

RESEARCH LETTER

Open Access



# Using an ensemble machine learning model to delineate groundwater potential zones in desert fringes of East Esna-Idfu area, Nile valley, Upper Egypt

Hesham Morgan<sup>1\*</sup> , Ahmed Madani<sup>1</sup>, Hussien M. Hussien<sup>2</sup> and Tamer Nassar<sup>1</sup>

## Abstract

The effects of climate change and rapid population growth increase the demand for freshwater, particularly in arid and hyper-arid environments, considering that groundwater is an essential water resource in these regions. The main focus of this research was to generate a groundwater potential map in the Center Eastern Desert, Egypt, using a random forest classification machine learning model. Based on satellite data, geological maps and field survey, fifteen effective features influencing groundwater potentiality were created. These effective features include elevation, slope angle, slope aspect, terrain ruggedness index, curvature, lithology, lineament density, distance from major fractures, topographic wetness index, stream power index, drainage density, rainfall, as well as distance from rivers and channels, soil type and land use/land cover. Collinearity analysis was used for feature selection. A 100 dependent points (57 water points and 43 non-potential mountainous areas) were labeled and classified according to hydrogeological conditions in the three main aquifers (Basement, Nubian and Quaternary Aquifers) in the study area. The random forest algorithm was trained using (70%) of the dependent points. Then, it was validated using (30%) and the hyper-parameters were optimized. Groundwater potential map was predicted and classified as good (5.1%), moderate (0.1%), poor (4.2%) and non-potentiality (90.6%). Sensitivity (92%), F1-score (94%) and accuracy (97%) are validation methods used due to the imbalanced dataset problem. The most important effective features for groundwater potential map were determined based on the random forest and the receiver operating characteristics curve. Groundwater management sustainability was discussed based on the predicted groundwater potential map and aquifer conditions. Therefore, the random forest model is helpful for delineating groundwater potential zones and can be used in similar locations all over the world.

**Keywords** Groundwater potential map, Imbalanced dataset, Random forest, Variable importance

## Introduction

Arid and hyper-arid environments suffer from water scarcity. Industrial development, rapid population growth and climate change compel governments worldwide, especially in the Middle East to explore sustainable water resources. By 2025, the majority of the world's countries will face a freshwater deficit (Amarasinghe and Smakhtin 2014). Groundwater is a vital water resource in these environments. Given that groundwater is an invisible natural resource, defining groundwater potential

\*Correspondence:

Hesham Morgan  
hesham.hashem.geo@cu.edu.eg

<sup>1</sup> Department of Geology, Faculty of Science, Cairo University, P.O.B. 12613, Giza, Egypt

<sup>2</sup> Geology Department, Desert Research Center, El Mataryia, P.O.B. 11753, Cairo, Egypt



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

zones is critical for socio-economic management, planning, and sustainable development. The availability and movement of groundwater are influenced by ecological, topographical, hydrological, atmospheric, and geological parameters (Oh et al. 2011).

Numerous studies on groundwater potential mapping (GWPM) have been conducted by researchers using various approaches. Early GWPMs were based on hydrogeological laboratory testing, sample drilling and field investigations. Although these earlier approaches provide precise identification of subsurface hydrogeological features, they can be time-consuming and expensive (Ganapuram et al. 2009; Nampak et al. 2014).

Conventionally, remote sensing and geographic information system (GIS) are integrated with knowledge-driven models that are effectively applied to delineate the groundwater prospect zone including weights of evidence (Elewa and Qaddah 2011; Lee et al. 2012; Pourtaghi and Pourghasemi 2014; Madani and Niyazi 2015; Tahmassebpour et al. 2016), analytical hierarchy process (Arulbalaji et al. 2019; Ramachandra et al. 2022). However, because the models utilized in these studies are dependent on expert opinion, the effectiveness of the groundwater assessment potential was subjective, mostly high bias and insufficiently accurate.

Recently, with the exponential increase in computing power and the advancements of algorithms, machine learning has continuously been utilized to solve several real-world issues including GWPM (Karpatne et al. 2019; Elmahdy et al. 2021). Machine learning is a subset of artificial intelligence that enables software applications to grow increasingly effective at predicting outcomes without explicitly programming them to do so, therefore, machine learning algorithms estimate new output values using previous data as input. The numbers of machine learning models have grown rapidly for GWPM, such as logistic regression (Park et al. 2017), K-nearest neighbor (Naghibi and Moradi Dashtpajardi 2017; Martínez-Santos and Renard 2020), Gaussian naive Bayes (Martínez-Santos and Renard 2020), decision tree (Naghibi et al. 2016; Chen et al. 2020; Patidar et al. 2021), random forest (Golkarian et al. 2018; Al-Fugara et al. 2020b; Prasad et al. 2020; El Bilali et al. 2021), support vector machine (Lee et al. 2017; Rizeei et al. 2019; Al-Fugara et al. 2020a), artificial neural network (Nguyen et al. 2020; Pradhan et al. 2020) and convolution neural network (Xu et al. 2020; Chen et al. 2021).

For creating groundwater potentiality maps, a variety of models have been created so far. According to a review of the literature, combining evolutionary algorithms with machine learning has produced better results (Naghibi and Moradi Dashtpajardi 2017; Al-Fugara et al. 2020b;

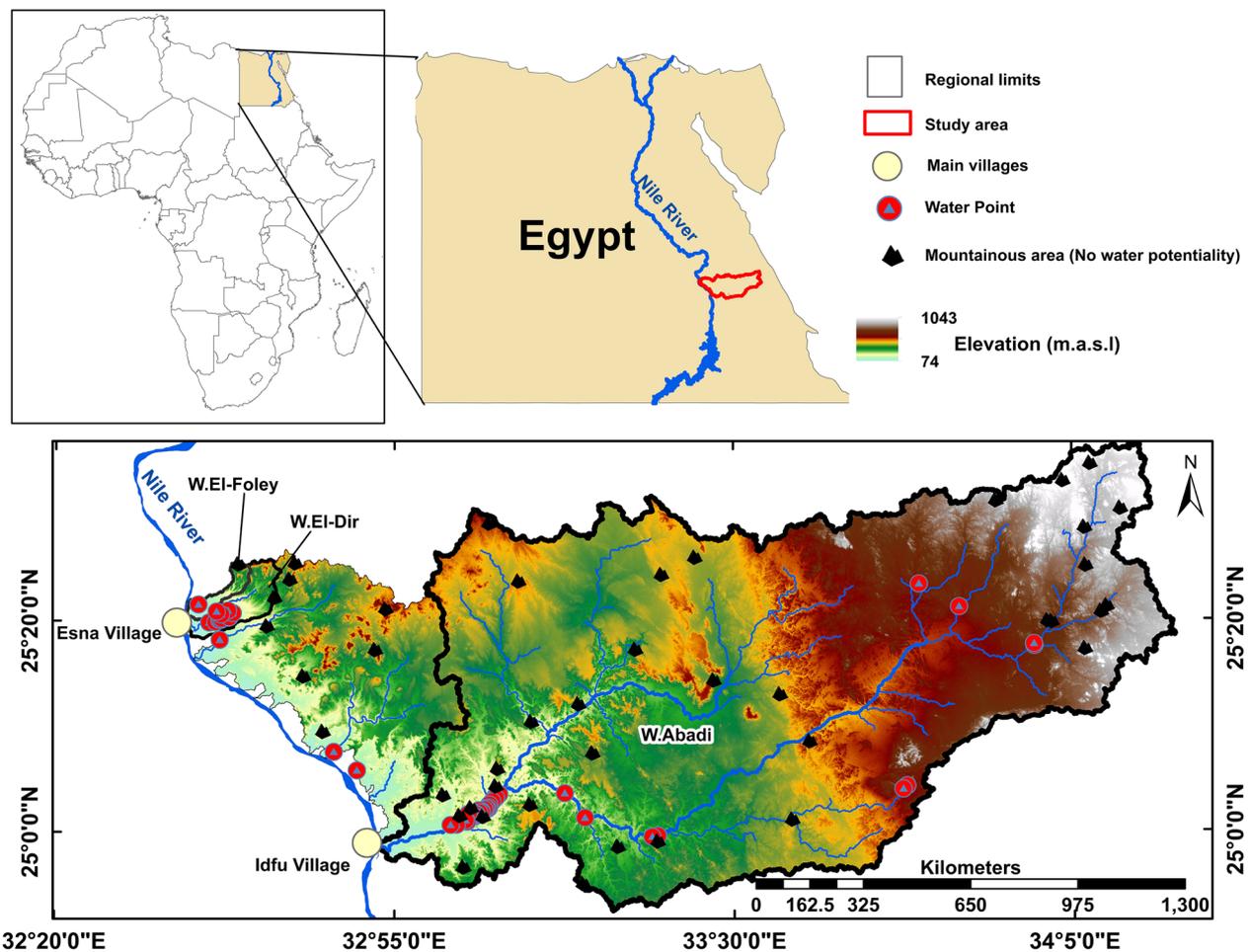
Pal et al. 2020). When machine learning algorithms are compared to each other in multi-models without sufficient consideration to the specifics and challenges linked to their structural characteristics, they cannot provide a suitable benchmark for researchers since not enough attention is devoted to them. As a result, understanding a model's specifics can greatly help in identifying its capabilities. The random forest (RF) model will be the focus of this research.

The RF algorithm is an ensemble machine learning model that has been used as a data-driven prediction for GWPM (Rahmati et al. 2016; Prasad et al. 2020). We have chosen the RF model in this study because: (a) it improves the decision tree accuracy by reducing overfitting; (b) it can deal with imbalanced data where water points concentrate in downstream in the main wadis; (c) performs well in high dimensionality data; (d) it is relatively strong against outliers and can overcome the "black-box" limitation of artificial neural networks (Palczewska et al. 2014) and offers a novel approach to GWPM by analyzing the relative importance of the groundwater effective features and determining the most important features; (e) results in higher prediction performance (Wiesmeier et al. 2011); (f) due to a wide number of trees, there is low bias and low variance; (g) acceptable error estimations using the model out of bag (OOB) error.

In this study, a cost-effective interdisciplinary research strategy comprising the integration of GIS, satellite images and RF model, as well as thematic layers produced from Arc GIS and field data, is used to determine GWPM in dry wadis in arid conditions in the East Idfu-Esna area as a case study in Egypt's Eastern desert.

## Study area

The research location lies in the Nile Valley in Upper Egypt, east of the villages of Idfu and Esna (Fig. 1). It extends across the center of the Eastern Desert in a NE direction. The Central Eastern Desert is a semi-arid area. The study area is bounded by latitudes 24°52' and 25°37'N and longitudes 32°33' and 34°15'E. It has a large land area of around 8000 km<sup>2</sup>. The elevation of the study area ranges from +1043 m in the upstream portion to +74 m in the downstream portion. It contains many wadis that end in the Nile River from Wadi Abadi in the southern part to Wadi El-Dir and El-Foley in the north part of the study area. Wadi Abadi has the largest drainage network, covering around 6700 km<sup>2</sup> and it stretches 200 km east crossing the Red Sea mountainous terrains. The study area contains about 57 water points that were collected data through a late field survey and from the previous study (Hammad et al. 2015). There are two main topographic zones in the study area: the first is made mainly



**Fig. 1** General map of Egypt shows the geographic location of the study area and a close-up view of SRTM DEM illustrates wadis dissecting the study area trending east-west direction associated with the location of water points and mountainous area

of basement rocks and is rough with high relief and the second zone is low relief and composed of sedimentary rocks. This zone descends gently westward towards the Nile and rises more steeply eastward into the basement range.

Geologically, sedimentary succession makes up the majority of the East Esna-Idfu region, which covers around 71% of the study area. The sedimentary succession ranges in age from upper Cretaceous to recent. Precambrian basement rocks cover about 29% and locate in the eastern part of the study area. They are composed of crystalline Neoproterozoic igneous and metamorphic rocks from the Arabian–Nubian shield, which range in age from 550 to 900 M (Sultan et al. 1990). Upper Cretaceous rocks are non-conformably found on top of the Precambrian basement rocks and are classified into four formations from bottom to top: Taref, Quseir variegated shale, Duwi, and Dakhla (Fig. 2).

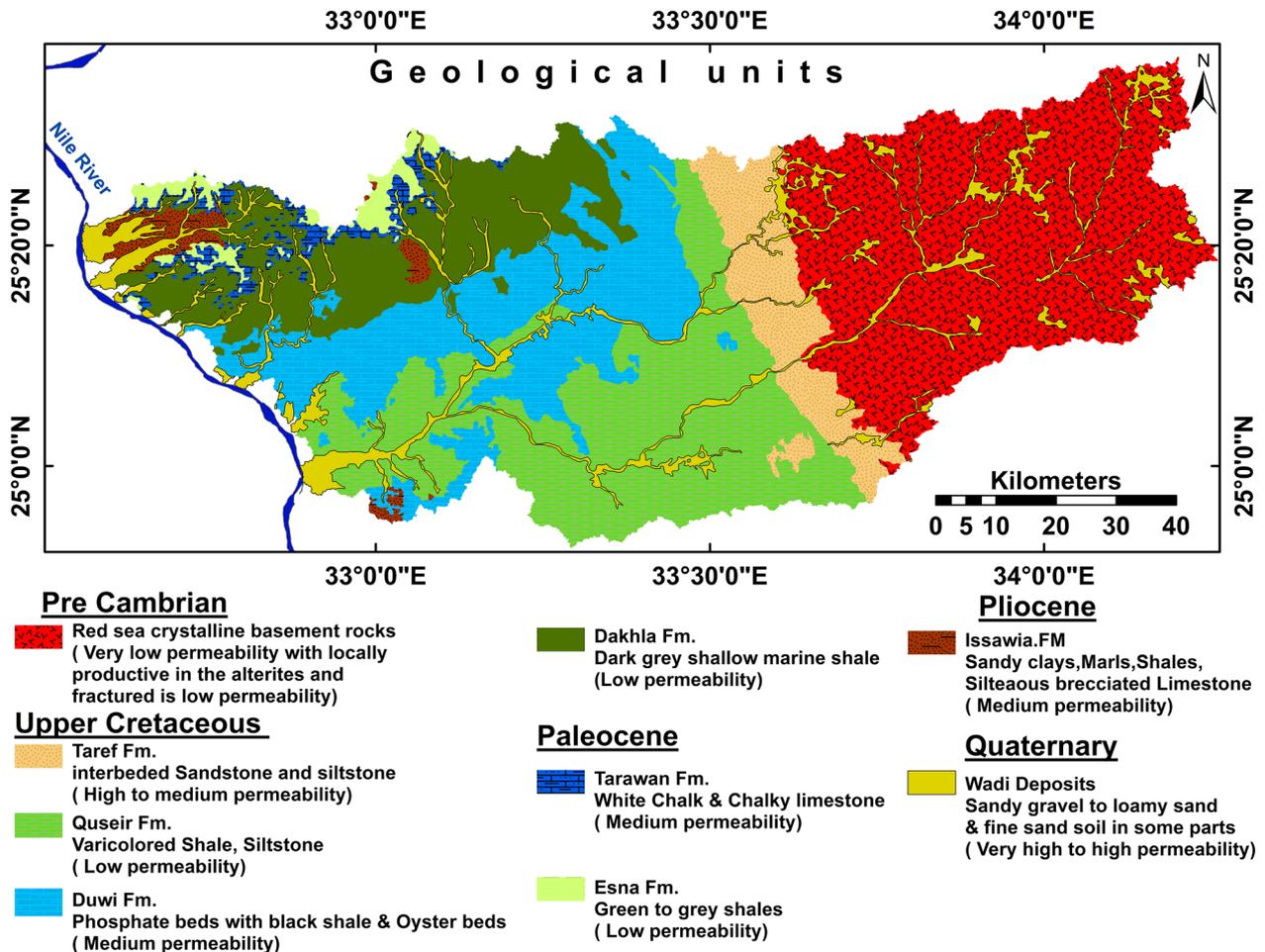
In terms of hydrogeology, three main aquifers have been identified: (a) an unconfined Quaternary alluvium aquifer near to Nile River, especially in the northern part at Wadi El-Dir and El-Foley; (b) a semi-confined Nubia Sandstone aquifer discovered in Wadi Abadi; and (c) a Precambrian fractured basement aquifer that consists of disconnected local aquifers. Permeability as a potentiality recharge relatively decreases from wadi deposits and Taref Sandstone to shale beds and Precambrian crystalline rocks that form the lowest permeability.

### Methodology

Material and methods used to utilize, enhance and evaluate RF classifier model for the prediction of the GWPM in the study area are presented as following.

### Data used and software

Various types of data were used in this investigation (Table 1). For dependent features, groundwater



**Fig. 2** The study area’s general geological units (Conco 1987) are associated with relative permeability

**Table 1** Data used for effective features creation and dependent feature preparation

Data used	Source data	Scale	Time period
Spring and well locations associated with groundwater information	(Hammad et al. 2015) Field survey	–	2015 September 2021
Rainfall	The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) <a href="https://power.larc.nasa.gov/data-access-viewer/">https://power.larc.nasa.gov/data-access-viewer/</a>	–	January 1981–December 2019
Digital elevation model (DEM)	Shuttle Radar Topography Mission (SRTM) <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a>	1 arc second, 30 m spatial resolution	23th September 2014
Satellite image	Multispectral satellite landsat8 (OLI) <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a>	30 m spatial resolution and 15 m panchromatic band	1st July 2021
	Sentinel-2A satellite image <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a>	10 m spatial resolution	29th June 2021
Geology	(Conco 1987)	1:500,000	1987
Infiltration test and sieve analysis for soil type recognition	Field survey	–	September 2021

information (number of water points, depth to water, aquifer type, etc.) is collected from 57 water points (wells and springs) in 2015 (Hammad et al. 2015) and 2021 through field survey. Forty-three points are selected in the mountainous area and high land to mark non-potential groundwater area. For effective features creation, different types of data are collected. Along with geologic maps and fieldwork data, four different types of satellite remote sensing data were collected for digital image processing.

Arc GIS Pro 2.8 software was used to create effective features. It uses python programming language associated with machine learning libraries such as Scikit-Learn and geospatial libraries such as Arc Py to run RF Algorithms. The SPSS statistics 20 software was used to calculate and draw receiver operating characteristics (ROC) curve and determine the most important effective features depending on the area under curve (AUC).

**Methods**

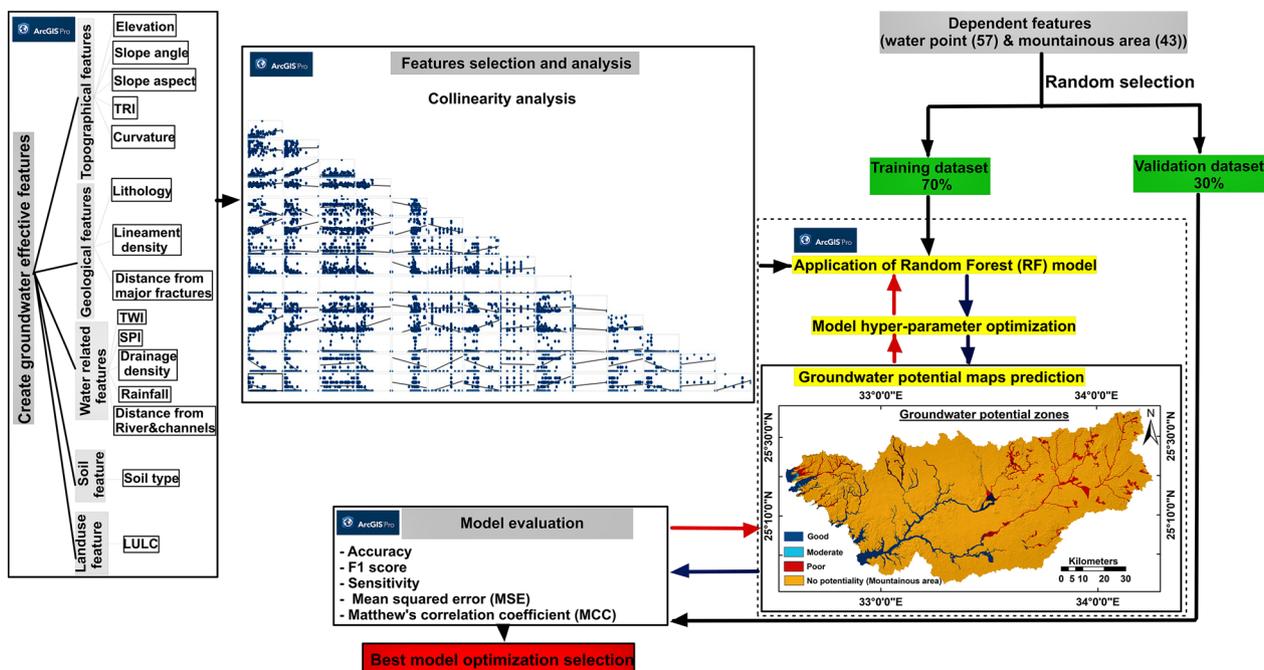
Knowledge extraction such as GWPM from data is made possible by machine learning through a mechanism known as "the Machine Learning Life Cycle" (Ashmore et al. 2021).

In Fig. 3, a complete cycle of RF classification algorithm flowchart is illustrated to predict GWPM performance and hydrogeological acceptable as following: (a)

dependent features preparation by labeling every water point (as good, moderate or poor) based on collected groundwater information associated with labeling all points of mountainous area as non-potential; (b) effective features creation (Table 2): create 15 features: topographical features (elevation, slope angle, slope aspect, terrain ruggedness index (TRI) and curvature), geological features (lithology, lineament density and distance from major fractures), water-related features (topographic wetness index (TWI), stream power index (SPI), drainage density, rainfall and distance from rivers and channels), soil features (soil type) and land use features (land use/land cover (LULC)); (c) feature selection and collinearity analysis; (d) random selection and splitting of dependent features as 70% using RF model training and 30% for model validation; (e) utilization of ensemble RF classification using by training 70% of dependent features on effective features; (f) model enhancement by optimize hyper-parameters according to performance resulting from validation and create GWPM based on the best model optimization; (g) model evaluation using equations in Table 2 and finally discuss the most important features.

**Results**

In this paper, the results of each part through machine learning life cycle to predict acceptable GWPM are illustrated as following.



**Fig. 3** A machine learning flowchart is developed to predict the groundwater potential map

**Table 2** The equations used to create some effective features (TRI, TWI and SPI) associated with the equations used for random forest model evaluation

The equations used for creating effective features		
S. n.	Index	Equation
1	TRI	$(Avg - Min)/(Max - Min)$
2	TWI	$\ln\left(\frac{Fa}{\tan\beta}\right)$
3	SPI	$Fatan\beta$
The equations used for the random forest model validation		
	Measure	Equation
4	$MSE_{OOB}$	$n^{-1} \sum_{i=1}^n (\gamma_i - \Delta_i)$
5	Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
6	Sensitivity	$\frac{TP}{TP+FN}$
7	Precision	$\frac{TP}{TP+FP}$
8	F1-score	$2 \times \frac{(Sensitivity \times Precision)}{(Sensitivity+Precision)}$
9	MCC	$\frac{(TP \times TN - FP \times FN)}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$

TRI: terrain ruggedness index; Avg, Max, and Min: averages, highest and lowest elevation values of pixels in nine rectangular neighborhoods of elevation values, respectively; TWI: topographic wetness index; Fa: flow accumulation ( $m^2 m^{-1}$ );  $\beta$ : slope angle measured in degrees; SPI: stream power index; MSE: mean square error; OOB: out of bag; TP: true positive; TN: true negative; FP: false positive; FN: false negative; MCC: Matthews correlation coefficient

### Dependent features preparation

Labeling and classifying dependent features are mandatory before running supervised machine learning classification models (Kotsiantis et al. 2006). Forty-three points in mountainous were labeled as non-potential because they are high land from the surrounding areas and do not prospect for any future water well drilling. Fifty-seven water points are classified into 3 classes (good, moderate and poor) groundwater potentiality based on (Table 3): (a) aquifer type; (b) aquifer name and lithology; (c) depth to water; (d) drawdown in water level through last 7 years (2015–2021). The Precambrian aquifer is unconfined of unconnected local aquifers that form from faults and fractures, so all water wells located in this aquifer are of poor potential. The Nubia aquifer is a semi-confined aquifer significant in the down and middle stream of Wadi Abadi. Drawdown of wells through the last 7 years in downstream of Wadi Abadi is very low and average transmissivity based on pumping test is  $346.3m^2/day$ . Although the depth to water in the new wells (well 8 and well 9) is moderate to deep (44–55 m) in the middle stream, overall productivity is  $140 m^3/h$  “personal contact” and the total penetrated thickness is about 360 m of fine-to-medium sandstone. Therefore, all water points in the Nubian aquifer are of good potential. The quaternary aquifer is an unconfined aquifer that recharges from rainfall and partially from the Nile River and it is significant in

Wadi El-Dir and El-Foley in the Esna area and along the Nile River. Water wells in Wadi El-Dir are classified into 3 classes: good (low drawdown and near to Nile River), moderate (moderate drawdown, and water depth) and poor (high drawdown reaches 15 m and deep in water depth).

### Preparation of effective features

Even though satellite data cannot see very far below the surface, it offers data on characteristics that may indicate the existence of groundwater (Díaz-Alcaide and Martínez-Santos 2019). 15 effective features used in this study (Figs. 4 and 5) were created based on different types of satellite data, geologic maps and field measurements. The following paragraphs go into great depth on how each feature was created and how it relates to groundwater potentiality.

#### Topographic features

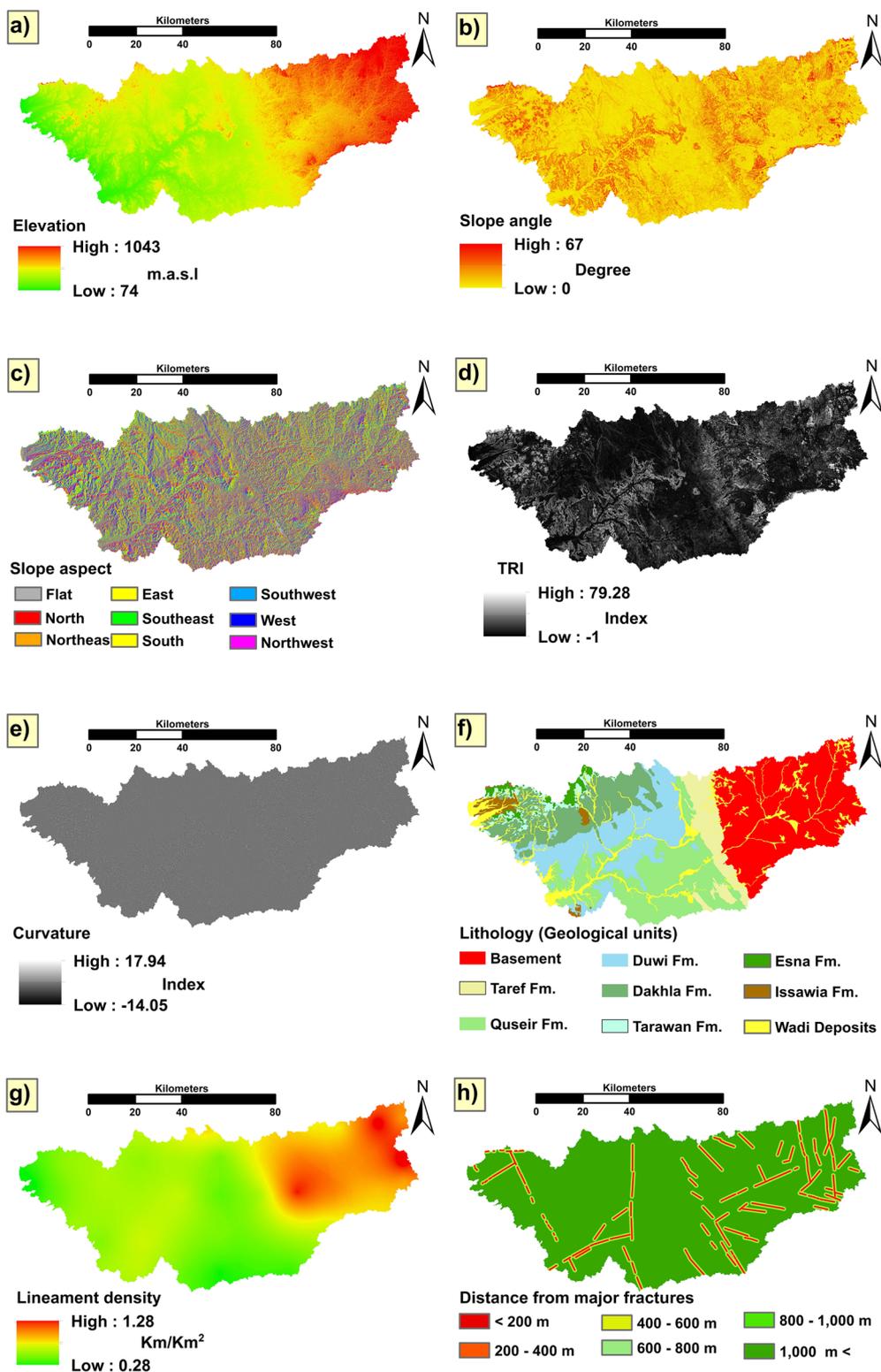
In the mountainous region, the topographical features serve as markers for determining groundwater conditions (Todd and Mays 2005; Das 2017). The potential for groundwater in a particular place is inversely related to elevation in an indirect manner. Elevation feature (Fig. 4a) has been created using SRTM-DEM data. SRTM-DEM data are processed in ArcGIS software using spatial analyst tools to establish the nature

**Table 3** Classifying water points based on groundwater information

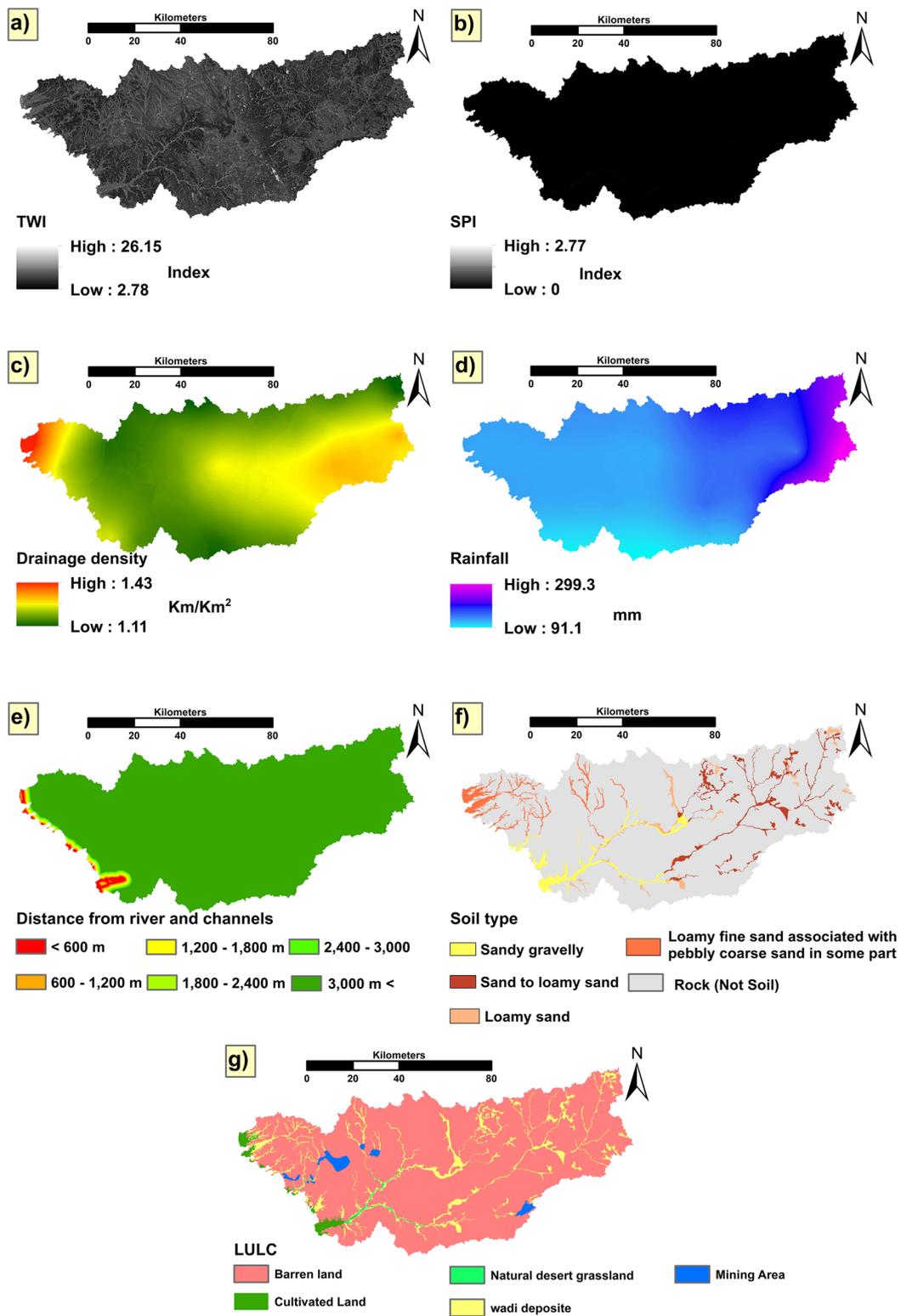
S. no.	Water point type	Basin name	Aquifer type	Aquifer name	Location	Longitude		Depth to water (2015–2021)		Potentiality
						Latitude	Longitude	2015	2021	
1	Hand-dug wells	Upstream of w. Abadi	Unconfined basement (local aquifer)	Precambrian aquifer	25.29305	34.01715	21	–	–	Poor
2					25.2962	34.016733	13.5	–	–	
3					25.35523	33.8881	11.2	–	–	
4					25.39138	33.8192	3.3	–	–	
5					25.07295	33.79751	–	36.3	–	
6					25.06789	33.79088	–	32.8	–	
7	Drilled wells	Middle stream of w. Abadi	Semiconfined aquifer	Nubia aquifer	24.99508	33.36889	–	60	–	Good
8					24.99331	33.35946	–	55	–	
9					25.02363	33.24399	–	44	–	
10					25.06227	33.20951	–	32.3	–	
11		Downstream of w. Abadi			25.06086	33.09681	8.5	–	–	
12					25.05564	33.08903	13	–	–	
13	Hand-dug well				25.05524	33.09096	8.5	10	1.5	
14	Drilled wells				25.04953	33.08483	10.5	–	–	
15					25.02464	33.04648	2	5.2	3.2	
16					25.04439	33.08006	10	–	–	
17					25.03958	33.07906	–	10.2	–	
18					25.03756	33.07654	–	10	–	
19					25.03856	33.07439	11.8	–	–	
20					25.03636	33.07169	11.9	–	–	
21					25.0367	33.07145	8.1	–	–	
22					25.03411	33.06639	10	11	1	
23					25.03537	33.062	–	4.2	–	
24					25.03233	33.06653	7.5	8	0.5	
25					25.03237	33.06243	–	15.8	–	
26					25.03108	33.05066	5.6	–	–	
27					25.0298	33.05705	7.2	8.5	1.3	
28					25.02908	33.05576	8	–	–	
29					25.02667	33.05197	8.3	11	2.7	
30					25.02619	33.0523	6	–	–	
31					25.02593	33.05221	7.2	–	–	
32					25.02555	33.05486	–	10	–	
33					25.02591	33.05341	–	12	–	
34					25.02426	33.04908	–	6	–	
35					25.02448	33.04635	2	–	–	

**Table 3** (continued)

S. no.	Water point type	Basin name	Aquifer type	Aquifer name	Location		Depth to water (2015–2021)			Potentiality
					Latitude	Longitude	2015	2021	Drawdown	
36	Hand-dug well				25.02999	33.0442	–	6.7	–	
37	Drilled well				25.0296	33.0443	–	7.8	–	
38	Hand-dug wells		Unconfined aquifer	Quaternary aquifer	25.0177	33.04155	3	4	1	
39					25.01735	33.04177	2.7	4.2	1.5	
40					25.01964	33.03901	–	4.1	–	
41					25.0128	33.02373	3	–	–	
42					25.01222	33.01346	1.5	1.7	0.2	
43	Drilled well	W. Al-Mafallis			25.09834	32.85235	–	2.9	–	
44	Spring	W. Hilal	Confined	Nubia	25.12756	32.81267	Flowing	–	–	Moderate
45		W. El-Sabil			25.30405	32.61621	36	–	–	Moderate
46					25.33117	32.59856	18	20	2	Good
47	Drilled well	W. El-Dir	Unconfined aquifer	Quaternary aquifer	25.33142	32.5979	18	20	2	Good
48					25.33219	32.60896	–	30	–	Moderate
49					25.33567	32.61537	–	40	–	
50					25.33562	32.61775	–	44.6	–	
51					25.33738	32.61988	39	47	8	
52					25.33968	32.6327	40	54	14	Poor
53					25.34631	32.6385	–	63	–	
54					25.35036	32.63036	45	60	15	
55					25.34415	32.61833	65	–	–	
56					25.34867	32.611	–	32	–	Moderate
57		W. El-Foley			25.35972	32.57956	–	20	–	Good



**Fig. 4** Effective features creation for GWPM: **a** elevation; **b** slope angle; **c** slope aspect; **d** TRI; **e** curvature; **f** lithology; **g** lineament density; **h** distance from major fractures



**Fig. 5** Effective features creation for GWPM: **a** TWI; **b** SPI; **c** drainage density; **d** rainfall; **e** distance from river and channels; **f** soil type; **g** LULC

of the slope of the entire area to produce the slope angle (Fig. 4b), slope aspect (Fig. 4c), terrain ruggedness index (TRI) (Fig. 4d) and curvature (Fig. 4e) maps. Low-slope areas are suitable for water accumulation and infiltration. Curvature is the derivative of elevation and defined as the rate of change of slope (Catani et al. 2013), it affects the acceleration and convergence of water runoff. TRI gives an objective quantification of topographic heterogeneity (Riley et al. 1999) influencing drainage. It is calculated in Eq. (1) in Table 2.

### **Geological features**

The groundwater is usually located in the pore spaces between grains in rocks and the secondary porosity such as faults and joints. Lithology is an important indicator of hydrogeological properties that defines the hydrogeological characteristics of aquifer materials (Hussien et al. 2017; Yidana et al. 2020). The interpretation of false color composite (FCC) of Landsat 8 band ratios (3/5, 1/4, 1/6) associated with published geological maps (Conco 1987) and field surveys were employed in lithological discriminating of distinct rock units (Fig. 4f).

Lineaments, which are considered secondary porosity, are a significant feature to be considered while investigating groundwater potentiality. Various researchers have used the relationship between groundwater potential and lineaments to emphasize that high lineament density closely correlates with high groundwater potentiality (Magowe and Carr 1999; Hung et al. 2005; Al-Ruzouq et al. 2019). Remote sensing data, such as the panchromatic band of Landsat 8 and the combination of Landsat 8 bands (7,5,3), were utilized in conjunction with a published geological map (Conco 1987) and field trip in order to visually extract structural lineaments and determine major linear fractures, using ArcGIS software to create lineament density (Fig. 4g) and distance from major fractures features (Fig. 4h).

### **Water-related features**

Various features are resulting from surface water runoff such as topographic wetness index (TWI), stream power index (SPI) and drainage density. Some significant features recharge the aquifers in the study area such as rainfall and distance from rivers and channels.

The TWI is a secondary topographic index that shows how topography affects the quantity of runoff generation and flow accumulation at any site within the catchment region (Gokceoglu et al. 2005). Recently, TWI (Fig. 5a) has been widely used for groundwater potential mapping creation (Prasad et al. 2020; Paryani et al. 2022). SPI (Fig. 5b) is a measure of how much water flow erodes.

TWI and SPI are calculated in Eqs. (2) and (3) (Moore et al. 1991) in Table 2.

The drainage density feature is a vital component in hydrogeological research. The drainage networks in the area under investigation are taken from SRTM-DEM data and analyzed using spatial analyst tools in ArcGIS software. The entire length of streams per square meter is known as drainage density. The research area is graded by 10 min of degree and divided into polygons, drainage density (Fig. 5c) is then calculated for each polygon, and a raster surface is interpolated from points using kriging ArcGIS software.

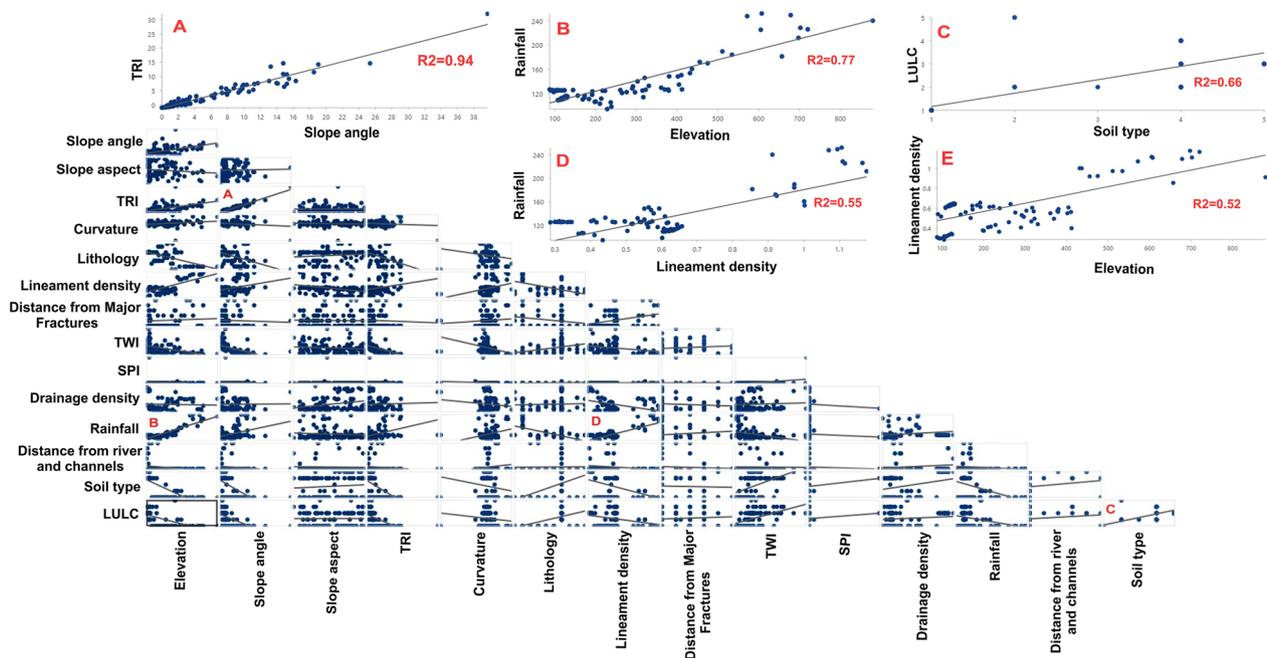
To measure the quantity of precipitation in the research region for the last four decades and produce the rainfall feature (Fig. 5d), MERRA-2 for precipitation data are employed. The monthly MERRA-2 cumulative rainfall data for 39 years (from January 1981 to December 2019) was used to create the rainfall thematic layer. Kriging ArcGIS software was used to interpolate a raster surface from the points. Groundwater recharge is also controlled by the distance from the surface channel network and the water body (Adeyeye et al. 2019). To extract the channel network, a visual interpretation approach based on sentinel-2A images validated by Google Earth satellite imagery was utilized (Fig. 5e).

### **Soil feature**

Soil types impact groundwater recharge by determining the quantity of water that may percolate into underlying formations (Das 2017). PCA is constructed using Landsat 8 satellite images to differentiate between distinct soil types in the research area's Quaternary deposits. Using the data derived from the PCA color composite image, the infiltration test and sieve analysis for soil samples from various places in Quaternary deposits were performed during the field survey (Fig. 5f). The infiltration capacity equilibrium based on infiltration test in sandy gravelly, sand to loamy sand, loamy sand and loamy fine sand soil are 13.8, 4.5, 2 and 0.53 mm/min, respectively.

### **Land use feature**

The types of land use/land cover (LULC) have an impact on groundwater recharge (Kaur et al. 2020). A visual interpretation method based on sentinel-2A that was validated by Google Earth satellite imagery and field trip were used to produce the LULC feature (Fig. 5g). Barren land is a LULC class that is not a prospect for groundwater potentiality because it is a mountainous area, as well as all water points and developments, are located within wadis.



**Fig. 6** Pairwise linear relationships associated with R-squared for all effective features

### Collinearity analysis (CA)

CA is a vital method in feature selection before machine learning model training (Chen et al. 2021; Víctor et al. 2021). It is a statistical technique for a linear relationship between two independent features.  $R$ -squared is a common and widely used in CA (Pradhan et al. 2020). Very high  $R$ -squared ( $>0.95$ ) leads to a major problem in the training dataset and creates inaccurate results (Daoud 2018).

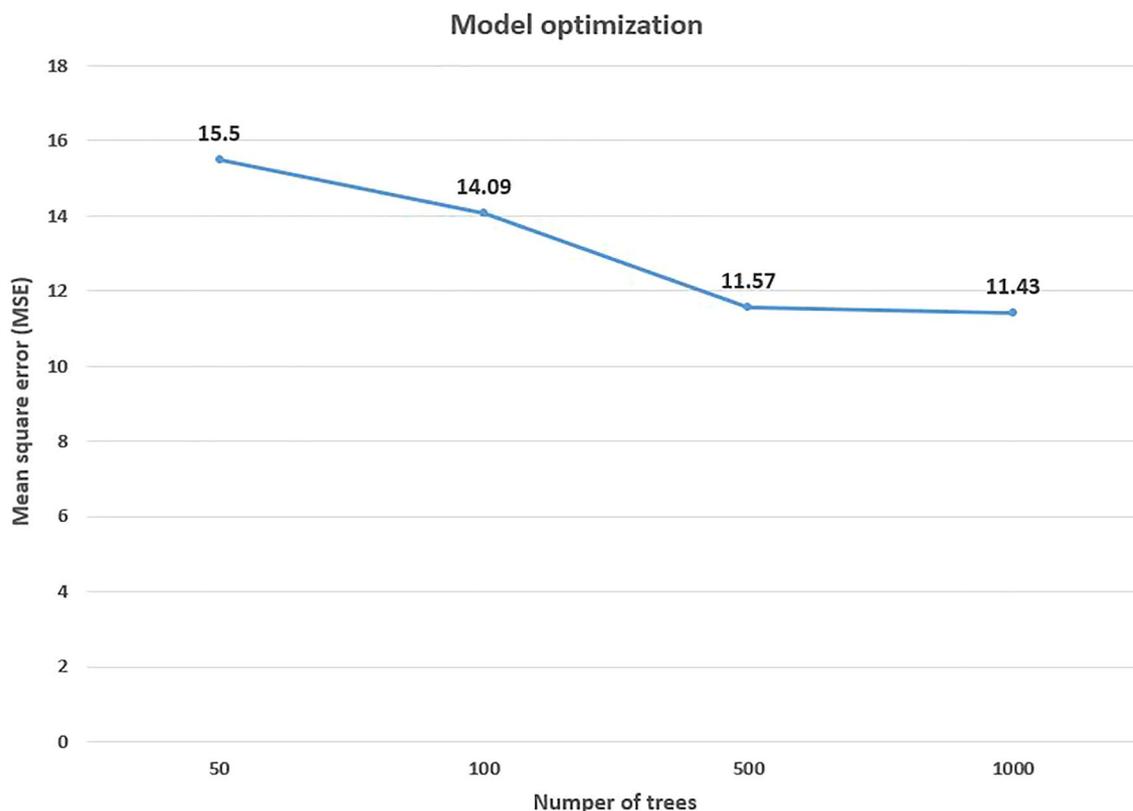
Figure 6 shows the linear relationship associated with  $R$ -squared between features. No significant very high  $R$ -squared ( $>0.95$ ) between features relationships. There are quite strong positive relationships between the following features: (a) TRI and slope angle ( $R^2=0.94$ ), both of them are important to express topography by different methods depending on DEM; (b) rainfall and elevation ( $R^2=0.77$ ), precipitation increases in high land like red sea mountainous area; (c) LULC and soil type ( $R^2=0.66$ ), most of water points and developments are located in soil material within wadis; (d) rainfall and lineament density ( $R^2=0.55$ ), both of them increase in Precambrian basement area in red sea mountainous area; (e) lineament density and elevation ( $R^2=0.52$ ), high elevations are high fractured and deformed Precambrian basement rocks. The other features are low  $R$ -squared.

**Table 4** RF model characteristics associated with validation

Model characteristics	
Number of trees	50
Leaf size	1
Mean tree depth	5
Number of randomly sampled variables	3
Percent of validated point	30
Model out of bag (OOB) error	
Mean square error ( $MSE_{OOB}$ )	15.5
MCC Matthews correlation coefficient	

### Utilization of RF classification model

RF was created as an extension of classification and regression trees (CART) to increase the model's prediction performance (Breiman 2001). The model construction procedure is similar to that of CART, with the exception that multiple trees are produced, resulting in some kind of a "forest of decision models". For classification, the RF model employs the resampling strategy that changes the predictive features randomly to maximize the diversity within every tree. This technique combines numerous decision trees to explain the spatial link between effective groundwater variables and dependent variables. Each decision tree is constructed from a



**Fig. 7** Number of trees as a hyper-parameter optimization

bootstrap sample of raw data, allowing for robust error quantification with the residual validation set, referred to the out of bag (OOB) sample. The mean square error ( $MSE_{OOB}$ ) of all trees is calculated in Eq. (4) in Table 2.

Table 4 summarizes the RF characteristics model used for training as well as  $MSE_{OOB}$  as a validated method.

#### Model hyper-parameter optimization

Using hyper-parameter optimization to enhance the RF model. The number of trees is the most important hyper-parameter in The RF model. With increasing number of trees from 50 to 1000 trees,  $MSE_{OOB}$  decreased from 15.5 to 11.4 (Fig. 7).

#### GWPM prediction

Figure 8 shows predicted GWPM based on the trained RF classification model after enhancement. The predicted GWPM was classified to no potentiality area (90.6%), poor (4.2%), moderate (0.1%) and good (5.1%). This model target is to delineate groundwater potentiality within wadis in the study area. In Wadi El-Dir and Wadi El-Foley (Fig. 8a), the quaternary aquifer is delineated as:

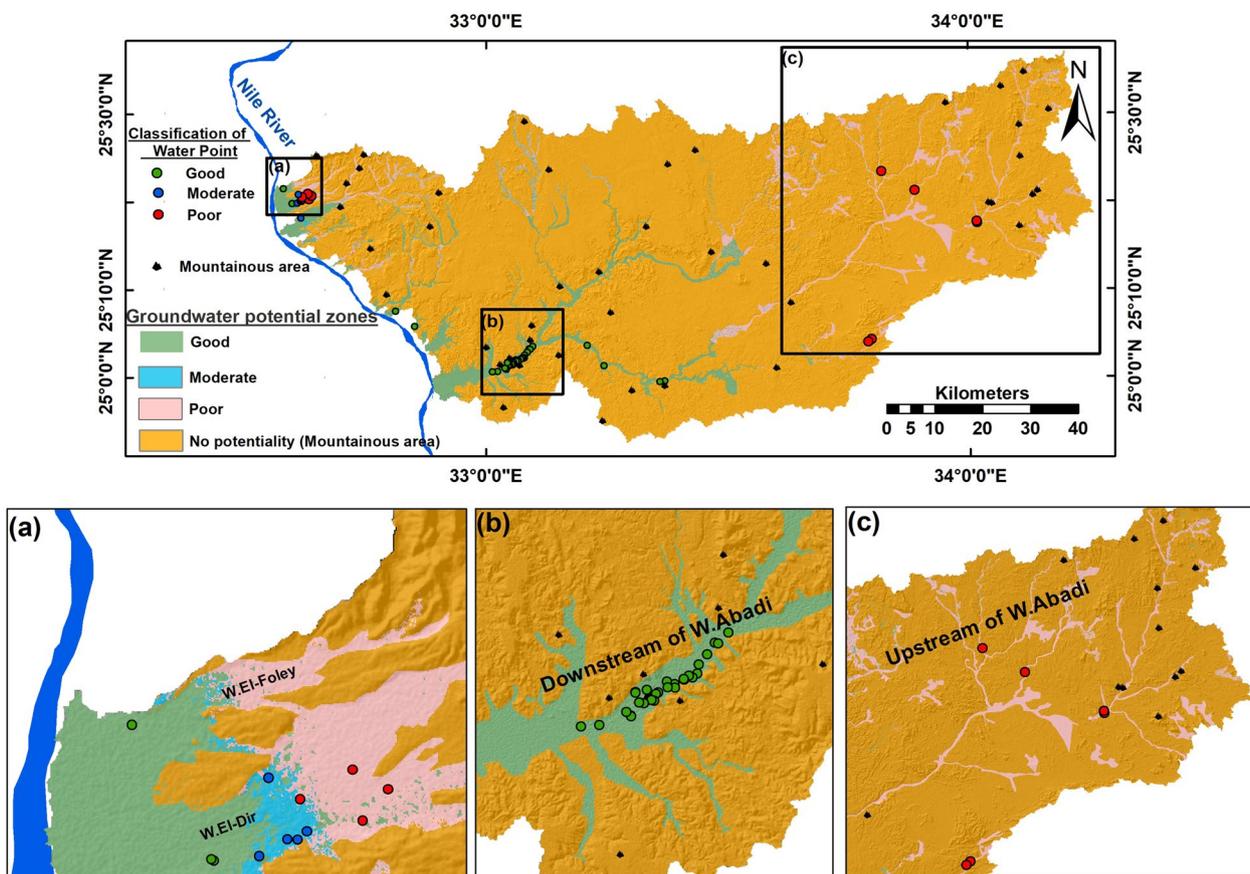
(a) good (near the Nile River); (b) moderate (appears only in this area as a transitional zone between good and poor zones); (c) poor (appears in upstream of the quaternary aquifer and in basement aquifer). In downstream and middle stream of Wadi Abadi (Fig. 8b), the Nubia aquifer is classified as a good potentiality. In upstream of Wadi Abadi (Fig. 8c), the basement aquifer is delineated as a poor potentiality. This predicted map is hydrogeological acceptable in this study area.

#### Discussion

This paper concerns the study of RF algorithm as an ensemble machine learning model taking into consideration the previous studies to predict GWPM. The outcomes of this work are discussed as follows.

#### Validation and performance

For the evaluation of the predicted GWPM, the model's validation methods are essential. Confusion matrix (CM) of the model can be visualized (Fig. 9). Due to imbalanced classification data set, accuracy cannot be used solely to evaluate model performance. The following calculations



**Fig. 8** GWPM derived from RF classification model associated with classification of dependent points: **a** zoom in Esna area; **b** zoom in downstream of Wadi Abadi; **c** zoom in upstream of Wadi Abadi

can be used in Eqs. (5, 6, 7 and 8) based on CM (Sokolova and Lapalme 2009; Chicco and Jurman 2020) in Table 2.

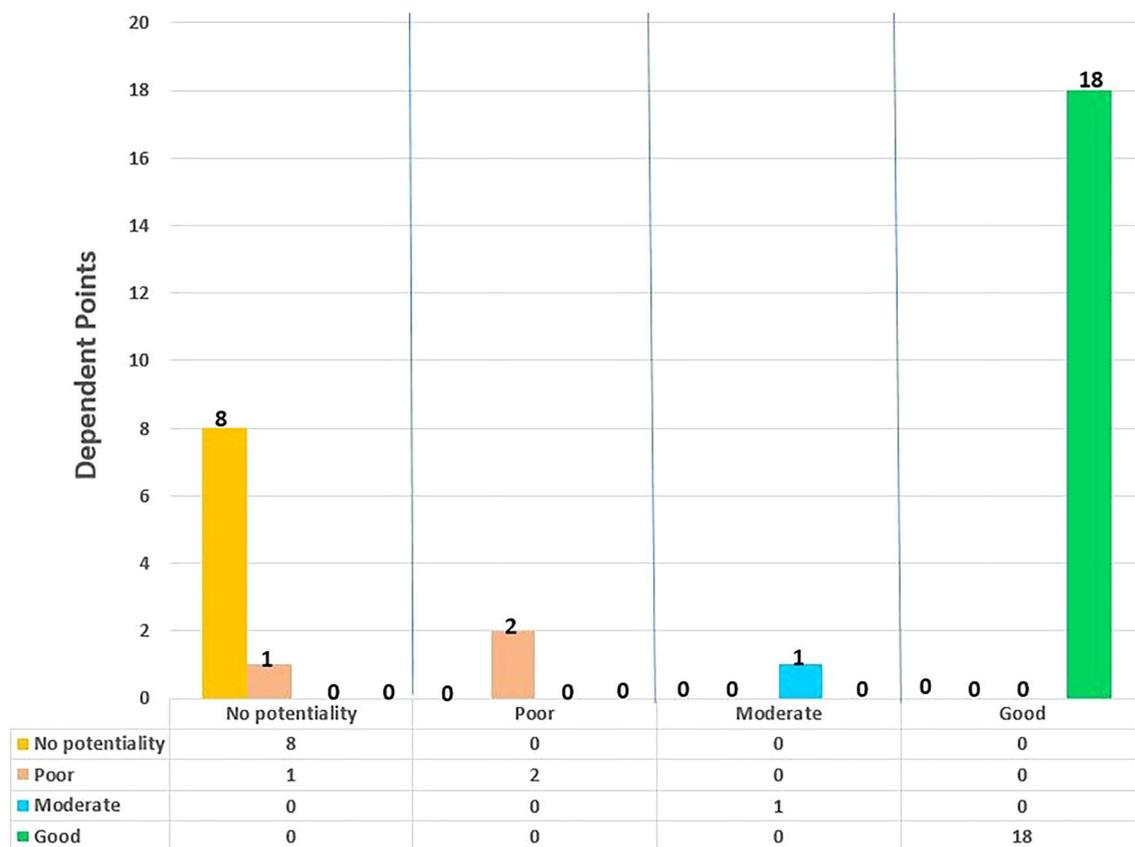
The results of different methods as summarized in Table 5 were used to evaluate RF model and prove that model is best fit with over all accuracy (97%) and sensitivity (92%) at the validation.

The RF is useful to predict high-accuracy GWPM. It proved its strength against knowledge-based methods (Al Saud 2010; Patra et al. 2018; Murmu et al. 2019; Andualem and Demeke 2019; Morgan et al. 2022) and many of data-driven methods (Rahmati et al. 2016; Rizeei et al. 2019; Chen et al. 2020). There is no requirement for statistical assumptions, or outlier removal previously.

**Effective features importance for GWPM**

The “variable importance” tool of the RF model was used to highlight the relative importance of the 15 groundwater effecting features. In this situation, soil type was the most important feature, followed by TWI, LULC,

lineament density and rainfall while slope aspect had the lowest importance (Fig. 10). Soil type is the highest effective variable since most of water points are located within wadis and consist of different types of soil with various infiltration rate control the groundwater potentiality recharge. No water points are located in the rock area (not soil area). TWI is another variable for GWPM. It affects flow accumulation and direction. LULC is an important variable due to no water points in barren mountainous area. Wadi deposits and natural desert grassland are very important recharging areas for GWPM. Lineament density is a very important factor in the study area. It built basement aquifer that covers about 30% of the study area and plays a partial role in the Nubian aquifer. In Precambrian basement aquifer, the presence of groundwater is primarily governed by secondary porosity (fractures, joints and weathered rocks) rather than the primary porosity. Rainfall is a vital factor to recharge the aquifers in the study area, it is the



**Fig. 9** Groundwater classes with respect to dependent points occurrence potential zones

**Table 5** RF model evaluation

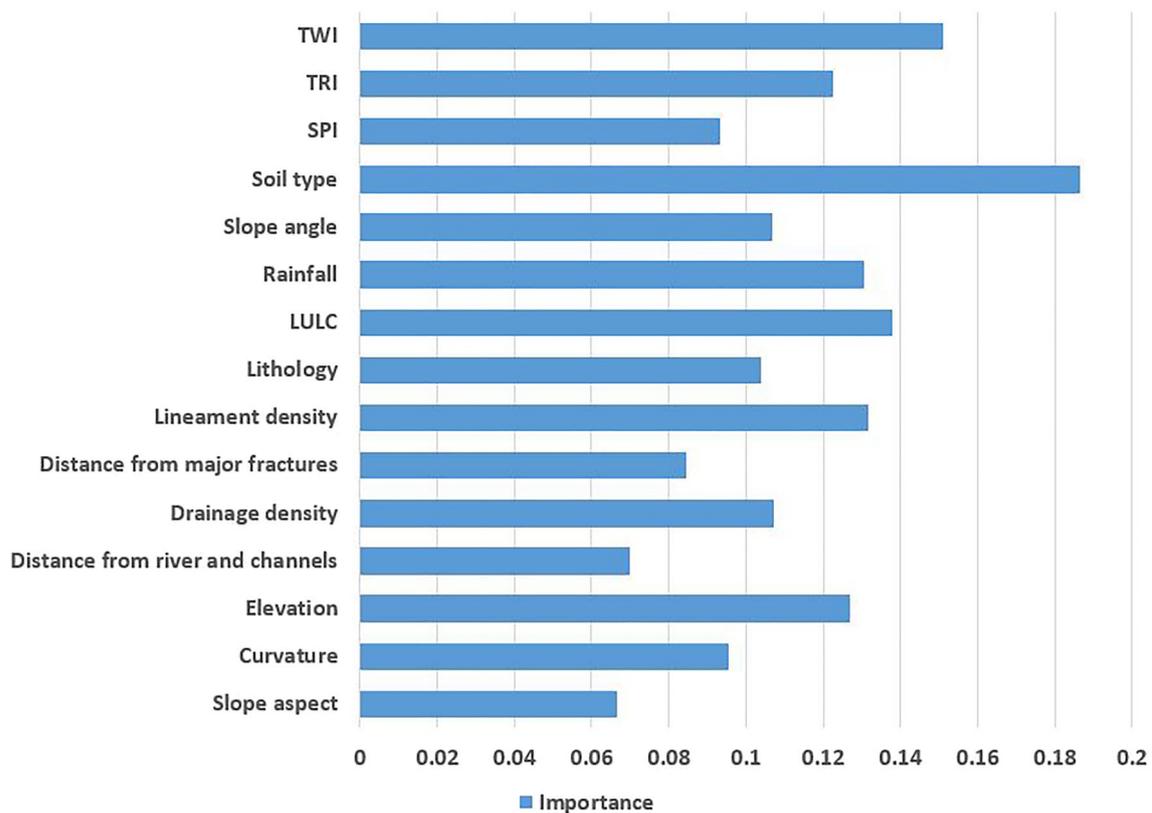
Classification diagnostics					
	F1-score	MCC	Sensitivity	Precision	Accuracy
Training data					
Overall	1.00	1.00	1.00	1.00	1.00
Validation data					
Non-potentiality	0.94	0.92	1.00	0.89	0.97
Poor	0.80	0.80	0.67	1.00	0.97
Moderate	1.00	1.00	1.00	1.00	1.00
Good	1.00	1.00	1.00	1.00	1.00
Overall	0.94	0.93	0.92	0.97	0.97

only recharging source for basement and Nubian aquifers and partially recharging source for quaternary aquifer (Mohallel et al. 2019). According to RF model in this study area, slope aspect plays the lowest role in groundwater potentiality because the direction of the slope has negligible importance on GWPM.

ROC curve is an another tool to determine the most important features to GWPM (Fig. 11). ROC agreed RF that LULC, soil type, TWI and lineament density features are the most important. Lithology and distance from major fractures have higher AUC values because they play an important role in groundwater potentiality. The rainfall feature has a low AUC value contrary to RF method.

**Groundwater management sustainability**

Groundwater sustainability can be discussed based on: predicted GWPM, condition of the aquifers, field survey and historical well data. Each aquifer in the study area is discussed as following: (a) Quaternary aquifer in Esna area, there are gradual remarkable drawdown rates in static water level increase eastward. In the moderate zone, the static water level drop (2–10 m) in the last 7 years while in poor zone water level dropped 14 m at the same period due to over pumping and farmers using flood irrigation methods as well as low recharge rate to aquifer. With this situation, Quaternary aquifer in Esna



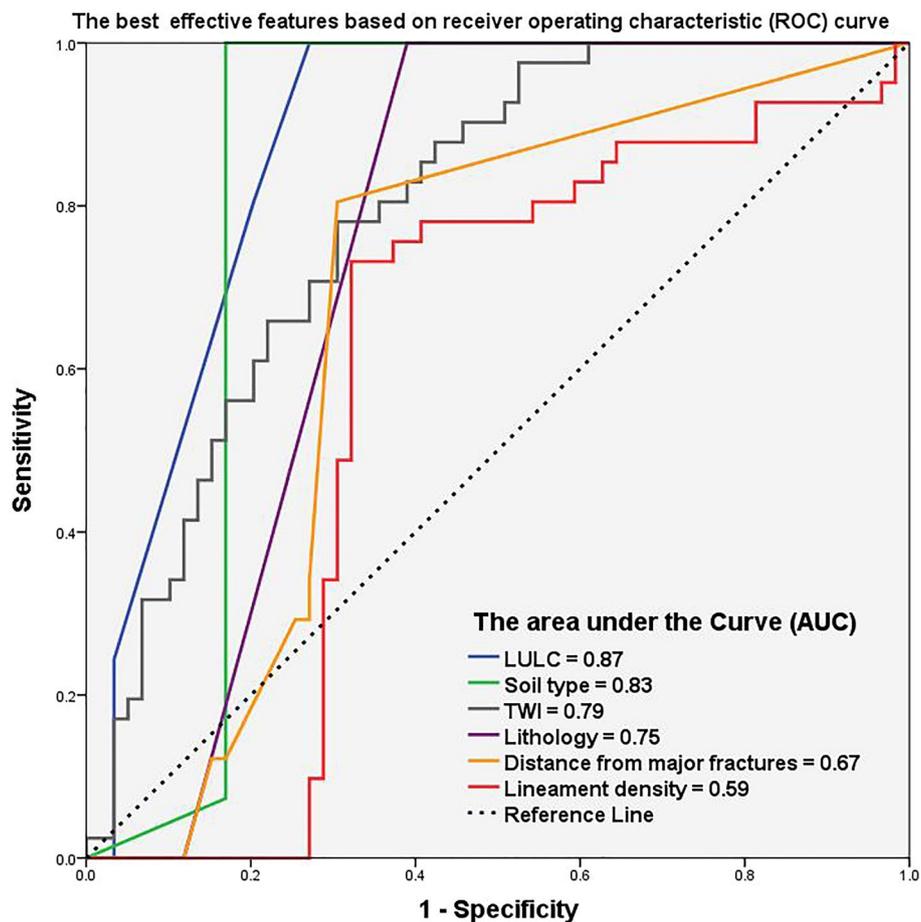
**Fig. 10** Effective features importance for GWPM prediction using the RF classification model

area will suffer from deterioration and drought will destroy the farms; (b) Nubian aquifer in down and middle stream of Wadi Abadi, it is a good potential aquifer, and this area is prospective for development and new land reclamation for agriculture, with modern types of irrigation methods for sustainability; (c) Precambrian basement aquifer, it is a poor potential aquifer with low recharge rate. It is composed mainly of isolated pockets of accumulated water that may be connected in some places through fractures.

### Conclusion

Although GWPM has been the subject of many research papers, it has become necessary to use well-developed machine learning algorithms in order to achieve high accuracy. Thus, in this paper, the random forest classifier model was used to produce GWPM using water points as dependent features associated with historical data for hydrogeological conditions and field survey measurements, splitting them randomly into training 70% for training the model and testing 30% for model evaluation. Fifteen effective features that influence groundwater potentiality were created. After hyper-parameters had been optimized to reach acceptable performance

results, then the GWPM was created. Due to imbalanced classification and spatial distribution of dependent variables, many validation methods were used besides accuracy. The validated methods in the acceptable test stage include accuracy 97%, selectivity (recall) 92%, F1-score 94%, MCC 93%. Based on “variable importance” analysis extracted from RF and ROC, it was found that soil type and LULC were the most important features for GWPM considering that most of the water points are located within wadies, but not in the mountainous area. Lineament density and distance from major fractures features are highly important because secondary porosity builds the Precambrian aquifer occupying about 30% of the study area. In the light of groundwater management sustainability based on predicted GWPM and hydrogeological conditions, the middle and downstream of Wadi Abadi are suitable for future development if modern methods of irrigation are used. The Quaternary aquifer in the Esna area is suffering from significant drop in static water levels over the last 7 years that needs water management to prevent aquifer deterioration. Finally, this study proves that machine learning, especially the random forest algorithm, is useful for GWPM and can be applied to similar regions worldwide.



**Fig. 11** The most important effective features based on the AUC of ROC curve

#### Acknowledgements

The authors appreciate the efforts of translation and interpretation office in Montreal, Quebec, Canada, for English review of the manuscript. We are thankful to the Sugar Factory's lounge where we were welcomed in this field trip. We are thankful to Cairo University for logistics.

#### Author contributions

Conceptualization, HM, HH and AM; methodology, HM and AM; software, HM and AM; validation, HM, HH, TN and AM investigation, HM, HH and TN; resources, HM and AM; data curation, HM; writing—original draft preparation, HM; writing—review and editing, HH, HM and AM; visualization, HM and HH; supervision, AM, HH and TN. All authors agreed to the published version of the manuscript. All authors read and approved the final manuscript.

#### Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). Funds, or other support was received.

#### Declarations

#### Competing interests

The authors declare no competing interests.

Received: 11 November 2022 Accepted: 3 January 2023

Published online: 09 February 2023

#### References

- Adeyeye OA, Ikpokonte EA, Arabi SA (2019) GIS-based groundwater potential mapping within Dengi area, North Central Nigeria. *Egypt J Remote Sens Space Sci* 22:175–181. <https://doi.org/10.1016/j.ejrs.2018.04.003>
- Al-Fugara A, Ahmadlou M, Al-Shabeeb AR et al (2020a) Spatial mapping of groundwater springs potentiality using grid search-based and genetic algorithm-based support vector regression. *Geocarto Int* 37:284–303. <https://doi.org/10.1080/10106049.2020.1716396>
- Al-Fugara A, Pourghasemi HR, Al-Shabeeb AR et al (2020b) A comparison of machine learning models for the mapping of groundwater spring potential. *Environ Earth Sci* 79:1–19. <https://doi.org/10.1007/s12665-020-08944-1>
- Al-Ruzouq R, Shanableh A, Yilmaz AG et al (2019) Dam site suitability mapping and analysis using an integrated GIS and machine learning approach. *Water*. <https://doi.org/10.3390/w11091880>
- Al Saud M (2010) Mapping potential areas for groundwater storage in Wadi Aurnah Basin, western Arabian Peninsula, using remote sensing and geographic information system techniques. *Hydrogeol J* 18:1481–1495. <https://doi.org/10.1007/s10040-010-0598-9>
- Amarasinghe UA, Smakhtin V (2014) Global water demand projections: past, present and future. *IWMI Res Rep* 156:1–24. <https://doi.org/10.5337/2014.212>
- Andualem TG, Demeke GG (2019) Groundwater potential assessment using GIS and remote sensing: a case study of Guna tana landscape, upper blue Nile Basin, Ethiopia. *J Hydrol Reg Stud* 24:100610. <https://doi.org/10.1016/j.ejrh.2019.100610>
- Arulbalaji P, Padmalal D, Sreelash K (2019) GIS and AHP techniques based delineation of groundwater potential zones: a case study from

- southern Western Ghats, India. *Sci Rep* 9:1–17. <https://doi.org/10.1038/s41598-019-38567-x>
- Ashmore R, Calinescu R, Paterson C (2021) Assuring the machine learning lifecycle. *ACM Comput Surv.* <https://doi.org/10.1145/3453444>
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazards Earth Syst Sci* 13:2815–2831. <https://doi.org/10.5194/nhess-13-2815-2013>
- Chen W, Li Y, Tsangaratos P et al (2020) Groundwater spring potential mapping using artificial intelligence approach based on kernel logistic regression, random forest, and alternating decision tree models. *Appl Sci* 10:1–23. <https://doi.org/10.3390/app10020425>
- Chen Y, Chen W, Chandra Pal S et al (2021) Evaluation efficiency of hybrid deep learning algorithms with neural network decision tree and boosting methods for predicting groundwater potential. *Geocarto Int* 0:1–21. <https://doi.org/10.1080/10106049.2021.1920635>
- Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21:1–13. <https://doi.org/10.1186/s12864-019-6413-7>
- Conco C (1987) Geological map of Egypt, scale 1: 500,000
- Daoud JI (2018) Multicollinearity and regression analysis. In: *J. Phys. Conf. Ser.* vol. 949. <https://doi.org/10.1088/1742-6596/949/1/012009>
- Das S (2017) Delineation of groundwater potential zone in hard rock terrain in Gangajalghati block, Bankura district, India using remote sensing and GIS techniques. *Model Earth Syst Environ* 3:1589–1599. <https://doi.org/10.1007/s40808-017-0396-7>
- Díaz-Alcaide S, Martínez-Santos P (2019) Review: advances in groundwater potential mapping. *Hydrogeol J* 27:2307–2324. <https://doi.org/10.1007/s10040-019-02001-3>
- El Bilali A, Taleb A, Brouziyne Y (2021) Comparing four machine learning model performances in forecasting the alluvial aquifer level in a semi-arid region. *J Afr Earth Sci* 181:104244. <https://doi.org/10.1016/J.JAFREARSCI.2021.104244>
- Elewa HH, Qaddah AA (2011) Groundwater potentiality mapping in the Sinai Peninsula, Egypt, using remote sensing and GIS-watershed-based modeling. *Hydrogeol J* 19:613–628. <https://doi.org/10.1007/s10040-011-0703-8>
- Elmahdy S, Ali T, Mohamed M (2021) Regional mapping of groundwater potential in ar rub al khali, arabian peninsula using the classification and regression trees model. *Remote Sens.* <https://doi.org/10.3390/rs13122300>
- Ganapuram S, Kumar GTV, Krishna IVM et al (2009) Mapping of groundwater potential zones in the Musi basin using remote sensing data and GIS. *Adv Eng Softw* 40:506–518. <https://doi.org/10.1016/j.advengsoft.2008.10.001>
- Gokceoglu C, Sonmez H, Nefeslioglu HA et al (2005) The 17 March 2005 Kuzulu landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity. *Eng Geol* 81:65–83. <https://doi.org/10.1016/J.ENGCEO.2005.07.011>
- Golkarian A, Naghibi SA, Kalantar B, Pradhan B (2018) Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS. *Environ Monit Assess* 190:1–16. <https://doi.org/10.1007/S10661-018-6507-8/TABLES/5>
- Hammad FA, El Fakharany MA, Shabana AR, Saleh AA (2015) Hydrogeological studies on Esna-Idfu area, East Nile valley, Eastern Desert, Egypt. In: *First Int Conf Fac Sci Benha Univ Role Appl Sci Dev Soc Serv* 5–6 Sept 2015, 1–21
- Hung LQ, Batelana O, De Smedt F (2005) Lineament extraction and analysis, comparison of LANDSAT ETM and ASTER imagery. Case study: Suoimuoi tropical karst catchment, Vietnam. *Remote Sens Environ Monit GIS Appl Geol V* 5983:59830T. <https://doi.org/10.1117/12.627699>
- Hussien HM, Kehew AE, Aggour T et al (2017) An integrated approach for identification of potential aquifer zones in structurally controlled terrain: Wadi Qena basin, Egypt. *CATENA* 149:73–85. <https://doi.org/10.1016/j.catena.2016.08.032>
- Karpatne A, Ebert-Uphoff I, Ravela S et al (2019) Machine learning for the geosciences: challenges and opportunities. *IEEE Trans Knowl Data Eng* 31:1544–1554. <https://doi.org/10.1109/TKDE.2018.2861006>
- Kaur L, Rishi MS, Singh G, Nath Thakur S (2020) Groundwater potential assessment of an alluvial aquifer in Yamuna sub-basin (Panipat region) using remote sensing and GIS techniques in conjunction with analytical hierarchy process (AHP) and catastrophe theory (CT). *Ecol Indic* 110:105850. <https://doi.org/10.1016/j.ecolind.2019.105850>
- Kotsiantis SB, Zaharakis ID, Pintelas PE (2006) Machine learning: a review of classification and combining techniques. *Artif Intell Rev* 26:159–190. <https://doi.org/10.1007/s10462-007-9052-3>
- Lee S, Kim YS, Oh HJ (2012) Application of a weights-of-evidence method and GIS to regional groundwater productivity potential mapping. *J Environ Manag* 96:91–105. <https://doi.org/10.1016/J.JENVMAN.2011.09.016>
- Lee S, Hong SM, Jung HS (2017) GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. *Geocarto Int* 33:847–861. <https://doi.org/10.1080/10106049.2017.1303091>
- Madani A, Niyazi B (2015) Groundwater potential mapping using remote sensing techniques and weights of evidence GIS model: a case study from Wadi Yalamlam basin, Makkah Province, Western Saudi Arabia. *Environ Earth Sci* 74:5129–5142. <https://doi.org/10.1007/s12665-015-4524-2>
- Magowe M, Carr JR (1999) Groundwater-2005-Magowe—relationship between lineaments and ground water occurrence in western Botswana. pdf. *Groundwater* 37:282–286. <https://doi.org/10.1111/J.1745-6584.1999.TB00985.X>
- Martínez-Santos P, Renard P (2020) Mapping groundwater potential through an ensemble of big data methods. *Groundwater* 58:583–597. <https://doi.org/10.1111/GWAT.12939>
- Mohallel SA, Abdella HF, Habibah AZ (2019) Hydrogeochemical assessment of groundwater quality at Wadi Abbadi, southern part of eastern desert. *Egypt Curr Sci Int* 8:422–438
- Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrol Process* 5:3–30. <https://doi.org/10.1002/HYP.3360050103>
- Morgan H, Hussien HM, Madani A, Nassar T (2022) Delineating groundwater potential zones in hyper-arid regions using the applications of remote sensing and GIS modeling in the eastern desert, Egypt. *Sustainability* 14:16942. <https://doi.org/10.3390/SU142416942>
- Murmu P, Kumar M, Lal D et al (2019) Delineation of groundwater potential zones using geospatial techniques and analytical hierarchy process in Dumka district, Jharkhand, India. *Groundw Sustain Dev* 9:100239. <https://doi.org/10.1016/j.gsd.2019.100239>
- Naghbi SA, Moradi Dashtpajardi M (2017) Evaluation de quatre méthodes d'apprentissage supervisé pour la cartographie du potentiel des sources d'eaux souterraines dans la région de Khalhal (Iran) à partir des fonctionnalités d'un SIG. *Hydrogeol J* 25:169–189. <https://doi.org/10.1007/s10040-016-1466-z>
- Naghbi SA, Pourghasemi HR, Dixon B (2016) GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ Monit Assess* 188:1–27. <https://doi.org/10.1007/s10661-015-5049-6>
- Nampak H, Pradhan B, Manap MA (2014) Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *J Hydrol* 513:283–300. <https://doi.org/10.1016/j.jhydrol.2014.02.053>
- Nguyen PT, Ha DH, Jaafari A et al (2020) Groundwater potential mapping combining artificial neural network and real adaboost ensemble technique: the Daknong province case-study, Vietnam. *Int J Environ Res Public Health.* <https://doi.org/10.3390/ijerph17072473>
- Oh HJ, Kim YS, Choi JK et al (2011) GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *J Hydrol* 399:158–172
- Pal S, Kundu S, Mahato S (2020) Groundwater potential zones for sustainable management plans in a river basin of India and Bangladesh. *J Clean Prod* 257:120311. <https://doi.org/10.1016/j.jclepro.2020.120311>
- Palczewska A, Palczewski J, Robinson RM, Neagu D (2014) Interpreting random forest classification models using a feature contribution method. *Adv Intell Syst Comput* 263:193–218. [https://doi.org/10.1007/978-3-319-04717-1\\_9/FIGURES/12](https://doi.org/10.1007/978-3-319-04717-1_9/FIGURES/12)
- Park S, Hamm SY, Jeon HT, Kim J (2017) Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using R and GIS. *Sustainability.* <https://doi.org/10.3390/su9071157>
- Paryani S, Neshat A, Pourghasemi HR et al (2022) A novel hybrid of support vector regression and metaheuristic algorithms for groundwater spring potential mapping. *Sci Total Environ* 807:151055
- Patidar R, Pingale SM, Khare D (2021) An integration of geospatial and machine learning techniques for mapping groundwater potential: a case

- study of the Shipra river basin, India. *Arab J Geosci* 14:1–16. <https://doi.org/10.1007/s12517-021-07871-0>
- Patra S, Mishra P, Mahapatra SC (2018) Delineation of groundwater potential zone for sustainable development: a case study from Ganga Alluvial Plain covering Hooghly district of India using remote sensing, geographic information system and analytic hierarchy process. *J Clean Prod* 172:2485–2502. <https://doi.org/10.1016/j.jclepro.2017.11.161>
- Pourtaghi ZS, Pourghasemi HR (2014) GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran. *Hydrogeol J* 22:643–662. <https://doi.org/10.1007/S10040-013-1089-6/TABLES/6>
- Pradhan AMS, Kim YT, Shrestha S et al (2020) Application of deep neural network to capture groundwater potential zone in mountainous terrain, Nepal Himalaya. *Environ Sci Pollut Res* 28:18501–18517. <https://doi.org/10.1007/s11356-020-10646-x>
- Prasad P, Loveson VJ, Kotha M, Yadav R (2020) Application of machine learning techniques in groundwater potential mapping along the west coast of India. *Geoscience Remote Sens* 00:735–752. <https://doi.org/10.1080/15481603.2020.1794104>
- Rahmati O, Pourghasemi HR, Melesse AM (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. <https://doi.org/10.1016/j.catena.2015.10.010>
- Ramachandra M, Babu KR, Kumar BP, Rajasekhar M (2022) Deciphering groundwater potential zones using AHP and geospatial modelling approaches: a case study from YSR district, Andhra Pradesh, India. *Int J Energy Water Resour*. <https://doi.org/10.1007/s42108-021-00169-7>
- Riley SJ, DeGloria SD, Elliot R (1999) *Terrain\_Ruggedness\_Index.pdf*. *Intermt J Sci* 5:23–27
- Rizeei HM, Pradhan B, Saharkhiz MA, Lee S (2019) Groundwater aquifer potential modeling using an ensemble multi-adoptive boosting logistic regression technique. *J Hydrol* 579:124172. <https://doi.org/10.1016/j.jhydrol.2019.124172>
- Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45:427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Sultan M, Chamberlain KR, Bowring SA et al (1990) Geochronologic and isotopic evidence for involvement of pre-Pan-African crust in the Nubian shield, Egypt. *Geology* 18:761–764. [https://doi.org/10.1130/0091-7613\(1990\)018%3C0761:GAEFI%3E2.3.CO;2%0A](https://doi.org/10.1130/0091-7613(1990)018%3C0761:GAEFI%3E2.3.CO;2%0A)
- Tahmassebi N, Rahmati O, Noormohamadi F, Lee S (2016) Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. *Arab J Geosci* 9:1–18. <https://doi.org/10.1007/S12517-015-2166-Z/TABLES/3>
- Todd DK, Mays LW (2005) *Groundwater hydrology*, 3rd edn. Wiley
- Víctor GE, Marie-Louise V, Elisa D et al (2021) Delineation of groundwater potential zones by means of ensemble tree supervised classification methods in the Eastern Lake Chad basin. *Geocarto Int* 0:1–28. <https://doi.org/10.1080/10106049.2021.2007298>
- Wiesmeier M, Barthold F, Blank B, Kögel-Knabner I (2011) Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil* 340:7–24. <https://doi.org/10.1007/s11104-010-0425-z>
- Xu H, Wang D, Ding Z et al (2020) Application of convolutional neural network in predicting groundwater potential using remote sensing: a case study in southeastern Liaoning, China. *Arab J Geosci* 13:1–12. <https://doi.org/10.1007/s12517-020-05585-3>
- Yidana SM, Dzikunoo EA, Aliou AS et al (2020) The geological and hydrogeological framework of the Panabako, Kodjari, and Bimbilla formations of the Voltaian supergroup—revelations from groundwater hydrochemical data. *Appl Geochem* 115:104533. <https://doi.org/10.1016/j.apgeochem.2020.104533>

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)