second tree, the second tree's errors are corrected by the third tree, and so on. The whole ensemble of trees that produces the prediction is used to create the prediction. BDTs are, in general, the simplest approaches for achieving top performance on a wide range of machine learning problems when correctly set. They are, nevertheless, memory-intensive learners. As a result, a BDT model may not be able to handle the huge datasets that linear learners can.

Finally, LG is a well-known statistical approach for predicting the probability of a given event, and it is particularly useful for classification problems. The algorithm predicts the probability of occurrence of an event by fitting data to a logistic function.

2.5. Error Statistics Used to Evaluate the Model Performances

Five error measures were used in this study to evaluate the performances of the ML models: accuracy (*ACC*), precision (*P*), recall (*R*), F-score (*F*), and area under the receiver operating characteristic curve (*AUC*). *ACC* is the proportion of cases that are predicted correctly by the model [54]. It is computed in terms of confusion matrix as

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

where, *TP* is the number of cases that predicted correctly as high potential, *FN* is the number of cases predicted incorrectly as high potential, *TN* is the number of cases predicted correctly as low potential, and *FP* is the number of cases predicted incorrectly as low potential. Precision is the value of overall positive real outcomes (high potential outcomes) and is defined as

$$Precision (P) = \frac{TP}{TP + FP} \quad (5)$$

The recall is the proportion of positive cases that are predicted correctly and is calculated as

$$Recall (R) = \frac{TP}{TP + FN} \quad (6)$$

The F-score is the weighted average of *P* and *R* and is computed as

$$F = \frac{2PR}{P + R} \quad (7)$$

Finally, the receiver operating characteristic curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system, as its discrimination threshold is varied. It is created by plotting the true positive rate (*TPR*) against the false positive rate (*FPR*) at various threshold settings. The *AUC* is a measure of the usefulness of a test in general, where a greater area means a more useful test. The *AUC* is computed as

$$FPR = \frac{FR}{FP + TN} \quad (8)$$

$$TPR = \frac{TP}{TP + FN} \quad (9)$$

$$AUC = \sum_{i \in (TP+FP+FN+TN)} \frac{(TPR_i + TPR_{i-1})(FPR_i + FPR_{i-1})}{2} \quad (10)$$

3. Results

Weka 3.8.3 software was used to evaluate the predictive capabilities of the used factors and based on the information gain ratio with the average merit using 10-fold cross-validation (Table 2). Aquifer saturated thickness had the highest average merit (0.31), followed by rainfall (0.238), hydraulic conductivity (0.189), depth to groundwater (0.185), specific yield (0.183), elevation (0.161), geology (0.062), fault density (0.058), drainage density (0.048), soil (0.048), LULC (0.046), and distance to faults (0.038). The average merit values for slope, TWI, and SPI are zero, and thus these factors were excluded because they were unlikely to play a role in regulating groundwater production in this area.

Based on the outcomes of the training process (Table 3) the DJ classifier had the highest *ACC* (0.94), followed by BDT (0.93), DF (0.93), LD-SVM (0.91), SVM and ANN models (both 0.86), BPM and LR (both 0.85), and AP (0.84). In terms of precision, all models performed well, with DJ presenting the highest performance (0.95), followed by ANN (0.92) and BDT (0.90). The BDT model showed the highest performance concerning the classification of high potential locations, having a recall index of 0.82, followed by LD-SVM (0.81), DJ (0.80), and DF (0.78). In terms of the F-score, the highest performance models were DJ (0.87), BDT (0.86), DF (0.83), and LD-SVM (0.83). Finally, for *AUC* values, DJ and BDT showed the

highest performance (0.95), followed by DF (0.930) and ANN (0.90). The other algorithms also performed well ($AUC > 0.80$).

Table 2. Feature selection with information gain ratio (10-fold cross-validation).

Attribute	Average Merit	Average Rank
Aquifer saturated thickness	0.310 ± 0.016	1.0 ± 0.00
Rainfall	0.238 ± 0.038	2.7 ± 1.42
Hydraulic conductivity	0.189 ± 0.013	3.3 ± 1.19
Depth to groundwater	0.185 ± 0.008	4.0 ± 0.63
Specific yield	0.183 ± 0.015	4.2 ± 0.75
Elevation	0.161 ± 0.007	5.8 ± 0.40
Geology	0.062 ± 0.004	7.3 ± 0.44
Fault density	0.058 ± 0.013	8.2 ± 0.98
Drainage density	0.048 ± 0.004	9.6 ± 1.02
Soil	0.048 ± 0.003	9.7 ± 1.10
LULC	0.046 ± 0.007	10.6 ± 1.02
Distance to faults	0.038 ± 0.005	11.6 ± 0.92
TWI	0.000 ± 0.000	13.1 ± 0.30
Slope	0.000 ± 0.000	13.9 ± 0.30
SPI	0.000 ± 0.000	15.0 ± 0.00

Table 3. Examining the performance of the models used using error measures.

Model	Error Measures Used									
	Training Phase					Testing Phase				
	ACC	AUC	Precision	Recall	F-Score	ACC	AUC	Precision	Recall	F-Score
AP	0.84	0.86	0.76	0.59	0.66	0.83	0.85	0.75	0.58	0.65
LR	0.85	0.87	0.82	0.54	0.65	0.84	0.86	0.81	0.53	0.64
BPM	0.85	0.87	0.78	0.62	0.69	0.84	0.86	0.77	0.61	0.68
NN	0.86	0.90	0.92	0.51	0.66	0.85	0.89	0.91	0.50	0.65
SVM	0.86	0.87	0.84	0.59	0.69	0.85	0.86	0.83	0.58	0.68
LD-SVM	0.91	0.91	0.85	0.81	0.83	0.90	0.90	0.84	0.80	0.82
DF	0.92	0.93	0.89	0.78	0.83	0.91	0.92	0.88	0.77	0.82
BDT	0.93	0.95	0.90	0.82	0.86	0.92	0.94	0.89	0.81	0.85
DJ	0.94	0.95	0.95	0.80	0.87	0.93	0.94	0.94	0.79	0.86

3.1. Results Validation

After training, the test dataset was passed to the algorithms (Table 3 and Figure 10). In terms of overall ACC, all models performed very well, with the best performance by DJ (0.93), BDT (0.92), DF (0.91), and LD-SVM (0.90). The remaining models also showed high ACC (>0.8). In the case of precision, the highest performance model was DJ (0.94), followed by ANN (0.91), BDT (0.89), DF (0.88), LD-SVM (0.84), SVM (0.83), LR (0.81), BPM (0.77), and AP (0.75). In terms of recall, the best performing model was BDT (0.81), followed by LD-SVM (0.80), DJ (0.79), and DF (0.77). The F-score was best for DJ (0.86), followed by BDT (0.85) and LD-SVM (0.83). The other models had F-scores between 0.60 and 0.70 and are thus classified as moderate-performance. Finally, all models had $AUC > 0.80$ (very good to excellent), with DJ and BDT having the highest values (0.94), followed by DF (0.92) and LD-SVM (0.90).

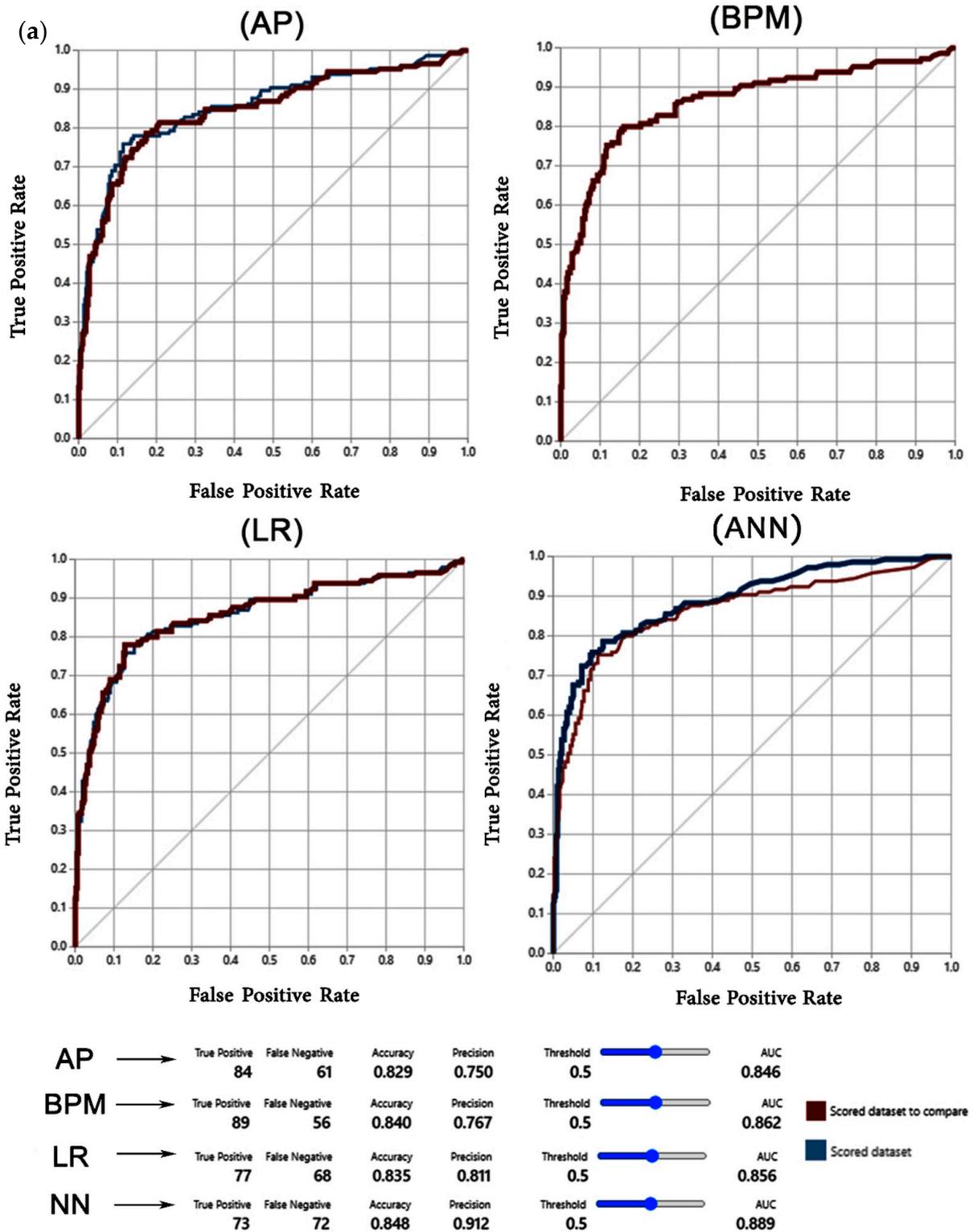


Figure 10. Cont.

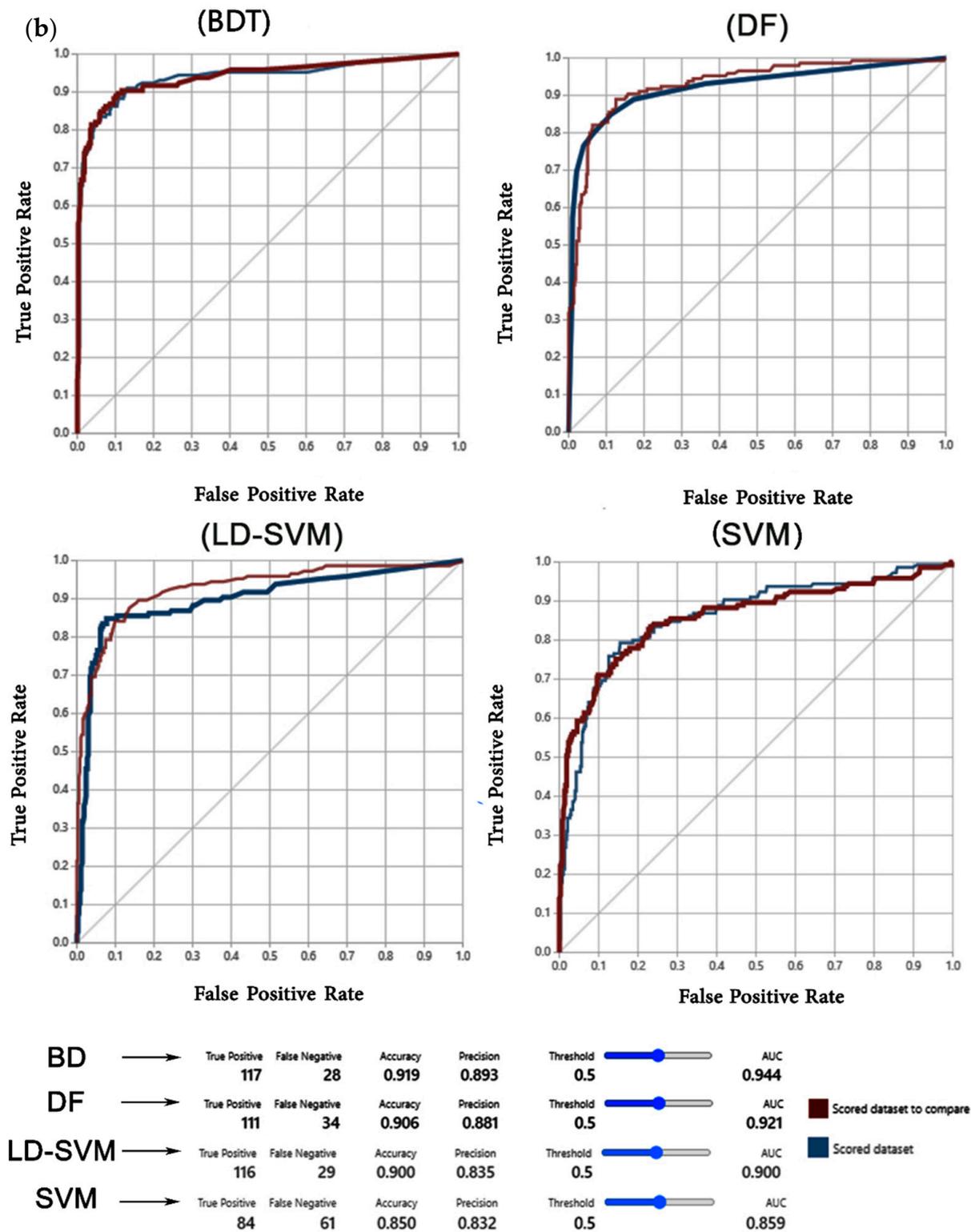


Figure 10. Cont.

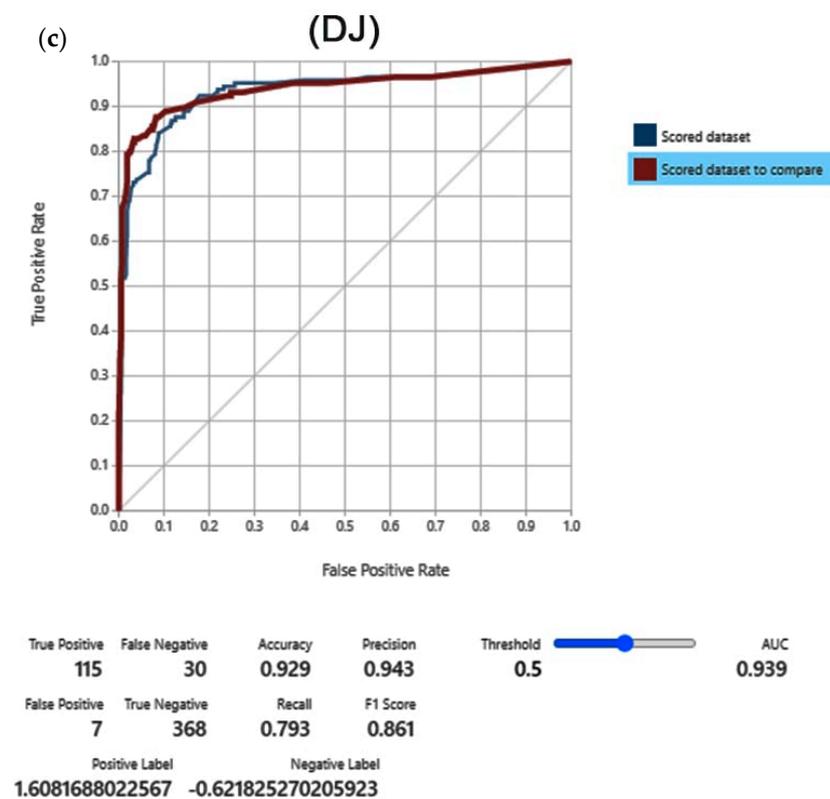


Figure 10. (a) Validation results (testing stage) for AP, BPM, LR, ANN. (b) Validation results (testing stage) for BDT, DF, LD-SVM, and SVM. (c) validation results (testing stage) for DJ model.

Training and testing revealed that the DJ, BDT, DF, and LD-SVM models performed best in demarcating the groundwater productivity in the study area. The probability values of these models for the training and testing stage were exported to ArcGIS 10.8.1 and interpolated to map groundwater potential based on the predicted probability of 0 (low) to 1 (high). We employed a natural break classification system in the GIS platform to reclassify the probability values into five groups—namely, very low, low, moderate, high, and very high (Figure 11)—to build a theme zoning map that could be more readily understood by end-users and policymakers. The natural break classification scheme is the most common system used in groundwater potential mapping [5,23,55–57]. We reduced the five groups to three zones (very low–low, moderate, and high–very high) and listed the zonal areas calculated by the best four models in Table 4. Overall, the spatial distributions of groundwater potential zones for the DJ, DF, and LD-SVM models were similar. The very low–low zone occupied 29.3–38% of the study area, mainly in the north and northeast. The moderate zone covered 31.8–33.8% of the study area, mainly in the south. The high–very high zone encompassed 34.0–36.9% of the study area, primarily in the middle. For the BDT model, the groundwater potential zones were somewhat different: the very low–low zone was distributed over large areas in the north, middle, and south (38% of the study area), the moderate zone covered 23% of the study area, and the high–very high zone encompassed 39%.

When it comes to developing strategies to manage the aquifer in the region, the discrepancy in the groundwater potential map between the second-best algorithm (BDT) and the remainder of the algorithms may cause some difficulty for hydrogeologists and decision makers. To reduce this difference (i.e., the uncertainty), the ensemble approach was used to produce three maps: the first map relied on all four best algorithms (i.e., DJ, BDT, DF, and LD-SVM), Figure 12a, while the second map was produced by the ensemble of algorithms that gave a similar pattern to the distribution of groundwater zones in the study area (DJ, DF, and LD-SVM), Figure 12b. The third map was produced by the ensemble

of all used models (Figure 12c). When comparing these two maps, it can be seen that the distribution of groundwater potential zones is very similar for the first two maps and slightly different from the third map. The very low–low zones occupied 77.6% (1976 km²) and 77.5% (1974.8 km²) of the study area, the moderate zone occupied 9.9% (251.6 km²) and 10.2% (258.4 km²), and the high–very high zones occupied 12.5% (319.4 km²), for the first and second ensemble models, respectively. In case of the third map (all used models), the very low-low zones were distributed over 72.8% (1852.9 km²) of the study area, the moderate zone covered 13.7 (348.3 km²), and the high-very high zones occupied 13.6% (345.8 km²). Perhaps, the reason for the result similarity is that all algorithms used performed well, especially in the training stage. When using the ensemble approach, the weaknesses of the algorithms that performed fairly low in the testing stage may be mitigated, which could explain why these maps show the same pattern and approximate occupied potential areas.

Table 4. Areas occupied by groundwater potential zones.

Groundwater Potential Zone	Algorithms Used							
	DJ		BDT		DF		LD-SVM	
	Area (%)	Area (km ²)	Area (%)	Area (km ²)	Area (%)	Area (km ²)	Area (%)	Area (km ²)
Very low–low	33.9	862	38	973	34.2	870	29.3	744
Moderate	32.1	816	23	576	31.8	806	33.8	856
High–very high	34.0	862	39	991	34.0	864	36.9	940

3.2. Source of Result Uncertainty

The uncertainty in the results can be attributed to either the physical model or the method used to integrate the data (i.e., the algorithm) [9]. The physical models here mean the groundwater-influencing occurrence factors. We used both surface and subsurface factors to delineate groundwater potential zones; thus, the maps produced can be used to quantify groundwater potentiality, but only for the shallow groundwater aquifer system in the study area. When attempting to assess the deep aquifers in comparison with the shallow ones, more detailed research should be conducted. Although there is no global agreement on the type and number of factors used in the analysis of groundwater potential, the inclusion of both surface and subsurface factors in this analysis makes it more realistic than using only surface factors, which only reflect the amount of renewable storage of the aquifer.

Another source of uncertainty may have resulted from the use of interpolation techniques, which may be caused by the type of interpolation technique (deterministic or stochastic), the number of samples taken, which is controlled by the nature of the field work, and the financial cost, which may vary depending on the factor used. To reduce the uncertainty in our instance, we used the stochastic kriging technique and its derivative Empirical Bayesian Kriging, which is believed to be more accurate than deterministic models, such as the inverse distance approach. Another significant source of uncertainty is attributed to the ML algorithms, particularly if there is no relationship between the groundwater influential factors and the borehole location target variable. Examining the overfitting (Table 3) revealed that there is no overfitting problem, allowing the findings of training and testing to be used to map the groundwater potential.

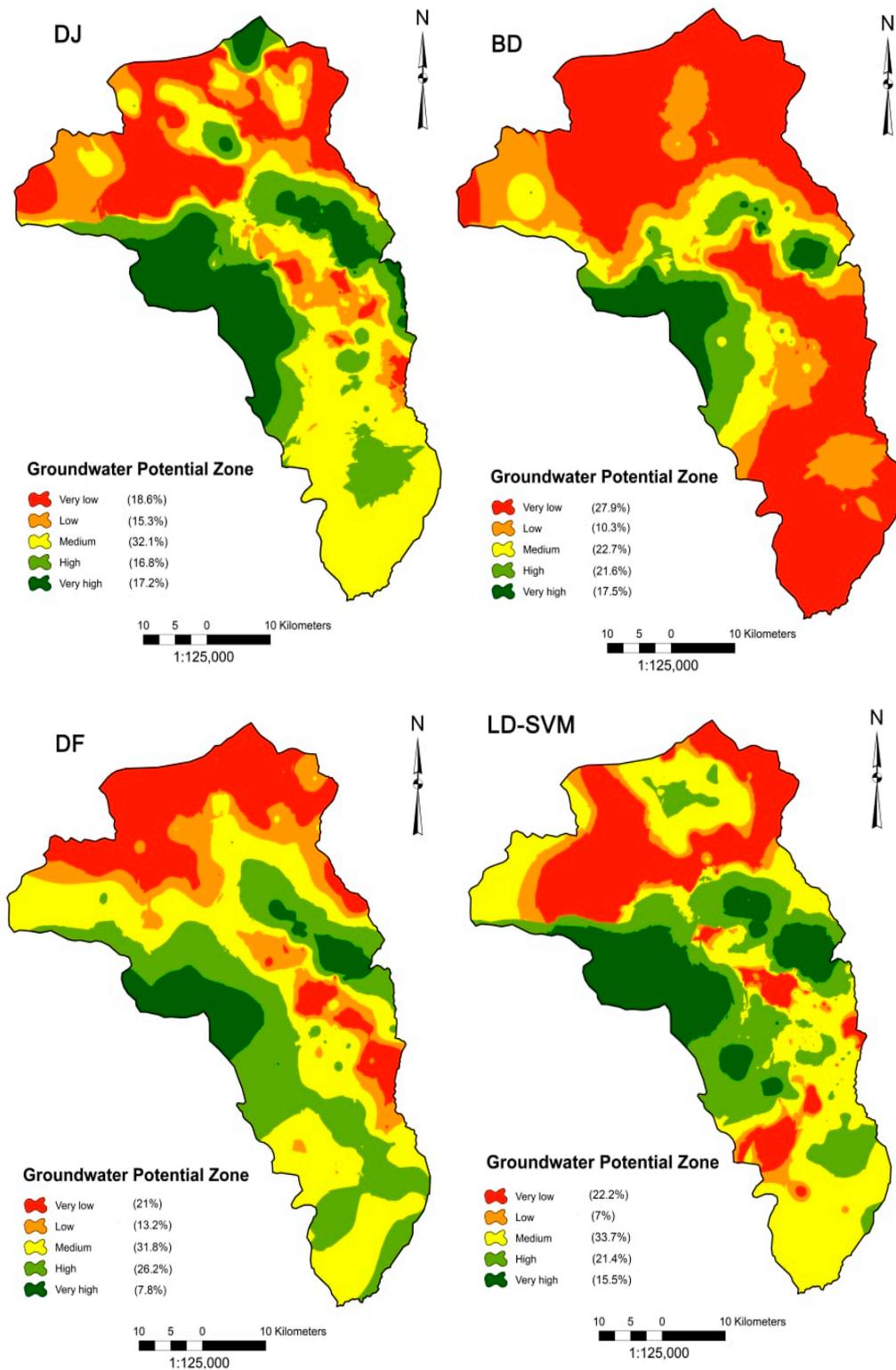


Figure 11. Groundwater potential for the best four models.

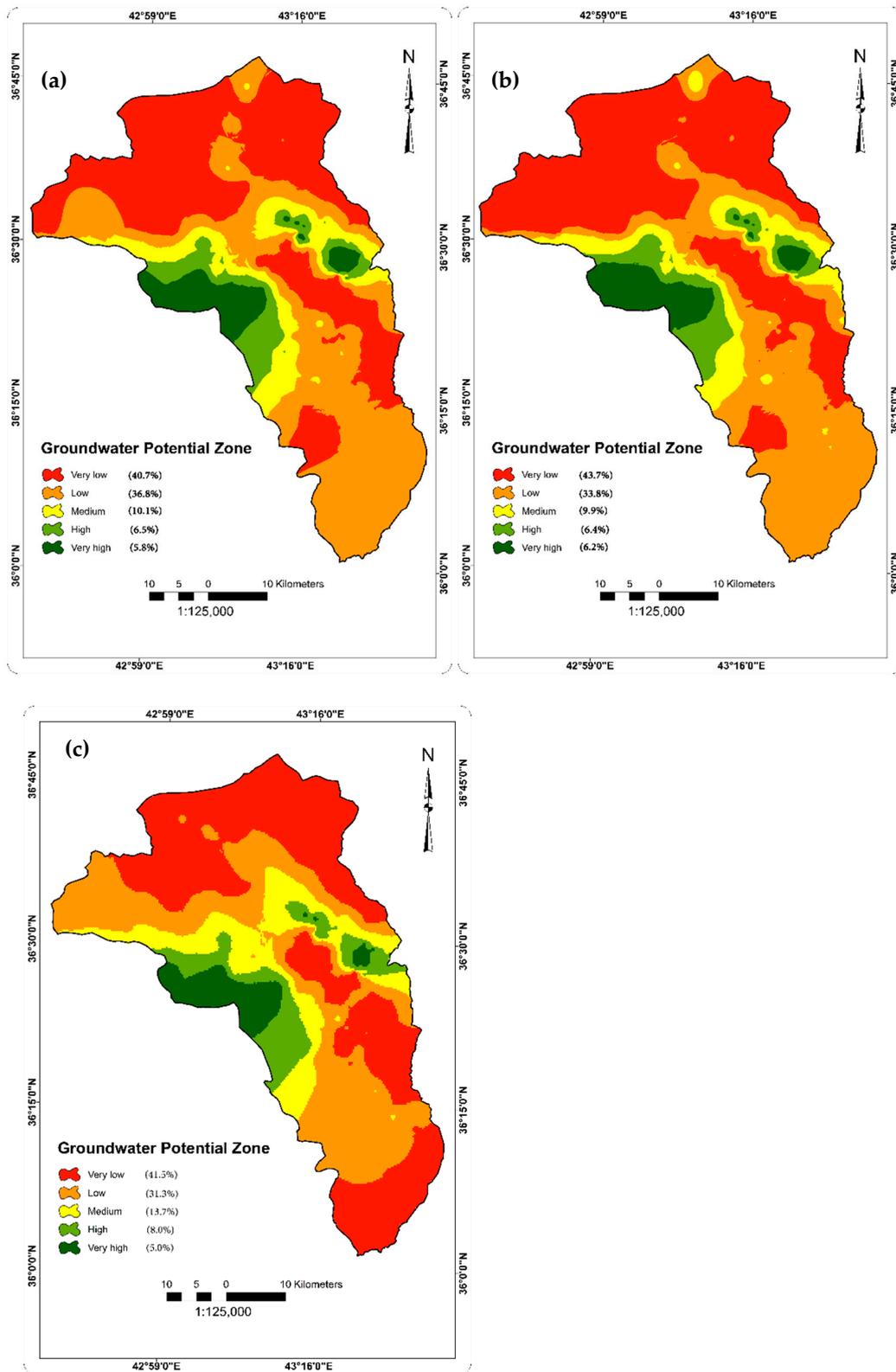


Figure 12. Groundwater potential maps by ensemble models: (a) all best performance models, (b) all similar pattern models, and (c) all used models.

4. Discussion

4.1. Contribution of Groundwater Factors in Developing Groundwater Potential Map

The findings of this research indicate that a combination of subsurface factors (aquifer saturated thickness, aquifer hydraulic conductivity, depth to groundwater, and aquifer specific yield) and surface factors (rainfall and elevation) contribute most to groundwater potential in the study area. Both transmissivity (the product of saturated thickness and hydraulic conductivity) and specific yield have a significant impact on aquifer groundwater storage fluctuation, regulating groundwater productivity and accessible groundwater for extraction. Elevation is a key morphological factor that influences the development of soil and vegetation profiles, which can affect infiltration rates, while rainfall can affect the quantity of runoff generation and percolation through the soil column [35,58]. As the only main source of recharge in this aquifer is rainfall, it is obvious why this factor was more important than other factors used in the analysis of groundwater potentiality. Geology, fault density, drainage density, soil, LULC, and distance to faults were less important (average merits < 0.1) in contributing to groundwater potential in the study area. TWI, SPI, and slope had no influence on the groundwater availability because the study area is nearly flat and there is a relatively smooth slope from northeast to southwest.

4.2. Validation of the Models

The results of the present study are in agreement with previous outcomes concerning groundwater potentiality mapping and the usage of machine learning models and their superiority in developing groundwater potential mapping [59–61]. In the present study, the DJ model achieved the highest performance for all error measures and was slightly better than BDT in both training and testing stages, followed by DF and LD-SVM. Other algorithms were good in terms of ACC, AUC, and accuracy, but in terms of recall and F-score, they had poor performance. The superiority of BDT in modeling groundwater potential and spring potential is well known [62]. However, to our knowledge, this is the first time the DJ technique has been used to model groundwater potential; therefore, this algorithm is promising in this field, as it has outperformed the master machine learning algorithms such as ANN and SVM, which are known for their predictive abilities.

4.3. Distribution of Groundwater Potential Zones

Mapping the probability of groundwater potential in the study area using the DJs, DF, and LD-SVM models showed the same pattern of potentiality distribution, and their results were different from results obtained by BDT. The ensemble technique [9,54] was used to reduce the difference, and three ensemble maps were produced: with all four best performance models, with only similar pattern algorithms, and with all used algorithms. The produced three ensemble maps revealed the same pattern and occupied approximately the same potential areas. The very low–low, moderate, and high–very high zones covered 12.4% (633.3 km²), 10.05 (510 km²), and 77.55% (3950.8 km²) of the study area, respectively.

Overall, the very low–low zones appeared in the areas where the saturated thickness of the aquifer and the hydraulic characteristics had low values compared with other parts of the study area, and this may have been why these portions of the aquifer have low production values, even though these areas receive more rainfall than the middle and southern parts. At the same time, these portions also have a high elevation range and steep slope, and thus, most of the falling rainfall may turn into surface runoff instead of infiltrating to enhance the groundwater storage. On the other hand, high and very high potentiality zones were distributed in the east and west of the middle of the study area. The moderate zone was distributed unevenly between the other zones, and a large part of it was concentrated in the southern part. In general, it can be said that the high potential occupied the middle part of the study area, the moderate zone encompassed the southern part, and the low potential zone covered the northern part. The distribution of potentiality zones was closely related to the subsurface factors that control the ease with which groundwater flows through the aquifer material and the capability of the aquifer

material to store groundwater, i.e., the aquifer transmissivity and storativity. These factors in most of the studies related to groundwater potential are neglected because, in most cases, it is difficult or impossible to collect enough data to define the spatial distribution of these thematic layers, especially in poor data countries. Thus, most research depends only on the surficial factors that only determine the renewable storage, which in most cases, especially in arid and semi-arid regions, represents only a small part of the storage available in the aquifer being studied. Therefore, the study of the aquifer itself and its characteristics is very important for mapping the groundwater potential, in addition to the factors that only control the renewable amount of water entering the aquifer system, such as soil, LULC, and topographical-related factors.

5. Concluding Remarks

In this study, the Microsoft Azure cloud computing service was used for modeling groundwater potential in the Nineveh Plain of northern Iraq using nine machine learning algorithms—namely, ANN, AP, SVM, LD-SVM, BPM, DFs, DJs, BDT, and LG. The following conclusions were drawn from this study. First, the most important factors in groundwater potential are aquifer saturated thickness, followed by rainfall, hydraulic conductivity, depth to groundwater, and specific yield. Most of these are subsurface factors that control the ability of the aquifer to transmit and store water, and it is important to incorporate these factors in developing groundwater potential maps. The other factors that have a role in controlling groundwater potentiality include elevation, geology, fault density, drainage density, soil, and LULC. Second, the DJ and BDT models were the best-performing ML models, followed by DF and LD-SVM. The DJ model, which does not appear to have been used for modeling groundwater potential mapping, outperformed well-known, highly efficient ML algorithms. Third, the high–very high groundwater potential zone in the study area mostly covers the middle part and is associated with high values of aquifer hydraulic characteristics. The ensemble technique is very efficient for reducing the uncertainty in groundwater potential analysis and can be used when the spatial distribution of groundwater potentiality zones produced by the different algorithms is different. Finally, using a cloud computing service offers an advanced environment in which to run and test multiple algorithms for modeling groundwater potential in a fast and inexpensive way.

Author Contributions: A.Z.A.-O.: methodology, formal analysis, writing—original draft, and data curation; A.M.A.-A.: conceptualization, supervision, writing—review and editing, and methodology; T.A.H.: methodology, data curation, and formal analysis; A.E.F.: supervision, writing—review and editing; B.P.: supervision, writing—review and editing and funding acquisition; K.N.A.M.: writing—review and editing; A.A.: writing—review and editing and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney. This research was also supported by Researchers Supporting Project number RSP-2021/14, King Saud University, Riyadh, Saudi Arabia. Additionally, support was provided by UKM—Industry Grant: ZF-2021-003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The publicly archived datasets used for the analysis are cited in the manuscript. The analysis was carried out at University of Basrah, Basrah, Iraq, and the derived data can be provided upon request to the corresponding author (Biswajeet.Pradhan@uts.edu.au).

Acknowledgments: Authors would like to express their gratitude to the editorial board and the respected reviewers for their critical comments, which helped us to revise the manuscript, particularly reviewer #1, whose comments were extremely helpful in bringing the manuscript to its current form.

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

References

- Zektser, I.; Everett, L. *Groundwater Resources of the World and Their Use, IHP-VI Ser*; International Groundwater Resources Assessment Centre: Delft, The Netherlands, 2004; Volume 6, 346p.
- Famiglietti, J.S.; Lo, M.; Ho, S.L.; Bethune, J.; Anderson, K.; Syed, T.H.; Swenson, S.C.; de Linage, C.R.; Rodell, M. Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophys. Res. Lett.* **2011**, *38*, 1–4. [[CrossRef](#)]
- Winter, T.C.; Harvey, J.W.; Franke, O.L.; Alley, W.M. *Ground Water and Surface Water. A Single Resource. USGS Circular*; Diane Publishing: Collingdale, PA, USA, 1999; Volume 1139.
- Díaz-Alcaide, S.; Martínez-Santos, P. Advances in groundwater potential mapping. *Hydrogeol. J.* **2019**, *27*, 2307–2324. [[CrossRef](#)]
- Al-Abadi, A.M.; Handhal, A.M.; Al-Ginamy, M.A. Evaluating the Dibdibba Aquifer Productivity at the Karbala–Najaf Plateau (Central Iraq) Using GIS-Based Tree Machine Learning Algorithms. *Nat. Resour. Res.* **2020**, *29*, 1989–2009. [[CrossRef](#)]
- 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](#)]
- 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](#)]
- Jha, M.K.; Chowdary, V.M.; Chowdhury, A. Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques. *J. Hydrol.* **2010**, *18*, 1713–1728. [[CrossRef](#)]
- Çelik, R. Evaluation of groundwater potential by GIS-based multicriteria decision making as a spatial prediction tool: Case study in the Tigris River Batman-Hasankeyf Sub-Basin, Turkey. *Water* **2019**, *11*, 2630. [[CrossRef](#)]
- Das, N.; Mukhopadhyay, S. Application of multi-criteria decision making technique for the assessment of groundwater potential zones: A study on Birbhum district, West Bengal, India. *Environ. Dev. Sustain.* **2020**, *22*, 931–955. [[CrossRef](#)]
- Singh, L.K.; Jha, M.K.; Chowdary, V. Assessing the accuracy of GIS-based Multi-Criteria Decision Analysis approaches for mapping groundwater potential. *Ecol. Indic.* **2018**, *91*, 24–37. [[CrossRef](#)]
- Al-Abadi, A.M. Modeling of groundwater productivity in northeastern Wasit Governorate, Iraq using frequency ratio and Shannon's entropy Models. *Appl. Water Sci.* **2017**, *7*, 699–716. [[CrossRef](#)]
- Al-Abadi, A.M.; Fryar, A.E.; Rasheed, A.A.; Pradhan, B. Assessment of groundwater potential in terms of the availability and quality of the resource: A case study from Iraq. *Environ. Earth Sci.* **2021**, *80*, 1–22. [[CrossRef](#)]
- Davoudi Moghaddam, 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](#)]
- Abd Manap, M.; Sulaiman, W.N.A.; Ramli, M.F.; Pradhan, B.; Surip, N. A knowledge-driven GIS modeling technique for groundwater potential mapping at the Upper Langat Basin, Malaysia. *Arab. J. Geosci.* **2013**, *6*, 1621–1637. [[CrossRef](#)]
- Arulbalaji, P.; Padmalal, D.; Sreelash, K. GIS and AHP techniques based delineation of groundwater potential zones: A case study from southern Western Ghats, India. *Sci. Rep.* **2019**, *9*, 1–17. [[CrossRef](#)]
- Kumar, T.; Gautam, A.K.; Kumar, T. Appraising the accuracy of GIS-based multi-criteria decision making technique for delineation of groundwater potential zones. *Water Resour. Manag.* **2014**, *28*, 4449–4466. [[CrossRef](#)]
- 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.* **2015**, *8*, 7059–7071. [[CrossRef](#)]
- 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](#)]
- Das, S. Comparison among influencing factor, frequency ratio, and analytical hierarchy process techniques for groundwater potential zonation in Vaitarna basin, Maharashtra, India. *Groundw. Sustain. Dev.* **2019**, *8*, 617–629. [[CrossRef](#)]
- 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](#)]
- Nguyen, P.T.; Ha, D.H.; Avand, M.; Jaafari, A.; Nguyen, H.D.; Al-Ansari, N.; Van Phong, T.; Sharma, R.; Kumar, R.; Le, H.V. Soft computing ensemble models based on logistic regression for groundwater potential mapping. *Appl. Sci.* **2020**, *10*, 2469. [[CrossRef](#)]
- 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](#)]
- Hayley, K. The present state and future application of cloud computing for numerical groundwater modeling. *Groundwater* **2017**, *55*, 678–682. [[CrossRef](#)]
- Armbrust, M.; Fox, A.; Griffith, R.; Joseph, A.D.; Katz, R.; Konwinski, A.; Lee, G.; Patterson, D.; Rabkin, A.; Stoica, I. A view of cloud computing. *Commun. ACM* **2010**, *53*, 50–58. [[CrossRef](#)]
- Ganeshkumar, M.; Ramesh, V. A Study on Digital India Programme Using Azure Cloud and Twitter Data. *Int. J. Comput. Intell. Res.* **2017**, *13*, 781–790.
- Hayley, K.; Schumacher, J.; MacMillan, G.; Boutin, L. Highly parameterized model calibration with cloud computing: An example of regional flow model calibration in northeast Alberta, Canada. *Hydrogeol. J.* **2014**, *22*, 729–737. [[CrossRef](#)]
- Hunt, R.J.; Luchette, J.; Schreuder, W.A.; Rumbaugh, J.O.; Doherty, J.; Tonkin, M.J.; Rumbaugh, D.B. Using a cloud to replenish parched groundwater modeling efforts. *Groundwater* **2010**, *48*, 360–365. [[CrossRef](#)]
- Wang, D.; Xu, H.; Shi, Y.; Ding, Z.; Deng, Z.; Liu, Z.; Xu, X.; Lu, Z.; Wang, G.; Cheng, Z. The groundwater potential assessment system based on cloud computing: A case study in islands region. *Comput. Commun.* **2021**, *178*, 83–97. [[CrossRef](#)]

30. Jassim, S.Z.; Goff, J.C. *Geology of Iraq*; Dolin, Prague and Moravian Museum: Brno, Czech Republic, 2006.
31. Driscoll, F.G. *Groundwater and Wells*; Johnson Screens; U.S. Geological Survey: Washington, DC, USA, 1986; 799p.
32. Al-Abadi, A.M.; Al-Temmeme, A.A.; Al-Ghanimy, M.A. A GIS-based combining of frequency ratio and index of entropy approaches for mapping groundwater availability zones at Badra–Al Al-Gharbi–Teeb areas, Iraq. *Sustain. Water Resour. Manag.* **2016**, *2*, 265–283. [[CrossRef](#)]
33. Wada, Y.; Van Beek, L.P.; Van Kempen, C.M.; Reckman, J.W.; Vasak, S.; Bierkens, M.F. Global depletion of groundwater resources. *Geophys. Res. Lett.* **2010**, *37*, 1–5. [[CrossRef](#)]
34. Al-Abadi, A.M.; Shahid, S. A comparison between index of entropy and catastrophe theory methods for mapping groundwater potential in an arid region. *Environ. Monit. Assess.* **2015**, *187*, 1–21. [[CrossRef](#)] [[PubMed](#)]
35. Oh, H.-J.; Kim, Y.-S.; Choi, J.-K.; Park, E.; Lee, S. GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *J. Hydrol.* **2011**, *399*, 158–172. [[CrossRef](#)]
36. Ozdemir, A. Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *J. Hydrol.* **2011**, *405*, 123–136. [[CrossRef](#)]
37. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrol. Sci. J.* **1979**, *24*, 43–69. [[CrossRef](#)]
38. Moore, I.D.; Grayson, R.; Ladson, A. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30. [[CrossRef](#)]
39. Naghibi, S.A.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Rezaei, A. Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran. *Earth Sci. Inform.* **2015**, *8*, 171–186. [[CrossRef](#)]
40. Alozeer, A.; Abdaki, M.A.; Al-Iraqi, A.; Al-Samman, S.; Al-Hammadi, N. Estimation of mean areal rainfall and missing data by using gis in nineveh, Northern Iraq. *IGJ* **2020**, 93–103. [[CrossRef](#)]
41. Subasi, A. *Practical Machine Learning for Data Analysis Using Python*; Academic Press: Cambridge, MA, USA, 2020.
42. Brownlee, J. *Master Machine Learning Algorithms: Discover how They Work and Implement Them from Scratch*; Machine Learning Mastery: Boston, MA, USA, 2016.
43. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378.
44. Wells, J.C.; Hung, T.T. Longman pronunciation dictionary. *RELC J.* **1990**, *21*, 95–97. [[CrossRef](#)]
45. Hassoun, M.H. *Fundamentals of Artificial Neural Networks*; MIT Press: Cambridge, MA, USA, 1995.
46. Yegnanarayana, B. *Artificial Neural Networks*; PHI Learning Pvt. Ltd.: New Delhi, India, 2009.
47. Priddy, K.L.; Keller, P.E. *Artificial Neural Networks: An Introduction*; SPIE Press: Bellingham, WA, USA, 2005.
48. Shanmuganathan, S. Artificial neural network modelling: An introduction. In *Artificial Neural Network Modelling*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 1–14.
49. Suthaharan, S. Machine learning models and algorithms for big data classification. *Integr. Ser. Inf. Syst* **2016**, *36*, 1–12.
50. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
51. Herbrich, R.; Graepel, T.; Campbell, C. Bayes point machines. *J. Mach. Learn. Res.* **2001**, *1*, 245–279.
52. Shotton, J.; Sharp, T.; Kohli, P.; Nowozin, S.; Winn, J.; Criminisi, A. Decision jungles: Compact and rich models for classification. In Proceedings of the NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 234–242.
53. Ghasemkhani, N.; Vayghan, S.S.; Abdollahi, A.; Pradhan, B.; Alamri, A. Urban Development Modeling Using Integrated Fuzzy Systems, Ordered Weighted Averaging (OWA), and Geospatial Techniques. *Sustainability* **2020**, *12*, 809. [[CrossRef](#)]
54. Arabameri, A.; Rezaei, K.; Cerda, A.; Lombardo, L.; Rodrigo-Comino, J. GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *Sci. Total Environ.* **2019**, *658*, 160–177. [[CrossRef](#)] [[PubMed](#)]
55. Razavi-Termeh, S.V.; Sadeghi-Niaraki, A.; Choi, S.-M. Groundwater potential mapping using an integrated ensemble of three bivariate statistical models with random forest and logistic model tree models. *Water* **2019**, *11*, 1596. [[CrossRef](#)]
56. 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](#)]
57. 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*, 1–27. [[CrossRef](#)] [[PubMed](#)]
58. Lee, S.; Hyun, Y.; Lee, S.; Lee, M.-J. Groundwater potential mapping using remote sensing and GIS-based machine learning techniques. *Remote Sens.* **2020**, *12*, 1200. [[CrossRef](#)]
59. Prasad, P.; Loveson, V.J.; Kotha, M.; Yadav, R. Application of machine learning techniques in groundwater potential mapping along the west coast of India. *GIScience Remote Sens.* **2020**, *57*, 735–752. [[CrossRef](#)]
60. 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](#)]

-
61. Kamali Maskooni, E.; Naghibi, S.A.; Hashemi, H.; Berndtsson, R. Application of advanced machine learning algorithms to assess groundwater potential using remote sensing-derived data. *Remote Sens.* **2020**, *12*, 2742. [[CrossRef](#)]
 62. Sachdeva, S.; Kumar, B. Comparison of gradient boosted decision trees and random forest for groundwater potential mapping in Dholpur (Rajasthan), India. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 287–306. [[CrossRef](#)]