

Research article

# Optimized Artificial Intelligence-Based Algorithm for Groundwater Potential Mapping in Trenggalek Regency, Indonesia

Muhammad Hidayatul Ummah<sup>1</sup>, Diyono Diyono<sup>2,\*</sup>

<sup>1</sup> Geomatics Engineering Master Program, Department of Geodetic Engineering, Universitas Gadjah Mada, Jln. Grafika No. 2, Yogyakarta (55281), Indonesia; <sup>2</sup> Department of Geodetic Engineering, Faculty of Engineering, Universitas Gadjah Mada, Jln. Grafika No. 2, Yogyakarta (55281), Indonesia.

<sup>\*</sup> Correspondence: diyono@ugm.ac.id

Citation:

Ummah, M. H., & Diyono, D. (2025). Optimized Artificial Intelligence-Based Algorithm for Groundwater Potential Mapping in Trenggalek Regency, Indonesia. *Forum Geografi*. 39(3), 398-419.

Article history:

Received: 11 September 2025  
Revised: 2 December 2025  
Accepted: 2 December 2025  
Published: 10 December 2025

## Abstract

Groundwater management is essential as more than one-third of the world's population relies on groundwater as a source of freshwater. Exploration of groundwater potential serves as a practical implementation to ensure accessibility to freshwater. Therefore, this study developed a machine learning (ML), deep learning (DL), and stacking learning (SL) based model for groundwater potential mapping in Trenggalek Regency, Indonesia. A total of 740 spring locations were used as training data, and 18 variables were considered in the modelling. The eighteen parameters were classified into geological, topographic, land cover, climatological, hydrological, and geophysical factors. We used several algorithms, including gradient boosting decision trees (GBDT), random forest (RF), recurrent neural network (RNN), convolutional neural network (CNN), SL GBDT-RF, SL CNN-RNN, and SL GBDT-RF-CNN-RNN. This study optimized each basic learning task through hyperparameter fine-tuning using a tree-Parzen structured estimator (TPE) method. Models were evaluated using four metrics: accuracy (Acc), Cohen's kappa (CK), Matthews correlation coefficient (MCC), and receiver operating characteristic (ROC) area under the curve (AUC). Of the seven models generated, SL GBDT-RF achieved the best performance on the test data, with Acc, MCC, CK, and AUC values of 0.957, 0.915, 0.915, and 0.990, respectively. The geological unit parameter has the highest relative contribution rate in all prediction models. Based on the best model, the study area is dominated by the low-potential class, accounting for 31.29%. This study contributes to providing a benchmark for the development of groundwater potential prediction using ML, DL, and SL algorithms across various case studies. In addition, this study can be used by concerned stakeholders in sustainable water resource planning, drought disaster management, and the prevention of inappropriate groundwater exploration.

Keywords: Groundwater potential mapping, machine learning, deep learning, stacking learning, and hyperparameter optimizations.

## 1. Introduction

Groundwater is one of the freshwater resources to fulfill common needs, as well as for agriculture and industry (Li *et al.*, 2020). According to AQUASTAT data published by the Food and Agriculture Organization (FAO), the global withdrawal rate of freshwater from groundwater in 2022 reached 23.06%. This places groundwater as the second-largest source of freshwater withdrawal in the world (FAO, 2022). Therefore, mapping groundwater potential is necessary for the sustainable management of water resources (Khosravi *et al.*, 2018), especially in Trenggalek Regency. This regency, located in East Java Province, Indonesia, is affected by recurrent annual drought (Regional Disaster Management Agency Of Trenggalek Regency, 2024). In addition, more than 60% of the Trenggalek population relies on groundwater as a freshwater resources (Statistical Centre Of East Java Province, 2024). Exploration of groundwater is a de facto solution for ASEAN countries such as Indonesia during emergencies such as droughts (Vrba & Salamat, 2007). The groundwater potential map shows locations that sufficiently supply freshwater resources. This information can be used to formulate strategies for emergencies such as drought.

Groundwater exploration using terrestrial methods such as drilling, hydrological, and geo-electricity surveys is less effective for large areas, even if it produces great accuracy (Tolche, 2021; Yariyan *et al.*, 2022). Determining groundwater potential zoning using terrestrial survey techniques requires high-end instruments and skilled labor. In addition, surveys are not viable in areas with limited accessibility (Abdelouhed *et al.*, 2021). Consequently, terrestrial exploration of groundwater potential is unlikely to be feasible for large areas such as regency within a limited time, especially given the highly diverse geomorphological conditions in the study area. Time and cost remain a consideration in socio-environmental studies (Muthu & Sudalaimuthu, 2021). Geographical information system (GIS) and remote sensing (RS) data can be used for groundwater potential identification using specific approaches to evaluate groundwater influence parameters (Andualem & Demeke, 2019). Expert judgement approaches, such as the analytical hierarchy process (AHP), have limitations on subjectivity to compute parameter weight influencing groundwater (Esen, 2023).



**Copyright:** © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons-mons.org/licenses/by/4.0/>).

Currently, the latest approach is artificial intelligence (AI). Both machine learning (ML) and deep learning (DL) have optimal performance in predicting groundwater potential in various locations by considering multiple factors such as geology, climatology, topography, hydrology, and land cover (Pokhrel *et al.*, 2024). These approaches are more effective than expert judgment-based approaches and linear statistics because they can overcome the complex and non-linear relationship between groundwater influencing factors and groundwater presence through certain statistical approaches (Nguyen *et al.*, 2024).

Various AI-based algorithms have been used for groundwater potential mapping. Several ML algorithms applied in this context include artificial neural network (ANN), logistic model tree (MLT), logistic regression (LR), alternating decision tree (ALT), k-nearest neighbor (KNN), random forest (RF), AdaBoost (ADB), support vector machine (SVM), decision stump (DS), cat-Boost (CB), extreme gradient boosting (XGBoost), and voting ensemble (VE). With these algorithms, they produce area under the curve (AUC) values of 0.720 to 0.993 (Chen *et al.*, 2020; Halder *et al.*, 2024; Nguyen *et al.*, 2024; Parasar *et al.*, 2025; Pham *et al.*, 2019; Sarkar *et al.*, 2024). In addition, Arabameri *et al.* (2020) conducted a study that integrates ML algorithms with multi-criteria decision making (MCDM). Their work produced a groundwater potential model with an AUC of 0.925 by integrating the frequency ratio (FR) technique, Vise Kriterijumska Optimizacija I Kompromisno Resenje (VIKOR), and RF. DL algorithms can recognize complex relationships between dependent and independent variables, making them suitable for modelling applications that require high complexity. DL algorithms used for groundwater potential mapping in previous studies include convolutional neural network (CNN), long short-term memory (LSTM), deep learning neural network (DLNN), deep learning tree (DLT), deep boosting (DB), and recurrent neural network (RNN). From these studies, the AUC values varied from 0.801 to 0.910 (Chen *et al.*, 2022; Hakim *et al.*, 2022; Moughani *et al.*, 2023; Ragragui *et al.*, 2024). There is no universal judgment on the best algorithm for mapping groundwater potential, either DL or ML. This arises because ML and DL-based modelling are highly dependent on the characteristics of the training data and on how the algorithm processes them.

Many previous studies have focused on comparing algorithms with pre-defined hyperparameters for groundwater potential mapping (Arabameri *et al.*, 2020; Chen *et al.*, 2020; Chen *et al.*, 2022; Parasar *et al.*, 2025; Pham *et al.*, 2019). On the other hand, algorithm optimization through hyperparameter fine-tuning plays an important role in improving the generalization and robustness of models (Seifi & Niaki, 2023). Hyperparameter tuning is essential for controlling the behavior of algorithms during the training process, thereby affecting the performance of ML and DL models (Wu *et al.*, 2019). The most common techniques used in the hyperparameter tuning process in spatial modelling applications are grid search (Kanwar *et al.*, 2025; Lei *et al.*, 2025; Li *et al.*, 2023; Nguyen *et al.*, 2023; Shams *et al.*, 2024). These techniques are relatively simple, but the computational cost increases with the complexity of the search space and the number of hyperparameters examined. Methods such as the tree-parzen structured estimator (TPE) can reduce this computational cost by using a Bayesian-based approach. The hyperparameter tuning process can proceed more efficiently (Lai, 2024). However, the use of TPE for hyperparameter optimization in spatial modelling applications, particularly for groundwater potential mapping, remains relatively limited, especially in Indonesia.

Each algorithm has its own weaknesses in performing the prediction process. Performance improvements to address these weaknesses, through ensemble learning between single algorithms (often referred to as basic learning) (Wang *et al.*, 2024). Several techniques have been developed to integrate basic learning methods, namely boosting, bagging, and stacking. Among these three ensemble techniques, stacking has outperformed boosting and bagging according to several studies in various case studies, including for predicting soil moisture, agribusiness product prices, landslide susceptibility, and high-frequency trading (Das *et al.*, 2022; Ferrouhi & Bouabdallaoui, 2024; Ribeiro & Coelho, 2020; Wu & Wang, 2022). The stacking technique is advantageous because it can identify the strengths and weaknesses of each basic learning model. Stacking learning can combine heterogeneous basic learners, which bagging and boosting cannot. Stacking Learning (SL) is a technique that integrates the output of each base learning tier (tier 1 output) into tier 2 using meta-learning. Previous research integrates multiple base learning using ensemble learning techniques such as voting, boosting, and bagging (Lv *et al.*, 2022). Explicit and concurrent evaluation of ML, DL, and stacking learning for groundwater potential mapping remains limited, especially in Indonesian case studies. The use of AI algorithms for groundwater potential mapping in Indonesia in previous studies used RF, SVM, ANN, and XGBoost, with accuracy values varying from 0.58 to 0.978 (Nugroho *et al.*, 2024; Nugroho *et al.*, 2024). Performing DL and stacking, which integrates multiple algorithms, has not yet been conducted in Indonesia.

Previous studies using DL and ML algorithms have mainly considered topography, land cover, geology, hydrology, and hydrogeology as influence factors for groundwater potential mapping. However, to the best of our knowledge, geophysical factors such as the mean shear-wave velocity at 30 m depth (vs30) and the complete Bouguer anomaly (CBA) have not been considered in studies mapping groundwater potential using ML and DL algorithms. Both parameters have an indirect influence on the presence of groundwater.

There are even studies that interpret groundwater potential from gravity measurements (Alshehri & Mohamed, 2023; Handayani *et al.*, 2018). The CBA parameter is a product of the gravity anomaly. According to Newton's law, gravitational force is a function of the mass distribution and spatial location within and on the surface of the Earth; changes in groundwater accumulation affect the water mass in a region. Thus, the gravity force value affects the presence of groundwater in an area. In addition, with the gravity force value, it is possible to interpret the porosity conditions of rock units that affect the groundwater transport system (Rustadi *et al.*, 2022). Vs30 correlates with geological maturity and primary lithology (Meneisy *et al.*, 2023). Vs30 values can be used as an indication of fluid saturation levels in a rock layer, particularly in early-stage, unconsolidated lithological units (Okay & Özacar, 2023).

Shear wave velocity provides information related to aquifer characteristics. It can provide information related to geological properties such as hydraulic conductivity, permeability, storability, and porosity (Khalilidermani & Knez, 2024). Despite CBA and VS30 indirectly affecting groundwater potential, these two geophysical parameters can enhance the performance of groundwater potential prediction. These geophysical factors provide subsurface information that complements other parameters. The model can identify groundwater potential characteristics in the study area by leveraging hidden geophysical controls, yielding a more reliable model.

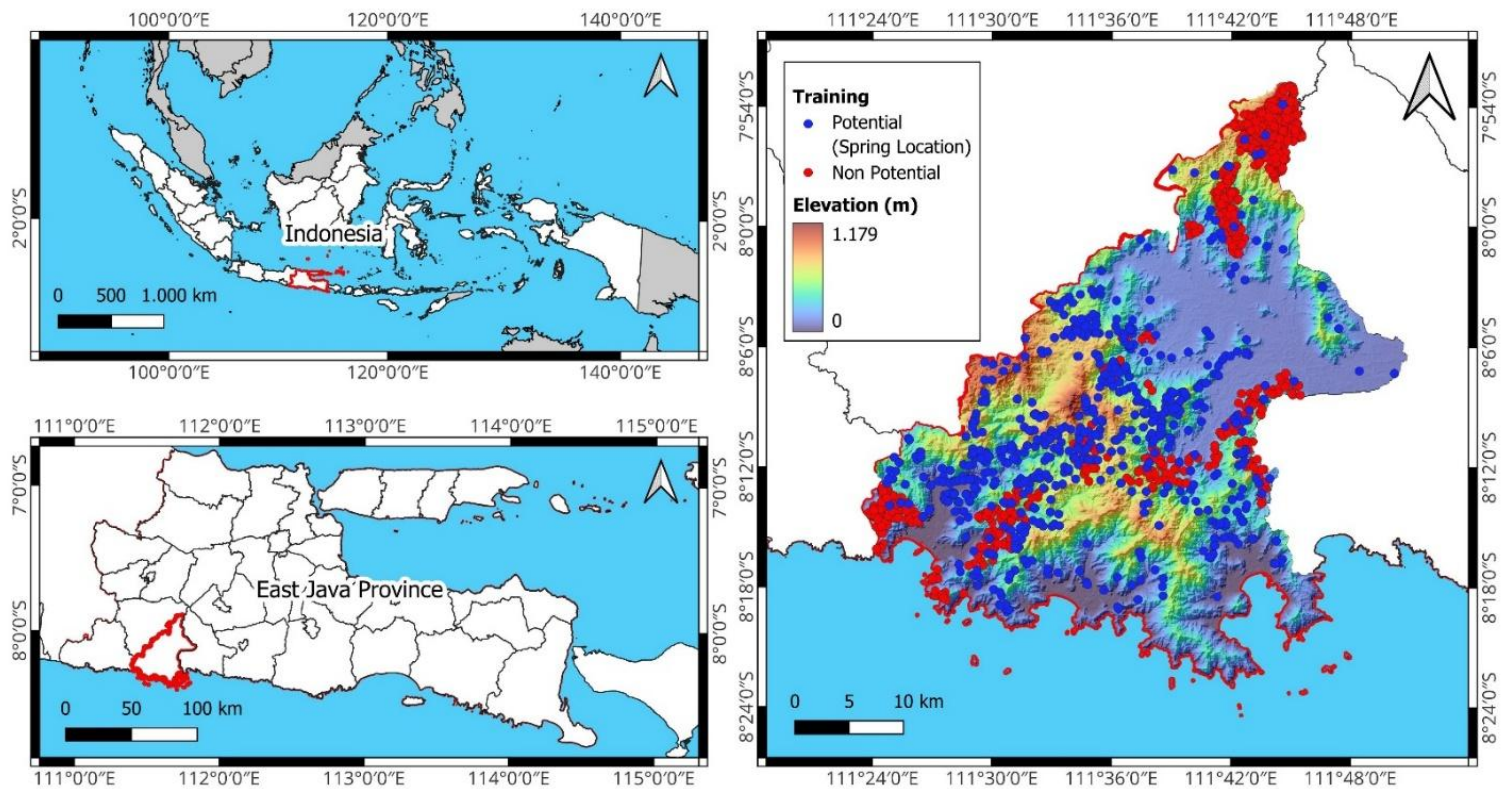
Therefore, to fill gaps in previous research, this study evaluates the implementation and integration of ML and DL algorithms for mapping groundwater potential in Indonesia, with particular focus on the Trenggalek Regency area. In addition, this study employs hyperparameter fine-tuning with TPE to optimize model performance. In this modelling, geophysical factors such as CBA and Vs30 are considered modelling parameters. The ML algorithms used are RF and GBDT. Meanwhile, the DL algorithms used are 1-D CNN and RNN. The ML and DL algorithms were chosen because, based on the literature review, they have high performance and can identify relationships between groundwater influencing factors and the potential presence of complex groundwater. There are three schemes for integrating the algorithms: integrating each DL and ML algorithm, and combining DL and ML. The integration technique used is stacking with logistic regression (LR) as the meta-learner. Therefore, this study will produce seven models that will be evaluated for performance using accuracy, Cohen's kappa, Matthews correlation coefficient, and receiver operating characteristic (ROC) AUC. This study then compares the performance of the nine models to obtain the optimal model for accurately identifying groundwater potential in Trenggalek Regency.

## 2. Research Methods

### 2.1. Study Area

Trenggalek Regency is one of the regions in East Java Province, Indonesia. Trenggalek is located at 111° 23' 31.2" to 111° 50' 48.95" E, and 8° 23' 18.46" to 7° 53' 18.16" S. Trenggalek Regency experiences hydrometeorological drought disasters that occur annually. There were 123 drought-related disasters in Trenggalek Regency throughout 2023, highlighting the water-scarcity problem. The population growth rate of Trenggalek Regency was 0.51% in 2023, in line with the increasing dependence on groundwater as a freshwater source, which reached 135,293 m<sup>3</sup>. Therefore, a groundwater potential map is urgently needed, as it can help in optimizing the fulfillment of clean water needs in Trenggalek Regency.

Trenggalek Regency has diverse geomorphological conditions. Two-thirds of Trenggalek Regency is mountainous. Trenggalek Regency has elevations ranging from 0 to 1,179 metres above sea level. Trenggalek Regency is a tropical region with an average rainfall of 1,879 to 2,756 mm/year and an average land surface temperature of 24°C to 31°C. The geology of Trenggalek Regency is dominated by the Mandalika formation, which is Neogene in age and consists of interbedded volcanic breccia, lava, tuff, and sandstone.



**Figure 1.** Location of Trenggalek Regency and the Distribution of Training Points.

## 2.2. Conditioning Factors of Groundwater Potential

The selection of conditioning factors can affect the quality of groundwater potential modelling results. The conditioning factors were determined based on a literature review and data availability. This study considered 18 parameters such as TWI, river density (RD), SPI, land cover (LC), normalized difference vegetation index (NDVI), annual rainfall (AR), evapotranspiration (ET), land surface temperature (LST), atmospheric pressure in the surface (SP), elevation, slope, slope aspect, plan curvature (PC), lineament density (LD), geological unit (GU), soil type (ST), complete bouguer anomaly (CBA), and average shear wave velocity to 30m (VS30). The 18 parameters are classified into six factors: hydrology, land cover, climate, topography, geology, and geophysics. Detailed conditioning factor information can be seen in Table 1.

**Table 1.** The groundwater potential conditioning factors used in the study are described, and their influence on groundwater potential is discussed.

Factor	Parameter	Role in Groundwater Potential	Source
Hydrology	Topographic Wetness Index	Control the ability of groundwater infiltration rate in an area	(Senapati & Das, 2021)
	River Density	Influence on surface runoff and water infiltration. Hydraulic connections can occur between surface water and groundwater in areas with high RD, which will affect the groundwater recharge process	(Abijith <i>et al.</i> , 2020)
Land Cover	Stream Power Index	Negative SPI values are easier to retain groundwater	(Echogdali <i>et al.</i> , 2022)
	Land Cover	Controls soil moisture level, penetration, and surface runoff rate. Therefore, it can directly affect the level of groundwater recharge	(Senapati & Das, 2021)
	NDVI	The higher the vegetation density, the greater the groundwater potential of the area	(Senapati & Das, 2021)
Climate	Annual rainfall	Main source of groundwater recharge that impacts surface runoff and depends on the duration, volume, and intensity of rainfall	(Ahmed <i>et al.</i> , 2020)
	Evapotranspiration	Increased ET can shift the proportion of the rainfall as a source of groundwater recharge	(Ghiat <i>et al.</i> , 2021)
	Land surface temperature	Groundwater quantity decreases when LST increases because water will evaporate faster from the soil	(Silver <i>et al.</i> , 2007)



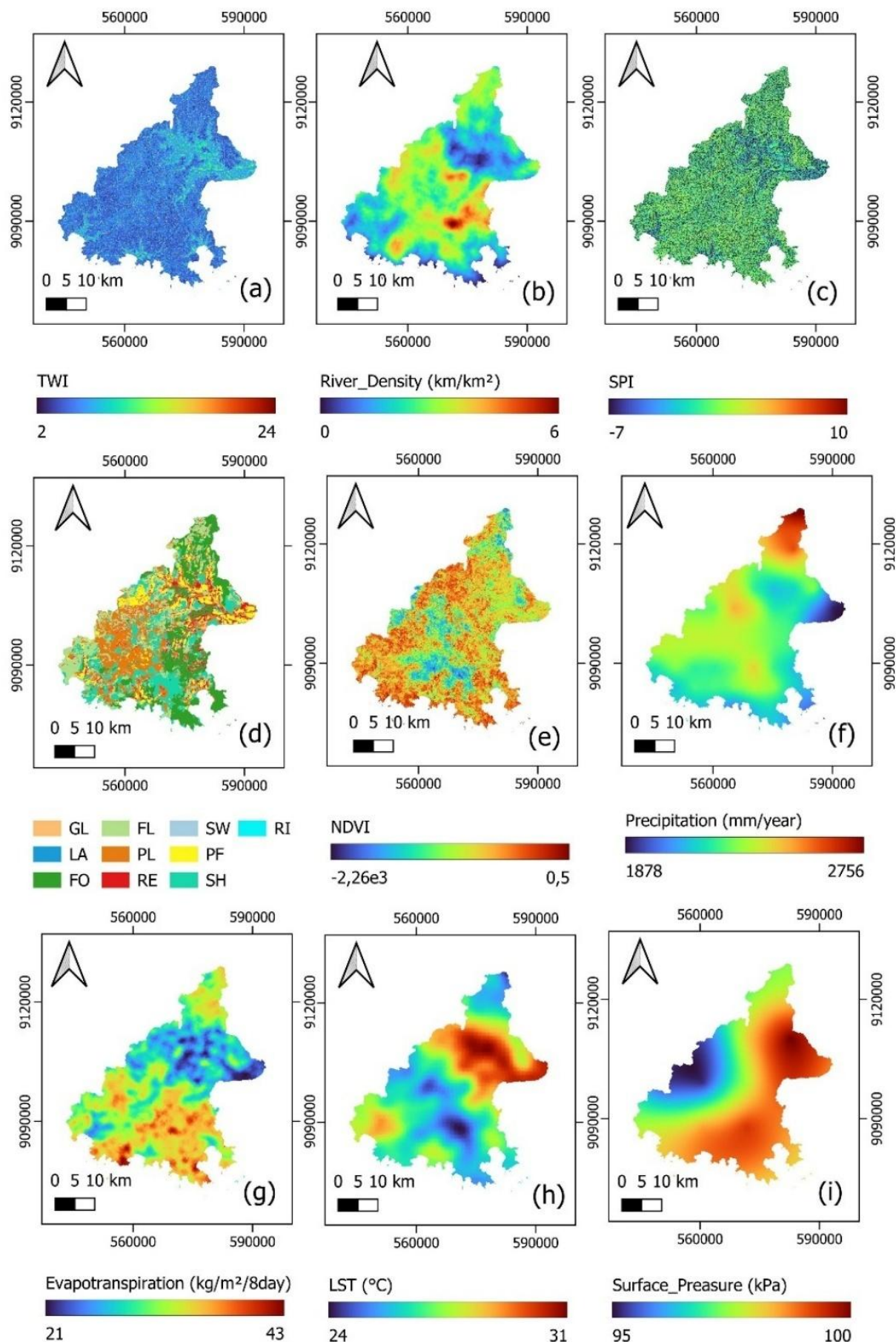
**Table 2.** (Continued).

Factor	Parameter	Role in Groundwater Potential	Source
Topographic	Atmospheric pressure at the surface	SP produces subsurface material compression, thereby narrowing rock pores. Therefore, it can influence the water absorption capacity of the aquifer	(Kramer <i>et al.</i> , <a href="#">2023</a> )
	Elevation	Influence the local rainfall frequency and processes.	(Berhanu & Hatiye, <a href="#">2020</a> )
	Slope	Affect the rate of water flow and infiltration in an area.	(Maskooni <i>et al.</i> , <a href="#">2020</a> )
Geology	Slope aspect	Influence the moisture level and vegetation density in an area.	(Benjmel <i>et al.</i> , <a href="#">2022</a> )
	Plan curvature	Curvature affects water flow convergence and divergence, as well as flow acceleration and declination.	(Al-Abadi <i>et al.</i> , <a href="#">2016</a> )
	Geological Type	Rock formation can influence the permeability and porosity characteristics of an aquifer.	(Muavhi <i>et al.</i> , <a href="#">2022</a> )
	Soil Type	Determine permeability in an area. In addition, texture, granule size, and soil composition influenced soil water content.	(Rehman <i>et al.</i> , <a href="#">2024</a> )
Geophysical	Lineament Density	It represents secondary porosity and is crucial in regulating the hydrometeorological conditions of groundwater flow and storage.	(Berhanu & Hatiye, <a href="#">2020</a> )
	VS30	VS30 can be used as an indicator of fluid saturation in a rock layer, especially in early unconsolidated lithological units.	(Rustadi <i>et al.</i> , <a href="#">2022</a> )
	CBA	CBA can interpret the porosity conditions of rock units that affect the groundwater transport system	(Okay & Özacar, <a href="#">2023</a> ; Khalilidermani & Knez, <a href="#">2024</a> )

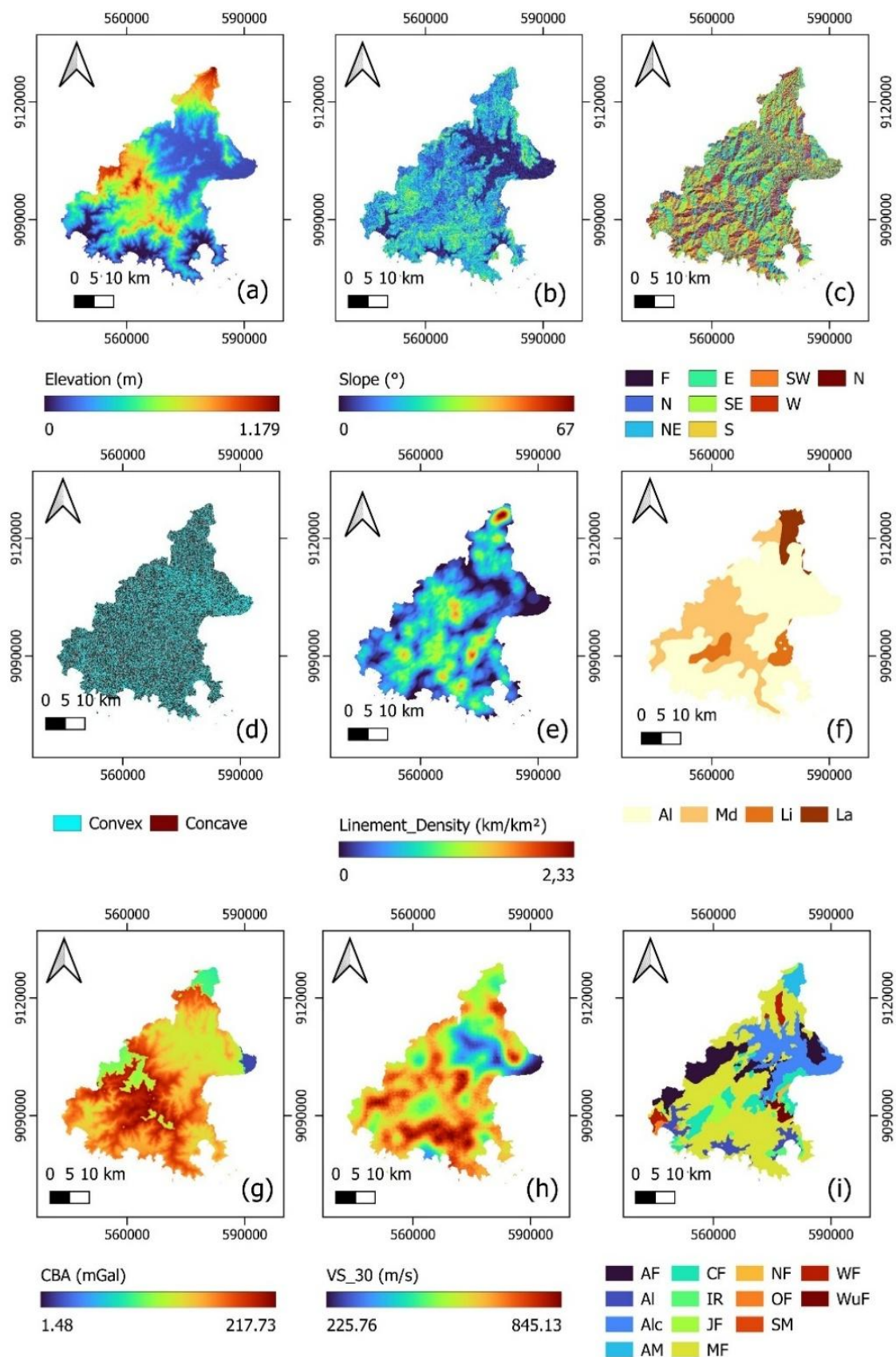
There were 1,840 samples used to train the model, consisting of 740 potential points and 740 non-potential points. The possible points used are spring locations with discharge rates of 10–100 liters/second. Meanwhile, non-potential points were random points located in GU with low porosity (0.2% - 8%). The distribution of training points can be seen in Figure 1. All parameters were resampled in raster format with a spatial resolution of 30m. The details of the data used in this study can be seen in Table 3. The data was processed with spatial data processing software to produce modelling parameters. The parameters can be shown in Figure 2 and Figure 3.

**Table 3.** Summary of data resources and derived parameters required in this study.

Data	Source	Spatial resolution/ scale	Derivative parameter(s)
Spring locations	Local government of Trenggalek Regency	1:50,000	Potential class training
Digital elevation model (DEM)	National Geospatial Information Agency of Indonesia (well known as BIG)	8m	Elevation, Slope, slope aspect, TWI, LD, and SPI
River density	BIG	1:25,000	RD
Land cover	BIG	1:25,000	LC
Landsat-8 imagery	United States Geological Survey (USGS), which is obtained from the Google Earth (GEE) platform	30 m	NDVI
Annual rainfall	Climate Hazard Group Infrared Precipitation with Station Data (CHIRPS) from the GEE catalogue	0.5 arc degree	AR
MODIS imagery	National Aeronautics and Space Administration (NASA) from the GEE catalogue.	500 m	ET and LST
Geological unit map	Volcanology and Geological Hazard Mitigation Center of Indonesia (well known as PVMBG)	1:100,000	GU and non-potential training
Soil type map	Local government of Trenggalek Regency	1:50,000	ST
Global Gravity Model Plus (GGm-Plus) Free-air gravity model	Murray lab	200 m	CBA
SRTM2Gravity terrain correction model	Murray lab	90 m	CBA
VS30	USGS	30 arc seconds	Vs30
ERA5-Land	European Center for Medium-Range Weather Forecast (ECMWF)	11 km	SP



**Figure 2.** Conditioning Parameters 1 to 9 for Groundwater Potential Mapping; (a) TWI; (b) RD; (c) SPI; (d) LC (GL: grass land; LA: lake; FO: forest; FL: field land; PL: plantation; RE: residential; SW: swamp; PF: paddy field; SH: shrub; RI: river); (e) NDVI; (f) AR; (g) ET; (h) LST; (i) SP.



**Figure 3.** Conditioning Parameters 10 to 18 for Groundwater Potential Mapping; (a) Elevation; (b) Slope; (c) Aspect (F: Flat; N: North; NE: Northeast; E: East; SE: Southeast; S: South; SW: Southwest; W: West; N: North); (d) PC; (e) LD; (f) ST (Al: Alluvium; Md: Mediterranean; Li: Litosol; La: Latosol); (g) CBA; (h) VS30; (i) GU (AF: Arjosari formation, porosity  $\pm 25\%$ -40%; AI: alluvium,  $\pm 25\%$ -40%; Alc: Alluvium and coastal deposits,  $\pm 25\%$ -55%; AM: Argohalangan morphocet,  $\pm 0.2\%$ -8%; IR: Intrusive rock,  $\pm 0.1\%$ -8%; JF: Janten formation,  $\pm 25\%$ -40%; MF: Mandalika formation,  $\pm 7\%$ -25%; NF: Nampol formation,  $\pm 45\%$ -55%; OF: Oyo formation,  $\pm 25\%$ -40%; SM: Sedudo morphonite,  $\pm 0.2\%$ -8%; WF: Wonosari formation,  $\pm 0.27\%$ -4.1%; WuF: Wuni formation,  $\pm 48\%$ -52%).



### 2.3. Pre-Processing

This study uses two types of data: discrete and continuous. To ensure data reliability in the modelling process, pre-processing was carried out. Discrete data are transformed into numerical form using the one-hot encoding (OHE) technique. OHE transforms each class into a binary vector, with 1 if it appears in a particular row and 0 if it does not. The OHE technique can be seen in equation 1 (Samuels, 2024). Meanwhile, for numerical data, the Yeo-Johnson transformer (YJT) and standard scaler are applied. The YJT is a technique that aims to transform the data distribution to near normal (Yeo & Johnson, 2000). On the other hand, the standard scaler or z-score standardization technique is a method to ensure model independence from specific feature scales. The standard scaler transforms data to a mean of 0 and a standard deviation of 1 (Demir & Sahin, 2024). The YJT and standard scaler can be seen in Equations 2 and 3.

$$v_i[j] = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

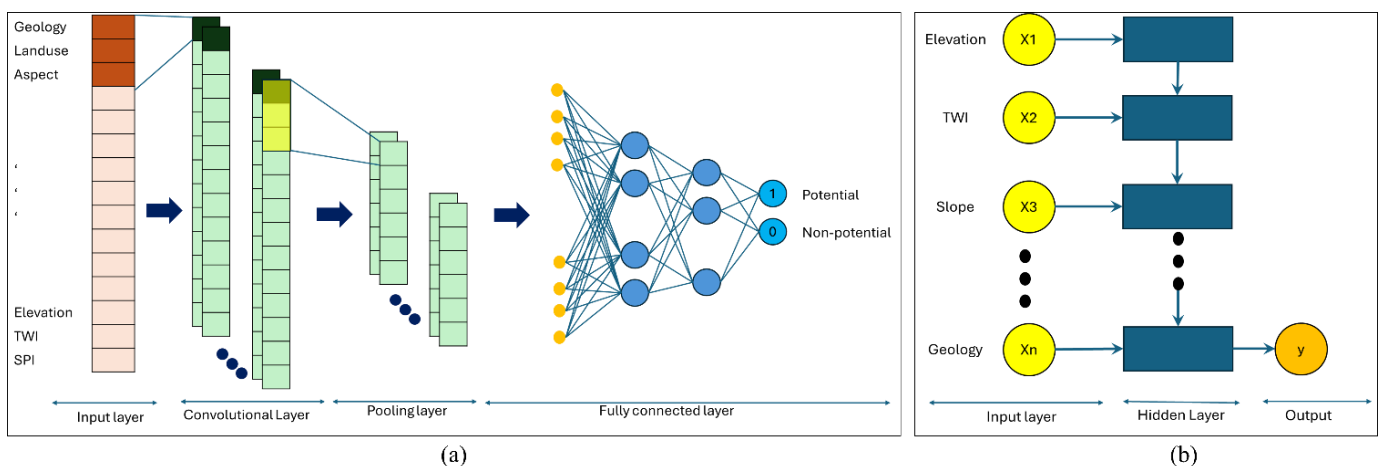
$$\psi(y, \lambda) = \begin{cases} \frac{(y+1)^\lambda - 1}{\lambda} & (\lambda \neq 0, y \geq 0) \\ \log(y+1) & (\lambda = 0, y \geq 0) \\ \frac{-[(-y+1)^{2-\lambda} - 1]}{2-\lambda} & (\lambda \neq 2, y < 0) \\ -\log(-y+1) & (\lambda = 2, y < 0) \end{cases} \quad (2)$$

$$Z = \frac{x - \bar{x}}{\sigma} \quad (3)$$

Where,  $v_i[j]$  Is OHE vector in parameter-I, class j;  $\psi(y, \lambda)$  is the result of the YJT transform;  $y$  is original value of the parameter;  $\lambda$  is a constant estimated by maximum likelihood estimation (MLE);  $Z$  is a standardized value;  $x$  is the original value before the standardized value;  $\bar{x}$  is the mean value; and  $\sigma$  is the standard deviation (Demir & Sahin, 2024; Samuels, 2024; Yeo & Johnson, 2000).

### 2.4. Modelling Process

This study evaluates the use of ML and DL for groundwater potential mapping in Trenggalek Regency, Indonesia. In addition to using each algorithm individually, this study used the stacking technique to integrate ML and DL models. Each algorithm performed hyperparameter tuning using a tree-structured Parzen estimator (TPE). The two ML algorithms used are RF and GBDT. RF is an ML algorithm that uses bootstrapping and bagging to form multiple DTs. Each DT is constructed with several n-parameters, and n-data are randomly selected. The result of the RF is the voting of the results of each single DT (Breiman, 2001). GBDT is a straightforward combination of DT and gradient descent. GBDT comprises two main processes: boosting and classification. These two processes can improve model performance and minimize overfitting through the regularization process within sequential steps to iterate on the loss function (Chen *et al.*, 2020). Loss function is the difference of prediction and the actual value (Yang *et al.*, 2024).



**Figure 4.** The Architecture of the DL Model: (a) 1D CNN; (b) RNN.

CNN and RNN were used as DL algorithms in this study. CNN can solve complex problems and consists of three main layers: convolutional layers (CL), pooling layers (PL), and fully connected layers (FCL) (Zhang *et al.*, 2022). CL is a convolutional filter (also known as a kernel) where the input features expressed by a matrix of dimension  $N \times 1 \times P$  are convolved with a filter of size



$n \times l \times p$ , where  $N > n$  and  $P > p$ , resulting in an output feature map. PL is a subsampling technique applied to the feature map generated by CL. After the PL layer, the feature map will be flattened and moved to the FCL layer. In this layer, each neuron is connected to all neurons from the previous hidden layer, so it is called a fully connected (FC) approach (Fang *et al.*, 2020). The overall architecture of the CNN used in this study is illustrated in Figure 4(a). RNNs are designed to process data sequentially by considering hidden states that capture information about previous inputs. RNNs have recurrent connections (Ji *et al.*, 2023), as shown in Figure 4(b). Such connections allow information to loop within the network. The network memorizes the input information of the previous state ( $t-1$ ) and then applies it to the calculation of the output at state  $t$ . The nodes between hidden layers are interconnected, and the input to the hidden layer includes not only the production of the previous input layer but also the output of the hidden layer from the previous state (Mienye *et al.*, 2024).

This study uses two strategies to optimize the prediction model: hyperparameter optimization for each algorithm and stacking. Stacking is a method of successfully integrating different ML or DL models (Sikora, 2017). In the stacking model, heterogeneous prediction models are used, where the output of all the algorithms is given as input to the meta-model to get a more stable model (Khoshkroodi *et al.*, 2024). Logistic regression (LR) is the meta-learner applied in this study, which utilizes the maximum likelihood method to discover the 'best fit' association between input and output (Sun *et al.*, 2018).

TPE is the technique used to tune each algorithm's hyperparameters. TPE is a Bayesian technique used for optimizing hyperparameters that utilizes a kernel density estimator to model the probability distribution of hyperparameters using an objective function (Ishii *et al.*, 2023). TPE imitates the density distribution function to optimize hyperparameters in both the optimal and worst-case settings. The iterative approach is used to select hyperparameters for sampling, evaluate model performance, and update probability density functions. The conditional probability theory  $p(x|y)$  is used, where  $x$  is the hyperparameter and  $y$  is the loss from using that hyperparameter. The first step in TPE optimisation is selecting the threshold loss ( $y^*$ ) given the available data, such as based on the median. The formation of probability density functions  $l(x)$  and  $g(x)$  is supported by the threshold loss as seen in equation 4 (Rong *et al.*, 2021). Afterwards, other hyperparameter combinations are determined by maximizing the likelihood between  $l(x)$  and  $g(x)$  using equation 5 (Islam *et al.*, 2024).

$$p(x|y) = \begin{cases} l(x) & \text{if } y < y^* \\ g(x) & \text{if } y > y^* \end{cases} \quad (4)$$

$$\operatorname{argmax} \left( \frac{l(x)}{g(x)} \right) \quad (5)$$

## 2.5. Model Evaluation and Inspection

Model evaluation is a validation process for the accuracy and reliability of a prediction model. Evaluation in this study was carried out on two datasets, namely training and testing, to determine whether each model suffered from overfitting. Model evaluation is carried out using four parameters. Model evaluation parameters can be seen in Table 4. Where TP (true positive) is a positive class that is predicted positive; FP (false positive) is a negative class that is predicted positive; FN (false negative) is a positive class that is predicted negative, and TN (true negative) is a negative class that is predicted negative (Stern, 2021; Tharwat, 2018).

**Table 4.** Model Evaluation Metrics.

Metric	Equation	Range value	Best Value
Accuracy (ACC)	$\frac{TN + TP}{TN + TP + FP + FN}$	0-1	1
Matthew's Coefficient Correlation (MCC)	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	(-1) -1	1
Cohen's kappa (CK)	$\frac{2(TP \times TN - FP \times FN)}{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}$	0-1	1
ROC-AUC	$\int TPR d(FPR)$	0-1	1

Model evaluation was performed using k-fold cross-validation (KFCV). KFCV divides the dataset into n-fold. 1/n fold is used to evaluate the model, and the other n-1 fold is used to train the model. This research uses five folds. This means the training-to-testing ratio is 80:20. The KFCV technique is used to detect overfitting by using the t-test with degrees of freedom (df) n-1, which was four. Model inspection improves our understanding of how to interpret the model. Two additional

approaches to model inspection in this study were permutation feature importance (PFI) and partial dependence plots (PDPs).

PFI approximates the relative importance of parameters in a predictive model by permuting each parameter's values. As accuracy varies significantly, the parameter has an important contribution to the model (Altmann *et al.*, 2010). The PDP employs the Shapley additive explanation (SHAP) value, which computes the Shapley value for every sample in the dataset. Shapley values offer a game-theoretic mechanism for reasonably apportioning the prediction output to individual input features. SHAP was initiated by Shapley, (1953) with game theory. It considers every possible combination of parameters in a row of data to calculate the marginal contribution of each parameter. This value is calculated for each parameter value. The SHAP value can be calculated using equation 6. In this equation,  $\phi_i v(x)$  is the SHAP value for feature  $i$  with the prediction model  $v(x)$ ;  $S$  is the feature subset dataset;  $F$  is the entire feature;  $S \subseteq F/\{i\}$  is a subset of feature  $F$  without considering parameter- $i$ ;  $v(x_{S \cup \{i\}})$  is the prediction output with subset  $S$  considering feature- $i$ ; and  $v(X_S)$  is the prediction output with subset  $S$  without considering  $i$ -feature (Lundberg & Lee, 2017).

$$\phi_i v(x) = \sum_{S \subseteq F/\{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [v(x_{S \cup \{i\}}) - v(X_S)] \quad (6)$$

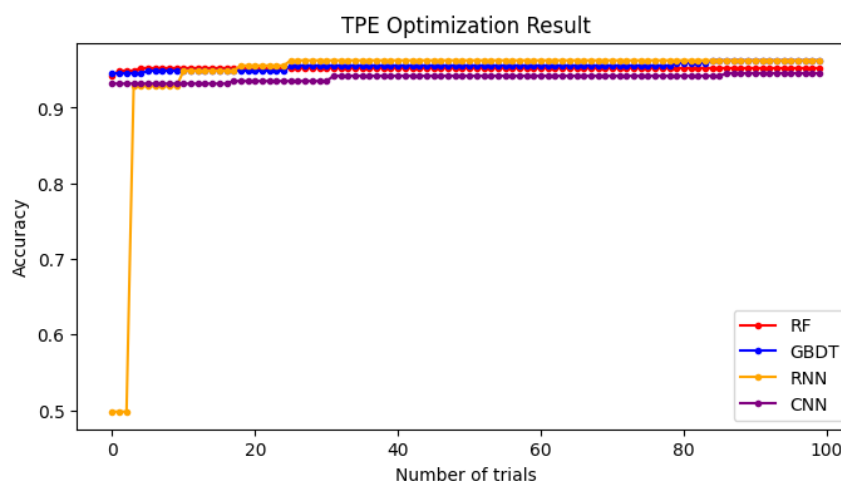
### 3. Results and Discussion

#### 3.1. Hyperparameter Optimization Result

Hyperparameter tuning is one of the crucial steps of constructing an optimal prediction model (Jafari *et al.*, 2023). TPE reduces computational cost by selecting the most appropriate pair using a Bayesian-based method, rather than evaluating all pairs of hyperparameters. The optimum hyperparameter configurations for each algorithm are presented in Table 5. In this study, the number of iterations is limited to 100 trials, with test-set accuracy as the objective value. Although there is no universal judgment on the number of trials that ensures maximum accuracy or convergence, several studies have used a setting of 100 trials and achieved convergent results (Fan *et al.*, 2022; Kilic *et al.*, 2024; Zhao *et al.*, 2025).

**Table 5.** Optimization Result of the Hyperparameter for Each Algorithm Using TPE.

Algorithm	Hyperparameter	Search spaces Range	Step	Distribution	Optimal value
RF	Max depth each DT	[1,32]	1	Uniform	5
	Max random feature for each DT	['sqrt', 'log2', 'None']	-	Choice	'sqrt'
	Min samples each leaf	[1,10]	1	Uniform	9
	Min samples can be split	[2,10]	1	Uniform	9
	Number of DT	[100, 2,000]	100	Uniform	1,500
GBDT	Number of boosting	[100, 2,000]	100	Uniform	100
	Max feature for best split consideration	['sqrt', 'log2', 'None']	-	Choice	None
	Max depth each DT	[1,32]	1	Uniform	13
	Min samples make the internal node	[2,10]	1	Uniform	4
	Min samples each leaf	[1,10]	1	Uniform	6
	Learning rate	[0.07,0.08]	-	Choice	0.07
	Number of subsamples	[0.5,0.6]	-	Choice	0.6
	Number of recurrent network layers (RNL)	[1,5]	1	Uniform	3
	Number of neurons in each RNL	[50,150]	50	Uniform	150
	Activation function in the output layer	['sigmoid', 'softmax']	-	Choice	'softmax'
CNN	Optimizer	['Adam', 'Adamax']	-	Choice	'Adam'
	Activation function in output layer	['sigmoid', 'softmax']	-	Choice	'Sigmoid'
	Number of FCL	[1,6]	1	Uniform	4
	Number of CL	[1,4]	1	Uniform	2
	Number of neurons each FCL	[64, 1,024]	2	Quniform	1,024
	Optimizer	['Adam', 'Adamax']	-	Choice	'Adam'



**Figure 5.** The TPE Optimization Objective Value in the Trial Process of Four Basic Learning Models.

Optimal hyperparameter optimization using TPE can be visualized in a line diagram, as shown in Figure 5. Figure 5 illustrates the relationship between trial  $n$  and the objective value. According to the figure, the RF, CNN, and GBDT algorithms have an accuracy of more than 0.9 at the initial state or trial to 0. Meanwhile, the RNN algorithm has an accuracy of 0.49 at the initial state. It happens because, in the initial state, the hyperparameter configuration is randomly selected. The best trial or convergence conditions for the RF, GBDT, CNN, and RNN algorithms are 4, 84, 25, and 86, respectively. The highest objective values in the RF, GBDT, CNN, and RNN algorithms are 0.952, 0.962, 0.962, and 0.945, respectively. Overall, GBDT and CNN have the highest objective values, but CNN finds convergent conditions earlier than GBDT. While it has a very low aim value at the initial state, the RNN algorithm experiences a drastic increase in the third trial.

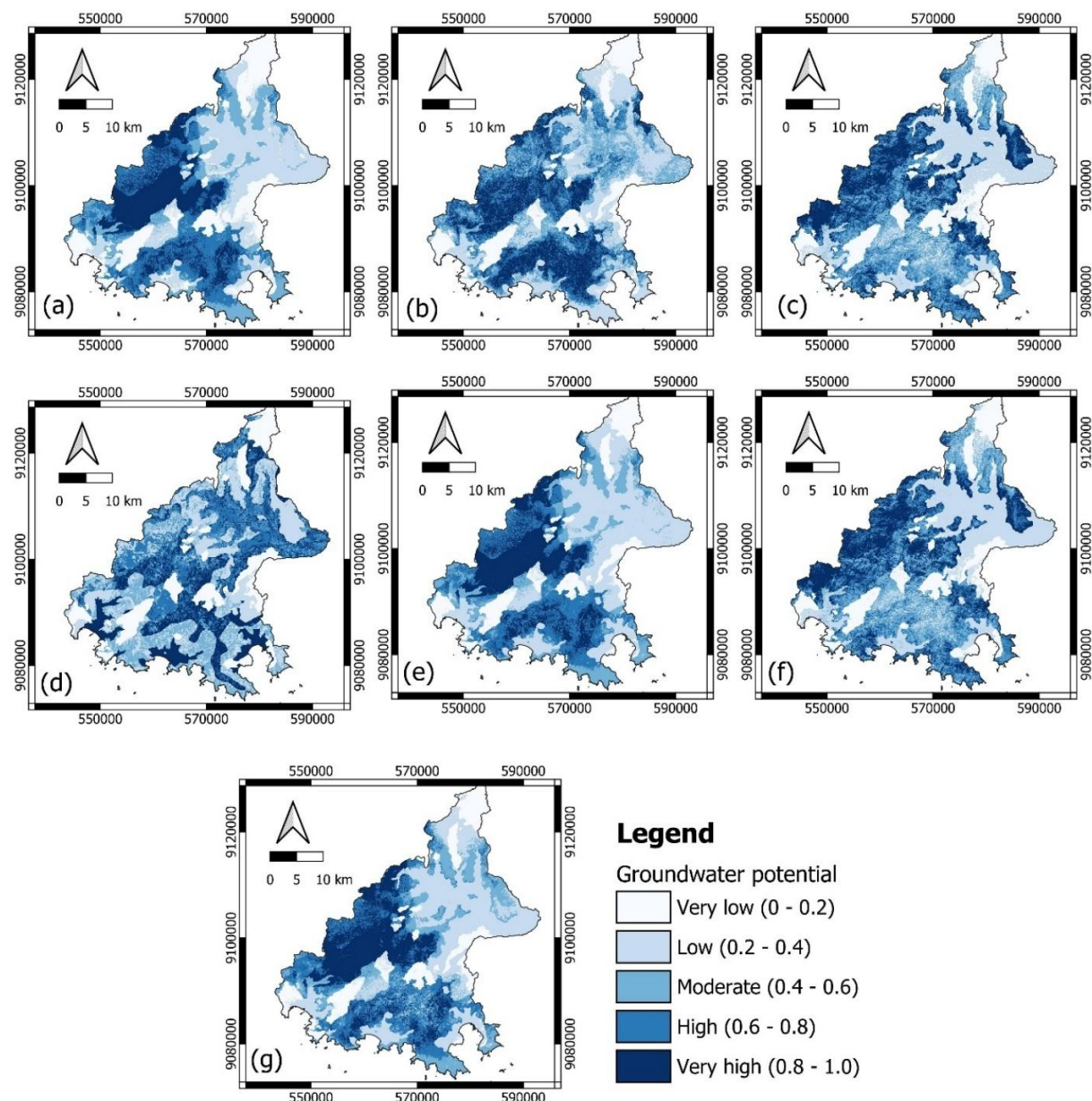
### 3.2. Groundwater Potential Map: Each Model

Groundwater potential maps are generated by applying each prediction model to pre-processed parameter data, including categorical data encoding, transformation, and standardization for numerical data. There are seven prediction models generated in this study, namely RF, GBDT, RNN, 1D CNN, RF-GBDT stacking, 1D CNN-RNN stacking, and RF-GBDT-RNN-1D CNN stacking. The output of the modelling process is the probability of groundwater potential. The range of probability values is 0 to 1. The closer the value is to 1, the higher the groundwater potential. Meanwhile, the closer the value is to 0, the lower the groundwater potential. The groundwater potential probability is classified into five classes: very low (0 - 0.2), low (0.2 - 0.4), medium (0.4 - 0.6), high (0.6 - 0.8), and very high (0.8 - 1). Equal intervals were used to demonstrate the consistency and interpretability of groundwater potential mapping results expressed as continuous probability values—equal intervals allowed for an objective comparison of groundwater potential class interpretations across prediction models.

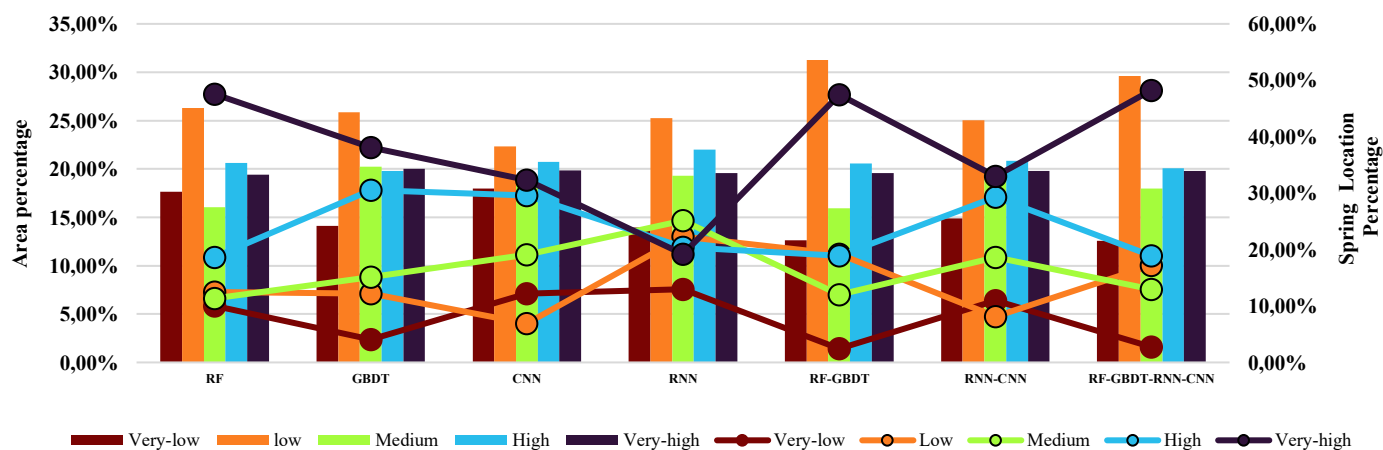
Natural and quantile classification techniques are highly dependent on the distribution of each prediction model result, leading to less objective comparative analysis. A few spatial modelling studies have used the equal interval technique to facilitate comparisons among models (Bi *et al.*, 2025; Hossen *et al.*, 2025; Nurwatik *et al.*, 2022; Prasad *et al.*, 2020). The groundwater potential map produced in this study can be seen in Figure 6. The map is visualized with graduated colors ranging from light blue to dark blue. Light blue represents very low potential, while dark blue represents very high potential. Overall, the probability distribution almost has the same pattern between prediction models, where high probability is found in the west of the study area. Meanwhile, low probabilities are found in the north, some in the southwest.

Figure 7 presents the percentage area of each potential class and the percentage of spring locations within each class. Each model indicates that the class with the largest area is the low class, with an area of 22.32% to 31.29%. The very low class is the smallest area, except for the RF model. The very low class has an area percentage ranging from 12.57% to 17.66%. For the RF model, medium is the smallest-area potential class, with 16.03%. The percentage of potential points ought to be related in a direct relationship to the groundwater potential class. The lowest potential class has the least percentage of points at 4.05%, and the highest percentage is in the very high class at 38.11%. Most models show an inverse trend between the low and medium classes, with more potential points in the class of low potential. However, in general all models showed that the class of very high potential contained the most significant proportion of potential points.





**Figure 6.** Groundwater Potential Maps Generated from Seven AI-based Algorithms: (a) RF; (b) GBDT; (c) CNN; (d) RNN; (e) RF-GBDT; (f) RNN-CNN; (g) RF-GBDT-RNN-CNN.



**Figure 7.** Area Percentages (Shown by Bar Chart) and Spring-Location Percentages (Shown by Line Chart) by Potential Class.

### 3.3. Model Performance

The predictive model that has been established was evaluated through the KFCV method and utilized the ACC, MCC, CK, and ROC-AUC assessment metrics. The evaluation results are represented through the average value and standard deviation. Standard deviation can be used to quantify the robustness of each metric across the five cross-validation iterations. The results are shown in Table 6. A low standard deviation for each metric indicates stable performance across all folding schemes. Overall, all models achieved high prediction accuracy ( $ACC > 0.9$ ; MCC and  $CK > 0.85$ ). The GBDT algorithm demonstrated better consistency and stability than other single algorithms, as evidenced by its lower performance variance. GBDT, which iterates on bias values, can produce more stable prediction models with the best performance. Although both are built on DT, GBDT is superior to RF because it minimizes prediction variance while placing greater emphasis on bias. However, the performance of GBDT and RF is not significantly different, with only a 0.01 difference on several parameters, and GBDT is relatively more stable than RF. The bagging technique, which randomly selects features, makes RF less stable than the boosting technique used in GBDT.

**Table 6.** Metric Evaluations Using KFCV and P-Value (PV) Results of the T-test Between the Training and Testing Datasets for Each Model.

Model	Training Acc	MCC	CK	Testing Acc	MCC	CK	PV Acc	MCC	CK
RF	0.950 ± 0.002	0.901 ± 0.004	0.900 ± 0.004	0.938 ± 0.019	0.877 ± 0.038	0.876 ± 0.038	0.244	0.247	0.244
GBDT	0.950 ± 0.003	0.902 ± 0.006	0.901 ± 0.006	0.938 ± 0.014	0.878 ± 0.028	0.876 ± 0.027	0.170	0.173	0.170
CNN	0.943 ± 0.005	0.888 ± 0.011	0.886 ± 0.011	0.930 ± 0.013	0.863 ± 0.023	0.859 ± 0.026	0.128	0.122	0.128
RNN	0.935 ± 0.005	0.877 ± 0.008	0.871 ± 0.009	0.936 ± 0.012	0.878 ± 0.021	0.872 ± 0.024	0.946	0.936	0.945
RF-GBDT	0.957 ± 0.004	0.915 ± 0.008	0.915 ± 0.008	0.957 ± 0.016	0.915 ± 0.032	0.915 ± 0.032	1.000	0.996	1.000
CNN-RNN	0.939 ± 0.005	0.915 ± 0.009	0.913 ± 0.010	0.939 ± 0.023	0.883 ± 0.043	0.877 ± 0.046	0.989	0.997	0.989
RF-GBDT-CNN-RNN	0.955 ± 0.003	0.910 ± 0.006	0.909 ± 0.007	0.953 ± 0.014	0.907 ± 0.028	0.906 ± 0.029	0.855	0.865	0.856

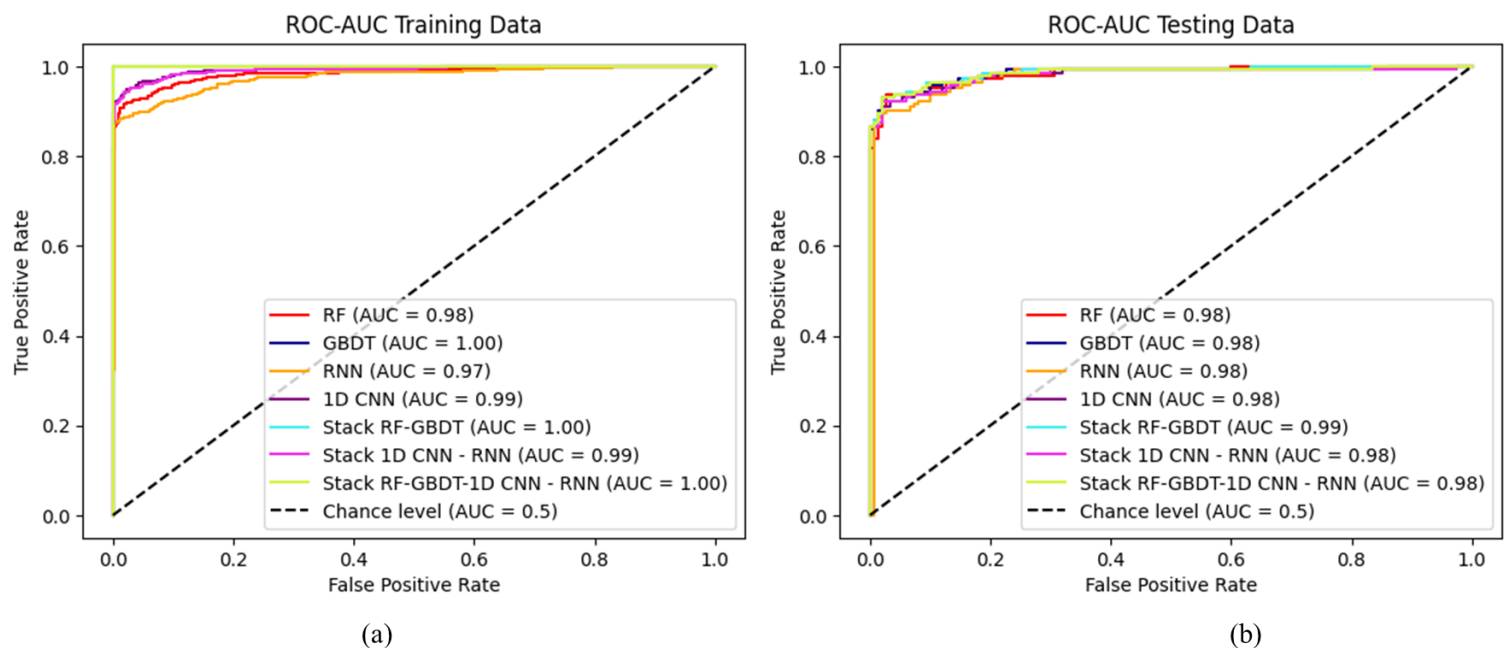
Stacking learning (SL-ML) models that integrate RF and GBDT achieve the best performance and generalization on the test data, thereby confirming the safe and beneficial combination of the two algorithms. The SL-ML technique outperformed because it complemented the weaknesses of each weak learner. The SL technique reduced the bias and variance of the prediction values produced by each weak learner. SL-ML had advantages over SL-DL and the combination of SL-ML-DL due to the nature of the data used in the groundwater potential prediction model. The DL algorithm outperforms ML when there is a highly complex relationship between dependent and independent variables. This study confirms that the relationship between dependent and independent variables in mapping groundwater potential in the study area is not too complex. DT-based algorithms such as RF and GBDT are better at capturing the relationship between dependent and independent variables in this case study.

In the DL group, RNN outperformed CNN. Although RNNs are designed to solve sequential problems such as time-series prediction, research confirms that RNNs also excel at non-sequential data. RNN builds internal hierarchical relationships among independent variables, enabling it to capture better relationships with dependent variables in the dataset used in this study. This technique outperforms CNN-based convolutional feature extraction in this study.

Model performance can be visualized using the ROC-AUC curve, which is shown in Figure 8. Several models show an AUC value of 1.00 (perfect) on training data, but decrease to 0.98–0.99 on testing data. These models are RF, GBDT, and SL-ML. An ROC-AUC value of 1.00 is standard in ensemble models based on decision trees such as RF, GBDT, and SL-ML. The same phenomenon has been observed in several studies using DT-based ensemble models (Jin *et al.*, 2024; Ouali *et al.*, 2023; Prasad *et al.*, 2025). However, it should be noted that the decreases are not significant, with the resulting gap ranging from 0.02 to 0.01. This phenomenon shows that prediction models can learn the relationship between dependent and independent variables well enough to generalize new data (testing) effectively. The lowest AUC in the training dataset is for RNN, at 0.97. The highest AUC value in the testing dataset is the stack ML model (0.99). The other models have an AUC value of 0.98. In general, for every model included: DL, DL-ML, GBDT, RNN, CNN, and Stack ML, the AUC values from prediction models using the training and testing data are all more than 0.95, which indicates that all the models have excellent performance.

The t-test is performed to assess whether there is a statistically significant difference between the two evaluations on the training and testing data. This was done to determine whether the model was overfitting. The  $H_0$  set was that the two datasets were not statistically different at a 95%

confidence level. The t-test result refers to the  $P$  – value (PV) which is the probability value to support  $H_0$ . The  $H_0$  is acceptable if the PV is higher than 0.05. The test results for each model were represented in Table 6 with the PV component. Based on the PV value, it can be inferred that none of the models overfit because the PV value is greater than 0.05. In general, all the prediction models performed well and did not exhibit overfitting.



**Figure 8.** ROC-AUC Performance Evaluation of the Seven AI-based Models during Training on the Training dataset (a) and the Testing Dataset (b).

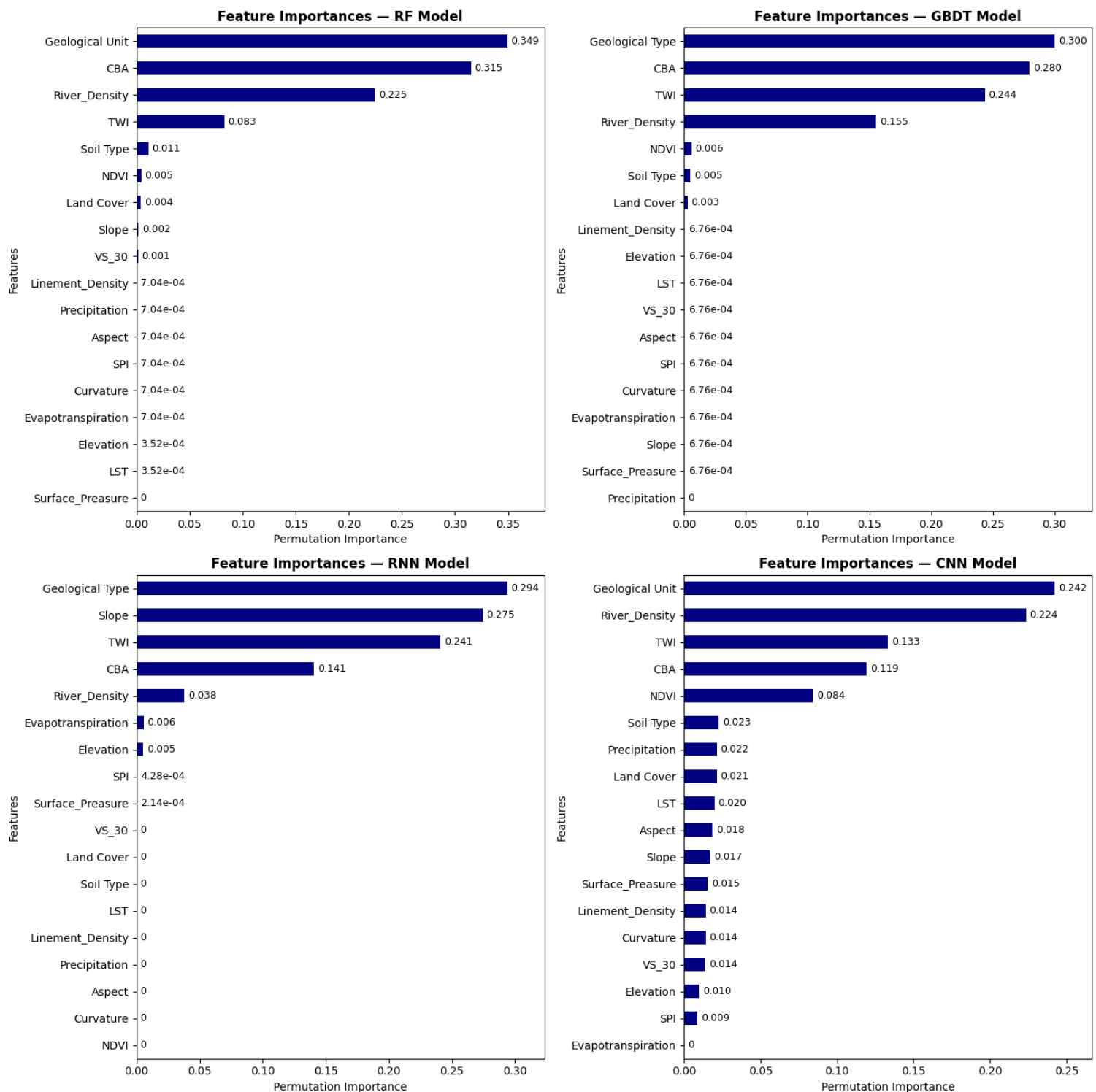
### 3.4. Model Inspections

PFI analysis identifies the most influential parameters in each prediction model. The results of the PFI analysis were conducted on all four basic learning models, as can be seen in Figure 9. **Error! Reference source not found..** The PFI values are presented in a bar chart, with the x-axis ranging from 0 to 1 and the y-axis showing the variable names, sorted in order of highest PFI values. According to these results, GU is the most influential parameter in all models with PFI values ranging from 0.242 to 0.349. Other parameters that significantly influence the four prediction models are CBA, RD, TWI, and slope. There are nine parameters with a PFI value of 0 in the RNN model: NDVI, ST, LC, AR, LST, slope aspect, VS30, LD, and PC. Meanwhile, the other three models each have one parameter with a PFI value of 0.

Parameters with PFI 0 in the CNN, RF, and GBDT models are ET, SP, and AR, respectively. Visually, other parameters have PFI 0 in the RF and RNN models. However, these parameters have values greater than 0 but less than 0.001. A value of 0 in PFI may be due to PFI's limitations in explaining agnostic models. PFI only calculates the decrease in accuracy when the value of a variable in the test data is randomized to assess its importance. A value of 0 in PFI does not represent that the variable has no theoretical effect on the prediction model. However, this condition occurs when randomizing a particular variable does not decrease the accuracy of the initial model. PFI is highly sensitive to variables that are correlated with each other (Kaneko, 2022). If two variables are highly correlated, randomizing one variable will not significantly affect the initial accuracy, because the other variable already captures the values required by the model. For instance, LST and NDVI have a correlation value of 0.4 according to the analysis results. This yields LST or NDVI PFI values of 0 or near 0. A feature selection process is needed to ensure that the variables are not redundant with one another when creating a prediction model.

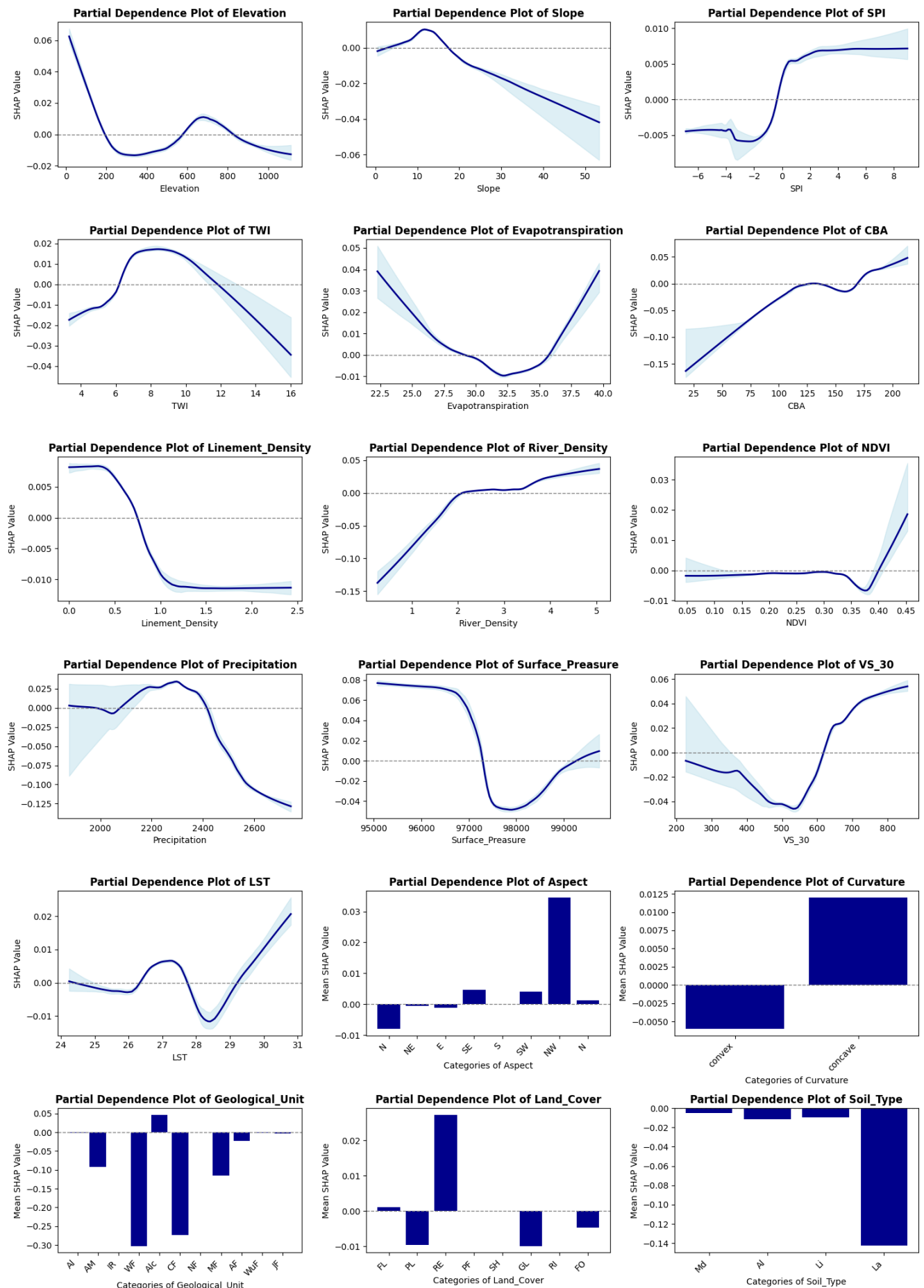
To identify the response characteristics of each parameter to the prediction model, this study uses PDP. PDP describes the relationship between a parameter's value and its SHAP value. PDP analysis is only performed on the best basic learning model, namely GBDT. The analysis results can be seen in Figure 10. **Error! Reference source not found..** Continuous parameters are represented by a line graph (dark blue line) with an uncertainty polygon (light blue color). The uncertainty polygon represents a confidence interval of 95%. Meanwhile, for discrete parameters, it is defined as the average SHAP value for each class, shown in a bar chart.





**Figure 9.** Feature Importance for Each Basic Learning Model.

The non-linear characteristics of each variable in relation to groundwater potential can be interpreted through the PDP. Topographical features such as elevation and slope have a negative relationship with groundwater potential. This indicates that the higher the surface elevation, the lower the groundwater accumulation. The higher the surface, the faster the runoff process in the area, resulting in less infiltration. It is supported by the condition that steeper slopes can increase lateral flow and reduce the infiltration process. The south-east to north slope direction shows a positive SHAP value due to the shorter solar exposure compared to the north to east direction. The duration of solar exposure can affect the level of evapotranspiration, resulting in a decrease in groundwater quantity in the area. Concave terrain has a positive SHAP value due to its morphological ability to accumulate water properly.



**Figure 10.** Partial Dependence Plots for GBDT Model Variables (continuous = line with confidence interval, categorical = mean SHAP by class).

Hydrological conditions such as SPI, TWI, and river density likely have a negative relationship with groundwater potential. SPI and TWI are hydrological parameters derived from topographical factors. Drainage connectivity is represented by river density. The higher the drainage connectivity and TWI, the greater the potential for surface flow accumulation, thereby increasing the probability of lateral infiltration phenomena. SPI represents flow strength; theoretically, the higher the flow strength, the lower the lateral infiltration. However, the PDP results suggest a different interpretation. The higher the SPI value, the higher the groundwater potential. This could occur when supported by permeable geological and soil conditions.

Climatological factors play a significant role in the capability of groundwater recharge and storage. AR with a range of 2,200–2,400 mm/year has a high SHAP value, whereas above this range, the SHAP value decreases significantly. This occurs because high rainfall typically occurs in areas with high topography and is often associated with steep slopes. This condition results in low infiltration rates despite high recharge potential due to high runoff processes. ET shows a negative correlation with SHAP values, meaning that higher ET is associated with greater potential for groundwater, surface water, and vegetation water loss. LST negatively correlates with SHAP, but SHAP values are highest when LST values are too. It is possible that in high-temperature areas, supported by high groundwater runoff (basin areas), with permeable geological conditions. SP negatively correlates with SHAP values, because higher atmospheric pressure can reduce porosity by narrowing the inter-rock gaps beneath.

Land cover and vegetation conditions influence the characteristics of groundwater dynamics. The NDVI parameter is positively associated with groundwater potential. This supports the statement that areas with high vegetation can increase the infiltration process. The type of land cover with the highest SHAP value is residential. Residential areas may reduce infiltration capacity. However, this can be supported by geological conditions with high permeability and recharge potential.

Subsurface conditions greatly influence groundwater transport systems. In the GU parameters, it has the highest range of SHAP values compared to other parameters. The negative SHAP value in GU is obtained by the GU class with low porosity, namely 0% - 8%. Meanwhile, positive SHAP values are obtained by the GU class with a high porosity of 25% to 55%. The soil type parameter in all classes has a negative SHAP value. The lowest negative value is Latosol, which has a low permeability value of less than 0.13 cm/hour. LD has a negative association with the SHAP value because sloping areas (where landforms are rare) reduce the runoff rate, resulting in a positive effect on groundwater recharge.

Both CBA and VS30 geophysical properties are positively associated with groundwater potential. This condition indicates that the higher the rock density, the higher the groundwater potential in the study area. This may occur because volcanic influences dominate the study area. The Mandalika geological formation dominates the study area. This formation has rock types with medium to high density but is prone to fracturing. The presence of these fractures allows surface water to be transported into groundwater.

### 3.5. Discussions

This study implemented AI-based algorithms for groundwater potential mapping. The utilization of AI-based algorithms can improve the accuracy and objectivity of groundwater potential mapping. RF and GBDT algorithms represent ML in this study. Meanwhile, RNN and CNN represent the DL algorithm. The integration of ML, DL, and ML-DL algorithms is carried out to optimize the prediction model with the SL approach with LR as meta-learning. Utilized 18 parameters, all prediction models performed well with ACC>0.90; CK>0.85; and MCC>0.85. An interesting finding on the performance of the prediction models was that by applying SL to integrate the algorithms, they can improve 2.04% to 4.45% in ML stacking, 0.44% to 2.32% in DL stacking, and 1.18% to 5.47% in ML-DL stacking compared to the performance of every single algorithm. The application of stacking techniques has been demonstrated to decrease the bias and variance of prediction outcomes significantly (Lu *et al.*, 2023).

Another important finding is that the CBA parameter, which is rarely used in modelling groundwater potential, shows a significant influence on all models in this study. Values of CBA variation represent lateral density characteristics of lithological units (Zaenudin *et al.*, 2020). The most influential variable in each model is GU. Rock characteristics such as porosity affect the absorption rate of surface water into groundwater. The GU variable, as the variable with the highest relative contribution, aligns with research by Bai *et al.* (2022), which had a relative contribution value of 0.95. Research by Nugroho *et al.* (2024), GU is in the second-highest position with a relative contribution rate of 0.16. However, other studies produced the opposite condition, where



the GU parameter had less effect on the prediction model, for example, in a study by Aslam *et al.* (2025), where GU was in the last position in the relative contribution level of the prediction model, with a value of 0.15. The less influential parameters in this study are mainly climatological factors; for instance, in the GBDT model, AR has the lowest PFI value, SP is the lowest PI in the RF model, and ET is the lowest PFI in the CNN model. This condition aligned with research by Hasan *et al.* (2025) where climatological factors such as rainfall had the lowest relative contribution, at 0.89%. The rainfall variable in the study by Prasad *et al.* (2020) was also lacking in significant influence in predicting groundwater potential with a contribution value of 10%. However, studies by Aslam *et al.* (2025) and Bai *et al.* (2022) had different results, where rainfall had a significant influence on the prediction model. This indicates that each case study has its own characteristics in indicating the potential presence of groundwater.

Compared to previous similar studies using a hybrid algorithm integration approach and AI for groundwater potential mapping (Arabameri *et al.*, 2020; Chen *et al.*, 2020; Chen *et al.*, 2022; Parasar *et al.*, 2025; Pham *et al.*, 2019), this study has advantages in performance optimization through hyperparameter tuning and stacking techniques. Both techniques can optimize the performance of the prediction model, resulting in good and stable evaluation values during cross-validation. In addition, this study also uses modelling parameters that are rarely used in similar studies, namely CBA and VS30. Both parameters have been proven to contribute significantly to modelling using AI-based algorithms through PFI values. Modelling parameters used in this study not only utilize surface phenomena but also subsurface phenomena through parameters on geological and geophysical factors. Hence, the model is more reliable and applicable for decision-making purposes related to freshwater resource management.

This study can provide implications for depicting areas lacking access to groundwater potential in Trenggalek Regency. All models indicate that the study area is dominated by low potential. Some parts of the area, however, have high potential. This information can be leveraged for decision-making in providing equitable access to freshwater resources in Trenggalek Regency during such emergency conditions as drought. The primary contribution of this study is to explore the ML, DL, and SL integration algorithms for groundwater potential. Moreover, this study bridges the black-box problem of the prediction model by applying PFI and PDP. Therefore, it is important to make the model more interpretable. The limitation of this study is not conducting feature selection, which may improve the performance of the prediction model and increase its effectiveness due to reduced computational cost. Further exploration of AI-based algorithms and other parameters is recommended, as the best prediction model in this study may not necessarily perform the following when applied to other study areas due to differences in geological and geomorphological characteristics.

#### 4. Conclusion

This study implemented and evaluated AI algorithm-based prediction models. This study utilizes two ML algorithms, namely RF and GBDT, two DL algorithms, namely RNN and CNN, an SL ML algorithm, an SL DL algorithm, and an SL ML-DL algorithm. Based on ACC, MCC, CK, and ROC-AUC, the SL ML model has the best performance. The mean evaluation values of the SL-ML model in ACC, MCC, CK, and AUC are 0.957, 0.915, 0.915, and 0.99, respectively. Nevertheless, all prediction models performed very well with  $ACC > 0.90$ ,  $MCC > 0.85$ ,  $CK > 0.85$ , and  $AUC > 0.95$ . According to the PFI value, GU was a parameter with the highest contribution to all prediction models. GU PFI values for the RF, GBDT, CNN, RNN, and GBDT models are 0.349, 0.300, 0.242, and 0.294, respectively. All prediction models indicate that Trenggalek Regency was dominated by the low-potential class, ranging from 22.32% to 31.29%. Based on this information, the government can serve as a basis for the sustainable management of clean water resources. In addition, this information will help reduce the cost of locating well drilling sites to meet the supply demands of various sectors.

Future research should consider validating groundwater prediction models derived from AI-based algorithms using ground-based data, such as geoelectric surveys. Geoelectric surveys can be conducted across multiple study areas, including low- and high-potential zones. By validating these models, the resulting predictions become more reliable for subsequent policy application. Additionally, feature selection can reduce redundancy in modelling parameters and provide an efficient process, leading to improved predictive model performance. The use of higher-resolution data is highly recommended to produce models suitable for more detailed regional planning. Exploration of variables affecting groundwater potential and various other algorithms is necessary. Testing different algorithms is essential to obtain optimal groundwater potential prediction models for other case studies with different geological and topographical conditions.

## Acknowledgements

We would like to express our gratitude to the Trenggalek Regency Government, the Indonesian Geospatial Information Agency, the Indonesian Volcanology and Geological Disaster Mitigation Center, the United States Geological Survey (USGS), the National Aeronautics and Space Administration (NASA), Murray Lab, and the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing access to data to support this study.

## Author Contributions

**Conceptualization:** Ummah, M. H., Diyono, D.; **methodology:** Ummah, M. H.; **investigation:** Ummah, M. H.; **writing—original draft preparation:** Ummah, M. H.; **writing—review and editing:** Diyono, D.; **visualization:** Ummah, M. H. All authors have read and agreed to the published version of the manuscript.

## Conflict of interest

All authors declare that they have no conflicts of interest.

## Data availability

Data is available upon Request.

## Funding

The research was funded by the Indonesia Endowment Fund for Education Agency (well known as LPDP or Lembaga Pengelola Dana Pendidikan) under the Ministry of Finance of Indonesia.

## References

- Abdelouhed, F., Ahmed, A., Abdellah, A., Yassine, B., & Mohammed, I. (2021). Using GIS and remote sensing for the mapping of potential groundwater zones in fractured environments in the CHAOUIA-Morocco area. *Remote Sensing Applications: Society and Environment*, 23, 100571. doi: 10.1016/j.rsase.2021.100571.
- Abijith, D., Saravanan, S., Singh, L., Jennifer, J. J., Saranya, T., & Parthasarathy, K. S. S. (2020). GIS-based multi-criteria analysis for identification of potential groundwater recharge zones-a case study from Ponnaniyar watershed, Tamil Nadu, India. *HydroResearch*, 3, 1–14.
- Ahmed, Z., Ansari, M. T., Zahir, M., Shakir, U., & Subhan, M. (2020). Hydrogeophysical investigation for groundwater potential through Electrical Resistivity Survey in Islamabad, Pakistan. *Journal of Geography and Social Sciences (JGSS)*, 2(2), 147–163.
- Al-Abadi, A. M., Al-Temmeme, A. A., & Al-Ghanimy, M. A. (2016). 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. *Sustainable Water Resources Management*, 2, 265–283.
- Alshehri, F., & Mohamed, A. (2023). Investigation of groundwater potential using gravity data in Wadi Fatimah and its surroundings, Western Saudi Arabia. *Frontiers in Earth Science*, 11, 1225992. doi: 10.3389/feart.2023.1225992.
- Altmann, A., Tološi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, 26(10), 1340–1347. doi: 10.1093/bioinformatics/btq134
- Andualem, T. G., & Demeke, G. G. (2019). Groundwater potential assessment using GIS and remote sensing: A case study of Guna tana landscape, upper blue Nile Basin, Ethiopia. *Journal of Hydrology: Regional Studies*, 24, 100610. doi: 10.1016/j.ejrh.2019.100610.
- Arabameri, A., Lee, S., Tiefenbacher, J. P., & Ngo, P. T. (2020). Novel Ensemble of MCDM-Artificial Intelligence Techniques for Groundwater-Potential Mapping in Arid and Semi-Arid Regions (Iran). *Remote Sensing*, 12(3), 490. doi: 10.3390/rs12030490.
- Aslam, B., Maqsoom, A., Hassan, U., Maqsoom, S., Alaloul, W. S., Musarat, M. A., Shahzaib, M., & Irfan, M. (2025). Comparison Between Machine Learning and Bivariate Statistical Models for Groundwater Recharge Zones. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 49(1), 853–876. doi: 10.1007/s40996-024-01721-1.
- Bai, Z., Liu, Q., & Liu, Y. (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. doi: 10.1007/s11053-022-10100-4.
- Benjmel, K., Amraoui, F., Aydda, A., Tahiri, A., Yousif, M., Pradhan, B., Abdelrahman, K., Fnais, M. S., & Abioui, M. (2022). A multidisciplinary approach for groundwater potential mapping in a fractured semi-arid terrain (Kerdous Inlier, Western Anti-Atlas, Morocco). *Water*, 14(10), 1553.
- Berhanu, K. G., & Hatiye, S. D. (2020). Identification of groundwater potential zones using proxy data: Case study of Megech Watershed, Ethiopia. *Journal of Hydrology: Regional Studies*, 28, 100676.
- Bi, B., Li, J., Luo, T., Wang, B., Yang, C., & Shen, L. (2025). Positive–Unlabeled Learning-Based Hybrid Models and Interpretability for Groundwater Potential Mapping in Karst Areas. *Water*, 17(10). doi: 10.3390/w17101422.
- Bora Okay, H., & Özacar, A. A. (2023). A New Vs30 Prediction Strategy Taking Topography, Geology, Terrain and Water Saturation into Account: Application to Türkiye and California. *EGU General Assembly Conference Abstracts*, EGU-7445.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Chen, T., Zhu, L., Niu, R., Trinder, C. J., Peng, L., & Lei, T. (2020). Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models. *Journal of Mountain Science*, 17(3), 670–685. doi: 10.1007/s11629-019-5839-3.
- Chen, W., Li, Y., Tsangaratos, P., Shahabi, H., Ilia, I., Xue, W., & Bian, H. (2020). Groundwater Spring Potential Mapping Using Artificial Intelligence Approach Based on Kernel Logistic Regression, Random Forest, and Alternating Decision Tree Models. *Applied Sciences*, 10(2), 425. doi: 10.3390/app10020425.
- Chen, Y., Chen, W., Chandra Pal, S., Saha, A., Chowdhuri, I., Adeli, B., Janizadeh, S., Dineva, A. A., Wang, X., & Mosavi, A. (2022). Evaluation efficiency of hybrid deep learning algorithms with neural network decision tree and boosting methods for predicting groundwater potential. *Geocarto International*, 37(19), 5564–5584. doi: 10.1080/10106049.2021.1920635.
- Das, B., Rathore, P., Roy, D., Chakraborty, D., Jatav, R. S., Sethi, D., & Kumar, P. (2022). Comparison of bagging, boosting and stacking algorithms for surface soil moisture mapping using optical-thermal-microwave remote sensing synergies. *CATENA*, 217, 106485. doi: 10.1016/j.catena.2022.106485.
- Demir, S., & Sahin, E. K. (2024). *The effectiveness of data pre-processing methods on the performance of machine learning techniques using RF, SVR, Cubist, and SGB: a study on undrained shear strength prediction*. Stochastic Environmental Research and Risk Assessment.
- Echogdali, F. Z., Boutaleb, S., Bendarma, A., Saidi, M. E., Aadraoui, M., Abioui, M., Ouchchen, M., Abdelrahman, K., Fnais, M. S., & Sajinkumar, K. S. (2022). Application of the analytical hierarchy process and geophysical method for groundwater potential mapping in the Tata basin, Morocco. *Water*, 14(15), 2393.
- Esen, H. (2023). Analytical Hierarchy Process Problem Solution. In *Analytic Hierarchy Process—Models, Methods, Concepts, and Applications [Working Title]*. Retrieved From <https://doi.org/10.5772/intechopen.1001072>.
- Fan, M., Xiao, K., Sun, L., Zhang, S., & Xu, Y. (2022). Automated Hyperparameter Optimization of Gradient Boosting Decision Tree Approach for Gold Mineral Prospectivity Mapping in the Xiong’ershan Area. *Minerals*, 12(12). doi: 10.3390/min12121621.
- Fang, Z., Wang, Y., Peng, L., & Hong, H. (2020). Integration of a convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Computers & Geosciences*, 139, 104470. doi: 10.1016/j.cageo.2020.104470.
- FAO. (2022). *Water withdrawal by source* [Data set]. Retrieved from [https://api.data.apps.fao.org/api/v2/bigquery?sql\\_url=https://data.apps.fao.org/catalog/dataset/945666e6-7803-4621-b8ef-cfd885a84596/resource/4a000a1b-24f0-4328-aab6-b9b525892090/download/query\\_en.sql](https://api.data.apps.fao.org/api/v2/bigquery?sql_url=https://data.apps.fao.org/catalog/dataset/945666e6-7803-4621-b8ef-cfd885a84596/resource/4a000a1b-24f0-4328-aab6-b9b525892090/download/query_en.sql).
- Ferrouhi, E. M., & Bouabdallaoui, I. (2024). A comparative study of ensemble learning algorithms for high-frequency trading. *Scientific African*, 24, e02161. doi: 10.1016/j.sciaf.2024.e02161.
- Ghiat, I., Mackey, H. R., & Al-Ansari, T. (2021). A Review of Evapotranspiration Measurement Models, Techniques, and Methods for Open and Closed Agricultural Field Applications. *Water*, 13(18), 2523. doi: 10.3390/w13182523.

- Hakim, W. L., Nur, A. S., Rezaie, F., Panahi, M., Lee, C.-W., & Lee, S. (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. doi: 10.1016/j.ejrh.2022.100990.
- Halder, K., Srivastava, A. K., Ghosh, A., Nabik, R., Pan, S., Chatterjee, U., Bisai, D., Pal, S. C., Zeng, W., Ewert, F., Gaiser, T., Pande, C. B., Islam, A. R. Md. T., Alam, E., & Islam, M. K. (2024). Application of bagging and boosting ensemble machine learning techniques for groundwater potential mapping in a drought-prone agriculture region of eastern India. *Environmental Sciences Europe*, 36(1), 155. doi: 10.1186/s12302-024-00981-y.
- Handayani, L., Wardhana, D. D., Hartanto, P., Delinom, R., Sudaryanto, Bakti, H., & Lubis, R. (2018). Gravity survey of groundwater characterization at Labuan Basin. *IOP Conference Series: Earth and Environmental Science*, 118, 012015. doi: 10.1088/1755-1315/118/1/012015.
- Hasan, Md. M., Rana, M. M. S. P., & Ferdous, M. T. (2025). Prediction of Groundwater Potential Zone Using Machine Learning and Geospatial Approaches for an Industry-Dominated Area in Narayanganj, Bangladesh. *Journal of the Indian Society of Remote Sensing*, 53, 3635–3651. doi: 10.1007/s12524-025-02177-0
- Hossen, S., Uddin, M. S., Ali, Y., & Rana, P. (2025). Integrated analysis of land use and land cover changes and landslide susceptibility: A machine learning approach in Rangamati Sadar, Bangladesh. *Natural Hazards*, 121(16), 19387–19408. doi: 10.1007/s11069-025-07580-x.
- Ishii, A., Yamamoto, A., & Yamanaka, A. (2023). DMC-TPE: tree-structured Parzen estimator-based efficient data assimilation method for phase-field simulation of solid-state sintering. *Science and Technology of Advanced Materials: Methods*, 3(1), 2239133. doi: 10.1080/27660400.2023.2239133.
- Islam, R., Sultana, A., & Tuhin, MD. N. (2024). A comparative analysis of machine learning algorithms with a tree-structured Parzen estimator for liver disease prediction. *Healthcare Analytics*, 6, 100358. doi: 10.1016/j.health.2024.100358.
- Jafari, A., Alesheikh, A. A., Rezaie, F., Panahi, M., Shahsavari, S., Lee, M.-J., & Lee, S. (2023). Enhancing a convolutional neural network model for land subsidence susceptibility mapping using hybrid meta-heuristic algorithms. *International Journal of Coal Geology*, 277, 104350. doi: 10.1016/j.coal.2023.104350.
- Jamell Ivor Samuels. (2024). *One-Hot Encoding and Two-Hot Encoding: An Introduction*. Retrieved From <https://doi.org/10.13140/RG.2.2.21459.76327>
- Ji, J., Zhou, Y., Cheng, Q., Jiang, S., & Liu, S. (2023). Landslide Susceptibility Mapping Based on Deep Learning Algorithms Using Information Value Analysis Optimization. *Land*, 12(6). doi: 10.3390/land12061125.
- Jin, B., Liu, S., Zeng, T., Li, Y., Wang, T., Gui, L., Zhao, B., & Yin, K. (2024). Spatio-temporal forecasting of landslide hazard in Chongqing National Transmission Protection Regions, China. *International Journal of Digital Earth*, 17(1), 2392843. doi: 10.1080/17538947.2024.2392843.
- Kamali Maskooni, E., Naghibi, S. A., Hashemi, H., & Berndtsson, R. (2020). Application of advanced machine learning algorithms to assess groundwater potential using remote sensing-derived data. *Remote Sensing*, 12(17), 2742.
- Kaneko, H. (2022). Cross-validated permutation feature importance considering correlation between features. *Analytical Science Advances*, 3(9–10), 278–287. doi: 10.1002/ansa.202200018.
- Kanwar, M., Pokharel, B., & Lim, S. (2025). A new random forest method for landslide susceptibility mapping using hyperparameter optimization and grid search techniques. *International Journal of Environmental Science and Technology*, 22(11), 10635–10650. doi: 10.1007/s13762-024-06310-3.
- Khalilidermani, M., & Knez, D. (2024). Shear Wave Velocity Applications in Geomechanics with Focus on Risk Assessment in Carbon Capture and Storage Projects. *Energies*, 17(7). doi: 10.3390/en17071578.
- Khoshkroodi, A., Parvini Sani, H., & Aajami, M. (2024). Stacking Ensemble-Based Machine Learning Model for Predicting Deterioration Components of Steel W-Section Beams. *Buildings*, 14(1). doi: 10.3390/buildings14010240.
- Khosravi, K., Sartaj, M., Tsai, F. T.-C., Singh, V. P., Kazakis, N., Melesse, A. M., Prakash, I., Tien Bui, D., & Pham, B. T. (2018). A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment. *Science of The Total Environment*, 642, 1032–1049. doi: 10.1016/j.scitotenv.2018.06.130.
- Kilic, K., Ikeda, H., Narihiro, O., Adachi, T., & Kawamura, Y. (2024). A soft ground micro TBM's specific energy prediction using an explainable neural network through Shapley additive explanation and Optuna. *Bulletin of Engineering Geology and the Environment*, 83(5), 175. doi: 10.1007/s10064-024-03670-5.
- Kramer, R., Lu, Y., & Bokelmann, G. (2023). Interaction of Air Pressure and Groundwater as Main Cause of Sub-Daily Relative Seismic Velocity Changes. *Geophysical Research Letters*, 50(7). doi: 10.1029/2022GL101298.
- Lai, Y. (2024). *Application and Effectiveness Evaluation of Bayesian Optimization Algorithm in Hyperparameter Tuning of Machine Learning Models*, 351–355. doi: 10.1109/PEEE63877.2024.00070.
- Lei, X., Liu, J., Du, Y., Liu, L., Yuan, Y., Zhao, X., Guo, Z., & Wu, X. (2025). Optimizing machine learning and bagging-based hybrid models for landslide susceptibility mapping: A case study in Chenggu County, China. Retrieved From <https://doi.org/10.1038/s41598-025-30136-9>
- Li, H., Fan, Y., Gong, Z., & Zhou, D. (2020). Water accessibility assessment of freshwater wetlands in the Yellow River Delta National Nature Reserve, China. *Ecohydrology & Hydrobiology*, 20(1), 21–30. doi: 10.1016/j.ecohyd.2019.02.006.
- Li, W., Finsa, M. M., Laskey, K. B., Houser, P., & Douglas-Bate, R. (2023). Groundwater Level Prediction with Machine Learning to Support Sustainable Irrigation in Water Scarcity Regions. *Water*, 15(19), 3473. doi: 10.3390/w15193473.
- Lu, M., Hou, Q., Qin, S., Zhou, L., Hua, D., Wang, X., & Cheng, L. (2023). A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting. *Water*, 15(7), 1265. doi: 10.3390/w15071265.
- Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. Retrieved From <https://doi.org/10.48550/arXiv.1705.07874>.
- Lv, F., Yu, J., Zhang, J., Yu, P., Tong, D., & Wu, B. (2022). A novel stacking-based ensemble learning model for drilling efficiency prediction in earth-rock excavation. *Journal of Zhejiang University-SCIENCE A*, 23(12), 1027–1046. doi: 10.1631/2023.A2200297.
- Meneisy, A. M., Hamed, A., Khalifa, M., Bohoty, M. E., Ghamry, E., & Taha, A. (2023). Geotechnical Zonation and Vs30 Classification of Soil for Seismic Site Effects Evaluation in the Western Extension of New Aswan City, Egypt. *Pure and Applied Geophysics*, 180(11), 3813–3833. doi: 10.1007/s00024-023-03370-3.
- Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. *Information*, 15(9). doi: 10.3390/info15090517.
- Moughani, S. K., Osmani, A., Nohani, E., Khoshtinat, S., Jalilian, T., Askari, Z., Heddami, S., Tiefenbacher, J. P., & Hatamifakouei, J. (2023). Groundwater spring potential prediction using a deep-learning algorithm. *Acta Geophysica*, 72(2), 1033–1054. doi: 10.1007/s11600-023-01053-0.



- Muavhi, N., Thamaga, K. H., & Mutoti, M. I. (2022). Mapping groundwater potential zones using relative frequency ratio, analytic hierarchy process and their hybrid models: Case of Nzhelele-Makhado area in South Africa. *Geocarto International*, 37(21), 6311–6330.
- Muthu, K., & Sudalaimuthu, K. (2021). Integration of Remote sensing, GIS, and AHP in demarcating groundwater potential zones in Pattukottai Taluk, Tamilnadu, India. *Arabian Journal of Geosciences*, 14(17), 1748. doi: 10.1007/s12517-021-08110-2.
- Nguyen, D., Chou, T., Hoang, T., & Chen, M. (2023). Flood susceptibility mapping using machine learning algorithms: A case study in Huong Khe District, Ha Tinh Province, Vietnam. *International Journal of Geoinformatics*, 19(7), 1–15.
- Nguyen, H. D., Nguyen, Q.-H., Dang, D. K., Nguyen, T. G., Truong, Q. H., Nguyen, V. H., Bretcan, P., Șerban, G., Bui, Q.-T., & Petrisor, A.-I. (2024). Integrated machine learning and remote sensing for groundwater potential mapping in the Mekong Delta in Vietnam. *Acta Geophysica*, 72(6), 4395–4413. doi: 10.1007/s11600-024-01331-5.
- Nugroho, J. T., Lestari, A. I., Gustiandi, B., Sofan, P., Suwarsono, Prasasti, I., Rahmi, K. I. N., Noviar, H., Sari, N. M., Manalu, R. J., Arifin, S., & Taufiq, A. (2024). Groundwater potential mapping using machine learning approach in West Java, Indonesia. *Groundwater for Sustainable Development*, 27, 101382. doi: 10.1016/j.gsd.2024.101382.
- Nugroho, W. H. A., Nurwatik, N., & Kusuma Widya, L. (2024). Groundwater Potential Mapping Using Random Forest and Extreme Gradient Boosting Algorithms. *IOP Conference Series: Earth and Environmental Science*, 1418(1), 012035. doi: 10.1088/1755-1315/1418/1/012035.
- Nurwatik, N., Ummah, M. H., Cahyono, A. B., Darminto, M. R., & Hong, J.-H. (2022). A Comparison Study of Landslide Susceptibility Spatial Modeling Using Machine Learning. *ISPRS International Journal of Geo-Information*, 11(12). doi: 10.3390/ijgi11120602.
- Ouali, L., Kabiri, L., Namous, M., Hssaisoune, M., Abdelrahman, K., Fnais, M. S., Kabiri, H., El Hafyani, M., Oubaassine, H., Arioua, A., & Bouchaou, L. (2023). Spatial Prediction of Groundwater Withdrawal Potential Using Shallow, Hybrid, and Deep Learning Algorithms in the Toudgha Oasis, Southeast Morocco. *Sustainability*, 15(5). doi: 10.3390/su15053874.
- Parasar, P., Moral, P., Srivastava, A., Krishna, A. P., Majumdar, S., Bhattacharjee, R., Mishra, A. P., Mustafi, D., Rathore, V. S., Sharma, R., & Mustafi, A. (2025). Integrating genetic algorithms and machine learning for spatiotemporal groundwater potential zoning in fractured aquifers. *Journal of Hydrology: Regional Studies*, 62, 102800. doi: 10.1016/j.ejrh.2025.102800.
- Pham, B. T., Jaafari, A., Prakash, I., Singh, S. K., Quoc, N. K., & Bui, D. T. (2019). Hybrid computational intelligence models for groundwater potential mapping. *CATENA*, 182, 104101. doi: 10.1016/j.catena.2019.104101.
- Pokhrel, A., Khanal, A., & Khanal, A. (2024). Groundwater potential mapping of Rupandehi District using geographic information system and remote sensing. *International Journal of Energy and Water Resources*, 9, 1067–1093, doi: 10.1007/s42108-024-00327-7
- Prasad, P., Gogoi, D., Gogoi, D., Kumar, T., & Chandra, P. (2025). Threshold-based inventory for flood susceptibility assessment of the world's largest river island using multi-temporal SAR data and ensemble machine learning algorithms. *Stochastic Environmental Research and Risk Assessment*, 39(1), 251–269. doi: 10.1007/s00477-024-02860-7.
- Prasad, P., Loveson, V. J., Kotha, M., & Yadav, R. (2020). Application of machine learning techniques in groundwater potential mapping along the west coast of India. *GIScience & Remote Sensing*, 57(6), 735–752. doi: 10.1080/15481603.2020.1794104.
- Ragragui, H., Aouragh, M. H., El-Hmaidi, A., Ouali, L., Saouita, J., Iallamen, Z., Ousmana, H., Jaddi, H., & El Ouali, A. (2024). Mapping and modeling groundwater potential using machine learning, deep learning and ensemble learning models in the Saiss basin (Fez-Meknes region, Morocco). *Groundwater for Sustainable Development*, 26, 101281. doi: 10.1016/j.gsd.2024.101281.
- Regional Disaster Management Agency Of Trenggalek Regency. (2024). *Trenggalek district disaster infographic* [Data set].
- Rehman, A., Islam, F., Tariq, A., Ul Islam, I., Brian J. D., Bibi, T., Ahmad, W., Ali Waseem, L., Karuppannan, S., & Al-Ahmadi, S. (2024). Groundwater potential zone mapping using GIS and Remote Sensing based models for sustainable groundwater management. *Geocarto International*, 39(1). doi: 10.1080/10106049.2024.2306275.
- Ribeiro, M. H. D. M., & dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, 86, 105837. doi: 10.1016/j.asoc.2019.105837.
- Rong, G., Li, K., Su, Y., Tong, Z., Liu, X., Zhang, J., Zhang, Y., & Li, T. (2021). Comparison of Tree-Structured Parzen Estimator Optimization in Three Typical Neural Network Models for Landslide Susceptibility Assessment. *Remote Sensing*, 13(22). doi: 10.3390/rs13224694.
- Rustadi, R., Darmawan, I., Nandi Haerudin, N. H., Agus Setiawan, A., & Suharno, S. (2022). Groundwater exploration using integrated geophysics method in hard rock terrains in Mount Betung Western Bandar Lampung, Indonesia. *Journal of Groundwater Science and Engineering*, 10(1), 10–18.
- Sarkar, S. K., Rudra, R. R., Talukdar, S., Das, P. C., Nur, Md. S., Alam, E., Islam, M. K., & Islam, A. R. Md. T. (2024). Future groundwater potential mapping using machine learning algorithms and climate change scenarios in Bangladesh. *Scientific Reports*, 14(1), 10328. doi: 10.1038/s41598-024-60560-2.
- Seifi, F., & Niaki, S. T. A. (2023). Extending the hypergradient descent technique to reduce the time of optimal solution achieved in hyperparameter optimization algorithms. *International Journal of Industrial Engineering Computations*, 14(3), 501–510. doi: 10.5267/j.ijiec.2023.4.004.
- Senapati, U., & Das, T. K. (2021). Assessment of basin-scale groundwater potentiality mapping in drought-prone upper Dwarakeswar River basin, West Bengal, India, using GIS-based AHP techniques. *Arabian Journal of Geosciences*, 14(11), 960.
- Shams, M. Y., Elshewey, A. M., El-kenawy, E.-S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, 83(12), 35307–35334. doi: 10.1007/s11042-023-16737-4.
- Shapley, L. S. (1953). *A value for n-person games*.
- Sikora, R. (2017). A modified stacking ensemble machine learning algorithm using genetic algorithms. In *Artificial intelligence: Concepts, methodologies, tools, and applications*, 395–405.
- Silver, P. G., Daley, T. M., Niu, F., & Majer, E. L. (2007). Active source monitoring of cross-well seismic travel time for stress-induced changes. *Bulletin of the Seismological Society of America*, 97(1B), 281–293. doi: 10.1785/0120060120.

- Statistical Centre Of East Java Province. (2024). *Percentage Distribution of Households by District/City and Source of Drinking Water in East Java Province, 2023* [Statistics table; Table]. Retrieved From <https://jatim.bps.go.id/id/statistics-table/3/YzBaMlduSlFVbTVrUnpWeU9YRTJka0pVTFkU1FUMDkjMw/distribusi-persentase-rumah-tangga-menurut-kabupaten-kota-dan-sumber-air-minum-di-provinsi-jawa-timur-2023.html?year=2023>.
- Stern, R. H. (2021). Interpretation of the area under the ROC curve for risk prediction models. *arXiv Preprint arXiv:2102.11053*.
- Sun, X., Chen, J., Bao, Y., Han, X., Zhan, J., & Peng, W. (2018). Landslide Susceptibility Mapping Using Logistic Regression Analysis along the Jinsha River and Its Tributaries Close to Derong and Deqin County, Southwestern China. *ISPRS International Journal of Geo-Information*, 7(11). doi: 10.3390/ijgi7110438.
- Tharwat, A. (2018). *Classification Assessment Methods: A detailed tutorial*. Retrieved From <https://doi.org/10.1016/j.aci.2018.08.003>.
- Tolche, A. D. (2021). Groundwater potential mapping using geospatial techniques: A case study of Dhugeta-Ramis sub-basin, Ethiopia. *Geology, Ecology, and Landscapes*, 5(1), 65–80. doi: 10.1080/24749508.2020.1728882.
- Vrba, J., & Salamat, A. (2007). *Groundwater for emergency situations*. UNESCO.
- Wang, W., Gu, M., Li, Z., Hong, Y., Zang, H., & Xu, D. (2024). A stacking ensemble machine learning model for improving monthly runoff prediction. *Earth Science Informatics*, 18(1), 120. doi: 10.1007/s12145-024-01544-8.
- Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, 17(1), 26–40. doi: 10.11989/JEST.1674-862X.80904120.
- Wu, X., & Wang, J. (2022). Application of bagging, boosting and stacking ensemble and EasyEnsemble methods to landslide susceptibility mapping in the Three Gorges Reservoir area of China. *EGU sphere*, 2022, 1–31.
- Yang, K., Niu, R., Song, Y., Dong, J., Zhang, H., & Chen, J. (2024). Dynamic Hazard Assessment of Rainfall-Induced Landslides Using Gradient Boosting Decision Tree with Google Earth Engine in Three Gorges Reservoir Area, China. *Water*, 16(12). doi: 10.3390/w16121638.
- Yariyan, P., Avand, M., Omidvar, E., Pham, Q. B., Linh, N. T. T., & Tiefenbacher, J. P. (2022). Optimization of statistical and machine learning hybrid models for groundwater potential mapping. *Geocarto International*, 37(13), 3877–3911. doi: 10.1080/10106049.2020.1870164.
- Yeo, I., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954–959.
- Zaenudin, A., Risman, R., Darmawan, I. G. B., & Yogi, I. B. S. (2020). Analysis of gravity anomaly for groundwater basin in Bandar Lampung city based on 2D gravity modeling. *Journal of Physics: Conference Series*, 1572(1), 012006. doi: 10.1088/1742-6596/1572/1/012006.
- Zhang, S., Bai, L., Li, Y., Li, W., & Xie, M. (2022). Comparing Convolutional Neural Network and Machine Learning Models in Landslide Susceptibility Mapping: A Case Study in Wenchuan County. *Frontiers in Environmental Science*, 10, 886841. doi: 10.3389/fenvs.2022.886841.
- Zhao, X., Xiong, Z., Karlshöfer, P., Tziolas, N., Wiesmeier, M., Heiden, U., & Zhu, X. X. (2025). Soil organic carbon estimation using spaceborne hyperspectral composites on a large scale. *International Journal of Applied Earth Observation and Geoinformation*, 140, 104504. doi: 10.1016/j.jag.2025.104504.