GROUNDWATER POTENTIAL ASSESSMENT IN GIA LAI PROVINCE (VIETNAM) USING MACHINE LEARNING, REMOTE SENSING AND GIS

Huu Duy NGUYEN [1], Van Trong GIANG [2]*, Quang-Hai TRUONG [2], Gheorghe ȘERBAN [3]
Alexandru-Ionut PETRIȘOR [4,5,6,7]

DOI: 10.21163/GT_2024.192.02

ABSTRACT:
Population growth, urbanization and rapid industrial development increase the demand for water resources. Groundwater is an important resource in sustainable socio-economic development. The identification of regions with the probability of the existence of groundwater is necessary in helping decision makers to propose effective strategies for the management of this resource. The objective of this study is to construct maps of potential groundwater, based on machine learning algorithms, namely deep neural networks (DNNs), XGBoost (XGB), and CatBoost (CB), in the Gia Lai province of Vietnam. In this study, 12 conditioning factors, namely elevation, aspect, curvature, slope, soil type, river density, distance to road, land use/land cover (LULC), Normalized Difference Vegetation Index (NDVI), Normal Difference Built-up Index (NDBI), Normal Difference Water Index (NDWI), and rainfall were used, along with 181 groundwater inventory points, to construct the models. The proposed models were evaluated using the receiver operating characteristic (ROC) curve, the area under the curve (AUC), root-mean-square error (RMSE), mean absolute error (MAE). The results showed that the predictions of groundwater potential were most accurate using the XGB model; CB came second, and DNN was performed the least well. About 4,990 km^2 of the study area was found to be in the category of very low groundwater potential; 3,045 km^2 was in the low category; 2,426 km^2 was classified as moderate, 2,665 km^2 as high, and 2,007 km^2 as very high. The methodology used in the study was effective in creating groundwater potential maps. This approach, used in this study, can provide valuable information on the factors influencing groundwater potential and assist decision-makers or developers in managing groundwater resources sustainably. It also supports the sustainable development of the territory, including tourism. This methodology can be used in other geographic regions with a small change of input data.

Key-words: *Groundwater potential, Deep neural network, XGBoost, CatBoost, Gia Lai, Vietnam.*

1. INTRODUCTION

Groundwater is indispensable to humanity. Some 2.5 billion people depend on this resource every day (Prasad, Loveson et al. 2020, Gómez-Escalonilla, Martínez-Santos et al. 2022). The demand for groundwater is increasing rapidly, due to population growth and industrial development (Dey, Abir et al. 2023, Morgan, Madani et al. 2023). Today, 1.9 billion people (27% of the world's population) live in areas that can potentially be affected by severe water scarcity (Boretti and Rosa 2019). In 2050, it is estimated that this figure will increase to 2.7-3.2 billion, according to differing climate change scenarios (Boretti and Rosa 2019, Morgan, Madani et al. 2023).

[1] *Faculty of Geography, National University of Vietnam, Hanoi, nguyenhuuduy@hus.edu.vn*
[2*] *Institute of Vietnamese Studies and Development Science, Vietnam National University, Hanoi, Corresponding author: tronggv@vnu.edu.vn, truongquanghai.ivides@gmail.com*
[3] *Faculty of Geography, Babes-Bolyai University, Cluj-Napoca, Romania; gheorghe.serban@ubbcluj.ro*
[4] *Doctoral School of Urban Planning, Ion Mincu University of Architecture and Urbanism, Bucharest, Romania, alexandru_petrisor@yahoo.com*
[5] *Department of Architecture, Faculty of Architecture and Urban Planning, Technical University of Moldova, Chisinau, Moldova*
[6] *National Institute for Research and Development in Constructions, Urbanism and Sustainable Spatial Development URBAN-INCERC, Bucharest, Romania*
[7] *National Institute for Research and Development in Tourism, Bucharest, Romania*

In Vietnam, survey data from the National Water Resources Planning and Investigation Centre show that the total national groundwater resource is around 91 billion cubic meters per year (250.7 million cubic meters per day), out of which the share of fresh water represents approximately 69 billion cubic meters per year (189.3 million cubic meters per day (Nguyen et al. 2024). Despite these figures, the country is currently facing severe pressure from decreasing water levels, and environmental pollution due to socio-economic growth and climate change. Thus, insufficient access to drinking water has become a major concern, and a key indicator of sustainable development according to international bodies (Kamali Maskooni, Naghibi et al. 2020, Prasad, Loveson et al. 2020). Therefore, the construction of groundwater potential maps plays an important role in optimizing water resources.

Several studies on the construction of such maps have been carried out using various approaches. The traditional method is largely based on drilling samples during field missions and hydrogeological testing in laboratories (Nguyen et al. 2024, Singh et al. 2024, Uddin et al. 2024). Although these methods can precisely identify regions with the probability of occurrence of potential groundwater, they are very expensive and time consuming and their application is limited to wide areas only. In recent years, with the development of remote sensing data and GIS technology, they have been effectively integrated into knowledge-based models such as AHP and weights-of-evidence used to delineate groundwater prospecting areas (Çelik et al. 2024, Diriba et al. 2024). However, these models depend on the opinion of experts, so such assessments are inherently subjective.

Currently, with increasing computing power and advancements in algorithms, data-driven models are widely used by researchers to construct groundwater potential maps. The statistical technique has been considered one of the most appropriate methods for constructing maps at scales of 1:20,000 or 1:50,000. Popular statistical models used include boosted regression tree (Naghibi et al., 2016, Sachdeva and Kumar 2021) and weights-of-evidence (Lee et al., 2012, Tahmassebipoor et al., 2016). However, these models do not consider nonlinear relationships, which can lead to errors in the identification of areas with probability of groundwater occurrence, particularly in the context of climate change. Machine learning models have been developed to solve these problems; these models include support vector machines (Lee et al, 2018, Anh et al., 2023), random forest (Naghibi et al., 2016, Rahmati et al., 2016), AdaBoost (Mosavi et al., 2021), and artificial neural networks (Nguyen et al., 2020, Tamiru and Wagari 2022). The goal of each study is to find the most effective model. Machine learning has the advantage of being able to mimic the complex, nonlinear relationships between groundwater points and environmental conditions, hydrology, climate and human activities without requiring an underlying understanding of physics. However, according to literature reviews, there are several machine learning models, and each model has advantages and disadvantages. Thus, selecting an appropriate model for specific data is a complex process. Additionally, extrapolation issues are considered a major challenge when using machine learning. This is because a machine learning model might not be able to make accurate predictions on data that differs significantly from the data used during training (Nguyen et al., 2024).

A number of studies have been carried out in recent years in Vietnam to assess groundwater potential. Bien et al. (2023) used five machine learning algorithms, namely partial decision trees (PART), Fuzzy Unordered Rule Induction Algorithm (FURIA), multilayer perceptron (MLP), Forest by Penalizing Attributes (FPA), and the DECORATE ensemble of learning techniques, to construct groundwater potential maps for the Central Highlands of Vietnam. The results showed that the DECORATE model performed better than the other ones. Nguyen et al. (2024) integrated deep neural networks (DNNs) with the optimization algorithms Adam, Flower Pollination Algorithm, Artificial Ecosystem-Based Optimization, Pathfinder Algorithm, African Vulture Optimization Algorithm (AVOA), and Whale Optimization Algorithm to predict groundwater potential in the Central region of Vietnam. The results indicated the DNN-AVOA model to be more effective than the others. Ha et al. (2021) applied machine learning ensemble models, namely Adaboost‑Quadratic Discriminant Analysis (ABQDA), Multiboost-Quadratic Discriminant Analysis, RealAdaboost-Quadratic Discriminant Analysis, to identify regions of probability of groundwater occurrence in Dak Nong province, Vietnam. The results showed that the ABQDA model was more accurate than the others.

Although there are several studies that have used machine learning to predict groundwater potential around the world and in Vietnam in particular, models have achieved differing levels of accuracy in different regions; there are no universally appropriate models. Therefore, it is necessary to find the appropriate algorithm(s) for each region to properly support decision makers and developers in the effective management of water resources, in order to better serve socio-economic development (Kumar et al., 2023; Masroor et al., 2023).

The objective of this study is to evaluate the potential of groundwater by applying machine learning algorithms, namely deep neural networks, CatBoost (CB), and XGBoost (XGB) in the province of Gia Lai in Vietnam. The novelty of this study is that for the first time these machine learning models are used to predict groundwater potential in this study area, where the problem of managing water resources is a great challenge, particularly in the context of climate change. This study also fills important knowledge gaps on groundwater potential investigations in Gia Lai Province, Vietnam. The results can support decision makers in proposing effective strategies in the management and optimization of water resources, not only in a specific region, but in other similar regions in the world (with minor changes in the input data).

2. STUDY AREA AND DATA USE

2.1. Study area

Gia Lai Province (12° 58’ 40” -14° 37’ 00” N, 107° 28’ 04” -108° 54’ 40” E) covers a natural area of 15,536.9 km2 in the north of Vietnam’ s Central Highlands. Gia Lai is located almost entirely west of the Truong Son range (**Fig. 1**).

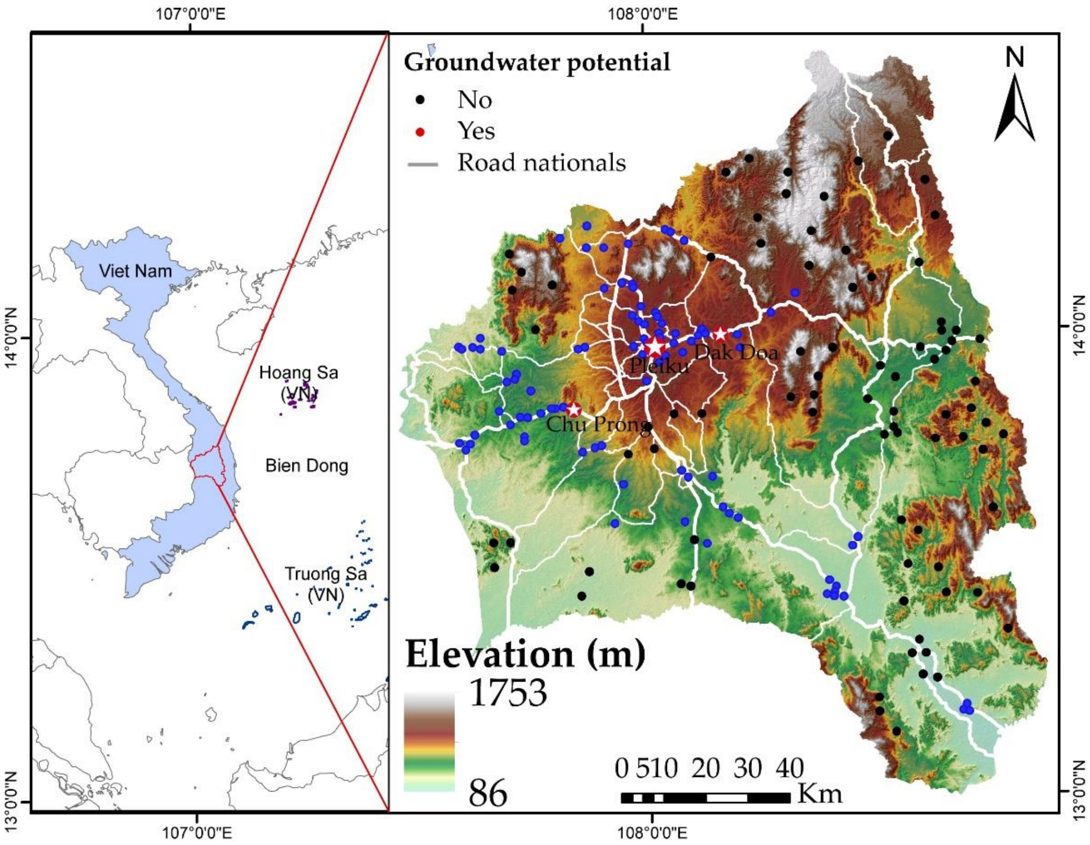

Fig. 1. Location of the study area.

The terrain gradually lowers from north to south and tilts from east to west, with quite complex alternating hills, plateaus and valleys. The average altitude is from 800 to 900 meters above sea level. The highest point is Kon Ka Kinh (1,748 m) in K'Bang district, while the lowest point is downstream of Ba River (100 m). The plateau is a popular and important terrain form of Gia Lai, with two plateaus made of basalt rock: Kon Ha Nung and Pleiku. Mountainous terrain accounts for 2/5 of the entire province's area, most of it located in the north, the mountain terrain is strongly separated, the surfaces of other terrain types of Gia Lai such as plateaus and valleys are also scattered with mountain. Relatively low-lying areas often form rivers when passing through sudden fault zones, typically An Khe valley in the East and Cheo Reo-Phu Tuc valley in the Southeast of the province.

Gia Lai has a provincial capital city, Pleiku, and two towns, An Khe, located in the West, and Ayunpa, located in the Southeast of the province. Pleiku City is located on Road 14, connecting with the provincial cities in the North and South, Kon Tum and Ban Me Thuat. Road 19 connects Pleiku City with the coastal city of Quy Nhon (Binh Dinh province) through An Khe town. Road 26 runs through Pleiku City, Ayunpa town to Tuy Hoa City (Phu Yen province). The province has an abundance of surface water, estimated at 23 billion m^3, distributed across the major river systems of the Ba and the Se San and tributaries of the Serepok. Groundwater potential is also considerable, and mainly concentrated in the basalt eruption water system, with total reserves of level A+B reaching 23,894 m3/day, level C1 61,065 m^3/day, and level C2 989,600 m^3/day. The surface water system also meets the community water needs of the province.

The climate is high-altitude tropical monsoon and is characterized by two distinct seasons: the rainy season lasts from May to October, and the dry season from November to April. The western Truong Son region enjoys average rainfall ranging between 2,200 and 2,500 mm, while in eastern Truong Son the figure is between 1,200- and 1,750-mm. Rainfall in the rainy season months accounts for about 75% of the total annual rainfall. Gia Lai province has 27 soil types, grouped into 7 main categories: red-brown and yellow-brown soils growing on basic and neutral igneous rocks (Rhodic Ferralsols), yellow-red soils on acidic magma soils (Ferralic Acrisols), red-yellow humus soils on mountains (Humic Acrisols), yellow-red soils changed by planting rice (Plinthic Acrisols), alluvial soils that are not deposited annually (Distric Fluvisols), eroded soils with bare stones (Lithic Leptosols).

Although Gia Lai province has abundant surface water sources, their distribution is uneven spatial and seasonal depending. In addition, the demand for domestic and irrigation water for industrial crops such as coffee, pepper, and cashew, as well as for fruit trees, is very high and increasing in recent years. En final, the Gia Lai province frequently experiences drought conditions. For example, in 2019, drought caused damage to thousands of acres of agricultural land and the damage to the economy was estimated at around 2 million dollars. Therefore, the construction of groundwater potential maps for the province is essential.

2.2. Data use

Well yields

Springs and wells are the points on land surfaces where groundwater is present. Spring and well data play a key role in the use of machine learning to predict groundwater potential. Thee is a complex relationships between groundwater points and natural and socioeconomic characteristics (Naghibi and Pourghasemi 2015, Prasad et al., 2020). Due to the high costs in the sampling process, spring and well data from previous studies were collected from the Department of Agriculture and Rural Development of Gia Lai Province and from a field mission, using GPS. 89 springs and well points were used as the final input data for the machine learning models.

This study used the binary model to identify areas with the probability of occurrence of groundwater potential. Therefore, the collection of non-spring and non-well samples was necessary to ensure the accuracy of the prediction model (Nguyen et al., 2024, Sharma et al., 2024, Singh et al., 2024). For the spatial model, several researchers have recommended that the number of spring and well points to be equal to the number of non-spring and non-well points.

Others have recommended that the number of non-spring and non-well points to be greater, if the study area is large (Arabameri et al., 2021). As Gia Lai is considered moderate in size, we collected an equal number of each of the two sets of points. 82 non-spring and non-well points were collected randomly throughout the study region with ArcGIS 10.6 software. A total of 181 data points were prepared to build the machine learning models. These points were identified as either 1 (for spring and well points) or 0 (for non-spring and non-well points).

Groundwater influencing factors

The selection of factors influencing the probability of groundwater occurrence is a difficult task, due to the complex and nonlinear nature of groundwater. There are no standard guidelines for their selection. The ambitious goal of this study was to integrate as many factors as possible. In the end, 12 influencing factors were selected, namely: elevation, curvature, aspect, slope (**Fig. 2a**); river density, distance to road, NDVI, NDBI (**Fig. 2b**); NDWI, rainfall, soil type and LULC (**Fig. 2c**).

Elevation, aspect, curvature, and slope were extracted from the ALOS PALSAR digital elevation model with a 12.5 m resolution. A topography map with a 1/50,000 scale was used to determine river density and distance to road with the ArcGis 10.6 Line Density and Euclidean Distance tool.

September 2023 Sentinel 2A imagery was used to extract NDVI, NDBI, and NDWI. 2023 LULC data was obtained from https://www.arcgis.com/apps/instant/media/index.html?appid=fc92d38533d440078f17678ebc20e8e2&fbclid=IwAR0V3ZEdUqhn79qN_JNPMtswxWfi2dE1_Gj-1ZD_XcN7oPyGMSn3-scE9KY

Annual rainfall from 2022 was accessed at https://chrsdata.eng.uci.edu/ . All these factors have been re-sampled at 12.5 m resolution using the ArcGIS 10.6 software.

Elevation and slope play an important role in the probability of the occurrence of groundwater resources, because altitude is directly linked to surface vegetation and recharge resources. In flat and low-lying regions, rainwater has more time to infiltrate and recharge groundwater. In the study region, the altitude ranges from 86 to 1753 m. The low-altitude regions are concentrated in the south and west (Ehsan et al., 2024, Oguntoyinbo et al., 2024). High altitude areas are distributed in the Kon Ka Kinh National Park area in the Northeast of the province.

Slope is another important topographical factor for a groundwater potential model because it is directly tied to the hydrological process and soil infiltration capacity. Regions with low slopes have the tendency to concentrate recharge resources (Raj et al., 2024). Areas with large slopes are concentrated in the eastern and northern mountains, and areas with small slopes are distributed in the Pleiku plateau area, located in the central and western areas of the province, and along the An Khe and Cheo valleys, Reo- Phu Tuc in the East and Southeast of the province.

Curvature was selected as a conditioning factor because it is directly related to the capacity of water accumulation and infiltration in the aquifer (Ray 2024). In Gia Lai, the curvature difference is not large by region, but tends to be small in river valleys, especially Ba River.

Aspect is linked to evaporation capacity and describes the flow direction, which strongly influences the recharge capacity of a region (Ehsan et al., 2024, Sharma et al., 2024). In the study area, the aspect is more complex in the northern mountainous areas and partly in the east.

LULC was chosen as it strongly influences hydrological processes, for example infiltration capacity, evaporation, and surface flow. The change of of land use has a significant influence on the recharge capacity of the groundwater resource. The increase in the surface area reduces the infiltration capacity of the soil, which leads to a reduction in the recharge capacity of the groundwater (Ray 2024, Tiwari et al., 2024). In Gia Lai, urban and construction lands are distributed in Pleiku city, An Khe and Ayunpa towns and along national and provincial highways; cultivated land is distributed mainly on the Pleiku and Kon Ha Nung plateaus, on river terraces; and natural and planted forests in mountain and hill areas. NDVI determines the density of vegetation in a region; therefore, it is considered an important factor in the probability of the occurrence of groundwater in a region.

NDBI measures the density of construction. Increasing the construction area reduces the infiltration capacity of the soil, which results in reducing the recharge capacity of groundwater (Huang et al., 2024, Rehman et al., 2024). High construction density is concentrated in Pleiku City, An Khe and Ayunpa towns.

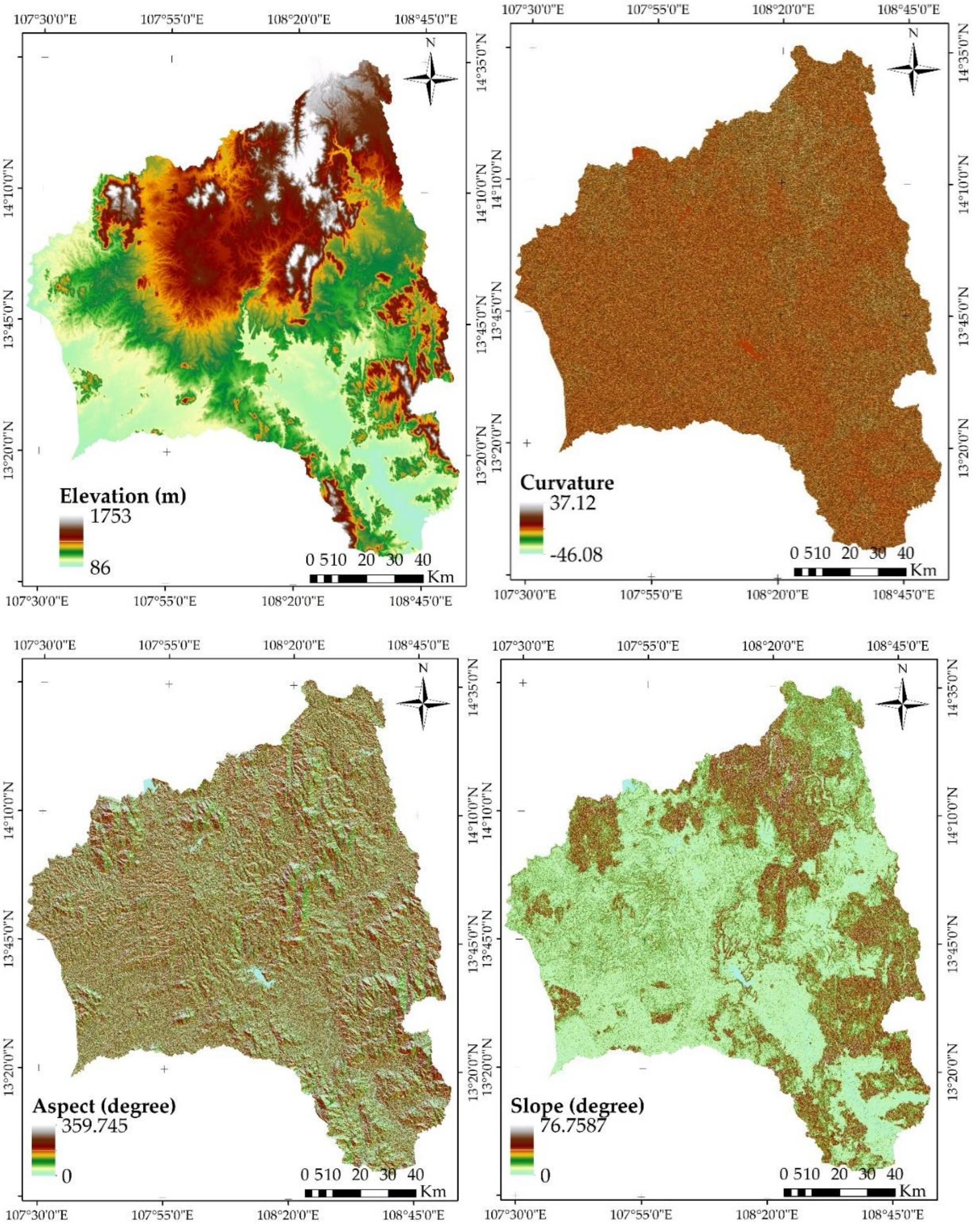

Fig. 2a. Conditioning factors used for the groundwater potential model: elevation, curvature, aspect, slope.

NDWI is considered an indispensable factor in identifying areas with the probability of groundwater occurrence because it is related to the groundwater recharge capacity in a region. The value of NDWI is proportional to the recharge capacity (Ghosh and Bera 2024). In Gia Lai, the NDWI index tends to be high in the Northeast, Southeast, Northwest and West; low in the center and south.

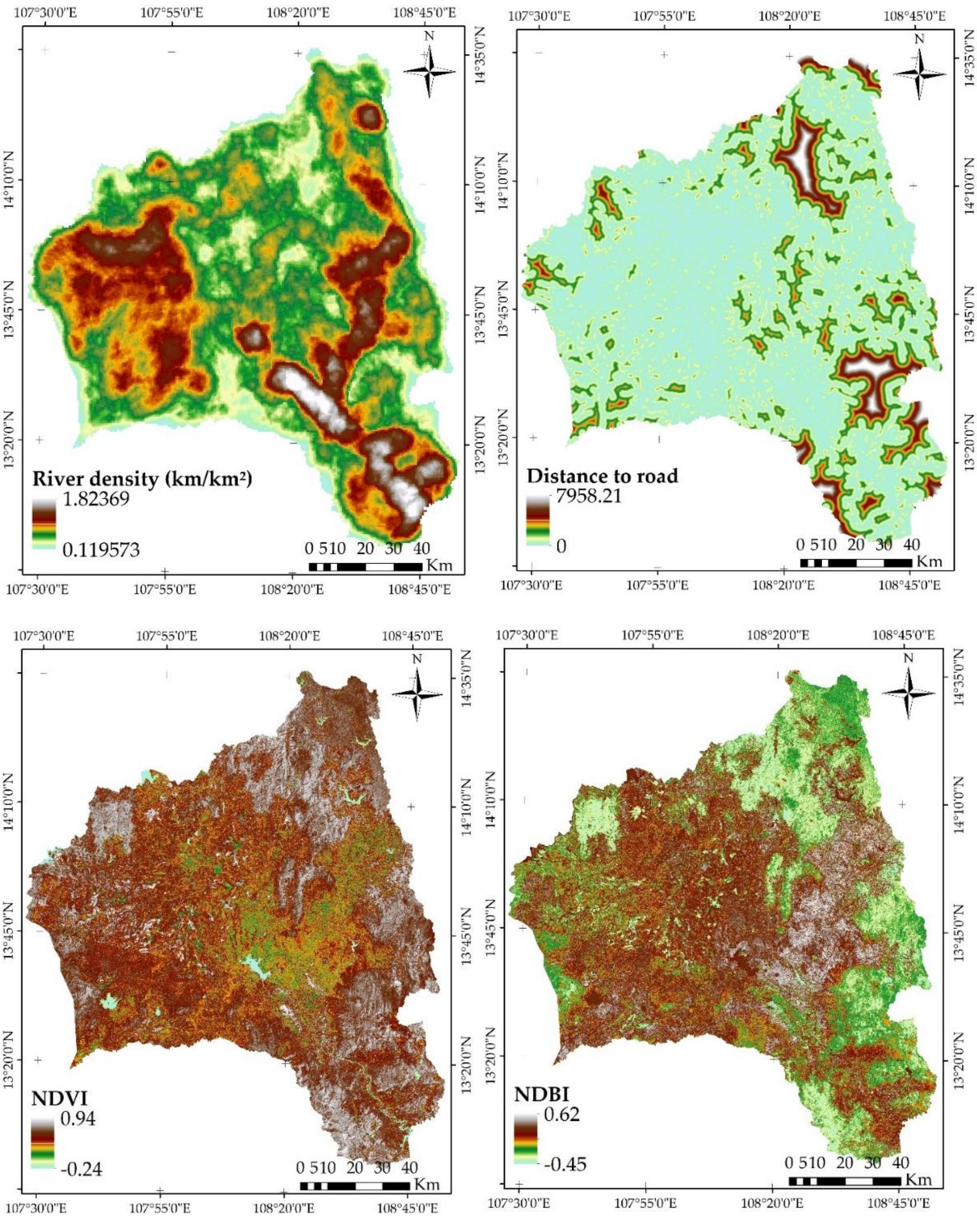

Fig. 2b. Conditioning factors used for the groundwater potential model: river density, distance to road, NDVI, NDBI.

Rainfall can increase groundwater recharge and so directly influences groundwater potential. Recharge depends on the amount of precipitation; precipitation value is proportional to groundwater recharge capacity (Raju et al. 2024). In the province, the rainfall is larger, about 1800 -2398 mm, concentrated in the northern region including mountains facing the wind; The Ba River valley area located in the Southeast has small rainfall, about 1432 mm -1600 mm.

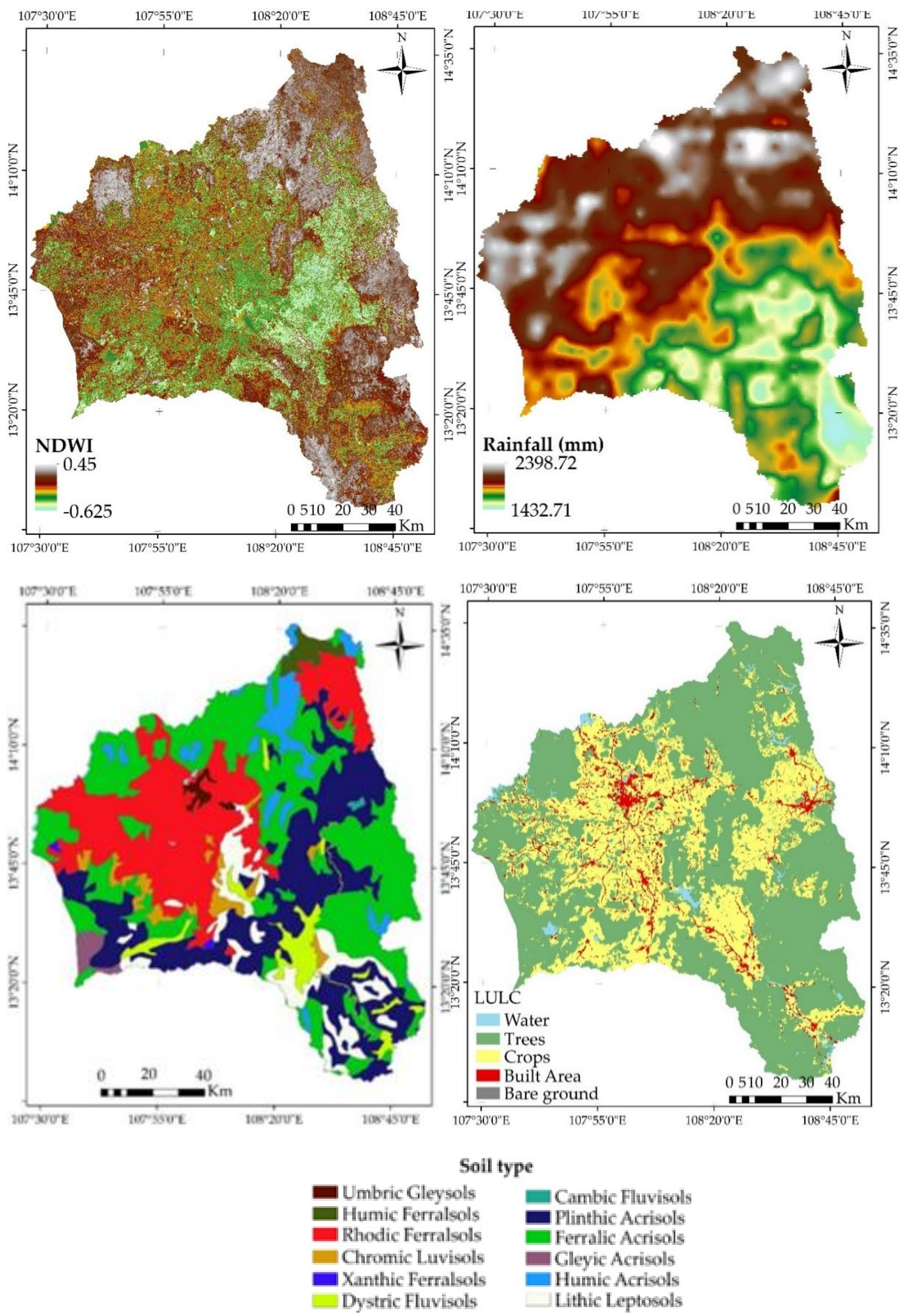

Fig. 2c. Conditioning factors used for the groundwater potential model: NDWI, rainfall, soil type, LULC.

Soil type plays an important role in predicting groundwater potential because it links to the permeability capacity of aquifer material and porosity of soil, which influence the groundwater recharge capacity (Dandapat et al., 2024, Nguyen et al., 2024). In Gia Lai, red-brown and yellow-brown soils develop on basic and neutral igneous rocks (Rhodic Ferralsols) distributed in the Pleiku Plateau and Kon Ha Nung Plateau. The soil absorbs water poorly but holds water well. Yellow-red soil on acid igneous soil (Ferralic Acrisols) has the ability to absorb water quickly, but poorer water retention ability. Humic Acrisols on the mountain have average water permeability and water holding capacity. Yellow-red soil changed due to rice cultivation (Plinthic Acrisols) has poor water absorption and retention capacity. Alluvial soils not deposited annually (Distric Fluvisols) have average water permeability and water holding capacity. Soil with eroded rocks (Lithic Leptosols) has very poor water absorption and water holding capacity. Alluvial soils not deposited annually (Distric Fluvisols) have average water permeability and water holding capacity. Soil with eroded rocks (Lithic Leptosols) has very poor water absorption and water holding capacity.

Rivers are, from a hydro-geological viewpoint, very important for controlling the movement and storage of potential groundwater. Areas with a high density of rivers and streams often have significant groundwater reserves (Ray 2024). In the study area, the river density is greatest along the Ba River valley in the east and the western edge of the Pleiku plateau.

Distance to road is considered another important factor because road construction can influence the infiltration capacity of the soil. Additionally, road construction can influence water drainage. All this can influence the water table recharge (Senapati and Das 2024).

3. METHODOLOGY

The construction of the groundwater potential map in Gia Lai province in Vietnam consisted of four main steps. The first was the preparation of input data, including the inventory map and conditioning factors; the second was the construction of groundwater potential models; the third was the validation of models, and the fourth was the analysis of groundwater potential map (**Fig. 3**).

i) Input data comprised the groundwater inventory map and 12 conditioning factors. The map was compiled using several sources: previous studies, data from the Department of Agriculture and Rural Development, and from a field mission. As conditioning factors were measured with different units, it was necessary to normalize the data with the assumption that the original values of all layers were retained, but the input database was standardized on similar ranges.

ii) Construction of groundwater potential model. The models DNN, CB, and XGB were used to predict groundwater potential. The DNN structure comprised three layers: the first was the input layer with 181 springs, well, non-spring, and non-well points, and 12 conditioning factors. These data was processed in the second layer, with 3 hidden layers and 11 neurons per layer. The weights in the hidden layers were optimized by Adam's optimization algorithm. In the end, the output layer consisted of two layers: groundwater and non-groundwater.

The performance of CB model depends on parameters such as iteration, depth, train direction, and loss function. In this study, the values of these parameters were 100, 3, logloss, and crossEntropy, respectively. The performance depended on *n*_estimators, max_depth, learning rate, subsample and colsample by tree. These parameter values were 200, 4, 0.004, 0.4 and 0.4, respectively.

It should be noted that the machine learning models were developed on Python languages using the Tensorflow libraries.

iii) Evaluation of the performance of the proposed models. Statistical indices were used to evaluate the performance of the proposed models, namely AUC, RMSE, MAE and R^2.

iv) Analysis of the groundwater potential map. After evaluating the proposed models, they were used to construct groundwater potential maps for Gia Lai province. This process was carried out by assigning the values of the 12 conditioning factors to all pixels in the entire study area. The pixels in the study area are then identified as either groundwater or non-groundwater.

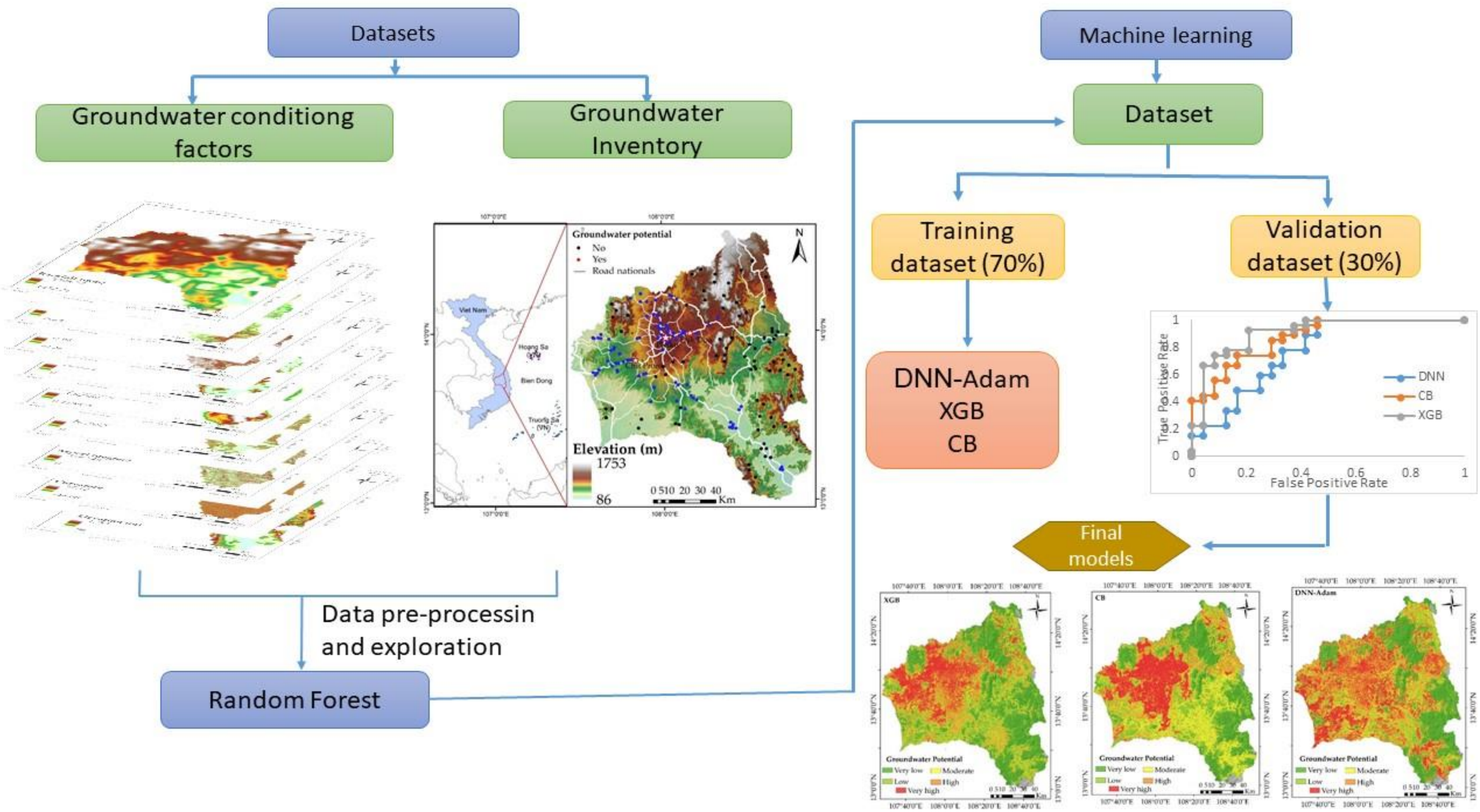

Fig. 3. Methodology used to create

3.1. Deep neural networks (DNN)

DNNs constitute a branch of machine learning that uses neural networks to solve complex, real-world, non-linear problems (Wang et al., 2022). DNNs have attracted the attention of researchers in recent years, particularly in Earth sciences, due to their ability to automatically learn complex data abstractions. The DNN model includes 3 layers: the input layer contains information related to groundwater prediction factors; this information is transmitted to and processed by the second layer, which includes one or more hidden layers; the output of the third layer is labeled groundwater and non-groundwater. With a large number of hidden layers, DNN models can solve many complex problems and are considered more powerful than simple neural networks (Bai et al., 2022, Hakim et al., 2022). The DNN model uses a backpropagation algorithm, which means that the output error is propagated to the hidden layers to update the weights. The DNN model computes the gradient of the loss function for each weight according to the chain rule and avoids redundant calculations from intermediate factors. During this operation, each neuron uses activation functions to process information in hidden layers. These activation functions are used for gradient training and are represented by rectified linear units (ReLU) (Hakim et al., 2022). There are many known optimization algorithms used to calculate weights of DNN models, such as SGD and AdaGrad. In this study, the weights of the DNN model were calculated using the adaptive moment (Adam). Adam is a stochastic optimization algorithm based on a first-order gradient. During the operation, Adam maintains the mean square of past slopes and also maintains the average of past slopes. Adam has the advantage of being simple to implement and requiring little memory. Previous studies have also shown Adam's optimization algorithm to be more accurate than other stochastic optimization models (Hakim et al., 2022).

3.2. CaTBoost (CB)

CB was first introduced by Yandex to solve both classification and regression problems using decision trees (Liang et al., 2023). CB is based on gradient boosting with the ideas of transforming weak learners into strong learners. It includes two main features: it works with categorical data (“Cat”) and it uses gradient boosting (“Boost”) (Tran et al., 2021). Gradient boosting is a process in which many decision trees are built iteratively (Koch et al., 2021). Each subsequent tree improves the result of previous tree, which optimizes the results. Each decision tree is then an evolution of an initial set of data. CB incorporates innovative techniques, such as target encoding and combining statistics from categorical variables, to leverage information present in categorical features and improve predictive performance (Abba et al., 2023, Yavuz Ozalp et al., 2023).

3.3. XgBoost (EXT)

XGB, or eXtreme Gradient Boosting, was first developed by Tianqi Chen in 2016 and can solve classification and regression problems (Chen and Guestrin 2016). The basic idea is to combine gradient descent with boosting methods in order to create a more powerful machine learning algorithm (Zounemat-Kermani et al., 2021). This involves assembling several algorithms with relatively low performance to create one that is much more efficient and satisfactory (Lim and Chi 2019, Tran et al., 2021). The result consists of the predictions of all the chained models. This method improves the model performance and stability while reducing its variance. Therefore, ensemble learning can achieve a much higher level of accuracy than using any of the individual models separately (Ghosh and Bera 2024).

3.4. Performance assessment

The models proposed in this study were evaluated using the statistical indices: AUC, RMSE, MAE. Details of information can be viewed in (Naghibi, Hashemi et al. 2020, Wang et al. 2023, Xiong, Guo et al. 2023):

$$\mathbf{AUC} = \sum \mathbf{TP} + \sum \frac{\mathbf{TN}}{\mathbf{P}} + \mathbf{N} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{predicted} - Y_{bserved})^2} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | Y_{predicted} - Y_{observed} | \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \ddot{y})^2}{\sum_{i=1}^{n}(y_i - \ddot{y})^2} \quad (4)$$

4. RESULTS AND DISCUSSIONS

4.1. Pre-processing results

Careful selection of conditioning factors plays an important role in the performance of any prediction model. This study used data-driven models to analyze the relationships between spring and well points and the explanatory variables and, therefore, factor selection can help models concentrate important factors and eliminate non-useful factors. This study used the RF technique to select the factors and assign a weight to each one. **Figure 4** illustrates the importance of each conditioning factor. The results showed that distance to road, soil type, NDBI, and slope were the most influential in identifying areas of probable groundwater. These factors are directly linked to the capacity of water infiltration and accumulation, all of which influences the recharge capacity of water table. In Gia Lai province, construction area has been growing at an increasing rate in recent years, as a result of population growth and economic development. This increase has been mainly concentrated in the low slope region.

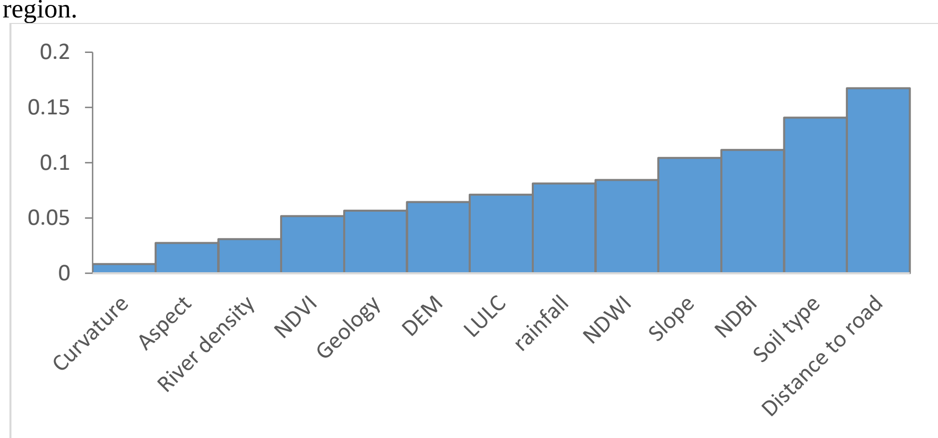

Fig. 4. The importance of groundwater conditioning factors, using RF.

NDWI, rainfall, LULC, and elevation were fifth to eighth most important, respectively. NDWI plays an important role in the occurrence of groundwater in a region because it is directly related to recharge capacity. In Gia Lai province, the region with a high NDWI value is concentrated in the eastern, southern and central regions of the province. Rainfall is important as it provides a source of water that can infiltrate the ground and recharge underground aquifers. Precipitation tends to accumulate in the region at low altitudes and, therefore, in these regions, conditions are conducive to strong recharge of underground aquifers. LULC reflects human activity in a region. The reduction of surface vegetation and increase of construction area led to the accumulation of water on the surface, which reduced the recharge capacity of underground aquifers.

Curvature, aspect, and river density appear to have less influence on the probability of occurrence of groundwater potential in Gia Lai province.

4.2. Modeling assessment results

This study used the AUC index to evaluate the performance of proposed models. Figure 5 shows the accuracy of proposed models when using training and validation data. The results show that the XGB model identified regions with a probability of occurrence of groundwater potential more accurately than the CB and DNN models. The DNN model was least effective.

For the validation data, the XGB model further identified regions with a higher probability of occurrence of groundwater potential than other models. The CB model ranked second. The DNN model performed worse than the XGB and CB models. The DNN model required a significant amount of data to provide more accurate results. Given the difficulties encountered in collecting data on spring and well points, only 82 points were available. Therefore, there may not have been enough data to properly train the DNN models (**Fig. 5**).

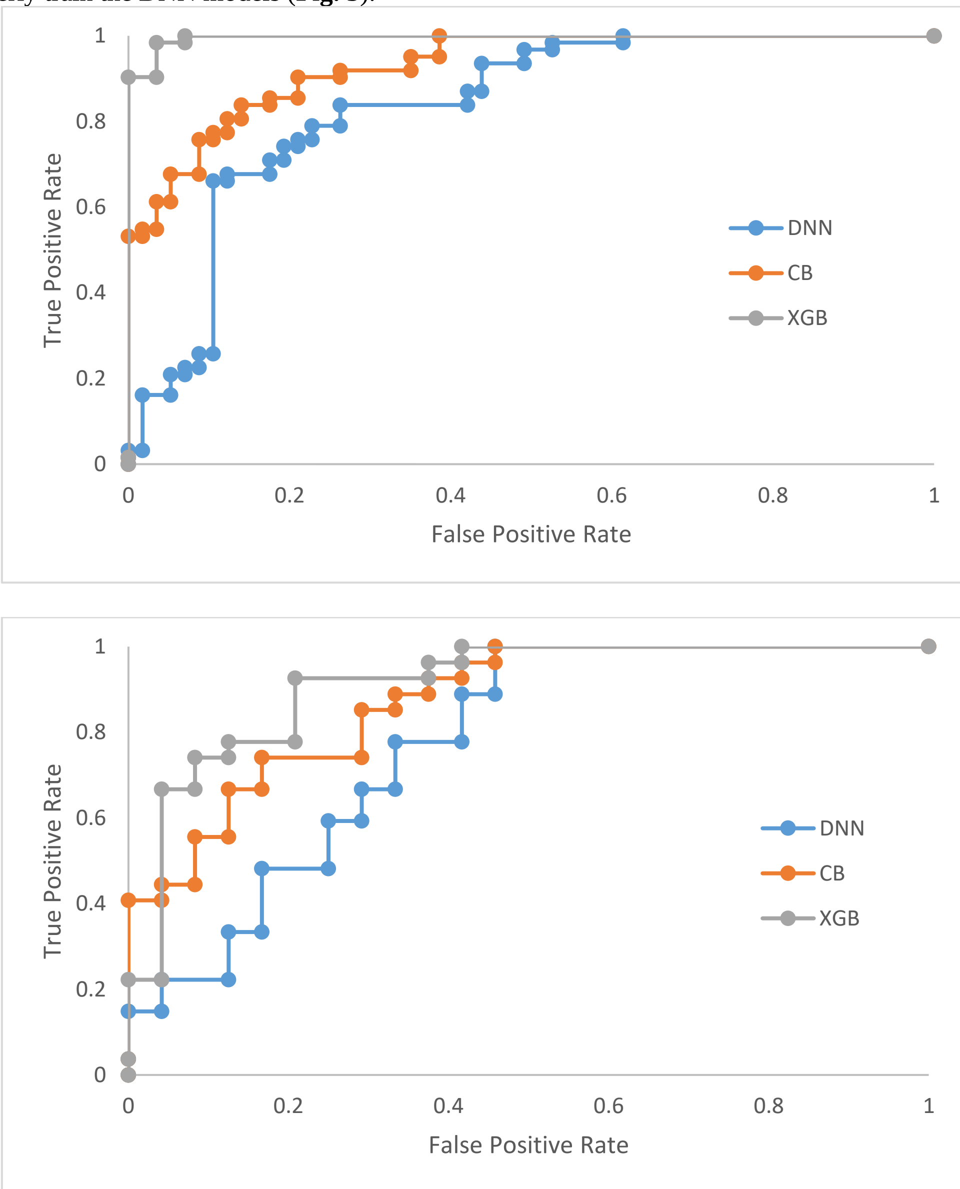

Fig. 5. AUC values for the training dataset (top) and validation dataset (bottom).

In addition to the AUC values, this study used the statistical indices RMSE, MAE, and R^2 to evaluate and compare the proposed models. The results show that the XGB model returned RMSE, MAE, and R^2 values higher than the other two models for the training data. For the validation data, XGB performed even better in terms of RMSE, MAE, and R^2. CB came second, and DNN performed least well.

In general, all proposed models were suitable for identifying regions with the probability of occurrence of groundwater potential in Gia Lai province. However, we recommend the use of XGB because several studies have pointed out that XGB has advantages in being able to solve problems with precision when data are not sufficient (**Table 1**).

Table 1.

Precision of DNN, XGB, and CB.

	Training dataset				**Validation dataset**			
	RMSE	**MAE**	**AUC**	**R^2**	**RMSE**	**MAE**	**AUC**	**R^2**
XGB	0.2	0.15	0.996	0.82	0.35	0.25	0.91	0.814
CB	0.4	0.36	0.929	0.80	0.41	0.373	0.87	0.796
DNN	0.41	0.373	0.835	0.78	0.42	0.383	0.77	0.779

4.3. Groundwater potential mapping

After the evaluation of the machine learning model, all the proposed models are performed to construct the groundwater potential map. The calculation of groundwater value in Gia Lai province is carried out by aggregating all pixels, which have 12 conditional factor values associated with them, in the machine learning model. The result of the model represents the groundwater value of the entire study area, on a scale from 0 to 1. The values were divided by five classes: very low, low, moderate, high and very high in using Break Natural methods. **Figure 6** shows the maps of groundwater potential produced by XGB, CB and DNN-Adam. The results reported that according to XGB, 4990 km2 of the province constituted the area of very low groundwater potential, with 3045 km2, 2426 km2, 2655 km2, and 2077 km2 were in the low, moderate, high and very high categories. For the CB model, the very low category covers 3325 km^2, low - 3573 km^2, moderate - 3489 km^2, high - 2308 km^2, and very high - 2489 km^2. According to the DNN-Adam model, approximately 3580 km^2 of the province is in the very low area of groundwater potential, 2198 km^2 - low, 2741 km^2 - moderate, 3377 km^2 - high, and 3298 km^2 - very high.

In general, although there are differences between the models, we found out that the regions of high and very high groundwater potential are located in the districts of Pleiku, Dak Doa, Ia Grai, and Chu Prong, as well as a small part of K'Bang. Although these areas are densely populated and have high construction density, the distance to the road is small, the rainfall quite large, slope is small, and the vegetation consists mainly of coffee and rubber, which helps to absorb and retain water well. Basalt soil is quite permeable and retains water well, providing a lot of water for the Pleistocene Basalt eruptive fissure aquifer.

The area with average groundwater potential is mainly distributed on the edge of the Pleiku plateau and Ba River valley in the south and southeast of the province. This area's terrain is less steep, there is not much rainfall, the road density is quite large, and the alluvial soil absorbs and retains water quite well, providing water for the aquifer, cracks and seams of Neogene sedimentary lagoon formations and modern alluvial porous aquifers.

Areas with low and very low groundwater potential are distributed mainly in the mountainous parts of Northeast, East and Northwest. Although there is heavy rainfall and forest cover, the construction density is small, but the altitude is high and slope steep, distance to the road long, and the weathered crust from metamorphic rocks has a weak ability to absorb and retain water, so the amount of surface water supplied to groundwater in the rainy season is very limited.

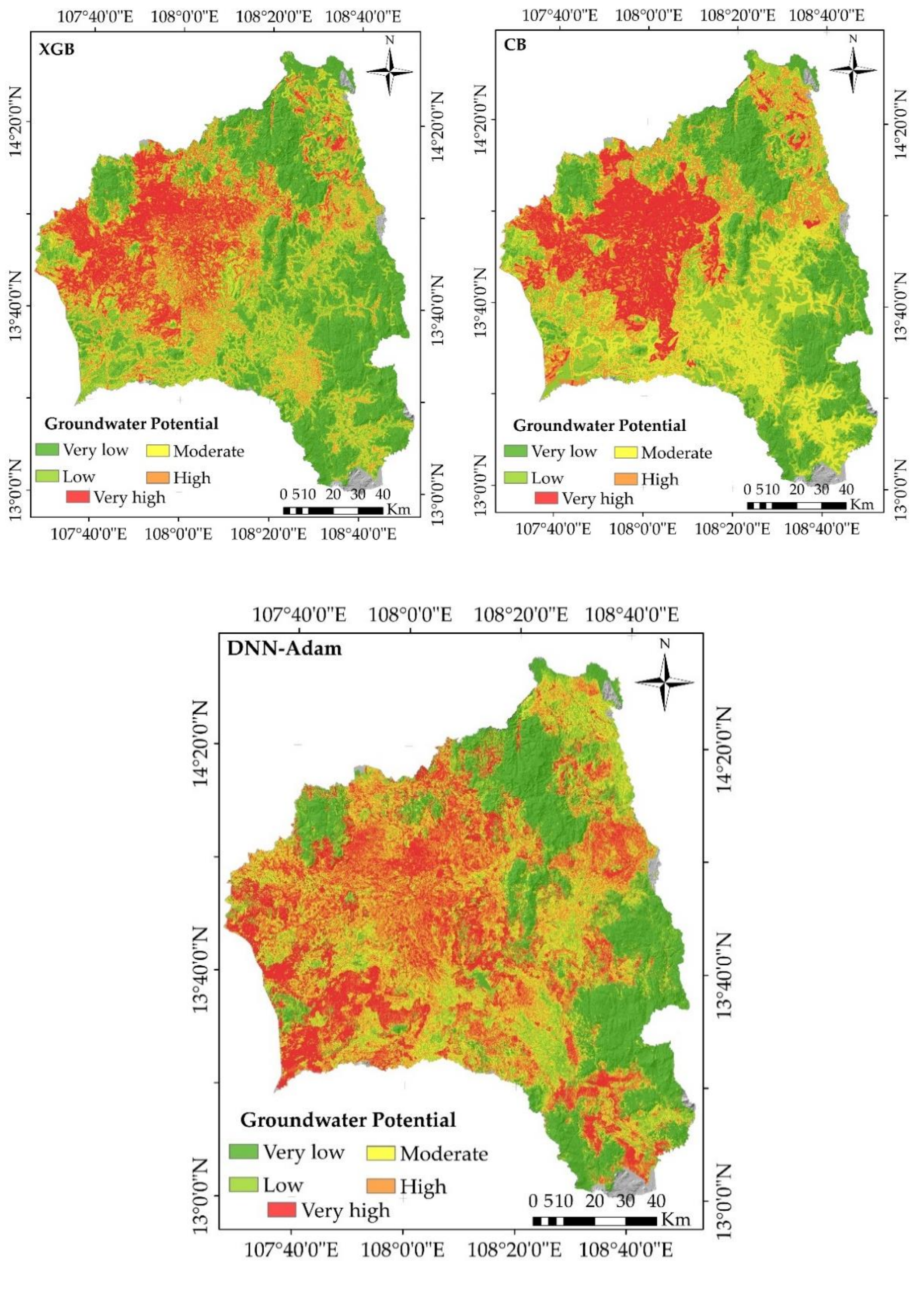

Fig. 6. Groundwater potential maping for the Gia Lai province.

5. DISCUSSION

Groundwater resources play an important role in the development of agriculture and industry all over the world (Pan et al., 2023, Huang et al., 2024). Construction of groundwater potential maps can be a successful method to support decision makers in effectively managing groundwater resources (Anh et al., 2023, Kumar et al., 2023, Vafadar et al., 2023, Wang et al., 2023). Many studies have used different methods to construct maps of groundwater potential.

Regional studies are still necessary to obtain appropriate information for water resource management in a specific locality. To this end, the objective of this study was to construct a groundwater potential map based on machine learning, namely DNN, XGB and CB, in Gia Lai province of Vietnam.

With the development of remote sensing and GIS, machine learning has received the attention of researchers in recent years, due to its application in spatial data modeling. One of is advantages is the ability to eliminate limitations linked to the lack of precise data. However, overfitting problems are considered a big challenge when using machine learning and each region has different characteristics, so selecting appropriate models is very important to build a groundwater potential map with accuracy (Anh et al., 2023, Wang et al., 2023, Nguyen et al., 2024).

Out of the three models used in this study, the XGB model was found to perform better than the other two. Several studies have also highlighted its ability to explain complex relationships between variables. Thus, it is considered an attractive choice for constructing a groundwater potential map. Additionally, XGB combines regularization techniques, which helps reduce overfitting issues. Additionally, it has the ability to handle missing data (Pan et al., 2023, Ngai et al., 2024). All these features allow the XGB model to yield more accurate predictions than other models.

The CB model came second. It can resolve missing data issues natively and is one of the most powerful algorithms in solving overfitting problems, by combining automatic regularization mechanisms (Gao et al., 2024, Raheja et al., 2024). The DNN model performed worst. In general, DNN models are suitable for big data studies. However, in this study, due to the difficulties encountered in data collection, only 181 data points were collected (Wang et al., 2024). The DNN model was not suitable for constructing groundwater potential maps.

The remaining question of this study is whether the proposed models can predict groundwater potential in the context of climate change, when precipitation and temperatures are not stable. This may not be an issue if the model can learn from climate change data. However, data collection is a big challenge, especially in developing countries like Vietnam, where data sharing policies are restrictive and financial resources limited (Nguyen et al., 2024, Nguyen et al., 2024, Nguyen et al., 2024).

The results of this study have also important implications for planners. The connection between water, particularly groundwater, and climate changes has been stressed out by previous studies (Dragoni and Sukhija, 2008, Earman and Dettinger, 2011, Amanambu et al., 2020). At the same time, other authors underlined the potential of planning to mitigate the impact of climate change (Wilson, 2006; Hurlimann and March, 2012, Petrişor et al., 2021). A good example of the relationship between water and planning comes from a World Bank study carried out in Romania, in order to develop specific guidelines for integrating the flood risk management in planning, with specific provisions related to groundwater (World Bank 2023). Therefore, accounting for the groundwater potential in planning may help increasing urban resilience to climate changes. In this regard, provided that other studies pointed out the presence of derogatory planning in Vietnam (Petrişor et al. 2020), i.e. making local exemptions from national planning provisions, our results plead for the need to enforce the planning provisions regarding underground water when planning for regions similar to the one investigated in our study. In addition, the important role in territorial or urban planning, the results of this study play an important role for tourism planning, because developing tourism, relies on tourist destinations that need to use water directly (food, accommodation, etc.) or indirectly. In these destinations water must be used in a way that balances tourist and resident populations, while ensuring water security. Therefore, developing new precise methods is very important to support decision-makers in tourism development too.

6. CONCLUSIONS

Groundwater resources play an important role in sustainable socio-economic development, particularly in the context of climate change and population growth. Therefore, the construction of precise groundwater potential maps is necessary to support those responsible for optimizing water resources. The objective of this study was to construct a potential groundwater map using the machine learning techniques DNN, XGB, and CB in the Gia Lai province of Vietnam. The conclusions of this study are stated below.

- The traditional machine learning model was more powerful than the deep learning model for the groundwater potential model, due to the collection of input data. The methodology in this study can be applied in other regions to construct groundwater potential maps.

- A comparison of the three proposed models showed XGB to have a better performance, with an AUC value of 0.91; second was CB (0.87), and third DNN (0.77). A successful implementation of these models can support decision makers in proposing effective strategies to optimize water resource management.

- The regions of high and very high probability of groundwater are concentrated in the west of the province, in the districts of Pleiku, Dak Doa, Ia Grai and Chu Prong, and a small part of K'Bang.

Although this study was successful in constructing groundwater potential maps, there were limitations relating to the selection of non-spring and non-well points. There are no universal guides for selecting these points. They were randomly selected from the study region. Inaccurate selection of areas without groundwater can lead to errors of the groundwater potential map. In addition, areas with groundwater potential are affected by surface temperature; regions with high surface temperatures are less likely to contain groundwater. Therefore, future research should try to integrate surface temperature into the groundwater potential model. Finally, the potential of groundwater is strongly influenced by climate change and human activities. Therefore, future studies should try to integrate these factors into a machine learning model to predict the future potential of groundwater.

The results of this study can support decision makers in identifying regions with high and very high probability of groundwater so they can implement appropriate infrastructure for developing agriculture and industry.

ACKNOWLEDGMENT

This research was funded by VNU Asia Research Center (ARC) from grants source by CHEY Institute for Advanced Studies, code CA.22.07A.

R E F E R E N C E S

Abba, S. I., M. A. Yassin, A. S. Mubarak, S. M. H. Shah, J. Usman, A. Y. Oudah, S. R. Naganna and I. H. Aljundi (2023). "Drinking Water Resources Suitability Assessment Based on Pollution Index of Groundwater Using Improved Explainable Artificial Intelligence." Sustainability 15(21): 15655.

Anh, D. T., M. Pandey, V. N. Mishra, K. K. Singh, K. Ahmadi, S. Janizadeh, T. T. Tran, N. T. T. Linh and N. M. Dang (2023). "Assessment of groundwater potential modeling using support vector machine optimization based on Bayesian multi-objective hyperparameter algorithm." Applied Soft Computing 132: 109848.

Arabameri, A., S. C. Pal, F. Rezaie, O. A. Nalivan, I. Chowdhuri, A. Saha, S. Lee and H. Moayedi (2021). "Modeling groundwater potential using novel GIS-based machine-learning ensemble techniques." Journal of Hydrology: Regional Studies 36: 100848.

Bai, Z., Q. Liu and Y. Liu (2022). "Groundwater potential mapping in hubei region of china using machine learning, ensemble learning, deep learning and automl methods." Natural Resources Research 31(5): 2549-2569.

Bien, T. X., A. Jaafari, T. Van Phong, P. T. Trinh and B. T. Pham (2023). "Groundwater potential mapping in the Central Highlands of Vietnam using spatially explicit machine learning." Earth Science Informatics 16(1): 131-146.

Boretti, A. and L. Rosa (2019). "Reassessing the projections of the world water development report." NPJ Clean Water 2(1): 15.

Çelik, M. Ö., L. Kuşak and M. Yakar (2024). "Assessment of Groundwater Potential Zones Utilizing Geographic Information System-Based Analytical Hierarchy Process, Vlse Kriterijumska Optimizacija Kompromisno Resenje, and Technique for Order Preference by Similarity to Ideal Solution Methods: A Case Study in Mersin, Türkiye." Sustainability 16(5): 2202.

Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.

Dandapat, K., U. Chatterjee, S. Das, A. Patra, H. Manna, A. Ghosh, S. C. Pal, A. R. M. Towfiqul Islam, R. Costache and E. Alam (2024). "Assessment of groundwater potential zones in data-scarce regions using GIS-based multicriteria decision making approach." Geocarto International 39(1): 2311202.

Dey, B., K. A. M. Abir, R. Ahmed, M. A. Salam, M. Redowan, M. D. Miah and M. A. Iqbal (2023). "Monitoring groundwater potential dynamics of north-eastern Bengal Basin in Bangladesh using AHP-Machine learning approaches." Ecological Indicators 154: 110886.

Diriba, D., S. Karuppannan, T. Takele and M. Husein (2024). "Delineation of groundwater potential zonation using geoinformatics and AHP techniques with remote sensing data." Heliyon.

Ehsan, M., H. Shabbir, A. M. F. Al-Quraishi, N. Al-Ansari, Z. Ahmad, K. Abdelrahman, M. T. Sohail, Z. Manzoor, A. Shafi and A. Elbeltagi (2024). "Groundwater delineation for sustainable improvement and development aided by GIS, AHP, and MIF techniques." Applied Water Science 14(2): 1-21.

Gao, W., M. Zang and G. Mei (2024). "Exploring influence of groundwater and lithology on data-driven stability prediction of soil slopes using explainable machine learning: a case study." Bulletin of Engineering Geology and the Environment 83(1): 2.

Ghosh, A. and B. Bera (2024). "Identification of potential dam sites for severe water crisis management in semi-arid fluoride contaminated region, India." Cleaner Water: 100011.

Ghosh, A. and B. Bera (2024). "Potentialities and development of groundwater resources applying machine learning models in the extended section of Manbhum-Singhbhum Plateau, India." HydroResearch 7: 1-14.

Gómez-Escalonilla, V., P. Martínez-Santos and M. Martín-Loeches (2022). "Preprocessing approaches in machine-learning-based groundwater potential mapping: an application to the Koulikoro and Bamako regions, Mali." Hydrology and Earth System Sciences 26(2): 221-243.

Ha, D. H., P. T. Nguyen, R. Costache, N. Al-Ansari, T. Van Phong, H. D. Nguyen, M. Amiri, R. Sharma, I. Prakash and H. Van Le (2021). "Quadratic discriminant analysis based ensemble machine learning models for groundwater potential modeling and mapping." Water Resources Management 35: 4415-4433.

Hakim, W. L., A. S. Nur, F. Rezaie, M. Panahi, C.-W. Lee and S. Lee (2022). "Convolutional neural network and long short-term memory algorithms for groundwater potential mapping in Anseong, South Korea." Journal of Hydrology: Regional Studies 39: 100990.

Huang, P., M. Hou, T. Sun, H. Xu, C. Ma and A. Zhou (2024). "Sustainable groundwater management in coastal cities: Insights from groundwater potential and vulnerability using ensemble learning and knowledge-driven models." Journal of Cleaner Production: 141152.

Kamali Maskooni, E., S. A. Naghibi, H. Hashemi and R. Berndtsson (2020). "Application of advanced machine learning algorithms to assess groundwater potential using remote sensing-derived data." Remote Sensing 12(17): 2742.

Koch, J., J. Gotfredsen, R. Schneider, L. Troldborg, S. Stisen and H. J. Henriksen (2021). "High resolution water table modeling of the shallow groundwater using a knowledge-guided gradient boosting decision tree model." Frontiers in Water 3: 701726.

Kumar, M., P. Singh and P. Singh (2023). "Machine learning and GIS-RS-based algorithms for mapping the groundwater potentiality in the Bundelkhand region, India." Ecological Informatics 74: 101980.

Lee, S., S.-M. Hong and H.-S. Jung (2018). "GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea." Geocarto international 33(8): 847-861.

Lee, S., Y.-S. Kim and H.-J. Oh (2012). "Application of a weights-of-evidence method and GIS to regional groundwater productivity potential mapping." Journal of Environmental Management 96(1): 91-105.

Liang, Z., W. Peng, W. Liu, H. Huang, J. Huang, K. Lou, G. Liu and K. Jiang (2023). "Exploration and comparison of the effect of conventional and advanced modeling algorithms on landslide susceptibility prediction: A case study from Yadong Country, Tibet." Applied Sciences 13(12): 7276.

Lim, S. and S. Chi (2019). "Xgboost application on bridge management systems for proactive damage estimation." Advanced Engineering Informatics 41: 100922.Masroor M, Sajjad H, Kumar P, Saha TK, Rahaman MH, Choudhari P, Kulimushi LC, Pal S, Saito O (2023) Novel ensemble machine learning modeling approach for groundwater potential mapping in Parbhani District of Maharashtra, India. Water 15:419.

Morgan, H., A. Madani, H. M. Hussien and T. Nassar (2023). "Using an ensemble machine learning model to delineate groundwater potential zones in desert fringes of East Esna-Idfu area, Nile valley, Upper Egypt." Geoscience Letters 10(1): 9.

Mosavi, A., F. Sajedi Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva and E. Rafiei Sardooi (2021). "Ensemble boosting and bagging based machine learning models for groundwater potential prediction." Water Resources Management 35: 23-37.

Naghibi, S. A., H. Hashemi, R. Berndtsson and S. Lee (2020). "Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors." Journal of Hydrology 589: 125197.

Naghibi, S. A. and H. R. Pourghasemi (2015). "A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping." Water resources management 29: 5217-5236.

Naghibi, S. A., H. R. Pourghasemi and B. Dixon (2016). "GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran." Environmental monitoring and assessment 188: 1-27.

Ngai, C. H., A. J. Kilpatrick and A. Ćwiek (2024). "Sound symbolism in Japanese names: Machine learning approaches to gender classification." Plos one 19(3): e0297440.

Nguyen HD, Nguyen Q-H, Bui Q-T (2024) Solving the spatial extrapolation problem in flood susceptibility using hybrid machine learning, remote sensing, and GIS. Environmental Science and Pollution Research:1-22.

Nguyen, H. D., D. K. Dang, N. Y. Nguyen, C. Pham Van, T. T. Van Nguyen, Q.-H. Nguyen, X. L. Nguyen, L. T. Pham, V. T. Pham and Q.-T. Bui (2024). "Integration of machine learning and hydrodynamic modeling to solve the extrapolation problem in flood depth estimation." Journal of Water and Climate Change 15(1): 284-304.

Nguyen, H. D., Q.-H. Nguyen and Q.-T. Bui (2024). "Solving the spatial extrapolation problem in flood susceptibility using hybrid machine learning, remote sensing, and GIS." Environmental Science and Pollution Research: 1-22.

Nguyen, H. D., V. H. Nguyen, Q. V. V. Du, C. T. Nguyen, D. K. Dang, Q. H. Truong, N. B. T. Dang, Q. T. Tran, Q.-H. Nguyen and Q.-T. Bui (2024). "Application of hybrid model-based machine learning for groundwater potential prediction in the north central of Vietnam." Earth Science Informatics: 1-21.

Nguyen, P. T., D. H. Ha, A. Jaafari, H. D. Nguyen, T. Van Phong, N. Al-Ansari, I. Prakash, H. V. Le and B. T. Pham (2020). "Groundwater potential mapping combining artificial neural network and real AdaBoost ensemble technique: the DakNong province case-study, Vietnam." International journal of environmental research and public health 17(7): 2473.

Oguntoyinbo, P. O., O. M. Iyiola and A. A. Komolafe (2024). "Integrated GIS-based and geophysical techniques in groundwater potential zonation: a case study of Jos North local government area." Sustainable Water Resources Management 10(2): 58.

Pan, Z., W. Lu, H. Wang and Y. Bai (2023). "Groundwater contaminant source identification based on an ensemble learning search framework associated with an auto xgboost surrogate." Environmental Modelling & Software 159: 105588.

Prasad, P., V. J. Loveson, M. Kotha and R. Yadav (2020). "Application of machine learning techniques in groundwater potential mapping along the west coast of India." GIScience & Remote Sensing 57(6): 735-752.

Raheja, H., A. Goel and M. Pal (2024). "A novel approach for prediction of groundwater quality using gradient boosting-based algorithms." ISH Journal of Hydraulic Engineering: 1-12.

Rahmati, O., H. R. Pourghasemi and A. M. Melesse (2016). "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran." Catena 137: 360-372.

Raj, S., K. S. Rawat, S. K. Singh and A. K. Mishra (2024). "Groundwater potential zones identification and validation in Peninsular India." Geology, Ecology, and Landscapes 8(1): 86-100.

Raju, A., R. P. Singh, P. K. Kannojiya, A. Patel, S. Singh and M. Sinha (2024). "Declining groundwater and its impacts along ganga riverfronts using combined Sentinel-1, GRACE, water levels, and rainfall data." Science of The Total Environment: 170932.

Ray, S. (2024). "Identifying the groundwater potential zones in Jamsholaghat sub-basin by considering GIS and multi-criteria decision analysis." International Journal of Environmental Science and Technology 21(1): 515-540.

Rehman, A., F. Islam, A. Tariq, I. Ul Islam, D. Brian J, T. Bibi, W. Ahmad, L. Ali Waseem, S. Karuppannan and S. Al-Ahmadi (2024). "Groundwater potential zone mapping using GIS and Remote Sensing based models for sustainable groundwater management." Geocarto International 39(1): 2306275.

Sachdeva, S. and B. Kumar (2021). "Comparison of gradient boosted decision trees and random forest for groundwater potential mapping in Dholpur (Rajasthan), India." Stochastic Environmental Research and Risk Assessment 35(2): 287-306.

Senapati, U. and T. K. Das (2024). "Delineation of potential alternative agriculture region using RS and AHP-based GIS techniques in the drought prone upper Dwarakeswer river basin, West Bengal, India." Ecological Modelling 490: 110650.

Sharma, Y., R. Ahmed, T. K. Saha, N. Bhuyan, G. Kumari, S. Pal and H. Sajjad (2024). "Assessment of groundwater potential and determination of influencing factors using remote sensing and machine learning algorithms: A study of Nainital district of Uttarakhand state, India." Groundwater for Sustainable Development 25: 101094.

Singh, G., J. Singh, O. A. Wani, J. C. Egbueri and J. C. Agbasi (2024). "Assessment of groundwater suitability for sustainable irrigation: A comprehensive study using indexical, statistical, and machine learning approaches." Groundwater for Sustainable Development 24: 101059.

Tahmassebipoor, N., O. Rahmati, F. Noormohamadi and S. Lee (2016). "Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing." Arabian Journal of Geosciences 9: 1-18.

Tamiru, H. and M. Wagari (2022). "Comparison of ANN model and GIS tools for delineation of groundwater potential zones, Fincha Catchment, Abay Basin, Ethiopia." Geocarto International 37(23): 6736-6754.

Tiwari, V., A. Kumar and M. Mukherjee (2024). "Spatiotemporal mapping of groundwater recharge potential zones for physical planning process−A case of Ajmer City, India." Ecohydrology & Hydrobiology.

Tran, D. A., M. Tsujimura, N. T. Ha, V. T. Nguyen, D. V. Binh, T. D. Dang, Q.-V. Doan, D. T. Bui, T. Anh Ngoc, L. V. Phu, P. T. B. Thuc and T. D. Pham (2021). "Evaluating the predictive power of different machine learning algorithms for groundwater salinity prediction of multi-layer coastal aquifers in the Mekong Delta, Vietnam." Ecological Indicators 127: 107790.

Uddin, M. G., M. H. Imran, A. M. Sajib, M. A. Hasan, M. T. M. Diganta, T. Dabrowski, A. I. Olbert and M. Moniruzzaman (2024). "Assessment of human health risk from potentially toxic elements and predicting groundwater contamination using machine learning approaches." Journal of Contaminant Hydrology 261: 104307.

Vafadar, S., M. Rahimzadegan and R. Asadi (2023). "Evaluating the performance of machine learning methods and Geographic Information System (GIS) in identifying groundwater potential zones in Tehran-Karaj plain, Iran." Journal of Hydrology 624: 129952.

Wang, X., Y. Zhang, B. He, J. Li, T. Yang, H. Sun, Q. Shao and C. Xu (2024). "Deep learning algorithms in predicting Cr (VI) removal performance of S-ZVI: Models building and optimal parameters prediction." Separation and Purification Technology 330: 125487.

Wang, Z., J. Wang and J. Han (2022). "Spatial prediction of groundwater potential and driving factor analysis based on deep learning and geographical detector in an arid endorheic basin." Ecological Indicators 142: 109256.

Wang, Z., J. Wang, D. Yu and K. Chen (2023). "Groundwater potential assessment using GIS-based ensemble learning models in Guanzhong Basin, China." Environmental Monitoring and Assessment 195(6): 690.

World Bank (2023). Proposed Guidance for Integrating Flood Risk Management into Urban and Spatial Planning Practices. World Bank, Washington, DC, USA.

Xiong, H., X. Guo, Y. Wang, R. Xiong, X. Gui, X. Hu, Y. Li, Y. Qiu, J. Tan and C. Ma (2023). "Spatial prediction of groundwater potential by various novel boosting-based ensemble learning models in mountainous areas." Geocarto International 38(1): 2274870.

Yavuz Ozalp, A., H. Akinci and M. Zeybek (2023). "Comparative analysis of tree-based ensemble learning algorithms for landslide susceptibility mapping: A case study in Rize, Turkey." Water 15(14): 2661.

Zounemat-Kermani, M., O. Batelaan, M. Fadaee and R. Hinkelmann (2021). "Ensemble machine learning paradigms in hydrology: A review." Journal of Hydrology 598: 126266.