Review

Review of machine learning algorithms used in groundwater availability studies in Africa: analysis of geological and climate input variables

Haoulata Touré[1] · Cyril D. Boateng[2,7] · Solomon S. R. Gidigasu[1] · David D. Wemegah[2] · Vera Mensah[1] · Jeffrey N. A. Aryee[3] · Marian A. Osei[3,4,5] · Jesse Gilbert[3] · Samuel K. Afful[6]

Abstract

Groundwater is crucial for Africa's potable water supply, agriculture, and economic development. However, the continent faces challenges with groundwater scarcity due to factors like population growth, climate change, and over-exploitation. Over the past ten years, machine learning has been increasingly and successfully used in groundwater availability studies across the world. This review paper explores the application of machine learning techniques in groundwater availability studies including groundwater level prediction and groundwater potential mapping studies by focusing on some of the studies conducted in Africa. The methodology involved downloading relevant papers, identifying and categorizing the machine learning algorithms employed, and quantifying their use. Geological and climatic variables were also identified, analyzed, and categorized to measure their usage frequency. The different algorithms and input variables extracted from each paper are graphically represented in this document highlighting the most employed ones. The findings suggest that more research needs to be conducted on the use of machine learning algorithms on this topic in Africa. In the reviewed papers Fuzzy-based algorithms are commonly used. The groundwater level prediction studies primarily focus on input variables related to hydrology/hydrogeology, while for potential mapping, geological aspects are the most investigated variables. In terms of climate, precipitation receives the most attention in the reviewed studies. The study highlights the potential of machine learning in improving water resource management and decision-making in the region.

Keywords Groundwater level prediction · Groundwater potential mapping · Machine learning · Africa

1 Introduction

Groundwater, which is the water occupying all the voids within a geological stratum is the main source of potable water for the majority of Africans [1]. According to the United Nations of Environment Programme (UNEP), 75% of Africans, mainly in Northern and Southern Africa, rely on groundwater as their primary drinking water supply [2]. It has been estimated that around 2 billion of the global population relies on groundwater as their primary water source [3]. This includes

✉ Cyril D. Boateng, cyrilboat@knust.edu.gh | ¹Department of Geological Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. ²Department of Physics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. ³Department of Meteorology and Climate Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. ⁴Centre for Ecology and Hydrology, Crowmarsh Gifford, Oxfordshire, UK. ⁵School of Earth and Environment, University of Leeds, Leeds, UK. ⁶Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. ⁷Caburu Company Ltd, Madina, P.O.BOX MD 2046, Accra, Ghana.

drinking water as well as various domestic and industrial purposes. The global population is increasing at a rate of 80 million people each year. This necessitates the identification of methods to augment the global water supply by an estimated 64 billion cubic meters annually [4]. Due to the increase in population growth and urbanization, the dependence and demand for groundwater are anticipated to grow significantly within the next decade [5]. This situation is expected to have a great impact on water availability and accessibility in the future in view of climate change challenges in Africa.

The amount of water that infiltrates and percolates into the underground water source during rainfall is greatly influenced by the type of land cover and other properties of the soil. Urbanization, often associated with major construction activities in Africa leads to changes in land cover. This can affect how water interacts with the land surface, by increasing impervious surfaces which prevents water from infiltrating into the ground, causing variations in the rate of water infiltration and leading to an increase in runoff [6]. Vörösmarty et al. [5] are of the view that changing land use practices with the potential of increasing soil impermeability also has the potential of reducing groundwater recharge resulting in a reduction in groundwater resources. According to [7] increasing urban areas may lead to a decrease in recharge in certain regions due to the increase in impervious surfaces created by developed infrastructures and soil compaction.

Furthermore, climate change and climate variability add another layer of complexity to this issue. Possible changes in precipitation patterns, leading to more intense rainfall events or prolonged droughts can impact groundwater availability [5]. Increased rainfall intensity can cause rapid runoff, preventing water from seeping into the ground. On the other hand, prolonged droughts can lower the water table and reduce the overall availability of groundwater. Water security in Africa can therefore be further compounded by climatic challenges and urbanization.

Figure 1, shows the ranking of African countries based on their water security. Out of 54 countries, only 13 (less than 25%) reached a modest level of water security in recent years, and around one-third are considered to have levels of water security below the threshold of 45. Water security is defined by the UNEP as "the capacity of a population to safeguard sustainable access to adequate quantities of acceptable quality water for sustaining livelihoods, human well-being, and socioeconomic development, for ensuring protection against waterborne pollution and water-related disasters, and for preserving ecosystems in a climate of peace and political stability" [8]. This suggests that there is insufficient access to safe and reliable water sources, potentially leading to challenges in meeting basic water needs and increased vulnerability to water-related risks in Africa. To help achieve the Sustainable Development Goals (SDGs), especially SDG 6, which seeks to ensure the availability and sustainable management of water and sanitation for all, there is the utmost need to monitor and quantify groundwater resources, facilitating groundwater level (GWL) modeling and predictive capabilities crucial for informed decision-making. This can be achieved by the adoption of dynamic and efficient methods in assessing the need and availability of groundwater resources especially in areas with high water scarcity and deteriorating quality [9].

Over the past ten years, machine learning (ML) has been increasingly and successfully used in groundwater availability studies across the world [10]. This review uses groundwater levels and groundwater potential maps as proxies for the measurement of groundwater availability. ML algorithms can analyze large datasets and identify complex patterns to accurately predict or map areas with high groundwater potential (GWP). The method has shown efficiency in various studies because of its capability to incorporate various environmental variables and factors that affect groundwater availability, such as topography, soil characteristics, and rainfall patterns. For instance, Sarkar et al., [11] employed a variety of input parameters including geological and climatic variables for groundwater mapping. To predict GWL, Gonzalez et al. [12], incorporated hydrology-related variables with geological and climates variables. These variable's importance for groundwater mapping has been highlighted in several studies [11–13].

It is crucial to accurately investigate groundwater availability for effective groundwater resource management. Monitoring groundwater provides valuable insights into short- and long-term variations in groundwater availability [10].

Fig. 1 Water security score of African countries [9]
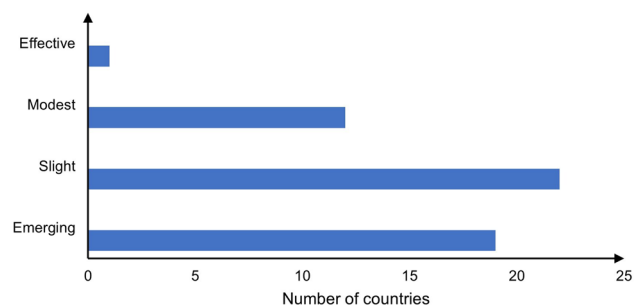
The influence of specific variables on groundwater availability in Africa is a topic of significant interest and importance. Understanding the relationship between these variables can help inform groundwater management and policy on the African continent [14, 15]. The aim of this review paper is twofold: firstly, to examine some of the current literature concerning groundwater availability in several African countries, with a specific emphasis on research utilizing ML methods for data analysis in groundwater level (GWL) prediction studies and groundwater potential mapping (GWP) studies. Secondly, to compile a comprehensive inventory of the diverse machine learning algorithms, as well as the climatic and geological variables employed by the researchers across Africa.

The rest of the paper is structured as follows, we present a methodology for our review, discuss ML case studies for groundwater availability in the section following the methodology and locate the various African studies that have used machine learning algorithms. We continue by discussing the different input variables (geological and climatic variables) used in the algorithms in the following section.

2 Methodology

In this study, a thorough analysis of some relevant African studies was conducted by downloading papers that focused on the topic. To gather the most relevant and recent data for this review, a comprehensive search was conducted using a combination of key terms including 'Groundwater Level Prediction', 'Groundwater Potential Mapping', 'Groundwater Level Prediction in Africa' 'Variables affecting Groundwater Level Prediction in Africa', 'Machine Learning Algorithms used to map Groundwater Potential' and other key terms related to the discussed topic. To narrow down the geographical focus, the term 'Africa' was added to most of the key terms used during the search step. This allowed us to capture papers that address the continent in their studies. The search was refined to include only peer-reviewed articles written in English. A variety of scholarly databases were utilized for this search, such as Google Scholar, and ScienceDirect. The geographical focus was narrowed down to Africa to align with the scope of this review.

Figure 2 shows the schematic diagram of the methodology used in this study. The different algorithms employed were identified and the most commonly used algorithms to their respective families (Table 1). The frequency of utilization for each algorithm was quantified to determine the number of times they were employed in studies on the African continent. This analysis provides valuable insights into the prevalence and popularity of specific algorithms within the research landscape. Additionally, a comprehensive analysis was carried out to determine the prevailing geological and climatic variables used as input variables.

By categorizing the different input variables into distinct groups, their frequency of usage was quantified. This quantitative assessment allowed for a visual illustration of the distribution and prominence of these variables in African research.

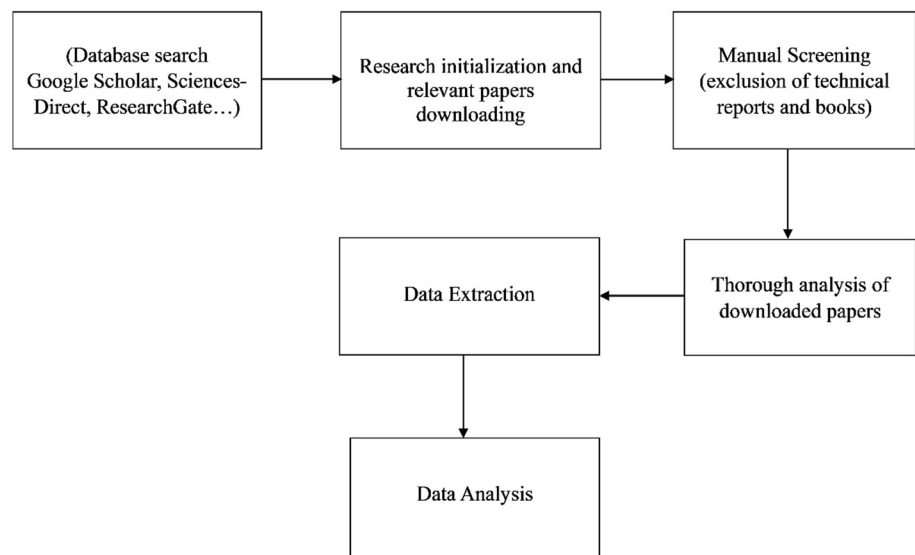Fig. 2 Schematic workflow for the study

Table 1 Different algorithms utilized in machine learning-based groundwater studies in Africa

Algorithm family	Utilized algorithms	References
Neural networks	Feedforward Neural Network with Multilayer Perceptron (FNN-MLP);	[16]
		[17]
	Feedforward Neural Network with Extreme Learning Machine (FNN-ELM);	[18]
		[19]
	Long Short-Term Memory Neural Network (LSTM-NN);	[20]
	Artificial neural network (ANN);	[21]
	Neural Network Autoregression Model (NNAR);	[22]
	Artificial Neural Network with Perceptrons Multilayers (ANN-PMC);	[23]
		[24]
	Discrete Wavelet Transform—Artificial Neural Network—Perceptrons Multilayers Combination (DWT-ANN-PMC);	[25]
		[26]
	Multiscale Feed Forward Neural Network	
	Deep Belief Network (DBN);	
	Input delay neural network (IDNN);	
	Recurrent Neural Network (RNN);	
	Generalized Radial Basis Function (GRBF)	
	Probabilistic Neural Network (PNN)	
Tree-based family	Decision Tree (DT);	[19]
	AdaBoost; Gradient Boosting Decision Tree (GBDT);	[22]
	Extra-tree;	[27]
	Random Forest Regression (RFR);	[28]
	Extreme Gradient Boosting (XGB);	[25]
	Gradient Boosting Tree (GBT);	[26]
	K-Nearest Neighbor-Random Forest (KNN-RF)	[12]
		[29]
Regression Algorithms	Nonlinear AutoRegressive (NARX);	[30]
	Deep AutoRegressive Models;	[21]
	Logistic Regression (LR);	[22]
	Multi Linear Regression (MLR);	[28]
	Support Vector Regression (SVR);	[25]
	Support Vector Machines (SVM);	[26]
	Multivariate Adaptive Regression Splines (MARS)	
Fuzzy-based Model	Fuzzy Logic (FL)	[31]
	Analytical Hierarchy Process (AHP)	[32]
	Fuzzy-Analytical Hierarchy Process (F-AHP)	[33]
		[34]
		[35]
		[36]
		[37]
		[38]
		[39]
		[40]
		[41]
		[42]
		[43]
		[44]
		[45]
		[46]

It has been observed that the researchers employed various terms to describe the same variables. To ensure clarity and facilitate analysis, it was necessary to emphasize and categorize these terms for better analysis of the variables.

3 Results and discussion

The majority of the studies analyzed in this paper are concentrated in the southern part of the continent. Table 1 provides an overview of the various studies conducted, which will be discussed. It showcases the different algorithms employed during these studies and categorizes them based on their respective family of algorithms.

These findings suggest that fuzzy-based algorithms, neural network algorithms, tree-based algorithms, and regression algorithms have been widely employed by various searchers in Africa. Their efficiency, already proven worldwide [10], justifies their selection by African researchers. These studies have ultimately confirmed the reliability of these algorithms. In the section below, a brief description of the case studies is provided. Following that, the geographical distribution of these case studies is presented.

3.1 Case studies

Due to their potential for being less time-consuming and their capacity to produce relevant findings, ML models are interestingly becoming alternatives to process-based models [22]. A large volume of literature is available on the applicability of machine learning algorithms to predict GWLs in different regions of the world [10]. This paper focuses on studies from Africa, where comprehensive details of the identified algorithms are presented (Table 1). Figure 3 presents the locations of the studies utilizing ML algorithms in the analyzed African papers in Table 1, as gathered from our comprehensive review. This section is divided into two subsections. The first part focuses on GWL prediction studies, while the second part describes GWP mapping studies.

3.1.1 Groundwater level prediction studies

In the Gondo aquifer in Burkina Faso, a study was carried out to predict GWL using ML algorithms [24]. The study aimed to identify ANN models that can effectively capture the complex dynamics of significant GWL variations using a relatively short length of data. Four different ANN models, including IDNN, RNN, GRBF, and PNN, were employed in the study. The input variables used were precipitation, water table depth (WTD), and temperature. The authors noted that these models have the advantage of providing accurate predictions even with limited GWL data, which is particularly beneficial in countries with limited monitoring capabilities. The findings suggested that predicting larger GWL variations with insufficient data can be more challenging. However, the RNN algorithm outperformed the other ANN models in predicting these variations, indicating its potential for predicting potential water shortages in the next 3 months. On average, IDNN and PNN demonstrated similar prediction performance.

In Ijebu-Jesa, southwestern Nigeria, the use of ANN was explored to predict GWL based on geoelectric parameters [17]. The piezometric head of all the accessible wells in the study area was obtained and GWL was derived from them. This in turn was used as the output parameter for the ANN model. Geoelectrical parameters were obtained through a geophysical investigation including vertical electrical soundings (VES). These parameters included aquifer resistivity (AQR), aquifer thickness (AQT), overburden resistivity (OR), overburden thickness (OT), and coefficient of anisotropy (COA). They were used as input variables for the predictive model. The authors aimed to leverage the nonlinear modeling capabilities of ANN for accurate GWL predictions. The validation results of the ANN model demonstrated its effectiveness in predicting GWL, with a mean square error (MSE) of 0.0014286 and a regression coefficient (R) of 0.98731. These findings indicate that the model produces reliable and accurate predictions. The success of the ANN technique in this study suggests its potential applicability in similar geological areas. The authors were able to achieve their objectives and found that the results were satisfactory and supported the conclusion that ANN is an effective tool for predicting GWLs, aiding in the planning and management of groundwater resources.

In another study conducted in West Africa precisely Ghana, the authors also used 2 algorithms from the ANN family namely the FNN-MLP and FNN-ELM, to predict GWR in a data-scarce region [16]. The models were trained using the input variables of effective rainfall, potential evapotranspiration (PET), and lagged GWR. The study found that the FNN-MLP model outperformed the FNN-ELM model in predicting GWR with a coefficient of determination (R^2) ranging from 0.97 to 0.99.

To predict GWL based on hydrogeological variables such as rainfall, evapotranspiration, and initial water table level (WTL), ANNs have been utilized by [18] in Nebhana aquifers (North-East Tunisia). The ANN architecture was composed of three layers: an input layer, which had some neurons equal to the number of input variables; a hidden layer, which
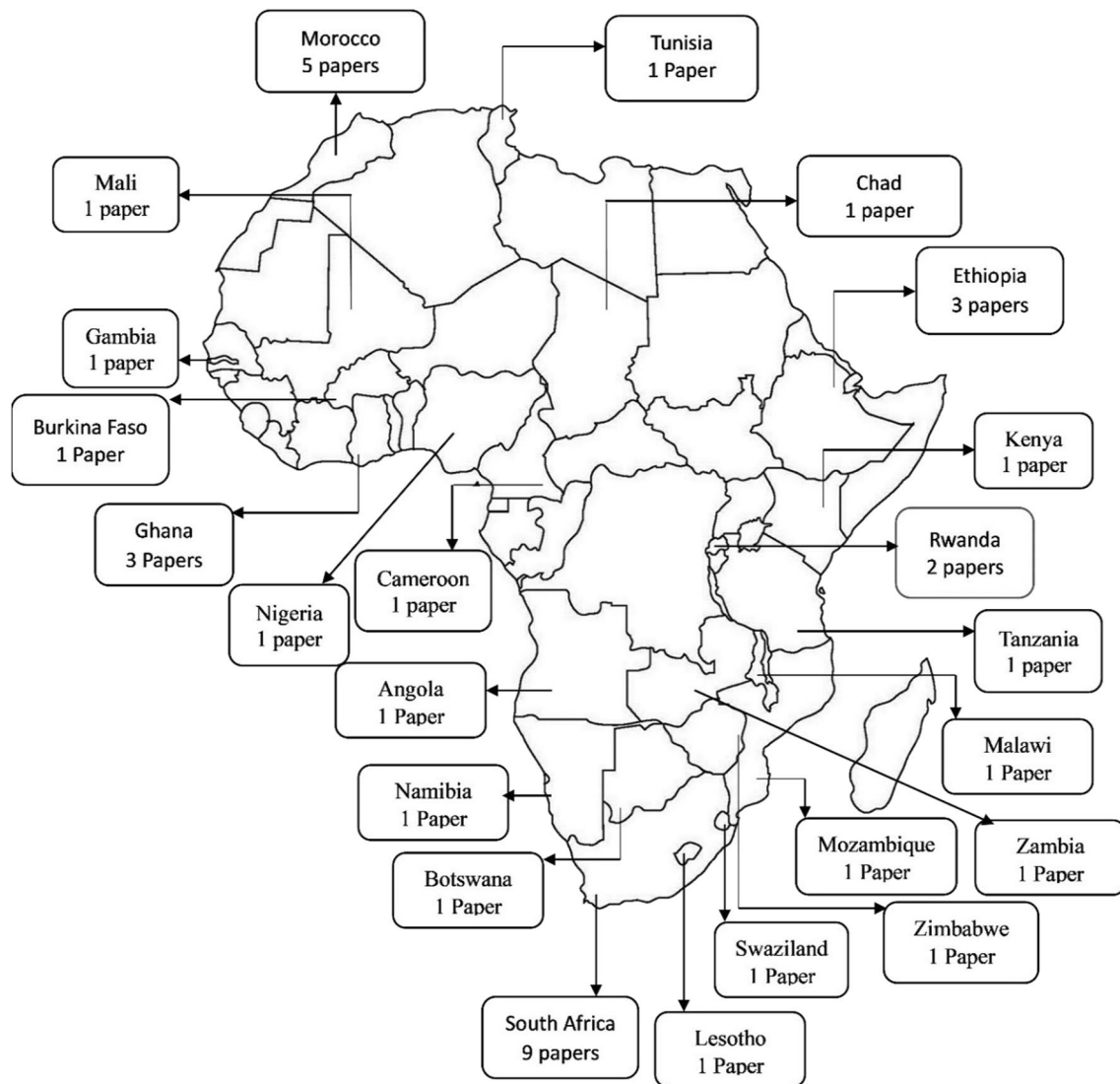
Fig. 3 Locations of studies on ML in groundwater availability studies across Africa

contains three neurons; and an output layer with one neuron. The performance of the designed ANN was evaluated using various metrics such as relative error, root mean square error, determination coefficient, and Nash–Sutcliffe efficiency coefficient. The study finally showed that the ANN algorithm has the potential to provide relatively good results with fewer input variables.

Gaffoor et al. [19] employed two ML algorithms, GBDT and LSTM-NN, to predict GWL variations in the Shire Valley Alluvial Aquifer (Southern South Africa). The algorithms were trained using hydro-climatic inputs and GWL changes from two boreholes (namely Ngabu and Nsanje). The authors set up experiments to train and test the algorithms to predict the change in the current month's GWL and the change in the following month's GWL. The algorithms were compared based on their R^2 scores, and the authors concluded that the LSTM outperforms the GBDT model, especially regarding slightly greater time series and extreme GWL changes.

In Aderemi et al. [30]'s research at the Karst belt in South Africa, they forecasted GWL using Regression Models such as SVM or LR, Deep Auto-Regressive models, and Nonlinear Autoregressive Neural Networks with External Input (NARX). These models were trained using four input variables, namely rainfall, temperature, groundwater usage, and precipitation. The findings showed that NARX and SVM have higher performance metrics and accuracy compared to the other models.

Kalu et al., [23] developed a ML modeling framework based on DBN to predict variation in monthly GWLs at 1–5-month time scales for 27 groundwater wells over the southern Africa region including Angola, Zambia, Malawi, Namibia, Botswana, Zimbabwe, Mozambique, South Africa, Lesotho, Swaziland. The predictor dataset used in the DBN network was

constituted from hydrological parameters, GWL estimates, and global climate indices. The DBN network was trained on the predictor dataset to forecast changes in GWLs up to 5 months lead times at most locations in the study region. The results highlighted how deep learning can help make informed decisions to lessen the effects of climate extremes on people and their properties. The authors found it to be a key tool for evaluating hydrological processes that could lead to extreme weather.

Gibson, [20] used the NNAR method to predict GWLs in the Steenkoppies compartment of the Gauteng and North West Dolomite Aquifer in South Africa. The input variables rainfall, temperature, groundwater usage, and spring discharge from the Maloney's Eye spring were used to train the model to learn the complex, interdependent relationships occurring in the groundwater system. The results indicate that the NNAR model was most accurate in predicting GWLs when the test data closely mirrored the training data. This is explained by the fact that ANNs, like NNAR, learn from the patterns in the training data. So, if the test data is too different, the model might struggle with predictions.

Ibrahimi et al. [21] adopted a more comprehensive approach, utilizing three distinct models for GWL prediction of the surface water table in the Saïss Plain (North of Morocco). They incorporated input variables such as precipitation, temperature, and average GWLs into their analyses. The first model was the ANN-PMC. This model uses the input variables to train the neural network, which then predicts the GWL based on the input data. The second model was DWT-ANN-PMC. This model used the discrete wavelet transform to extract features from the input data, which were then used to train the neural network. The trained neural network then predicted the GWL based on the input data. The third model was MLR. This model used the input variables to create a linear equation that predicted the GWL based on the input data. The performance of the three models was evaluated using statistical metrics such as mean absolute error (MAE), RMSE, and R2. The results indicate that the DWT-ANN-PMC model outperforms the other two models in predicting GWLs.

To also predict GWL in the Grootfontein Aquifer (South Africa) Kanyama et al. [20] employed five different data-driven techniques, including SVR, GBDT, DT, RFR, and FFNN algorithms. The chosen input variables were discharge, precipitation, and temperature. These variables were considered as model inputs for the four boreholes in the aquifer. The performance of the models was evaluated using two metrics: RMSE and R^2. These metrics were used to assess the accuracy and fit of the models. The GB algorithm performed the best among the five algorithms tested, achieving an R^2 score of up to 0.75 for one of the borehole sites. The FFNN algorithm also performed well, achieving the highest R^2 score of 0.77 for one of the sites. The results obtained suggested that the model performance is data-dependent and the variable responsible for that dependence was found to be the discharge rate.

In Rwanda a regression algorithm namely KNN-RF was employed to predict the GWL of a permeable aquifer [27]. In this study, the performance and capacity of the ensemble KNN-RF regression approach for predicting seasonal GWL in a fractured aquifer with limited data were examined. GWL data and important meteorological factors such as solar radiation, temperature, and precipitation from Mukarange in eastern Rwanda for our analysis were used as input variables. The experimental analysis revealed that the KNN-RF ensemble approach demonstrated stability, enhanced generalization competence, and improved prediction accuracy. Additionally, the KNN-RF model effectively captured the time-based changes in groundwater table depths. The authors concluded that incorporating solar radiation as a substitute for evapotranspiration led to further improvements in prediction accuracy. They also suggested that the KNN-RF model is well-suited for forecasting seasonal variations in groundwater depths even with limited samples.

In another study conducted in South Africa precisely in the Upper Crocodile Sub-Basin, GBR, and SVR algorithms were employed to predict GWL [28]. The study aimed to demonstrate that monthly GWLs can be predicted by antecedent GWLs and rainfall data. The authors chose these input variables based on the results of correlation analysis. Lag times were determined using cross-correlations for rainfall and autocorrelations for GWLs. The algorithms were trained using data from January 2011 to April 2018 and validated using data from May 2018 to September 2020. The GB algorithm showed better performance for the predicted and observed GWLs during both the calibration and validation periods of the algorithms and the authors concluded that GB can be used to predict future GWLs based on projected rainfall and previous GWLs.

The approach utilized by [25] consisted of employing as input variables 174 groundwater satellite images in five machine-learning algorithms to predict full groundwater images in the southern part of the African continent. These algorithms are XGB, MLR, RF, MLP, and SVR suggesting that the study is employing ANN, regression, and Tree-based algorithms. The initial 149 groundwater satellite images served to train the algorithms, while the remaining ones were reserved for validation purposes. The performance of the algorithms was assessed using RMSE and MAE and SVR outperformed the other algorithms. The research findings indicate that using suitable machine-learning techniques can lead to much more precise predictions.

Several ML algorithms, including MLR, MARS, ANN, RFR, and GBR, were used to predict the depth of the water table in the Bilate watershed located in southern Ethiopia [26]. The variables used in this study are static water level, elevation, soil type, and climate variables (i.e., precipitation, specific humidity, wind speed, land surface temperature (LST), and NDVI. The findings suggested that GBR consistently outperformed other models. On average, it achieved an impressive R^2 value of 0.77 and an MAE of 19 m across multiple experiments. The final output which is a map of predicted water levels was created using the best-performing algorithm. The authors demonstrated in their study that using a combination, of methods allows for a more robust and comprehensive understanding of the investigated parameter. The different ML algorithms have different strengths and weaknesses, and by testing a variety of classifiers, researchers can determine which ones perform the best for their specific dataset and research objectives. By using a diverse set of classifiers, the studies aimed to ensure robustness and accuracy in predicting groundwater potential (GWP). This approach helps to minimize the potential bias and limitations associated with relying on a single classifier, providing a more robust and comprehensive analysis.

3.1.2 Groundwater potential mapping studies

To map the GWP of the Middle Atlas plateaus, Morocco, an approach combining FL, GIS, and RS was employed [31]. To achieve this, thematic layers which will serve as inputs in the algorithms, were created using topographic maps, thematic maps, field data, and satellite images. These layers, including lithology, slope, karst degrees, land cover, lineament, and drainage density, were then prepared, classified, weighted, and integrated into a geographic information system (GIS) environment. The FL approach was employed to assign fuzzy membership values to the different thematic layers based on their classification and their significance in groundwater. The research findings indicate that combining thematic layers provides valuable information to local authorities and planners regarding suitable areas for groundwater exploration in the Middle Atlas plateaus of Morocco. This approach proves to be a straightforward and effective tool for addressing water resource-related topics.

Benjmel et al. [33] investigated the GWP of the Ighrem region in Morocco using the approach combining geospatial techniques of GIS and RS. This study instead employed AHP for the weightages of the thematic layers based on their importance for the occurrence of groundwater in the region. The variables used in this study are fault density (FD), drainage density (DD), distance from drainage (DFD), lineaments density (LD), distance from lineaments (DFL), lithology, slope, terrestrial water index (TWI), plan curvature, and profile curvature and node density (DN). The GWP map was created by the superposition of these layers in the GIS environment. Borehole yield was superimposed on the output GWP map for a validation purpose. In the research paper, the authors observed that the success rate of implanted drilling gradually improved in the most productive areas, indicating the effectiveness of the approach used. According to their findings, the authors concluded that the presented approach serves as a valuable tool for making informed decisions in hydrogeological prospecting.

Still in the southwest of Morocco, another most recent study employed a similar approach for GWP mapping [38]. The authors explored the process of determining GWP in the Akka basin by combining geospatial techniques and geological data. Two multi-criteria approaches, the geometric average and expected value, were used to integrate various factors that influence groundwater presence. The weights for each factor were assigned using the FL method, which transforms factor values into a range of 0 to 1. The input layers considered in the study include LD, DD, distance from rivers (DFR), DFL, permeability, slope, TWI, plan curvature, and profile curvature. Using this information, a GWP map was generated in a GIS environment. To determine the most efficient model, well data within the basin were used for assessment and comparison. The fact that the high-flow wells align perfectly with the high potential values in the GWP map for the expected value model further validates its reliability. The expected value model, with a value of 1.86 significantly outperformed the geometric average model, which had a value of 0.96. The authors concluded by the fact this comparison indicates that the expected value model is the superior choice for identifying target areas with high GWP.

In this study conducted in the transboundary watershed of the Chott-El-Gharbi (Algerian–Moroccan border), the authors aimed to identify the GWP zones by combining RS and GIS methods [43]. Eight thematic layers, including geology, rainfall, WTL, LD, slope, DD, elevation, and LULC were used as input variables. To assess the GWP, the AHP technique was employed, which allowed to map the potentiality effectively. This technique involved assigning ranks and weights to each factor based on their relative importance in terms of GWP. The output map was classified into five classes from Excellent for good potential areas to very poor. The map was then validated using existing borehole data. The accuracy of the map with the boreholes demonstrates the method's reliability. This is confirmed by the success rate of implanted

drilling, where the outcome provides that 72.41% of groundwater inventory data agree with the corresponding GWP zone classifications. This in turn proves the use of AHP is a reliable technique for GWP mapping.

Gómez-Escalonilla et al. [13] used a total of 20 machine learning classifiers, trained and tested them on a large borehole database to find a meaningful correlation between the presence or absence of groundwater and the explanatory variables in the Koulikoro and Bamako regions (Mali). The performance of the classifiers was assessed using various machine learning metrics, including accuracy, F1 score, and area under the curve (AUC).

The same method has been applied in a region of eastern Chad [29]. In this case, the performance of the classifiers was evaluated using metrics such as AUC curve, test scores, and balanced scores. The most relevant explanatory variables were identified based on the performance of the classifiers using these metrics. In the two cases, the best-performing algorithms in identifying potential groundwater areas correlating with borehole data were found to be tree-based algorithms, including decision tree, random forest, AdaBoost classifier, gradient boosting, and extra trees [13], random forest and extra-trees [29].

This study conducted by [42] focused on delineating potential groundwater zones in the Voltaian basin (Ghana) as an alternative source of potable water. Various datasets, including geological, geophysical, hydrological, and topographical data, were integrated using GIS to generate a GWP map. The fuzzy AHP was used to assign weights to the evidential layers based on their relevance. The GPM's effectiveness was evaluated using tests such as the Nash–Sutcliffe efficiency (NSE) and the Index of Agreement. The NSE test resulted in a value of 0.9996, while the Index of Agreement yielded a value of 0.9999. These findings demonstrate that the fuzzy AHP-based model is highly effective and reliable for groundwater mapping in the Voltaian basin.

The central region of Ghana is generally subject to unsuccess rate of drilling and water quality issues, particularly along the coast. To address this issue, [37] has decided to map the GWP zones of the region for various domestic purposes, a selection of hydrogeological and hydro chemical variables were used with geospatial techniques namely RS and GIS. The variables consisted of soil type, geology, rainfall, LULC, LD, and hydrochemistry data. The weight of each variable was determined through the Fuzzy AHP method and overlaid in the GIS software to generate the GWP map. The output map was validated using a set of boreholes yielded by the area under the curve (ROC), which was calculated to be 0.869, showing the effectiveness of the produced map. The authors of this study have concluded that their research plays a crucial role in enhancing water security and promoting sustainable development and management of groundwater resources in the Central Region of Ghana.

To make it easier to explore and use groundwater resources, conducted an analysis focused on identifying GWP zones in the Umuahia areas of the Niger Delta Basin in Nigeria [45]. By considering factors like rainfall, soil type, geology, DD, LD, slope, and LULC, seven input layers were generated using RS data and a GIS environment. AHP was used to prioritize these layers and integrate them into a single thematic layer using the weighted overlay tool in ArcGIS. The final map was obtained showing the potential areas where groundwater can be found. This indicates the influence of each factor on groundwater occurrence. These findings highlight the issue of random borehole failures in some parts of the Umuahia area. The detailed map of GWP zones produced by this study can be used for the effective management of aquifers to meet the region's water needs sustainably.

In the very western part of West Africa, most precisely in the Greater Banjul Area (GBA) Gambia, the same approach as the previous studies discussed which is the use of RS and GIS was employed to map groundwater zones [32]. This study used seven variables as input (geology, LULC, S, DD, soil, groundwater fluctuation, and aquifer transmissivity). To normalize the weights of these variables AHP technique was utilized. The authors were able to delineate groundwater zones using this method and found their results to be accurate when comparing the final output with each of the variables.

The Adamawa region located in the heart of central Africa (Cameroon), was subject to an investigation of its GWP to overcome the failure of the region's water system [41]. The main goal of this study was to elaborate a comprehensive map of the spatial distribution of GWP in the region using the fuzzy algebraic model. To achieve this, the study combined six important variables: aquifer depth, resistivity, thickness, transverse resistivity, transmissivity, and hydraulic conductivity. A total of fifty vertical electrical soundings (VES) points were interpreted. The authors following this approach, and using the fuzzy model to overlay the different variables, were able to demarcate the potential zones of the area. The findings suggested that the use of the fuzzy algebraic model is effective in mapping groundwater zones, but the method needs to be combined with prior geophysical and hydrological surveys for the characterization of hydro-parameters to be processed.

N'gwijabagabo et al., [44]' study focuses on analyzing the GWP recharge in Rwanda's Eastern Province using GIS technology. Several factors influence these zones, such as geology, rainfall, S, soil, LULC, and normalized derived vegetation index (NDVI). By assigning weights to each layer using the AHP, the study generated a map of potential recharge zones.

The findings revealed that a significant portion of the area had good to moderate groundwater potential. The accuracy of the map was validated using borehole yield data, providing valuable insights for sustainable groundwater management in the region.

Moving toward the part eastern of the continent, (Melese et al., 2022) studied the GWP of the Muga watershed, Abay basin in Ethiopia. The main goal of the study was to use geospatial techniques, and the AHP method, to assess and define the GWR zones of the area. Input layers, including geology, rainfall, S, soil, curvature, topography wetness index (TWI), elevation, DD, LULC, and LD, were prepared using RS satellite images and corresponding data. These layers were integrated using a multicriteria evaluation technique, with each parameter being ranked through weighted overlay index analysis (WOIA). The AHP technique was employed to assign weights to each thematic layer. The reliability of the results was assessed by calculating the consistency index and consistency ratio, which were found to be reasonably acceptable. To validate the findings, groundwater well locations were considered in the validation datasets. The prediction accuracy was explored using the ROC curve and the area under the curve, which yielded a value of 82.9%.

In the study conducted by Haile, [36] different methods were explored to map GWP in the Guder watershed, Ethiopia. Geospatial and FL techniques were employed on various thematic layers derived from factors that significantly influence groundwater occurrence. These factors include DD, S, rainfall, geology, LULC, LD, soil, and geomorphology. By assigning membership values to these factors based on expert opinions, previous studies, and research, the authors determined the recharge potential for different areas. A GWP map indicating areas with very high to low groundwater potential was then generated and validated by comparing them with each of the contributing input variables maps and found strong associations. These findings can assist decision-makers and policymakers in implementing suitable groundwater recharge strategies. FL algorithm, when combined with GIS, has proven to be a valuable tool for estimating groundwater recharge locations stated the authors.

In the arid areas of the Ewaso Ng'iro—Lagh Dera basin in Kenya, there is a growing demand for fresh water. A study conducted by Ghintji et al., [39] aimed to identify and map areas with potential for groundwater recharge. The variables affecting groundwater occurrence in the area were identified namely geology, geomorphology, slope, soil, DD, LD, and LULC. By integrating these variables using the AHP and Fuzzy-AHP in ArcGIS, the study generated a map indicating the zones of potential GWR. The final output's accuracy was validated using the location of freshwater boreholes. Both methods were tested using receiver operating characteristics (ROC) to evaluate their effectiveness. Fuzzy-AHP demonstrated higher accuracy (93.8%) compared to AHP (87.9%). The results indicate that both methods are suitable for delineating potential GWR zones in the study area. However, Fuzzy-AHP offers better accuracy as it doesn't rely solely on expert knowledge, overcoming a limitation of AHP. The study suggests protecting the identified recharge zones to ensure sustainable groundwater resource utilization. Policymakers can utilize these findings for the sustainable development and management of groundwater resources in the Ewaso Ng'iro—Lagh Dera basin.

To address the uncertainty in borehole drilling locations in Mpwapwa District (Tanzania), Ally et al., [46]'s study utilized RS and a GIS-based F-AHP to simulate GWP zones. By reclassifying, weighting, and ranking various thematic maps such as lithology, soil types, DD, lineament, magnetic intensity, slope, and elevation, the F-AHP model generated an overall GWPZ map in a GIS environment. The accuracy of the resulting map was validated using the overlaying method and the area under the curve (AUC) method. The GWP map revealed the most suitable areas for drilling. The accuracy of the generated map was found to be 72% using the overlaying method and 93% using the AUC method. The authors suggested that future research in the area should concentrate on examining how the groundwater storage in the aquifer system is influenced by physical and climatic changes in the environment.

In South Africa, a hybrid approach of geology, geophysics, geomorphology, and geoinformatics methods was employed to generate a GWP map of the Buffalo River catchment [34]. The different variables considered in the study are surficial lithology, rainfall distribution, LD, DD, TWI, LULC, and land surface temperature (LST). ML methods AHP was utilized to assign relevant weights to the input variables. The map obtained after this process was validated using borehole yield data. The calculated coefficient of determination (R^2) and correlation (R) demonstrated a remarkable assessment accuracy of more than 90%. According to the authors, this not only confirmed the suitability of the AHP computation but also acknowledged the excellence of the selected influencing factors and the proficiency in assigning pairwise weights. This study contributed to benefit the decision-makers, stakeholders, and the host community at large.

In [35] a similar approach was employed. The study aimed to assess GWP in the Kabompo catchment a study site within the Zambezi River Basin, in Southern Africa, by employing AHP, remote sensing, GIS techniques, and various factors influencing groundwater occurrence and movement. These factors were utilized to generate seven thematic maps, which were then weighted and scaled using an AHP tool based on their impact on groundwater occurrence and movement.

Fig. 4 Most frequently used ML algorithms in Africa
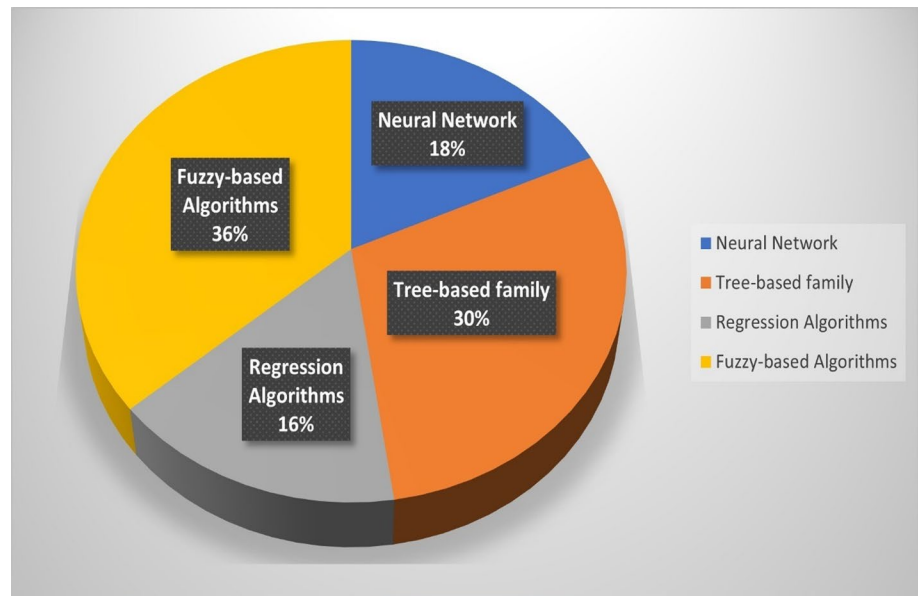
A weighted GWP map was created which was validated using existing boreholes. This revealed that 89% of them were located in moderate to very good potential zones, demonstrating the novelty and usefulness of this approach for assessing groundwater resources and integrated water management in the basin. The authors concluded that the use of the developed GWP map, along with local knowledge, could help promote the sustainable utilization and management of groundwater in the region. The findings of this research provide a foundation for future investigations on the effects of climate change on groundwater resources and the analysis of associated risks and uncertainties.

3.1.3 Frequency of algorithms applied

Fuzzy-based algorithms represent the most employed ML algorithms based on the studies reviewed in this study. Figure 4 shows the most frequently used ML algorithms in Africa. These algorithms widely employed over the continent account for 36% mostly in GWP mapping studies. They are followed by tree decision algorithms (GBDT, XGBoost, RFR, DT, KNN-R, and others). Regression, classification, and ranking tasks respond well to these traditional ML methods [17]. They emerge as the second most frequently employed algorithm, accounting for 30% of the usage. Following that, the Neural-Network algorithm family (Table 1) is utilized at a rate of 18%. The NN algorithm family is, in general, an ensemble of algorithms using a method for data processing that is largely inspired by the neural systems of humans and other animals. It is a non-linear statistical data analysis model between the input and output variables [47]. In the analyzed papers, the regression family algorithms (Table 1), account for a total of 16%. Interestingly, these algorithms were found to be less frequently utilized compared to other algorithms in the studies.

this study, we proved that at least two models should be compared using two different functions to select the best model presenting better target areas for further exploitation.

3.2 Location of studies

The review revealed that ML algorithms are increasingly being utilized in the field of groundwater availability in Africa. Specifically, there were seven studies in West Africa (Mali, Gambia, Burkina Faso, Ghana, Nigeria), seven in Eastern Africa (Tanzania, Kenya, Ethiopia, Rwanda), two in Central Africa (Cameroon, Chad), seven in North Africa (Morocco, Tunisia), and nine in Southern Africa (Angola, Zambia, Malawi, Namibia, Botswana, Zimbabwe, Mozambique, South Africa, Lesotho, Swaziland). In total 32 studies were reviewed. Figure 3 illustrates this geographical pattern.

Based on the analyzed literature in this paper, a notable concentration of studies from South Africa is observable. This may indicate that South Africa has emerged as a prominent contributor to the field within the continent. Exploring the reasons behind this concentration, such as available resources, research institutions, or specific initiatives, could

provide valuable insights into the advancements made in South Africa. Furthermore, investigating if these findings and methodologies can be applied to other regions within Africa would contribute to a more comprehensive understanding of groundwater prediction across the continent.

Fuzzy-based algorithms were predominantly used in study areas typically covering less than 4000 km^2 (for instance in [28, 32, 33, 40]) with some exceptions [39, 42, 43]. These exceptions include recent studies which successfully employed fuzzy-based algorithms in investing areas covering more than 10,000km^2. This could suggest these algorithms are proving their worth when it comes to covering vast areas and particularly dealing with large databases. Other ML algorithms such as tree-based algorithms are also increasingly being employed in studies covering large areas including large datasets [27, 30]. For instance a study conducted by [23], have taken a broader approach by covering multiple countries in the southern part of Africa. This serves as a great example for researchers to expand their focus and include larger parts of the countries in their studies. By doing so, a more comprehensive understanding can be gained of groundwater availability and better cater to the water needs of larger populations. This expanded coverage would greatly contribute to sustainable water resource management efforts in Africa. Additionally, studies that investigate small areas, based on the similarity of geology and other variables affecting groundwater availability, can apply their efficient methods to cover larger areas and provide more information on water availability throughout the country. The findings of this review also show that Morocco is making progress in the applicability of ML to cover a large part of the country.

This preference for alternative algorithms in larger areas could be due to their ability to handle complex spatial relationships and capture broader patterns. It is important to note that the choice of algorithm should be carefully considered based on the scale and characteristics of the study area. This revealed an interesting pattern in the choice of ML algorithms for further research to be done.

3.3 Prediction variables

The different algorithms explored by authors in different regions of Africa were trained using different sets of variables. The input variables used in the studies depend mostly on the data availability of the area and also on the understanding of the study area's geologic setting [48].

3.3.1 Geological variables

The geology of an area plays a crucial role when it comes to hosting groundwater as well as surface water infiltrating into an aquifer system, through the features of porosity and permeability which are lithological in studying the assessment of GWP in the Zambezi River Basin concluded that groundwater is normally found in the fissures, faults, and fractured zones within a geological formation [35]. The presence and flow of groundwater are mainly determined by the porosity and permeability of the surface and subsurface rock types. The same type of rock can form different geomorphic structures, leading to variations in porosity and permeability. This, in turn, alters the potential of groundwater [49]. As also stated by [50] the presence of groundwater in a geological formation and the potential for its use is largely dependent on the formation's porosity. Areas with high elevation and sharp inclines lead to greater runoff, while regions with topographical low points enhance infiltration. Moreover, a region with a high density of drainage paths boosts surface runoff in comparison to an area with a less dense drainage network.

This highlights how understanding the geology of the investigated area is crucial for groundwater study purposes. A list of the geological variables used by some of the machine learning case studies in Africa is given in Table 2.

The TWI indicates the effect of topography on an area [51] and helps to approximate moisture levels by identifying surface-saturation zones and the spatial distribution of soil moisture. DD inversely affects water absorption, while DFL impacts water penetration because faults give opportunities for water to penetrate the subsurface. Slope influences the opportunity for surface water to infiltrate permeable soils. It is a physical indicator, which approximates the areas of surface-saturation spot and the spatial distribution of soil moisture. Waterways affect runoff, which, discourages water retention at locations along the drainage path and thus affects infiltration into the surface. LULC affects the occurrence and availability of water to contribute to soil moisture and groundwater supply. As for altitude, it influences the direction and velocity of surface runoff and groundwater motion. In conclusion, lower altitudes, and gentler slopes, can be associated with permeable materials presence, which will give a greater infiltration system, which is conducive to a greater likelihood of groundwater presence.

Table 2 Geological variables used by the different case studies

References	Case study	Input variables	Output Parameters
[16]	Ghana	Recharge Runoff	GWR
[24]	Burkina Faso	Water Table Depth (WTD)	GWL
[27]	Rwanda	WTD	GWL
[22]	South Africa	Discharge point of groundwater Monthly GWL	GWL
[28]	South Africa	GWL	GWL
[26]	Ethiopia	Elevation Soil type Static water level (SWL)	GWL
[20]	South Africa	Spring recharge GWL	GWL
[21]	Morocco	GWL	GWL
[31]	Morocco	Lithology Slope Karst degree Land Cover Drainage density	GWR
[33]	Morocco	Fault density (FD) Drainage density (DD) Distance from drainage (DFD) Lineament density (LD) Distance from lineaments (DFL) Lithology Slope Terrestrial water index (TWI) Plan curvature Profile curvature Node density (DN)	GWP
[38]	Morocco	LD DD DFR DFL Permeability Slope TWI Plan curvature Profile curvature	GWP
[43]	Morocco	Geology WTL LD Slope DD Elevation Land use/Landcover (LULC)	GWP
[42]	Ghana	Fracture Density (FD) Lithology Soil type Stream power index (SPI) DD Slope TWI Relief	GWP
[37]	Ghana	Soil type Geology LULC LD	GWP

Table 2 (continued)

References	Case study	Input variables	Output Parameters
[45]	Nigeria	Soil type Geology DD LD Slope LULC	GWP
[32]	Gambia	Geology LULC Slope DD Soil type Aquifer transmissivity GW fluctuations	GWP
[41]	Cameroon	Aquifer depth Resistivity Transverse resistance Transmissivity Hydraulic conductivity	GWP
[44]	Rwanda	Geology Slope Soil type LULC	GWP
[40]	Ethiopia	Geology Slope Soil type Curvature TWI Elevation DD LULC LD	GWR
[36]	Ethiopia	DD Slope Geology LULC LD Soil Geomorphology	GWP
[39]	Kenya	Geology Geomorphology Slope Soil DD LD LULC	GWR
[46]	Tanzania	Lithology Soil Lineament Slope Elevation	GWP
[34]	South Africa	Lithology LD DD TWI LULC	GWP

Table 2 (continued)

References	Case study	Input variables	Output Parameters
[35]	South Africa	Lineaments Lithology LULC Soil Slope DD	GWP

In some cases, authors found some vegetation-related indices such as the Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) to be necessary for their GWL prediction studies [12, 29]. High NDVI values from natural vegetation may potentially suggest the likelihood of groundwater availability [48] as well as the NDWI which can detect water content and the presence of water bodies or moisture in the landscape.

LULC integration is frequently employed in studies that map GWP. This approach recognizes that land use changes, primarily driven by human activities, can significantly impact groundwater resources [12]. Geomorphology can be valuable in identifying characteristics that have the potential to facilitate groundwater infiltration and storage [12]. The depth of the water table is a valuable factor in mapping water tables as it helps identify the primary zones where aquifers are recharged and discharged [12]. Curvature describes the concavity or convexity of an area. Concave areas, particularly those found at lower elevations, can indicate the distribution of depressions where groundwater is likely to be present [48]. In terms of groundwater, the distance to sapping features can be important for understanding their impact on groundwater flow and availability. The closer a sapping feature is to a location, the more likely it is to affect the groundwater dynamics in that area. Monitoring the distance to these features can help in assessing the potential for groundwater infiltration and storage in the surrounding region.

Figures 5 and 6 respectively show the prevalence of geological inputs in GWL prediction and GWP mapping studies.

Figure 5, derived from our findings, shows the prevalence of hydrology-related variables at 79%. This also comprises hydrogeology variables which were the main variables utilized in the majority of the studies examined. Geomorphology accounts for 7% of the usage. Geology variables informing about soil were also utilized by a minority and made up 14%. This suggests that hydrology is the most studied geological variable, followed by soil type revealing aspects, while geomorphology aspects are less employed as input variables when investigating GWL prediction studies.

Based on our findings, Fig. 6 illustrates the distribution of geological, geomorphological, and hydrological variables in GWP mapping studies. The data shows that geological variables account for the largest portion at 41%, followed by geomorphology at 32%, and hydrology at 27%. This indicates that in these studies, the geology aspect of the area is the most extensively studied aspect, followed by geomorphology, while hydrology variables are used to a slightly lesser extent as input variables.

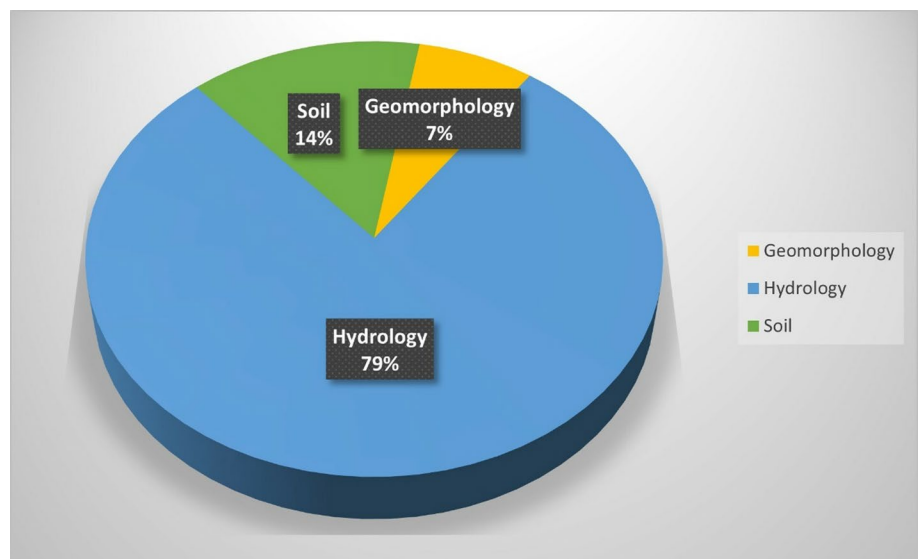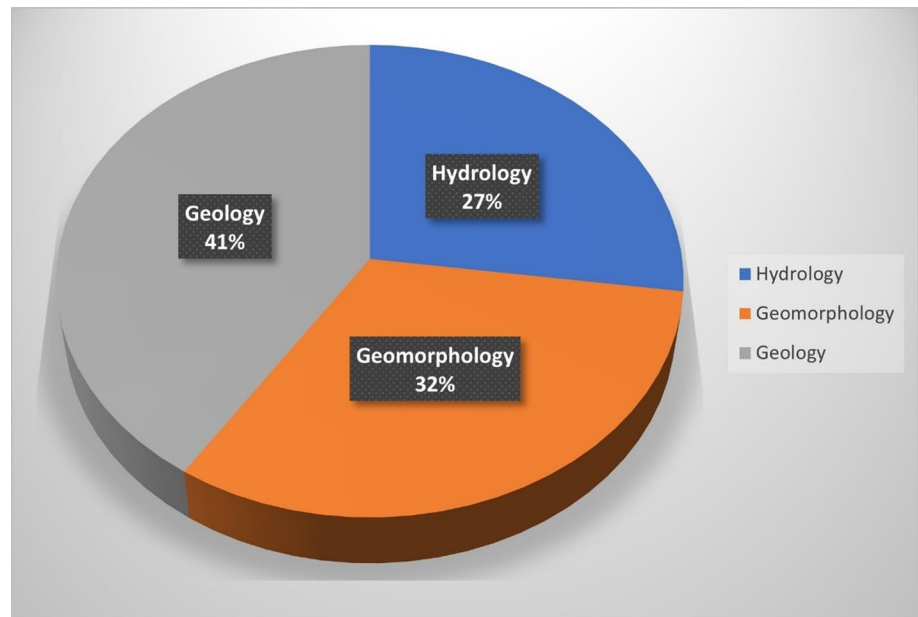**Fig. 5** Frequently utilized geological input variables in ML studies across Africa (GWL prediction case studies)

Fig. 6 Frequently utilized geological input variables in ML studies across Africa (GWP mapping case studies)

3.3.2 Climatic variables

Several studies [52–54] examined the influence of climatic variables on GWLs in Africa and confirmed the serious impact of variations in precipitation, temperature, and evapotranspiration on groundwater recharge. Table 3 shows the different climatic variables used in the different studies conducted in some African countries.

Precipitation (mainly rainfall and snow) acts as the primary source of recharge for the aquifer [22]. Areas with high precipitation and low temperature had higher GWLs, while areas with low precipitation and high temperature had lower GWLs.

Changes in climate variables such as temperature, precipitation, and evapotranspiration can have a significant impact on the amount of water that seeps into the ground to replenish groundwater resources. For instance, increased temperatures and changes in precipitation patterns can lead to reduced GWR, while, increased evapotranspiration rates can further exacerbate this issue. Also, climate change can directly or indirectly impact groundwater resources [55]. As such, understanding variation in climate, and GWR is crucial for sustainably managing freshwater resources. The main source of GWR is rainfall. Recharge is influenced by the rate of precipitation as well as surface and subsurface elements that permit or prohibit infiltration [12].

A large number of studies show that GWL variation is indeed sensitive to variations in temperature [53, 56]. A rise in temperature causes accelerated evaporation, lowering the recharge rate to the groundwater resource and leading to a drop in the groundwater table [53]. In Africa, temperature can have a significant impact on groundwater due to the continent's diverse climate and geography. Temperature variations can affect the rate of infiltration and GWR. Higher temperatures can increase evapotranspiration rates, which decreases the amount of water available for GWR. Based on this knowledge, we infer that with increased evapotranspiration and decreased precipitation, the impact of climate change will result in declining GWLs, which would cause some wells to become dry while others would become less productive due to the loss of available drawdown. GWP is largely influenced by recharge and recharge depends on five main factors which are climate (e.g. precipitation, temperature, and potential evapotranspiration (PET)), soils (e.g. texture, soil moisture), land cover (NDVI), geomorphology (e.g. landform surface slope and drainage density) and hydrology (e.g. streamflow and WTD) [12].

Our results enable us to visualize the ranks of different climatic variables based on their frequency of usage in studies analyzed (Fig. 7). Precipitation tops, with 54% of usage, followed by Temperature accounting for 27%. Evapotranspiration factors constitute 8% whereas Global Climate Indices made up 3% of the studies. Other temperature-related variables such as solar radiation, specific humidity, and wind speed were gathered together to not overwrite the temperature factor. This group accounted for 10%. This implies that Precipitation is the most commonly studied climate variable. Temperature receives a fair amount of attention too. However, evapotranspiration and global climate indices, are studied less frequently, suggesting these areas might be ripe for further exploration in future research.

Table 3 Climatic variables used by the different case studies

References	Case study	Input variables	Output Parameters
[16]	Ghana	PET Precipitation Temperature	GWR
[24]	Burkina Faso	Precipitation Temperature	GWL
[27]	Rwanda	Temperature Solar Radiation Precipitation	GWL
[22]	South Africa	Temperature Precipitation	GWL
[19]	South Africa	Evapotranspiration Precipitation Temperature	GWL
[30]	South Africa	Rainfall Temperature	GWL
[23]	Southern Africa Region	Precipitation Temperature Global Climate Indices	GWL
[28]	South Africa	Rainfall	GWL
[18]	Tunisia	Rainfall Evapotranspiration	GWL
[20]	South Africa	Rainfall Temperature	GWL
[21]	Morocco	Precipitation Temperature	GWL
[26]	Ethiopia	Precipitation Temperature Specific Humidity Wind Speed	GWL
[43]	Morocco	Rainfall	GWL
[37]	Ghana	Rainfall	GWL
[45]	Nigeria	Rainfall	GWL
[44]	Rwanda	Rainfall	GWL
[40]	Ethiopia	Rainfall	GWR
[36]	Ethiopia	Rainfall	GWL
[34]	South Africa	Rainfall	GWL
[35]	South Africa	Rainfall	GWL

4 Conclusion

This review aims to conduct an extensive examination of the current literature concerning research utilizing ML methods for quantifying groundwater availability and to compile a comprehensive inventory of the diverse ML algorithms, as well as the climatic and geological variables employed by some researchers across Africa. The study, identified several essential elements in the existing literature on investigation methods used, including the algorithms used, the input variables, and the target variables. The major finding was that the most utilized algorithms for ML studies in groundwater investigation are FL algorithms. In the studies we concentrated on, they predominantly demonstrated superior performance compared to other methods. Secondly, it is found that the most common variables used for machine learning studies in groundwater investigation are the hydrology-related variables for geological inputs in GWL prediction studies and geological variables for GWP mapping studies. For the climatic inputs, precipitation is the most used algorithm in the studies reviewed. These findings have significant implications for the understanding of large-scale groundwater availability and how climate will affect groundwater resources going forward.

Fig. 7 Frequently utilized climatic input variables in ML studies across Africa
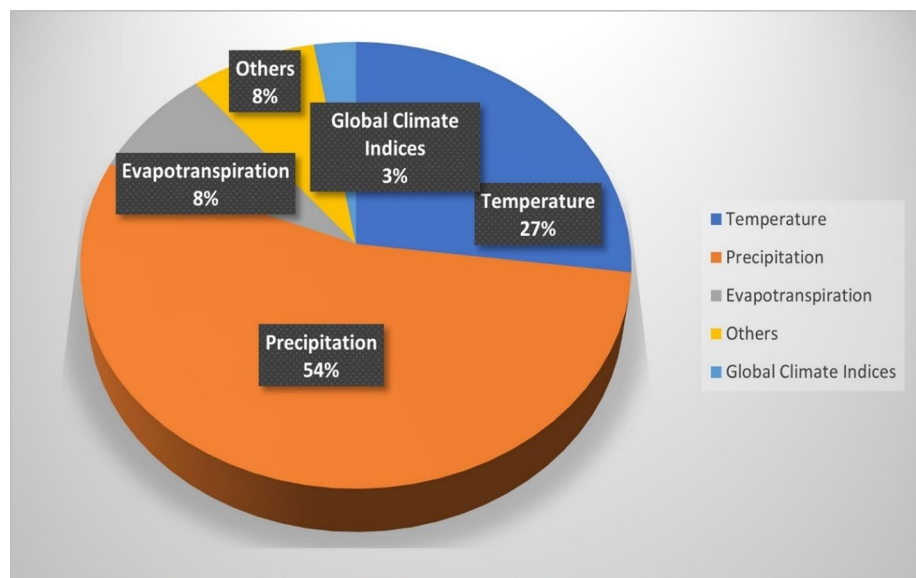
One thing to note is that during the research, there was a noticeable difference in the number of studies conducted in Africa compared to the rest of the world. This could suggest a lack of utilization of machine learning in the continent, despite its proven effectiveness and the water-related challenges in Africa. However, further investigations need to be conducted to clarify this. Their widespread use and proven efficiency worldwide suggest that Africa could greatly benefit from relying on them in the field of groundwater. Also, based on the studies addressed in this review available literature in Africa on the use of ML algorithms in GWL prediction and GWP mapping, there appears to be a gap in terms of comprehensive studies that cover a larger geographic scope. Most of the studies are concentrated in a small portion of the country, with limited coverage in other areas. This indicates a need for more research that encompasses a wider range of countries and regions within Africa.

Author contributions Conceptualization of the review was done by C.D.B. and H.T. Data curation was done by H.T., C.D.B. and S.S.R.G. The formal analysis was done by H.T. The investigation was done by C.D.B., S.S.R.G., D.D.W., J.N.A.A. and M.A.O. The methodology was established by C.D.B., H.T. The project administration was done by C.D.B. The resources were provided by C.D.B. The work was supervised by C.D.B., S.S.R.G., D.D.W., J.N.A.A., M.A.O. The validation was done by H.T., C.D.B., J.N.A.A., M.A.O., D.D.W., S.S.R.G., V.M., J.G., S.K.A. The Visualization was done by H.T., C.D.B., S.S.R.G. The original draft has been written by H.T. The written review was edited by H.T., C.D.B., J.N.A.A., M.A.O., D.D.W., S.S.R.G., V.M., J.G., S.K.A.

Availability of data and materials No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publications Not applicable.

Competing interests The authors declare no competing interests.

References

1. MacDonald AM, Bonsor HC, Dochartaigh BÉÓ, Taylor RG. Quantitative maps of groundwater resources in Africa. Environ Res Lett. 2012;7(2):024009. https://doi.org/10.1088/1748-9326/7/2/024009.
2. Wang H, Wang T, Zhang B, Li F, Toure B, Omosa IB, et al. Water and wastewater treatment in Africa—current practices and challenges. Clean: Soil, Air, Water. 2014;42(8):1029–35. https://doi.org/10.1002/clen.201300208.
3. Famiglietti JS. The global groundwater crisis. Nat Clim Change. 2014;4(11):945–8.
4. Chakkaravarthy DN. Water scarcity-challenging the future. Int J Agric Environ Biotechnol. 2019. https://doi.org/10.30954/0974-1712.08.2019.2.
5. Vörösmarty CJ, Green P, Salisbury J, Lammers RB. Global water resources: vulnerability from climate change and population growth. Science. 2000;289(5477):284–8. https://doi.org/10.1126/science.289.5477.284.
6. Olarinoye T, Foppen JW, Veerbeek W, Morienyane T, Komakech H. Exploring the future impacts of urbanization and climate change on groundwater in Arusha, Tanzania.
7. Sharp J. The impacts of urbanization on groundwater systems and recharge Sharp, Jack. (2010). The impacts of urbanization on groundwater systems and recharge. Aqua Mundi. 1. 51-56. https://doi.org/10.4409/Am-004-10-0008.
8. Bigas H. Water security and the global water agenda: a UN-water analytical brief. Hamilton, Ont.: United Nations University—Institute for Water, Environment and Health; 2013.
9. Oluwasanya G, Perera D, Qadir M, Smakhtin V. Water security in africa: a preliminary assessment. UNU-INWEH; 2022. https://collections.unu.edu/view/UNU:8718. Accessed 14 Sep 2023.
10. Tao H, Hameed MM, Marhoon HA, Zounemat-Kermani M, Heddam S, Kim S, et al. Groundwater level prediction using machine learning models: A comprehensive review. Neurocomputing. 2022;489:271–308.
11. Sarkar SK, Talukdar S, Rahman A, Shahfahad RSK. Groundwater potentiality mapping using ensemble machine learning algorithms for sustainable groundwater management. Front Eng Built Environ. 2022;2(1):43–54. https://doi.org/10.1108/FEBE-09-2021-0044/full/html.
12. Gonzalez RQ, Arsanjani JJ. Prediction of groundwater level variations in a changing climate: a Danish case study. ISPRS Int J Geo-Inf. 2021;10(11):792.
13. Gómez-Escalonilla V, Martínez-Santos P, Martín-Loeches M. Preprocessing approaches in machine-learning-based groundwater potential mapping: an application to the Koulikoro and Bamako regions, Mali. Hydrol Earth Syst Sci. 2022;26(2):221–43.
14. Altchenko Y, Awulachew SB, Brida B, Diallo HA, Mogbante D, Pavelic P, et al. Management of ground water in africa including transboundary aquifers: implications for food security, livelihood and climate change adaptation.
15. Cuthbert MO, Taylor RG, Favreau G, Todd MC, Shamsudduha M, Villholth KG, et al. Observed controls on resilience of groundwater to climate variability in sub-Saharan Africa. Nature. 2019;572(7768):230–4.
16. Siabi EK, Dile YT, Kabo-Bah AT, Amo-Boateng M, Anornu GK, Akpoti K, et al. Machine learning based groundwater prediction in a data-scarce basin of Ghana. Appl Artif Intell. 2022;36(1):2138130. https://doi.org/10.1080/08839514.2022.2138130.
17. Adiat KAN, Ajayi OF, Akinlalu AA, Tijani IB. Prediction of groundwater level in basement complex terrain using artificial neural network: a case of Ijebu-Jesa, southwestern Nigeria. Appl Water Sci. 2020;10(1):8. https://doi.org/10.1007/s13201-019-1094-6.
18. Derbela M, Nouiri I. Intelligent approach to predict future groundwater level based on artificial neural networks (ANN). Euro-Mediterr J Environ Integr. 2020;5(3):51. https://doi.org/10.1007/s41207-020-00185-9.
19. Gaffoor Z, Pietersen K, Jovanovic N, Bagula A, Kanyerere T, Ajayi O, et al. A comparison of ensemble and deep learning algorithms to model groundwater levels in a data-scarce aquifer of Southern Africa. Hydrology. 2022;9(7):125.
20. Gibson K. The application of machine learning for groundwater level prediction in the steenkoppies compartment of the Gauteng and North West dolomite aquifer, South Africa.
21. Ibrahimi A, Baali A, Couscous A, Kamel T, Hamdani N. Comparative study of the three models (ANN-PMC), (DWT-ANN-PMC) and (MLR) for prediction of the groundwater level of the surface water table in the Saïss Plain (North of Morocco). Int J Intell Eng Syst. 2017;10(5):220–30.
22. Kanyama Y, Ajoodha R, Seyler H, Makondo N, Tutu H. Application of Machine Learning Techniques In Forecasting Groundwater Levels in the Grootfontein Aquifer. In: 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC). Kimberley, South Africa: IEEE; 2020 [cited 2023 Aug 15]. p. 1–8.
23. Kalu I, Ndehedehe CE, Okwuashi O, Eyoh AE, Ferreira VG. A new modelling framework to assess changes in groundwater level. J Hydrol Reg Stud. 2022;43:101185.
24. Coulibaly P, Anctil F, Aravena R, Bobée B. Artificial neural network modeling of water table depth fluctuations. Water Resour Res. 2001;37(4):885–96. https://doi.org/10.1029/2000WR900368.
25. Hussein EA, Thron C, Ghaziasgar M, Bagula A, Vaccari M. Groundwater prediction using machine-learning tools. Algorithms. 2020;13(11):300.
26. Li W, Finsa MM, Laskey KB, Houser P, Douglas-Bate R. Groundwater level prediction with machine learning to support sustainable irrigation in water scarcity regions. Water. 2023;15(19):3473.
27. Kombo O, Kumaran S, Sheikh Y, Bovim A, Jayavel K. Long-Term Groundwater Level Prediction Model Based on Hybrid KNN-RF Technique. Hydrology. 2020;7(3):59.

28. Tladi TM, Ndambuki JM, Olwal TO, Rwanga SS. Groundwater level trend analysis and prediction in the upper crocodile sub-basin, South Africa. Water. 2023;15(17):3025.

29. Gómez-Escalonilla V, Vogt ML, Destro E, Isseini M, Origgi G, Djoret D, et al. Delineation of groundwater potential zones by means of ensemble tree supervised classification methods in the Eastern Lake Chad basin. Geocarto Int. 2022;37(25):8924–51. https://doi.org/10.1080/10106049.2021.2007298.

30. Aderemi BA, Olwal TO, Ndambuki JM, Rwanga SS. Groundwater levels forecasting using machine learning models: a case study of the groundwater region 10 at Karst Belt, South Africa. Syst Soft Comput. 2023;5:200049.

31. Aouragh MH, Essahlaoui A, El Ouali A, El Hmaidi A, Kamel S. Groundwater potential of Middle Atlas plateaus, Morocco, using fuzzy logic approach, GIS and remote sensing. Geomat Nat Hazards Risk. 2017;8(2):194–206. https://doi.org/10.1080/19475705.2016.1181676.

32. Jallow AG, Diongue DM, Emvoutou HC, Mama D, Faye S. Groundwater Recharge Zone Mapping Using GIS-based Analytical Hierarchy Process and Multi-Criteria Evaluation: Case Study of Greater Banjul Area. Am J Water Res. 2020;8(4):182–90.

33. Benjmel K, Amraoui F, Boutaleb S, Ouchchen M, Tahiri A, Touab A. Mapping of groundwater potential zones in crystalline terrain using remote sensing, GIS techniques, and multicriteria data analysis (case of the Ighrem region, Western Anti-Atlas, Morocco). Water. 2020;12(2):471.

34. Owolabi ST, Madi K, Kalumba AM, Orimoloye IR. A groundwater potential zone mapping approach for semi-arid environments using remote sensing (RS), geographic information system (GIS), and analytical hierarchical process (AHP) techniques: a case study of Buffalo catchment, Eastern Cape, South Africa. Arab J Geosci. 2020;13(22):1184. https://doi.org/10.1007/s12517-020-06166-0.

35. Ndhlovu GZ, Woyessa YE. Integrated assessment of groundwater potential using geospatial techniques in Southern Africa: a case study in the Zambezi river basin. Water. 2021;13(19):2610.

36. Haile MM. Approaches for delineating groundwater recharge potential zone using fuzzy logic model. Geofluids. 2022;2022:1–15.

37. Osiakwan GM, Gibrilla A, Kabo-Bah AT, Appiah-Adjei EK, Anornu G. Delineation of groundwater potential zones in the Central Region of Ghana using GIS and fuzzy analytic hierarchy process. Model Earth Syst Environ. 2022;8(4):5305–26. https://doi.org/10.1007/s40808-022-01380-z.

38. Echogdali FZ, Boutaleb S, Kpan RB, Ouchchen M, Bendarma A, El Ayady H, et al. Application of fuzzy logic and fractal modeling approach for groundwater potential mapping in semi-arid Akka Basin, Southeast Morocco. Sustainability. 2022;14(16):10205.

39. Githinji TW, Dindi EW, Kuria ZN, Olago DO. Application of analytical hierarchy process and integrated fuzzy-analytical hierarchy process for mapping potential groundwater recharge zone using GIS in the arid areas of Ewaso Ng'iro—Lagh Dera Basin, Kenya. HydroResearch. 2022;5:22–34.

40. Melese T, Belay T. Groundwater potential zone mapping using analytical hierarchy process and GIS in Muga Watershed, Abay Basin, Ethiopia. Glob Chall. 2022;6(1):2100068. https://doi.org/10.1002/gch2.202100068.

41. Aretouyap Z, Asfahani J, Abdulrazzaq ZT, Tchato SC. Contribution of the fuzzy algebraic model to the sustainable management of groundwater resources in the Adamawa watershed. J Hydrol Reg Stud. 2022;43:101198.

42. Amponsah TY, Danuor SK, Wemegah DD, Forson ED. Groundwater potential characterisation over the Voltaian basin using geophysical, geological, hydrological and topographical datasets. J Afr Earth Sci. 2022;192:104558.

43. Derdour A, Benkaddour Y, Bendahou B. Application of remote sensing and GIS to assess groundwater potential in the transboundary watershed of the Chott-El-Gharbi (Algerian–Moroccan border). Appl Water Sci. 2022;12(6):136. https://doi.org/10.1007/s13201-022-01663-x.

44. Ngwijabagabo H, Tsinda A, Dusengimana A, Minani M, Tuyishime CSP, Hirwa H, et al. Groundwater potential mapping using geospatial and AHP techniques in Eastern province of Rwanda. Rwanda J Eng Sci Technol Environ. 2023. https://doi.org/10.4314/rjeste.v5i1.4.

45. Amos-Uhegbu C, Jokotagba S, Mmeka J, Nwokoma E. Assessment of groundwater potential zones using multicriteria decision analysis: a case study of Umuahia areas of Niger delta basin, Nigeria. Niger J Phys. 2023;32(3):110–26.

46. Ally AM, Yan J, Bennett G, Lyimo NN, Mayunga SD. Assessment of groundwater potential zones using remote sensing and GIS-based fuzzy analytical hierarchy process (F-AHP) in Mpwapwa district, Dodoma, Tanzania. Geosyst Geoenviron. 2024;3(1):100232.

47. Chen Y, Chen W, Chandra Pal S, Saha A, Chowdhuri I, Adeli B, et al. Evaluation efficiency of hybrid deep learning algorithms with neural network decision tree and boosting methods for predicting groundwater potential. Geocarto Int. 2022;37(19):5564–84. https://doi.org/10.1080/10106049.2021.1920635.

48. Sahour H, Sultan M, Abdellatif B, Emil M, Abotalib AZ, Abdelmohsen K, et al. Identification of shallow groundwater in arid lands using multi-sensor remote sensing data and machine learning algorithms. J Hydrol. 2022;614:128509.

49. Shahid S, Nath SK, Roy J. Groundwater potential modelling in a soft rock area using a GIS. Int J Remote Sens. 2000;21(9):1919–24. https://doi.org/10.1080/014311600209823.

50. Waikar ML, Nilawar AP. Identification of groundwater potential zone using remote sensing and GIS technique. Int J Innov Res Sci Eng Technol. 2007;3(5):12163.

51. Prasad P, Loveson VJ, Kotha M, Yadav R. Application of machine learning techniques in groundwater potential mapping along the west coast of India. GIScience Remote Sens. 2020;57(6):735–52. https://doi.org/10.1080/15481603.2020.1794104.

52. Al-Gamal SA, Dodo AK, Observatory S. Impacts of climate changes on water resources in Africa with emphasis on groundwater. Environ Hydrol. 2009;17:1.

53. Chen Z, Grasby SE, Osadetz KG. Relation between climate variability and groundwater levels in the upper carbonate aquifer, southern Manitoba, Canada. J Hydrol. 2004;290(1–2):43–62.

54. Wu WY, Lo MH, Wada Y, Famiglietti JS, Reager JT, Yeh PJF, et al. Divergent effects of climate change on future groundwater availability in key mid-latitude aquifers. Nat Commun. 2020;11(1):3710.

55. Wang SJ, Lee CH, Yeh CF, Choo YF, Tseng HW. Evaluation of climate change impact on groundwater recharge in groundwater regions in Taiwan. Water. 2021;13(9):1153.

56. Hassan WH, Hussein HH, Nile BK. The effect of climate change on groundwater recharge in unconfined aquifers in the western desert of Iraq. Groundw Sustain Dev. 2022;16:100700.