Science of the Total Environment 957 (2024) 177693

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Fluoride contamination in African groundwater: Predictive modeling using stacking ensemble techniques

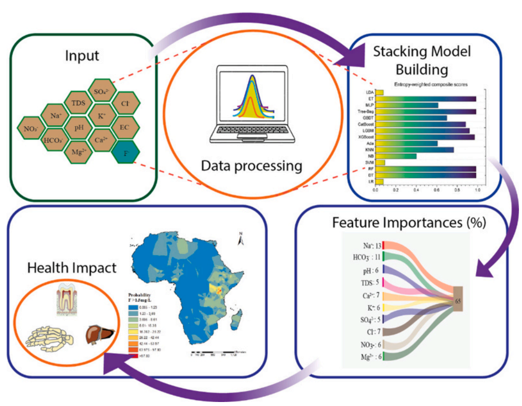
Usman Sunusi Usman [a], Yousif Hassan Mohamed Salh [a], Bing Yan [a,*], Jean Pierre Namahoro [b], Qian Zeng [a], Ismaila Sallah [a]

[a] School of Environmental Studies, China University of Geosciences Wuhan, 388 Lumo Road, Wuhan 430074, China
[b] School of Economic and Management, China University of Geosciences Wuhan, China

HIGHLIGHTS

- Application of the stacking ensemble learning method for African fluoride prediction.
- An AUC of 0.86 was realized, thus improving the effectiveness of fluoride prediction over traditional models.
- Na^+, HCO_3^-, Ca^{2+}, and Cl^- were their model's potential key predictors of fluoride contamination.
- The findings will help inform public health with targeted fluoride mitigation in high-risk areas.
- Provide a scalable framework for addressing global water contamination issues.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Deyi Hou

Keywords:
Groundwater
Fluoride
Stacking ensemble learning
Physiochemical parameters
Africa

ABSTRACT

Fluoride contamination of groundwater is a severe public health problem in Africa due to natural factors that include geological weathering of fluoride-bearing minerals and climatic conditions characterized by high evaporation rates that highly elevate fluoride levels. Anthropogenic activities further aggravate the problem and have affected millions of people in countries such as; South Africa, Tanzania, Nigeria, Ethiopia, Ghana, Kenya, Mauritania, Botswana, and Egypt. High fluoride levels of up to 10 mg/L have been encountered in parts of the East African Rift Valley, above the WHO's recommended limit of 1.5 mg/L, causing serious dental and skeletal fluorosis among the affected people. In this study, the distributions of F^- in groundwater of Africa were forecast using an advanced stacking ensemble learning model based on 11 crucial groundwater physiochemical variables and 6270 accessible statistics of observed concentrations. The enhanced algorithm incorporates randomized trees, Tree-Bag, RF, DT, XGB, and ET Machine as base trainees, with a simple Naïve Bayes as the meta-analyzer. The model's AUC score of 0.86 accurately represented the uneven distributions of groundwater fluoride. The results showed that 20–35 % of the continent's eastern part and 10 % of its western region are at risk of having fluoride levels exceeding WHO limits, with an expected population of around 80 million. Regionally, fluoride contamination ranges from 0.1 to 3 mg/L in West Africa was range from 0.0 to 13.29 mg/L, 0.01–588 mg/L in

* Corresponding author.
E-mail address: yanbing@cug.edu.cn (B. Yan).

https://doi.org/10.1016/j.scitotenv.2024.177693
Received 10 September 2024; Received in revised form 15 November 2024; Accepted 19 November 2024
0048-9697/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

East Africa, 0.04–65.9 mg/L in South Africa, and 0.1–10.5 mg/L in North and 0.01–1.9 mg/L in Central Africa. Na^+ and HCO_3^- are Africa's leading primary causes of fluoride contamination, with Ca^{2+} and Cl^- contributing to fluoride influence in some parts of the continent. This study helped identify health concerns linked to groundwater fluoride and offered guidance on assessing health risks in areas with sparse sample sizes.

1. Introduction

Groundwater is regarded as one of the most significant natural resources which is essential to the ecosystem's stability, public health, as well as environmental safety (Ezike and Chukwunonyerem, 2022). Groundwater is utilized for industrial and agricultural uses in Africa, where it is the main supply of drinking water (Kut et al., 2016; Gosomji et al., 2022). It supplies approximately 15 % of Africa's renewable water resources. Drinking water is significant in almost all of the continent's rural areas, particularly during the dry seasons and in the main arid zones (Kut et al., 2016). Groundwater is potentially influenced by a wide range of geoenvironmental factors, including topography, geological structure, lithology, landform, drainage pattern, land use, land cover, and climate (Subba Rao et al., 2022, 2024). Advanced new monitoring and management technology is required to prevent pollution and ensure water resources' sustainability (Rao et al., 2021). Groundwater in Africa is chemically suitable for most applications and has high microbiological purity. However, due to its hydrochemical characteristics, issues can occur, including excessive sulfate in mudstones and worn basements, as well as hardness in sandstones or limestone aquifers (Sharma and Kumar, 2020).

However, as a result of anthropogenic and geogenic sources, fluoride in groundwater is penetrating the water supply and damaging its quality in Africa (Andesikuteb Yakubu et al., 2020; Balogun et al., 2022; Nakayama et al., 2022). Natural sources include minerals that contain fluoride, such as hornblende and fluorite, found in volcanic, gneissic, and granite rocks and sediments (Kut et al., 2016; Chowdhury et al., 2019; Zango et al., 2019; Tolera et al., 2020; Sunkari et al., 2022 Shube et al., 2023; Sunkari et al., 2023). High fluoride levels in drinking water ($\geq$1.5 mg/L (WHO, 2023)) are the cause of human dental and skeletal fluorosis, and they are harmful to human health in many regions of the globe (Podgorski, and & Michael, B., 2022; Shaji et al., 2024), such as in Asia, Europe; North America, Australia, and South America (Al-Amry et al., 2020; Sawangjang and Takizawa, 2020; Vitor Portugal Gonçalves et al., 2020; Parrone et al., 2020; Amiri and Berndtsson, 2020; Aullón Alcaine et al., 2020; Bhandari et al., 2021; Lopes et al., 2022; Iqbal et al., 2023; Morales-deAvila et al., 2023; Mustafa et al., 2023; Alharbi and El-Sorogy, 2023; Sarkar et al., 2023; Shaji et al., 2024). In this case, we implement modeling to calculate the amount of fluoride in the groundwater. Therefore, an accurate assessment of groundwater is required.

The advancement of remote sensing and GIS technology worldwide has led to the development of various techniques for mapping groundwater potential (Prapanchan et al., 2024; Rajan et al., 2024). These methods have outperformed the traditional geophysical, geological, and hydrogeological techniques (Satheeshkumar et al., 2024). Nowadays, artificial intelligence (AI) technology has achieved credit in analyzing and quantifying groundwater pollution over a wide range (Sun et al., 2022; Priyadarshini et al., 2022; Mia et al., 2023). Recently, machine (ML) intelligence techniques have been regarded as a critical idea in hydrology investigation since they can tackle complicated problems (Sahoo et al., 2017; Hussein et al., 2020). Artificial Neural Networks (ANNs) have been shown effective in modeling groundwater, evaluating water quality, and supporting various hydrologic applications in several regions around the world, such as in Logan, USA, Faisalabad, Pakistan (Anandhi et al., 2022). A recent study using machine learning approaches has effectively identified fluoride as one of groundwater's chemical constituents and features (Nafouanti et al., 2023). Arsenic (Tan et al., 2020; Lombard et al., 2021), nitrate (Alkindi et al., 2022), using groundwater chemistry records in various hydrogeologic settings currently exist.

Stacked ensemble models represent a lately trendy machine learning technique in Earth sciences and water quality prediction, considering the models' advantage in applying several base learners in parallel, which finally enhances the accuracy and robustness of such a model. These models address issues with traditional single models, overfitting issues for complex nonlinear data interactions such as those typical in groundwater quality assessment. They provided more reliable spatial distribution patterns for contaminants such as fluoride. This is furthered through different environmental applications that enable the making of accurate pollutant predictions for a wide range of conditions in hydrogeology (Divina et al., 2018; Cui et al., 2021; Ribeiro et al., 2022). Therefore, the stacked ensemble model is adopted for this study to predict fluoride contamination in African groundwater, and to enhance predictive methodologies in Earth sciences to support sustainable water management.

Geoscientists often use artificial intelligence (AI) algorithms, such as generalized linear models (GLM), logistic multiple regression (LMR), generalized additive models (GAM), artificial neural networks (ANN), support vector machines (SVM), and extreme learning machines (ELM), to predict unknown events using complex datasets and historical data (Zhang et al., 2015; Barzegar et al., 2017; Najwa Mohd Rizal et al., 2022; Adnan et al., 2023). This is especially true when predicting water quality. However, these approaches have several drawbacks, including poor prediction when the testing data range is beyond the training dataset, as ANN illustrates (Najwa Mohd Rizal et al., 2022; Yaghoubi et al., 2024). Among this model, several models have also been used in forecasting water quality; for instance, the RF (Ouedraogo et al., 2019) model is an ensemble learning method that aggregates individual findings and creates several classification trees (for example, a single random forest with 1000 trees), and has been shown via actual research to outperform its members (Sakaa et al., 2022). Xgboost (Xia et al., 2022; Abba et al., 2023), LGBM (Chen et al., 2024). Recently, many researchers have combined multiple machine-learning techniques to estimate groundwater pollution (Li et al., 2020) by combining extreme learning machine (ELM) with particle swarm optimization (PSO) (Singh, 2023), using the Hybrid of hierarchical K-means Clustering and ANN model (HCA-ANN), Optimization using gray-wolf methods (GWO) (Ghobadi et al., 2022) and GWO-MLPANN, a multi-layer perceptron artificial neural network (MLPANN).

The stacking ensemble model typically prescribes several vital steps. To guarantee data diversity and representativeness, preprocessing, and preparation of the data were done first. The dataset was then split into training and testing sets. Second, several base learners who were either homogeneous or varied were chosen to be the primary learners (Divina et al., 2018; Cui et al., 2021; Ribeiro et al., 2022). On the training set, all learners were fitted independently, and regarding the test phase, they each produced their predictions (Arabameri et al., 2022). Thirdly, to aggregate the predictions from the primary learners, a meta-classifier, also referred to as the second-level trainee, was built (Yaseen, 2023). The meta-classifier learned to forecast a target feature by using the actual features and the predictions made by the base learners as input. Finally, the performance of the ensemble model was estimated by measuring its prediction accuracy and generalization magnitude on the training set. Because it integrates varied base learners skillfully, the stacking ensemble model effectively addresses individual constraints and lowers the risk of overfitting. Unlike a conventional single model, this method predicts distribution patterns with more accuracy.

However, a previous study by Malago (2017) He is focused on the comprehensive review of the distribution of fluoride levels across various regions in Africa. Such thorough assessments can offer insights into the environmental and spatial parameters influencing fluoride distribution when incorporated into the model development process. Thus, the model's capacity to anticipate fluoride pollution more precisely and consistently is improved by this background information, which will eventually help improve groundwater management practices throughout the continent.

In Africa, groundwater is used mainly for home water supplies. However, the safety of water resources has always been in jeopardy due to high F^- concentrations in groundwater. A thorough, large-scale geographical forecast of groundwater F^- distribution is desperately needed. Traditional single-machine learning methods cannot meet this requirement. Thus, this study distinguish to the existing literature as follows: (1) to establish the stacking ensemble learning model by fine-tuning base learner selection on the case of F^- contamination groundwater. (2) to create hazard maps illustrating the spatial issue of F^- in underground water throughout Africa. (3) evaluate the factors influencing groundwater F^- distributions, considering the predictor variables' relative contributions. Additionally, the study will demonstrate how well classification analysis performs when estimating water parameters. As a result, twenty-four African countries struggled with excessive fluoride in drinking water over WHO's recommended limit of 1.5 mg/L. The East Africa Rift Valley, which starts in the Jordanian Valley and stretches to Northeast Africa, Eritrea, Ethiopia, Kenya, Tanzania, and Malawi, is well known for having a high fluoride level in its groundwater. To the best of our knowledge, this study represents the first application of an ensemble learning stacking model at a continental scale for the prediction of fluoride contamination in groundwater, hence filling the gap in understanding fluoride distribution across the African continent.

Although this work addresses urgent environmental and population health issues, it contributes significantly to attaining certain SDGs initiated by the United Nations (UN). More specifically, this supports SDG 6: Clean Water and Sanitation, from the United Nations (Li, 2020); SDG 3: Good Health and Well-being, through the WHO, (2023), since the reduced health risks due to fluoride contamination; and SDG 13: Climate Action (Li, 2020). This attainment realizes three of its achievements toward sustainable development by attaining improved health, ensuring effective and accountable institutions, and ensuring inclusive and equitable quality education. Investigating the impacts of climate-induced high evaporation rates on water quality ensures that greater global efforts are being put into sustainable development, which could greatly benefit policymakers and public health authorities.

2. Materials and methods

2.1. Geology and hydrogeology of Africa

Africa is the second-largest continent after Asia, covers 30.3 million km^2 and has 54 countries. Its geological history includes seven Archaean cratonic nuclei, forming Southern, West, and Central Africa. The continent's Paleoproterozoic geology between 2.3 and 1.8 Ga. is characterized by tectonism, volcanism, and sedimentation, documented as the Eburnia Africa cratons in the Southern, Central, and Western Hemispheres (Thomas Schlüter, 2006; Begg et al., 2009). The Old Precambrian rocks cover 60 % of Africa's surface, with newer sedimentary basins in the north and west. Some of the world's oldest rock formations in the African Shield region are almost three billion years old. The continent's geological history includes the collapse of Gondwana during the Mesozoic period, leading to the Rift Valley system in East Africa (Thomas Schlüter, 2006). Africa also has large sedimentary basins, particularly in the Congo, Niger, and Nile deltas, containing significant hydrocarbons and mineral resources (Fig. s3).

African consists of three aquifers: consolidated and unconsolidated sedimentary aquifers, volcanic aquifers, and fractured basement aquifers. In limestone regions, karst aquifers can also be found. Significant sedimentary aquifers can be found in the Kalahari Basin, Karoo Basin, Lake Chad Basin, Niger Basin, and sedimentary basins in North Africa (Nijsten et al., 2018; Awaleh et al., 2020; Gebeyehu et al., 2022). The deep aquifer lies within the East African Humid Period (Fig. 1), having formed between 12,500 and 8700 years ago (late Pleistocene/early-to-mid Holocene) (Awaleh et al., 2020), In which 40 % of the continent was covered with 72 transboundary aquifers, mostly in arid or semi-arid and where 33 % of people reside (Nijsten et al., 2018). For example, Ethiopia's primary aquifers comprise metamorphic basement rocks, secondary felsic pyroclastic deposits, and bare volcanic rocks. Those aquifers are overlain by regolith, eluvial soil, and alluvial-lacustrine sediments cover these aquifers (Gebeyehu et al., 2022).

However, the Basement aquifers are common in the African Shield nations of Botswana, Zimbabwe, the Northern part of Nigeria, and South Africa. Groundwater is stored in porous zones that are produced by weathering. Intergranular flow occurs in sedimentary aquifers and the worn zone, but fracture flow is more common in crystalline foundations. Many cities rely on freshwater from coastal aquifers despite their susceptibility to saltwater intrusion (Titus et al., 2009; Mukwati, 2017; Bura et al., 2018).

2.2. Data gathering and acquisition

In this study, 6254 groundwater samples data were gathered from various sources (table s4), which were collected by hand pumps, boreholes, and hand-dug wells. Depending on the availability of sources, the sampling sites were methodically set up at 200-m intervals. The samples were gathered from different parts of Africa between 1996 and 2024. Sampling in each of the studies was done according to the respective procedures; also, careful monitoring of the possible sources of contamination was considered beforehand to ensure dependable and contaminant-free samples of groundwater (Daffi et al., 2022). The samples included data from different countries including Ghana, Kenya, Nigeria, and Ethiopia (Table s4).

The techniques for laboratory analysis, such as those for pH, conductivity (EC), temperature, fluoride concentration, and a suite of major and minor elements, are described in the individual studies from which the data were sourced. This information has not been reproduced here to maintain consistency, as in each study multiple methods and instruments were used. Full details of the methodologies and instruments used in each of the analytical techniques can be found in the referenced studies (Table s4). The groundwater physicochemical parameters are summarized in the descriptive statistical analysis presented in Table s5.

2.3. Handling missing values

Missing values were fixed during the preparation phase to guarantee the dataset's integrity for additional analysis. The mean value of the dataset's numerical columns was imputed using a Simple Imputer (DataWig Documentation, 2022). Specifically, this imputation approach was used for numerical characteristics since these columns are essential for the machine learning model that follows. However, Missing values is a step resulting in skewed findings, which can lead to low model accuracy and incorrect predictions, therefore managing them is an essential stage in the preparation of data (Shadbahr et al., 2023). Therefore, Numerical characteristics with missing values can be attributed to preserving dataset continuity and preventing the loss of crucial data during analysis (equ 1). For each missing value x_i in the feature X, if x_i is missing, replace it with mean $\widehat{X}$ of the observed values in feature X.

$$\widehat{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad (1)$$

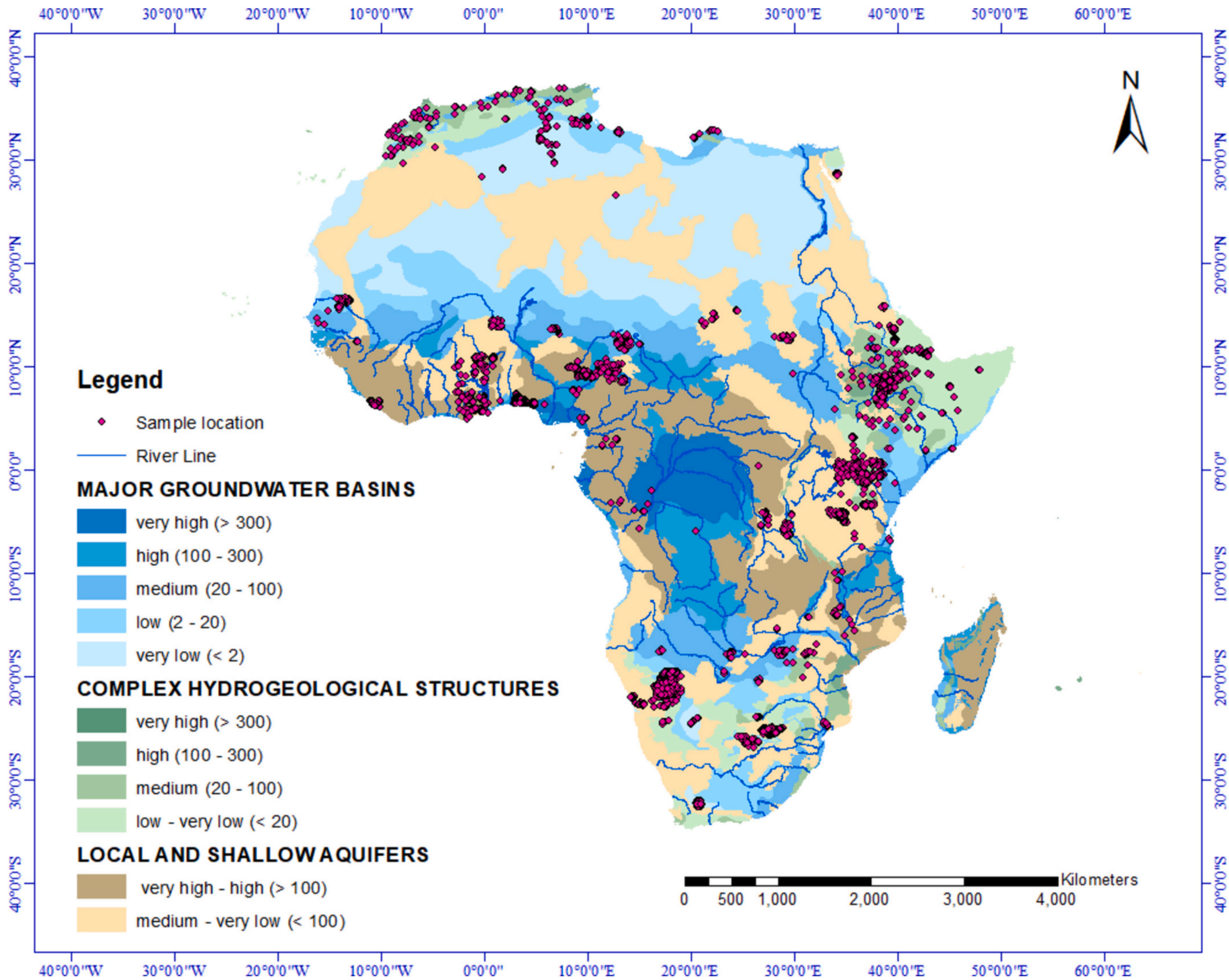


Fig. 1. A map of the African continent showing the sample Locations and groundwater potential (m^3/day).

Where X_i is the observed value in feature X, $\widehat{X}$ is the mean of observed value X, n is the number of observed values and x_iIs the specific value missing.

2.4. Model building

The dataset was arbitrarily split into training and testing phases in an 8:2 ratio. The division ensured that both models had comparable rates exceeding the 1.5 probability threshold, allowing for a consistent comparative assessment of model performance throughout the set. Fifteen potential fundamental learners were estimated using six performance metrics: Area Under the Curve (AUC), Accuracy, Precision, Recall, F1-score, and Kappa index. After 1000 iterations of each algorithm, the six metrics' average values were calculated (Table s1). The entropy weighting method was employed in this study to nominate weights to each of the six indicators objectively. The internal variabilities of six different measurements were used to calculate the weight coefficients (Table 1). Any performance parameter was multitudinous by the equivalent weight coefficient to arrive at the entropy-weighted composite score for predicting the F^- groundwater model (Fig. 5).

Consistent with slight model divergence inside clusters and huge model variance across clusters, the candidate fundamental learners who perform the best, as indicated by the entropy-weighted composite score, are included in the Stacking framework as base learners. Tree-Bag, RF, Decision tree, Extra-tree, and XGBoost were chosen as base/fundamental learners for this study's F^- forecasting models, (see Fig. 2).

Table 1
Weighting results of performance metrics computed through an entropy-based approach.

Performance Metrics	Information Entropy Value	Information Utility Value	Weight (%)
AUC	0.86	0.14	25.220
Accuracy	0.86	0.14	25.122
Precision	0.79	0.21	25.727
Recall	0.76	0.24	25.909
F1-score	0.79	0.21	25.228
Kappa	0.68	0.32	25.221

Base learners should be varied to gather various data components and guarantee strong performance. To prevent overfitting, the meta-learner, which aggregates the predictions of base learners, should function well while keeping a basic structure (Chatzimparmpas et al., 2021b; Chatzimparmpas et al., 2021a; Daza et al., 2023). By choosing a sample data set to verify its probabilistic computations, the Naive Bayes is a mathematical technique that finds the optimum classification for data in various problem domains with assured accuracy (Yang, 2018).

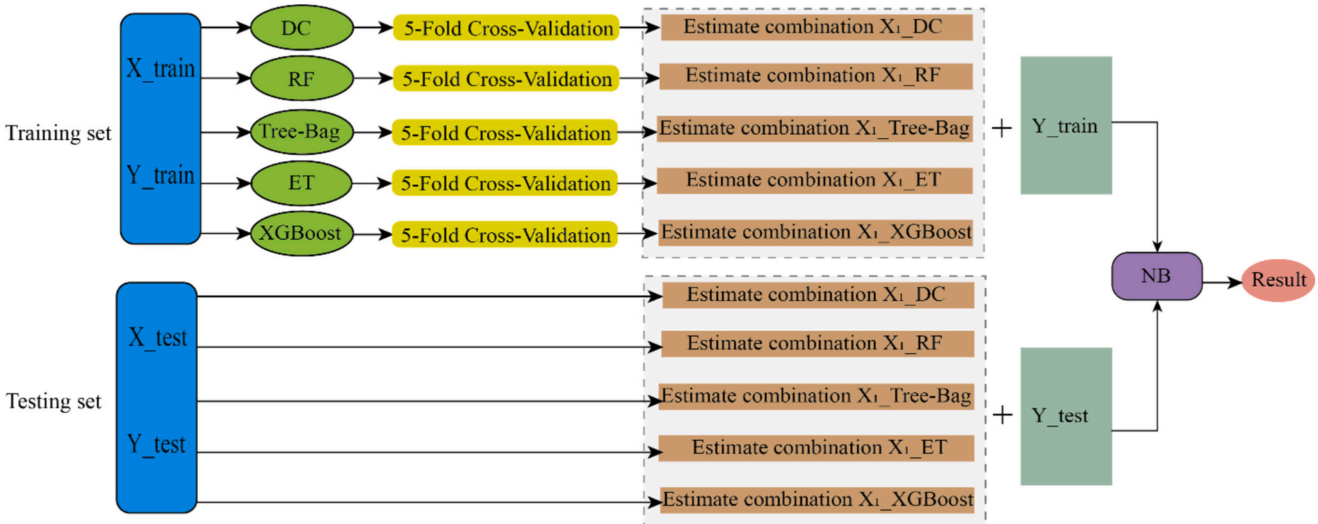


Fig. 2. The stacking ensemble model flowchart.

Thus, by combining Tree-Bag, RF, Decision tree, Extra-tree, and XGBoost as the fundamental learners and Naïve Bayes as the meta-learner, the enhanced stacking ensemble learning model was produced. A quintuple-fold cross-validation procedure was made possible by the approach, which involved splitting the training dataset into five subgroups. The predictions were combined through repeated training and forecasting of the base learners. The ensemble was created by constructing a new test dataset and calculating the average of these five rounds' predictions. Every constituent base learner underwent this process repeatedly. The second-layer meta-learner then used the unique prognostic outputs from each base learner's training phase and the outputs from the test phase as input data. The stacking ensemble model's final result was generated by training this meta-learner. Finally, eleven predictor variables physically adjacent to each other were subjected to the stacking ensemble model. However, much geoengineering research has shown that the stacking model integrates multiple machine-learning models, making it a better option than the approaches listed (Cao et al., 2023).

2.5. Predicted values

2.5.1. Pre-training data set

A statistical examination of the groundwater F^- contents revealed a wide range, from 0.01 to 588 mg/L. The median fluoride content was between 0.63 and 3.22 mg/L. A significant health risk was indicated by drinking water fluoride levels exceeding the WHO limit of 1.5 mg/L in 32 % of the 6270 samples. Significant geographical variations in groundwater fluoride concentrations were identified with a 541.9 variation coefficient, meaning substantial variability across the samples. Binary target variables were frequently used in modeling to create probability because of the highly heterogeneous distribution of dissolved fluoride. 1.5 mg/L was the intermediate threshold for the recorded concentrations, which were transformed into binary target variables as 0 and 1 and used as the dependent variable for additional evaluation and modeling. It was coded as 0 for samples whose aggregations were at or below the threshold and 1 for those with concentrations above it. Concentrations below or equal to the threshold are represented by samples encoded as 0, and concentrations beyond the threshold are represented by samples encoded as 1.5. The eleven components (Table s2) that comprise the primary physiochemical factors governing the enrichment of F^- in groundwater were comprehensively investigated in this work.

2.5.2. Predicted data

To identify locations at higher risk of fluoride contamination, the model used the preprocessed data and the major physiochemical parameters to forecast the possibility of fluoride concentrations surpassing the WHO limit. The model included multiple cutting-edge machine-learning algorithms to guarantee accuracy and consistency in its forecasts. The model's accuracy in capturing the regional variability and possible health hazards associated with fluoride in groundwater was confirmed by validating the projected data against observed fluoride concentrations. This predictive method offers vital insights into groundwater quality and public health consequences, which helps improve management and mitigation measures, particularly in areas with small sample sizes.

2.6. Models predictions

The stacking ensemble learning framework's effectiveness depended on selecting base learning models that aligned with the ideas of higher performance and algorithmic diversity (Yaseen, 2023). These models were integral to the framework. Consequently, a total of fifteen effective ML models (Fig. 3) were selected as candidate base learners for this study: k-Nearest Neighbors (KNN), Multi-layer Perceptron (MLP), Logistic Regression (LR), Linear Discriminant Analysis (LDA), support vector machines (SVM), Light-Gradient Boosting (LGBM), Naïve Bayes (NB), Decision Tree (DT), CatBoost, Extreme Gradient Boosting (XGBoost), Gradient Boosting Decision Tree (GBDT), Bagging Decision Tree (Tree-Bag), Random Forest (RF), Extremely Randomized Trees (ET), and Adaptive Boosting Classifier (AdaBoost). With information about groundwater F^- concentrations as response variables and an extensive collection of 11 physiochemical parameters as predictor variables, various prediction models aimed for groundwater samples that show after F^- concentrations of >1.5 mg/L were developed.

3. Results and discussion

3.1. Hydrogeochemical characteristics of Africa

The continent's varied climate and geology influence the hydrogeochemistry and elevated fluoride of Africa's subsurface water. The combination of volcanic, sedimentary, and metamorphic rocks that make up the continent's geological structure has varying effects on subsurface water mineralization. Minerals like fluorite (CaF_2), which contain fluoride, dissolve in volcanic environments, and there is a noticeable increase in fluoride levels in some areas, such as the East African Rift Valley, interactions between water and rock lead to higher fluoride concentrations in groundwater (Podgorski and Michael, 2022).

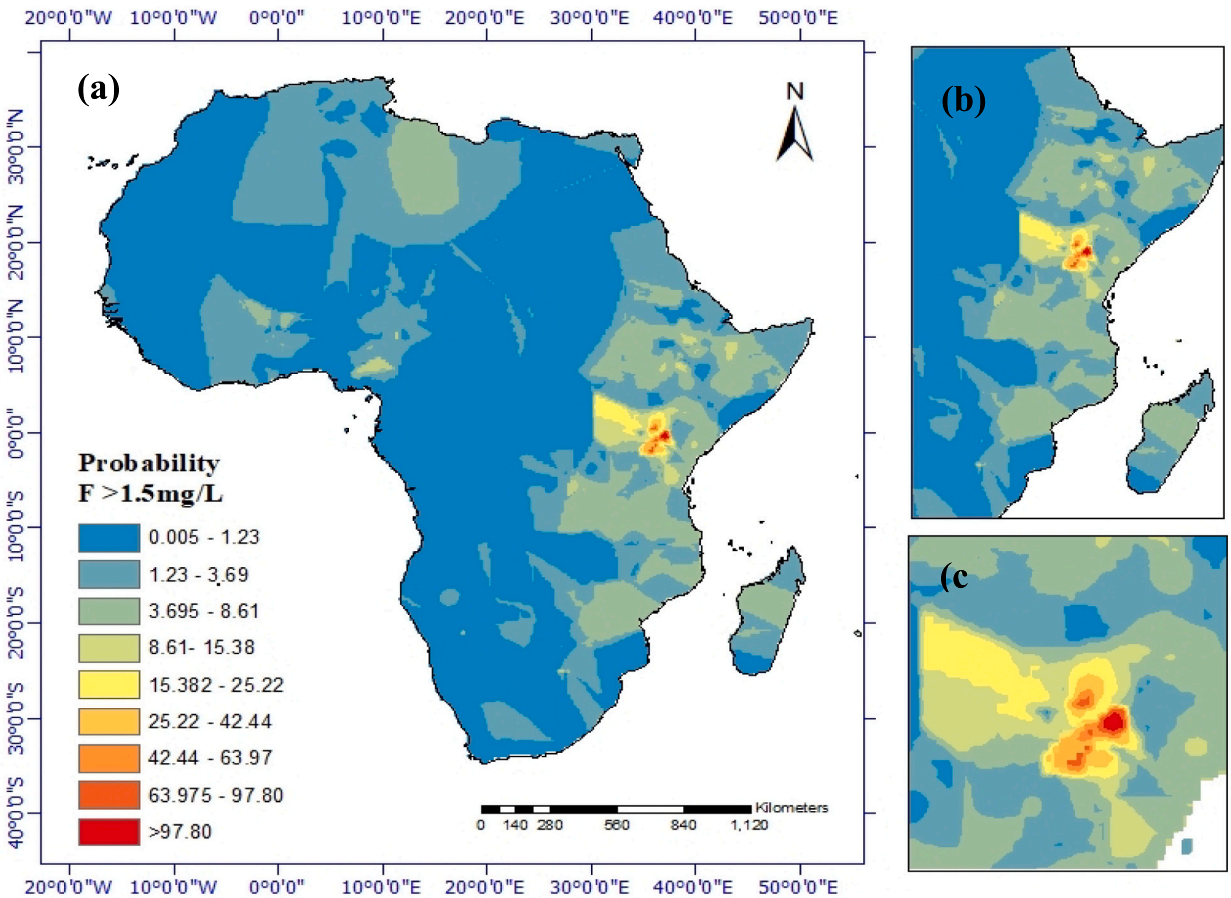


Fig. 3. Africa's continuous probability map for F^- concentrations >1.5 mg/L, using ordinary kriging interpolation.

In arid and semi-arid regions, soil alkalinity mobilizes additional fluoride ions, increasing their concentration in groundwater (Ijumulana et al., 2021; Ligate, 2023). Fluoride mobilization into subsurface water systems is highly dependent on the rate of water-rock interactions, which are influenced by climate differences from desert to tropical locations (Podgorski and Michael, 2022).

The pH and composition of the soil, in addition to anthropogenic sources like mining and agriculture, are essential determinants of the hydrogeochemistry of African groundwater. In East Africa, especially in the East African Rift Valley, fluoride concentrations are in general above 10 mg/L as a consequence of volcanic activity (Mohamud Hersi, 2003; Näslund and Snell., 2005; Jesse, 2023). In West Africa, the F^- levels in general lie between 0.5 and 10.3 mg/L, though studies in Nigeria reported an average of about 1.2 mg/L (Ayenew, 2008; Goyit et al., 2018). Likewise in South Africa, the concentration ranges from approximately 0.2 to 65.9 mg/L, levels as high as 5 mg/L have been recorded in some mining areas (Mccaffrey and Willis, 1998). In North Africa 50 % of the affected areas with a fluoride level usually at 3.19 mg/L, especially in arid regions where due to high evaporation rates concentrate fluoride (El-Said et al., 2010).

Since alkaline soils are prevalent in dry and semi-arid areas, they increase fluoride ion mobility, which raises groundwater concentrations (Podgorski and Michael, 2022). Seasonal and long-term fluctuations in the water table concentrate fluoride much more during dry seasons. A mix of naturally occurring and human-influenced factors necessitate customized management strategies to mitigate the health risks of high fluoride intake (Ijumulana et al., 2021; Ligate, 2023). Alkaline conditions in arid and semi-arid regions of Africa aggravate fluoride pollution in Nigeria, especially as a result of anthropogenic activities such as industrial discharge and landfill operations (Goyit et al., 2018; Andesikuteb Yakubu et al., 2020). This results in concentrations higher than the 1.5 mg/L WHO recommended limit, posing severe concerns to public health, particularly in places with high population density. These factors further reflect the need for tailored mitigation strategies to fix fluoride pollution across different climatic and geological settings (Shube et al., 2023). In countries like Tanzania, Kenya, Ethiopia, Nigeria, Malawi, and Djibouti, conditions like skeletal and dental fluorosis are common and can lead to joint discomfort and discolored teeth (Smedley et al., 2002; Rango et al., 2012; Kut et al., 2016; Shube et al., 2023).

Research has indeed indicated that the quality of groundwater can alter with seasons and is relatively more pronounced in arid and semi-arid areas compared to other regions (Araya et al., 2022; Podgorski and Michael, 2022). The sampling during the dry season for the determination of fluoride levels is essential for an appropriate determination because it gives a better definition of the effect of evaporation on the chemistry of the groundwater (Subba Rao et al., 2022). Therefore, samples collected during this period can be of great use and importance regarding not only the dynamic behavior of fluoride contamination but also its potential health implications on populations that depend on groundwater for drinking. Most of the samples in our study were collected during the dry season to make more reliable estimations of fluoride concentration and to provide critical insights into the geochemical processes involved.

3.2. Evaluation of the model performance

Metrics like area under the curve (AUC), accuracy, precision, and recall were used to assess the model's performance, these metrics give insight into the model's performance, in terms of distinguishing between correct and incorrect predictions, putting particular emphasis on correctly identifying high fluoride concentration. The basis of such calculations originally comes from the results of the confusion matrix (Table 2). AUC values of 0.868 were found for groundwater F^- forecasts, suggesting a solid ability to discriminate. Compared to the existing models in such similar studies, the given AUC score indicates a high improvement in predictive performance for the geochemical data (Islam and Shahjalal, 2019; Kouadri et al., 2021). The accuracy numbers indicated a reasonable overall correctness of the model since they were above 0.8. The accuracy and recall of the groundwater F^- forecasting were 0.792 and 0.769, respectively, demonstrating a replacement between capturing all constructive data and producing precise constructive predictions. According to the performance metrics discovered, the stacking ensemble learning model shown remarkable achievement in predicting the issue of F^- in groundwater.

3.3. Expectation of F^- mapping

Predicted probability maps, with a geographical resolution of 1 km^2, showed the various probabilities of high fluoride concentrations in groundwater throughout different regions of Africa. It indicated that probabilities of having high levels of fluoride are high only in eastern and western parts of Africa, which corresponds with the data as seen in Fig. 3a. The large area known as the high F^- risk zone, which covered roughly 20–35 % of the Eastern region and 10 % in the Western region, is at risk of having fluoride levels above WHO recommendations for the entire Continent when a likelihood threshold of 1.5 ($F^- \geq 1.5$ mg/L) is applied. Significant regions exceeding a likelihood threshold of 1.5 for exceptionally F^- in groundwater were prominently displayed by countries like South Africa, Tanzania, Nigeria, Ethiopia, Ghana, Kenya, Mauritania, Botswana, and Egypt. In contrast, the central domains of the continent displayed relatively mild predicted probabilities (<1.5). Large areas of Kenya and Ethiopia, with the northernmost capital, were expected to have low possibilities of elevated F^- in groundwater. The F^- hazard model's projections were combined to establish the population assessment that could be discovered with extreme F^- in groundwater (Fig. 3b & c). The map displays the likelihood of groundwater F^- concentrations surpassing 1.5 mg/L in different parts of Africa. The nations of South Africa, Tanzania, Nigeria, Ethiopia, Ghana, Kenya, Mauritania, Botswana, and Egypt have been shown to have significant fluoride concentrations. These places typically correlate with some of the continent's most populous areas. The map shows notable places in these high-fluoride zones where the population density is higher than 100 km^2. According to a study conducted in Africa, drinking water with excessive F^- levels puts approximately 80 million people in danger (Onipe et al., 2020; Ijumulana et al., 2021). >8.5 million individuals in Ethiopia alone suffer from skeletal and dental fluorosis brought on by excessive groundwater F^- levels (Tuinhof et al., 2011), and >3 million in the Meru and Arusha regions, Tanzania (Ijumulana et al., 2021). This calls for urgent public health intervention and better management of the groundwater in these high-risk areas.

3.4. Impact evaluation of the predictor factors

Eleven predictor variables were shown to have a significant effect on the groundwater F^- distribution, including HCO_3^-, Ca^{2+}, Mg^{2+}, SO_4^{2-}, K^+, EC, Na^+, NO_3^-, TDS, pH, and Cl^-, Stacking Ensemble model classifier was used to regulate which chemical factors were most important in forecasting particular water pollution conditions in our F^- contamination groundwater research. When utilized as a feature in the Stacking model, Fig. 4 shows the average reduction in tree impurity for each chemical parameter, where alteration in Na^+ levels are highly predictive parameters in infiltration water contamination in the study area, Na^+ levels are the most relevant factor in this regard (13 %) because of their relationships to geological origins and water hardness, HCO_3^- (11 %), Ca^{2+}, Cl^-, (7 %), respectively. Mg^{2+}, NO_3^-, K^+, EC (6 %), respectively, and SO_4^{2-}(5 %) also have essential roles in categorization, contributing 65 % to groundwater fluoride distribution in the study area. Likewise, TDS, pH, and Cl^- have the lowest impact on model decisions (Gupta et al., 2021) (Fig. 4). This is because their correlations with expected conditions are weaker or their variability is more minor (Sakaa et al., 2022).

TDS, pH, and Cl^- play significant yet minor roles when evaluating water quality. It was revealed that the primary origin of dissolved F^- is dissolved F^-bearing minerals, which are mainly leached by groundwater to produce dissolved F^- enrichment (Onipe et al., 2021). This implies that the visibility of TDS, pH, and Cl^- in the groundwater does not elevate fluoride in the study area. (Smedley et al., 2002; Kut et al., 2016; Abba et al., 2017; Goyit et al., 2018) Reveal that the study area consisted of fluoride-containing minerals such as apatite, cryolite, and fluorite, which dissolve and release fluoride into the water, which is commonly the reason for high Na^+, HCO_3^-, Cl^-, Mg^{2+}, SO_4^{2-}, and K^+ levels in groundwater. In particular, higher quantities of calcium (Ca^{2+}) are attributed to fluorite or calcium fluoride. Its origins may include other geological materials and geochemical processes occurring in the regions (Rango et al., 2012; Podgorski and Michael, 2022). This finding also calls for emphasizing those geochemical indicators directly associated with fluoride-bearing minerals in assessing groundwater quality in Africa.

3.5. Distribution of fluoride in groundwater across the African continent

Groundwater across many regions of the African continent naturally contains fluoride due to weathering, geology, high rates of evaporation, infrequent rainfall recharge, and human activity. Due to geological factors that elevated fluoride in groundwater, those countries include Ethiopia (Ayenew, 2008; Assefa Alemu et al., 2015; Van Landtschoote,

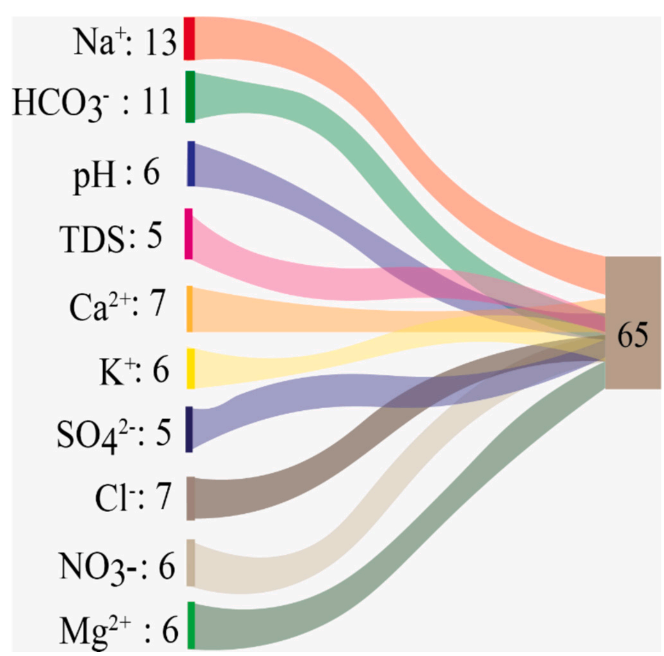


Fig. 4. The groundwater F^- prediction model's candidate base learners' entropy-weighted composite scores.

Table 2
Confusion matrix for forecast outcomes.

All condition		Actual classes	
		True	False
Predicted classes	True	TP	FP
	False	FN	TN

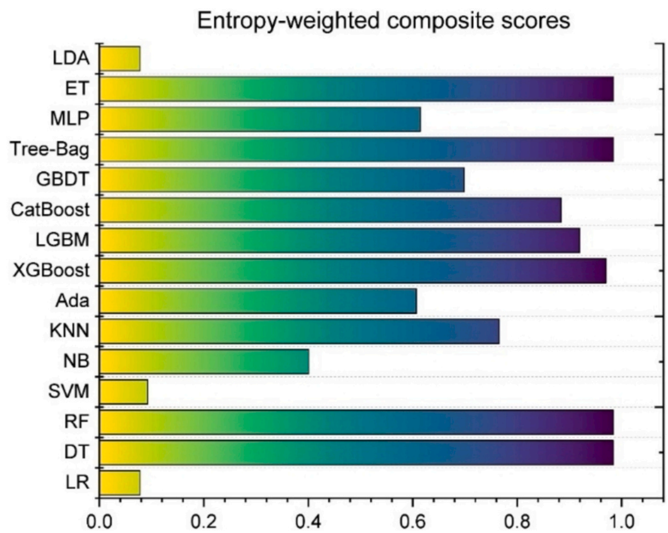


Fig. 5. The groundwater F^- prediction model's candidate base learners' entropy-weighted composite scores.

2016; Ayele et al., 2023; Shube et al., 2023), Nigeria (Dibal et al., 2017; Ankidawa et al., 2019; Mukkafa et al., 2019; Jamiu, 2020), Tanzania (Smedley et al., 2002; Nakayama et al., 2022; Said et al., 2022; Nakaya et al., 2023), and South Africa (Mccaffrey and Willis, 1998). The release and mobility of fluoride in groundwater are closely related to geological formations, weathering processes, and specific environmental conditions. Fluoride in groundwater has a close relation with geological formation, weathering process, and particular environmental condition (Smedley et al., 2002; Rango et al., 2012). Due to silicate weathering and erosion, high fluoride levels have been witnessed in countries like Ghana and Kenya (in certain areas), where the underground water is undersaturated with calcite and fluorite and supersaturated with albite (Tay et al., 2017; Wadira, 2020; Zango et al., 2021; Sunkari et al., 2018, 2023).

Likewise, excessive evaporation rates caused high fluoride levels in groundwater in nations including Mauritania, Botswana, and Egypt. (El-Said et al., 2010; Kane et al., 2012; Tshepo, 2017). However, the map also revealed that over 20–35 % of the continent's population may be at risk of fluoride contamination levels more than the WHO standard limit ($\geq$1.5 mg/L), as seen in Fig. 3, with the majority of those affected living in the Eastern part of the continent (including; Kenya, Ethiopia, Malawi, Tanzania, etc.), only around 10 % was affected in some Western part of the continent such as Ghana, and Nigeria. Previous research shows that the fluoride concentration in some parts of the continent was strongly correlated with the two main kinds of groundwater types, Ca-Mg-HCO_3 and Na-HCO_3. (Ayenew, 2008; Dibal et al., 2016; Tay et al., 2017; Kambuku et al., 2018; Okwir et al., 2023), Na-HCO_3 enrichment and Mixed Ca-Na-HCO_3 (Tay et al., 2017; Sunkari et al., 2018; Tolera et al., 2020; Said et al., 2022), which the elevation of fluoride in the study area causes severe dental fluorosis in children and young people, especially in the African Rift Valley, in the absence of gender discrimination, *Genu Valgum* in children.

3.6. Regularities distribution of geogenic fluoride groundwater in Africa

Groundwater F^- levels and distribution are controlled by the geological formations found across the continent (Fig. s3). Crucial geological elements that are important for groundwater chemistry are shown on the map, including cratons, sedimentary basins, and volcanic zones. The elevated fluoride is found in groundwater in the East African Rift Valley, where volcanic activity dissolves fluorite and other minerals containing fluoride (Ligate, 2023). According to Awaleh et al. (2020) and Gebeyehu et al. (2022), sedimentary basins in the Congo and Niger display unique hydrogeochemical features that influence fluoride levels. However, Fluoride contamination's spatial distribution (Fig. 3) can be explained by the continent's diverse geological formations, which include more recent sedimentary deposits and older Precambrian rocks. The African Rift Valley and other regions with high F^- concentrations provide serious health hazards (Ijumulana et al., 2021; Podgorski and Michael, 2022). Usually, this is notably the case in East and West Africa. Therefore, understanding the geological context is essential for assessing and managing fluoride risks in groundwater across Africa (Rango et al., 2012; Nijsten et al., 2018). Future research should focus on the integration of geochemical modeling into socio-economic assessment, to come up with the contribution of fluoride contamination given its public health and environmental impacts across Africa (Nijsten et al., 2018).

4. Conclusion

The elevated fluoride in groundwater was observed in African nations like South Africa, Tanzania, Nigeria, and Ethiopia, due to the combination of geogenic and anthropogenic factors including; weathering, geology, high evaporation rates, occasional rainfall, and anthropogenic activity; such as mining and industrial activities. Approximately 10 % of people in Western areas of the continent, such as Ghana and Nigeria, and over 20–35 % of people in Eastern regions, such as Kenya, Ethiopia, Malawi, and Tanzania, may be at risk of fluoride contamination above the WHO guideline level. In this study, we implemented an advanced Stacking Ensemble Learning model to map fluoride contamination in the groundwater of Africa. The performance of the model, as obtained by an AUC value of 0.86, reflects the high prediction capability for the areas with high levels of fluoride. Our results found that high fluoride concentrations affected approximately 80 million people, which increases the risk of dental and skeletal fluorosis, especially in semi-arid and arid regions.

The four critical physiochemical parameters were found, including Na^+, HCO_3^-, Ca^{2+}, and Cl^-. Due to their proximate connection to the geochemical processes influencing fluoride solubility and mobility in groundwater, these values have considerable explanatory value. Elevated TDS and Cl^- generally indicate elevated fluoride concentrations, particularly in regions with fluoride-containing minerals and sediments. Fluoride solubility is also influenced by Mg^{2+} concentration through intricate geological interactions. The study highlighted the significant roles of the continent's climate and geological diversity in shaping and influencing the hydrogeochemistry of African subsurface water; volcanic environments raise pH, soil composition, water-rock interactions, and fluoride levels. In dry and semi-arid regions, alkaline soils have increased fluoride ion mobility, which calls for specialized management strategies to reduce health risks. The SEL model was quite informative, however, the interaction of environmental and human variables in a dynamic way makes comprehension of the contamination of groundwater quite challenging. At the same time, the model takes into consideration the impacts of stress levels, whereby it identifies the populations that are most vulnerable because of contamination in the groundwater. This can be a useful tool regarding targeted preventive measures. These projections can be utilized by policymakers to decide on priorities of interventions in the affected areas as far as resources for the management of groundwater and public health protection are concerned. To focus on collecting primary data. More emphasis needs to be placed on finding more efficient deep learning and neural network models that scale up the performance of the model. Advanced techniques using these may improve predictive accuracy and perhaps allow a more informed analysis of nonlinear interactions in environmental contexts underpinning groundwater management in a sustainable way across Africa.

CRediT authorship contribution statement

Usman Sunusi Usman: Writing – review & editing, Writing –

original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yousif Hassan Mohamed Salh:** Validation, Conceptualization. **Bing Yan:** Writing – review & editing, Supervision. **Jean Pierre Namahoro:** Software. **Qian Zeng:** Resources, Formal analysis. **Ismaila Sallah:** Validation.

Consent to participate

All authors reviewed and approved the final manuscript.

Consent to publish

All authors approved this for publication.

Ethics approval

Not applicable.

Funding

This work was financially supported by the National Natural Science Foundation of China (Grants No. 41807201).

Declaration of competing interest

The authors have no relevant financial or non-financial interests to disclose.

Acknowledgments

The authors gratefully acknowledge the help of editors and reviewers in the future.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.177693.

Data availability

Data will be made available on request.

References

Abba, H.T., Wan Hassan, W.M.S., Saleh, M.A., Aliyu, A.S., Ramli, A.T., 2017. Terrestrial gamma radiation dose (TGRD) levels in northern zone of Jos plateau, Nigeria: statistical relationship between dose rates and geological formations. Radiat. Phys. Chem. 140, 167–172.

Abba, S.I., Yassin, M.A., Mubarak, A.S., Shah, S.M.H., Usman, J., Oudah, A.Y., Naganna, S.R., Aljundi, I.H., 2023. Drinking water resources suitability assessment based on pollution index of groundwater using improved explainable artificial intelligence. Sustainability 15 (21), 1–23. https://doi.org/10.3390/su152115655.

Adnan, R.M., Dai, H.L., Mostafa, R.R., Islam, A.R.M.T., Kisi, O., Heddam, S., Zounemat-Kermani, M., 2023. Modelling groundwater level fluctuations by ELM merged advanced metaheuristic algorithms using hydroclimatic data. Geocarto Int. 38 (1). https://doi.org/10.1080/10106049.2022.2158951.

Al-Amry, A.S., Habtoor, A., Qatan, A., 2020. Hydrogeochemical characterization and environmental impact of fluoride contamination in groundwater from Al-Dhala Basin, Yemen. Electronic Journal of Univerisity of Aden for Basic and Applied Sciences (EJUA-BA) 1 (1), 30–38.

Alharbi, T., El-Sorogy, A.S., 2023. Health risk assessment of nitrate and fluoride in the groundwater of Central Saudi Arabia. Water 15 (12), 1–12. https://doi.org/10.3390/w15122220.

Alkindi, K.M., Mukherjee, K., Pandey, M., Arora, A., Janizadeh, S., Pham, Q.B., Anh, D. T., Ahmadi, K., 2022. Prediction of groundwater nitrate concentration in a semiarid region using hybrid Bayesian artificial intelligence approaches. Environ. Sci. Pollut. Res. 29 (14), 20421–20436. https://doi.org/10.1007/s11356-021-17224-9.

Amiri, V., Berndtsson, R., 2020. Fluoride occurrence and human health risk from groundwater use at the west coast of Urmia Lake. Iran. *Arabian Journal of Geosciences* 13 (18), 1–23.

Anandhi, A., Karunanidhi, D., Sankar, G.M., Panda, S., Kannan, N., 2022. A framework for sustainable groundwater management. Water (Switzerland) 14 (21), 1–13. https://doi.org/10.3390/w14213416.

Andesikuteb Yakubu, A., Giwa Yakubu, C., Mohammed Tukur, Y., Lekwot Vivan, E., 2020. Analysis of groundwater quality around some dumpsites in Jos Metropolis, plateau state, Nigeria. Kaduna State University Environmental Science Journal (KESJ) 1 (2), 1–18. https://www.researchgate.net/publication/344074349.

Ankidawa, B.A., Audu, D.S., Inuwa, A.M., Vanke, I., 2019. Assessment of fluoride concentration in groundwater in Hong and environs, Adamawa state, northeastern Nigeria. Nigeria Journal of Engineering Science and Technology Research 5 (2), 28–39.

Arabameri, A., Santosh, M., Moayedi, H., Tiefenbacher, J.P., Pal, S.C., Nalivan, O.A., Costache, R., Ahmed, N., Hoque, M.A.-A., Chakrabortty, R., Cerda, A., 2022. Application of the novel state-of-the-art soft computing techniques for groundwater potential assessment. Arab. J. Geosci. 15 (10), 1–24. https://doi.org/10.1007/s12517-021-09005-y.

Araya, D., Podgorski, J., Kumi, M., Mainoo, P.A., Berg, M., 2022. Fluoride contamination of groundwater resources in Ghana: country-wide hazard modeling and estimated population at risk. Water Res. 212, 1–10.

Assefa Alemu, Z., Derso Mengesha, S., Assefa Alemayehu, T., Gizaw Serte, M., Weldetinsae Kidane, A., Tesfaye Teklu, K., 2015. Retrospective study of fluoride distribution in Ethiopian drinking water sources. Asian Journal of Applied Science and Engineering 4 (2), 126–136 doi:10.18034.

Aullón Alcaine, A., Schulz, C., Bundschuh, J., Jacks, G., Thunvik, R., Gustafsson, J.P., Mörth, C.M., Sracek, O., Ahmad, A., Bhattacharya, P., 2020. Hydrogeochemical controls on the mobility of arsenic, fluoride and other geogenic co-contaminants in the shallow aquifers of northeastern La Pampa Province in Argentina. Sci. Total Environ. 715, 1–17. https://doi.org/10.1016/j.scitotenv.2020.136671.

Awaleh, M.O., Boschetti, T., Adaneh, A.E., Daoud, M.A., Ahmed, M.M., Dabar, O.A., Soubaneh, Y.D., Kawalieh, A.D., Kadieh, I.H., 2020. Hydrochemistry and multi-isotope study of the waters from Hanlé-Gaggadé grabens (Republic of Djibouti, east African rift system): A low-enthalpy geothermal resource from a transboundary aquifer. Geothermics 86, 1–19.

Ayele, S., Workneh, Z.G., Meshesha, B.T., 2023. Water challenges of Dire Dawa city of Ethiopia: an assessment of urban growth and groundwater supply. Environmental Challenges 12, 1–14.

Ayenew, T., 2008. The distribution and hydrogeological controls of fluoride in the groundwater of central Ethiopian rift and adjacent highlands. Environ. Geol. 54 (6), 1313–1324. https://doi.org/10.1007/s00254-007-0914-4.

Balogun, M.A., Anumah, A.O., Adegoke, K.A., Maxakato, N.W., 2022. Variable water quality of domestic wells emphasizes the need for groundwater quality monitoring and protection: Stinkwater, Hammanskraal. Gauteng. Water SA 45 (2), 216–224. https://doi.org/10.1007/s10661-022-09805-z.

Barzegar, R., Asghari Moghaddam, A., Adamowski, J., Fijani, E., 2017. Comparison of machine learning models for predicting fluoride contamination in groundwater. Stoch. Env. Res. Risk A. 31 (10), 2705–2718. https://doi.org/10.1007/s00477-016-1338-z.

Begg, G.C., Griffin, W.L., Natapov, L.M., O'Reilly, S.Y., Grand, S.P., O'Neill, C.J., Hronsky, J.M.A., Djomani, Y.P., Swain, C.J., Deen, T., Bowden, P., 2009. The lithospheric architecture of Africa: seismic tomography, mantle petrology, and tectonic evolution. Geosphere 5 (1), 23–50. https://doi.org/10.1130/GES00179.1.

Bhandari, P., Banjara, M.R., Singh, A., Kandel, S., Rawal, D.S., Pant, B.R., 2021. Water quality status of groundwater and municipal water supply (tap water) from bagmati river basin in Kathmandu valley. Nepal. Journal of Water Sanitation and Hygiene for Development 11 (1), 102–111.

Bura, B., Goni, I.B., Sheriff, B.M., Gazali, A.K., 2018. Occurrence and distribution of fluoride in groundwater of Chad formation aquifers in Borno state. Nigeria. International Journal of Hydrology 2 (4), 528–537. https://doi.org/10.15406/IJH.2018.02.00121.

Cao, W., Fu, Y., Cheng, Y., Zhai, W., Sun, X., Ren, Y., Pan, D., 2023. Modeling potential arsenic enrichment and s using stacking ensemble learning in the lower Yellow River plain, China. J. Hydrol. 625, 1–15.

Chatzimparmpas, A., Martins, R.M., Kucher, K., Kerren, A., 2021a. Empirical study: Visual analytics for comparing stacking to blending ensemble learning. In: Proceedings - 2021 23rd International Conference on Control Systems and Computer Science Technologies, CSCS 2021, pp. 1–8. https://doi.org/10.1109/CSCS52396.2021.00008.

Chatzimparmpas, A., Martins, R.M., Kucher, K., Kerren, A., 2021b. StackGenVis: alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. IEEE Trans. Vis. Comput. Graph. 27 (2), 1547–1557.

Chen, K., Liu, Q., Yang, T., Ju, Q., Zhu, M., 2024. Risk assessment of nitrate groundwater contamination using GIS-based machine learning methods: A case study in the northern Anhui plain, China. J. Contam. Hydrol. 261, 1–10. https://doi.org/10.1016/j.jconhyd.2024.104300.

Chowdhury, A., Adak, M.K., Mukherjee, A., Dhak, P., Khatun, J., Dhak, D., 2019. A critical review on geochemical and geological aspects of fluoride belts, fluorosis and natural materials, and other sources for alternatives to fluoride exposure. J. Hydrol. 574, 333–359.

Cui, S., Yin, Y., Wang, D., Li, Z., Wang, Y., 2021. A stacking-based ensemble learning method for earthquake casualty prediction. Applied Soft Computing Journal 101, 1–16. https://doi.org/10.1016/j.asoc.2020.107038.

Daffi, R., Ahuchaogu, I.I., Lagasi, J.E., 2022. Assessment of the Physico-chemical and bacteriological parameters of groundwater for domestic consumption in parts of Langtang north, plateau state, Nigeria. Nigerian Research Journal of Engineering and Environmental Sciences 7 (3), 476–485.

Daza, A., Arroyo-Paz, Bobadilla, J., Apaza, O., & Pinto, J., 2023. Stacking ensemble learning model for predict anxiety level in university students using balancing methods. Informatics in Medicine Unlocked 42, 1–18. https://doi.org/10.1016/j.imu.2023.101340.

Dibal, H., Schoeneich, K., Lar, U., Garba, I., Lekmang, I., Daspan, R., 2016. Hydrogeochemical appraisal of fluoride in groundwater of Langtang area, plateau state. Nigeria. Global Journal of Geological Sciences 14 (1), 1–23. https://doi.org/10.4314/gjgs.v14i1.3.

Dibal, H.U., Dajilak, W.N., Lekmang, I.C., Nimze, L.W., Yenne, E.Y., 2017. Seasonal variation in fluoride content in Groundwaters of Langtang area. Northcentral Nigeria. Contemporary Trends in Geoscience 6 (1), 11–27. https://doi.org/10.1515/ctg-2017-0002.

Divina, F., Gilson, A., Goméz-Vela, F., Torres, M.G., Torres, J.F., 2018. Stacking ensemble learning for short-term electricity consumption forecasting. Energies 11 (4), 1–31. https://doi.org/10.3390/en11040949.

Documentation, DataWig, 2022. Machine learning for missing value imputation. Retrieved from. https://github.com/awslabs/datawig.

El-Said, G.F., El-Sadawy, M.M., Moneer, A.A., 2010. Incorporation of fluoride and boron into surface sediments along the Egyptian Mediterranean coast. Egypt. J. Aquat. Res. 36 (4), 569–583.

Ezike, E.C., Chukwunonyerem, S.C., 2022. Hydrogeochemistry and groundwater quality assessment in Ozalla, Nigeria. Irish Journal of Environment and Earth Sciences 6 (3), 14–26.

Gebeyehu, A., Ayenew, T., Asrat, A., 2022. Hydrogeochemistry of the groundwater system of the transboundary basement and volcanic aquifers of the Bulal catchment, southern Ethiopia. J. Afr. Earth Sci. 194, 1–18.

Ghobadi, A., Cheraghi, M., Sobhanardakani, S., Lorestani, B., Merrikhpour, H., 2022. Groundwater quality modeling using a novel hybrid data-intelligence model based on gray wolf optimization algorithm and multi-layer perceptron artificial neural network: a case study in Asadabad plain, Hamedan. Iran. Environmental Science and Pollution 29, 8716–8730. https://doi.org/10.1007/s11356-021-16300-4/Published.

Gosomji, A.D., Sha'ato, R., Adebisi, Q.B., Nanbol, G., 2022. Evaluation of the water quality status of raw and conventionally treated water from three sources in plateau state Nigeria. Chemistry Research Journal 2022 (1), 49–54. www.chemrj.org.

Goyit, M.P., Solomon, O.A., Kutshik, R.J., 2018. Distribution of fluoride in surface and ground water: a case study of Langtang north, plateau state, Nigeria. Int. J. Biol. Chem. Sci. 12 (2), 1057–1067. https://doi.org/10.4314/ijbcs.v12i2.33.

Gupta, D., Hazarika, B.B., Berlin, M., Sharma, U.M., Mishra, K., 2021. Artificial intelligence for suspended sediment load prediction: a review. Environ. Earth Sci. 80 (9), 1–39. https://doi.org/10.1007/s12665-021-09625-3.

Hussein, E.A., Thron, C., Ghaziasgar, M., Bagula, A., Vaccari, M., 2020. Groundwater prediction using machine-learning tools. Algorithms 13 (11), 1–16. https://doi.org/10.3390/a13110300.

Ijumulana, J., Ligate, F., Irunde, R., Bhattacharya, P., Maity, J.P., Ahmad, A., Mtalo, F., 2021. Spatial uncertainties in fluoride levels and health risks in endemic fluorotic regions of northern Tanzania. Groundw. Sustain. Dev. 14, 1–14.

Iqbal, J., Su, C., Wang, M., Abbas, H., Baloch, M.Y.J., Ghani, J., Ullah, Z., Huq, M.E., 2023. Groundwater fluoride and nitrate contamination and associated human health risk assessment in South Punjab. Pakistan. Environmental Science and Pollution Research 30 (22), 61606–61625. https://doi.org/10.1007/s11356-023-25958-x.

Islam, R., Shahjalal, M.A., 2019. Soft voting-based ensemble approach to predict early stage DRC violations, *2019 IEEE 62nd international Midwest symposium on circuits and systems (MWSCAS)*. Dallas, TX, USA 1081–1084. https://doi.org/10.1109/MWSCAS.2019.8884896.

Jamiu, A.A., 2020. Geochemistry of Groundwater of Shallow Coastal Aquifers of Eastern Dahomey Basin, Southwestern Nigeria [PhD thesis].. University of Strathclyde Engineering. http://doi:10.48730/ypyp-dd46.

Jesse, A., 2023. Assessment of Temporal and Spatial Variability of Groundwater in the Mbagathi River Catchment [MSc thesis].. University of Nairobi http://erepository.uonbi.ac.ke/handle/11295/164198.

Kambuku, D., Tsujimura, M., Kagawa, S., Mdala, H., 2018. Corroborating stable isotopic data with pumping test data to investigate recharge and groundwater flow processes in a fractured rock aquifer, Rivirivi catchment. Malawi. Environmental Earth Sciences 77 (6), 1–21. https://doi.org/10.1007/s12665-018-7403-9.

Kane, C.H., Diene, M., Fall, M., Sarr, B., Thiam, A., 2012. Reassessment of the resources of a deep aquifer system under physical and chemical constraints: the Maastrichtian aquifer. Journal of Water Resource and Protection 04 (04), 217–223.

Kouadri, S., Elbeltagi, A., Islam, A.R.M.T., Kateb, S., 2021. Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). Appl Water Sci 11 (12), 1–20. https://doi.org/10.1007/s13201-021-01528-9.

Kut, K.M.K., Sarswat, A., Srivastava, A., Pittman, C.U., Mohan, D., 2016. A review of fluoride in African groundwater and local remediation methods. Groundw. Sustain. Dev. 2, 190–212. https://doi.org/10.1016/j.gsd.2016.09.001.

Li, J., Lu, W., Wang, H., Fan, Y., Chang, Z., 2020. Groundwater contamination source identification based on a hybrid particle swarm optimization-extreme learning machine. J. Hydrol. 584, 1–23.

Li, X., 2020. Green Civilization Human Consensus on Global Collaboration for Sustainable Development, first ed. Beijing Normal University Press. https://doi.org/10.1007/978-981-15-7812-0.

Ligate, F. Josephat, 2023. Hydrogeochemistry of Arsenic, Fluoride, and Other Trace Elements in Groundwater in Northern Tanzania: Occurrence, Distribution, and Impacts on Drinking Water Quality [PhD thesis].. KTH Royal Institute of Technology https://kth.diva-portal.org/smash/get/diva2:1742725/FULLTEXT01.pdf.

Lombard, M.A., Bryan, M.S., Jones, D.K., Bulka, C., Bradley, P.M., Backer, L.C., Focazio, M.J., Silverman, D.T., Toccalino, P., Argos, M., Gribble, M.O., Ayotte, J.D., 2021. Machine Learning Models of Arsenic in Private Wells throughout the Conterminous United States As a Tool for Exposure Assessment in Human Health Studies. *Environ. Sci. Technol. 55* (8), 5012–5023. https://doi.org/10.1021/acs.est.0c05239.

Lopes, D., Barbosa, S., Cruz-Pérez, N., García-Gil, A., Marazuela, M.Á., Rodríguez-Alcántara, J.S., Santamarta, J.C., 2022. Excess of naturally occurring fluoride in groundwater discharge in Macaronesia: Brava Island. Cape Verde. Water 14 (21), 1–8.

Malago, J., 2017. Fluoride levels in surface and groundwater in Africa: A review. American Journal of Water Science and Engineering 3 (1), 1–17. https://doi.org/10.11648/j.ajwse.20170301.11.

Mccaffrey, L.P., Willis, J.P., 1998. Distribution and Causes of High Fluoride Groundwater in the Western Bushveld Area of South Africa [PhD thesis].. University of Cape Town http://hdl.handle.net/11427/9597.

Mia, M.Y., Haque, M.E., Islam, A.R.M.T., Jannat, J.N., Jion, M.M.M.F., Islam, M.S., Siddique, M.A.B., Idris, A.M., Senapathi, V., Talukdar, S., Rahman, A., 2023. Analysis of self-organizing maps and explainable artificial intelligence to identify hydrochemical factors that drive drinking water quality in Haor region. Sci. Total Environ. 904, 1–22. https://doi.org/10.1016/j.scitotenv.2023.166927.

Mohamud Hersi, A., 2003. Hydrogeochemistry of Groundwater in Nairobi Area [MSc thesis].. University of Nairobi http://erepository.uonbi.ac.ke/handle/11295/26431.

Morales-deAvila, H., Gutiérrez, M., Colmenero-Chacón, C.P., Júnez-Ferreira, H.E., Esteller-Alberich, M.V., 2023. Upward trends and lithological and climatic controls of groundwater arsenic, fluoride, and nitrate in Central Mexico. Minerals 13 (9), 1–12.

Mukkafa, S., Ishaku, J.M., Kwami, I.A., Seli, A.B., Abubakar, U.A., Bello, A.M., Abubakar, U., Bappa, M.U., 2019. Occurrence of fluoride in the groundwater of Kaltungo area and environs, north eastern Nigeria. Journal of Geoscience and Environment Protection 07 (08), 167–179. https://doi.org/10.4236/gep.2019.78012.

Mukwati, B.T., 2017. Hydrogeochemistry of the Kasane Thermal Spring. Botswana [MSc thesis].. University of Botswana https://ubrisa.ub.bw/handle/10311/1991.

Mustafa, O., Mahmmud, R., Sracek, O., Seeyan, S., 2023. Geogenic sources of arsenic and fluoride in groundwater: examples from the Zagros Basin, the Kurdistan region of Iraq. Water (Switzerland) 15 (11), 1–19.

Nafouanti, M.B., Li, J., Nyakilla, E.E., Mwakipunda, G.C., Mulashani, A., 2023. A novel hybrid random forest linear model approach for forecasting groundwater fluoride contamination. Environ. Sci. Pollut. Res. 30 (17), 50661–50674. https://doi.org/10.1007/s11356-023-25886-w.

Najwa Mohd Rizal, N., Hayder, G., Mnzool, M., Elnaim, B.M.E., Mohammed, A.O.Y., Khayyat, M.M., 2022. Comparison between regression models, support vector machine (SVM), and artificial neural network (ANN) in river water quality prediction. Processes 10 (8), 1–11. https://doi.org/10.3390/pr10081652.

Nakaya, S., Takada, R., Yasumoto, J., Masuda, H., Yoshitani, J., Shinjoh, R., Lugodisha, I., Komakech, H., 2023. Effect of groundwater residence time on geogenic fluoride release into groundwater in the Mt. Meru slope area, Tanzania, the great Rift Valley, East Africa. J. Contam. Hydrol. 253, 1–22. https://doi.org/10.1016/j.jconhyd.2022.104125.

Nakayama, H., Yamasaki, Y., Nakaya, S., 2022. Effect of hydrogeological structure on geogenic fluoride contamination of groundwater in granitic rock belt in Tanzania. J. Hydrol. 612, 1–16. https://doi.org/10.1016/j.jhydrol.2022.128026.

Näslund, J., Snell., 2005. GIS-Mapping of Fluoride Contaminated Groundwater in Nakuru & Baringo District. Kenya [MSc thesis].. Lulea University of Technology https://urn:nbn:se:ltu:diva-56091.

Nijsten, G.J., Christelis, G., Villholth, K.G., Braune, E., Gaye, C.B., 2018. Transboundary aquifers of Africa: review of the current state of knowledge and progress towards sustainable development and management. Journal of Hydrology: Regional Studies 20, 21–34. https://doi.org/10.1016/j.ejrh.2018.03.004.

Okwir, G., Kumar, S., Pramod, K.S., Gao, H., Njau, K.N., 2023. Conceptualization of groundwater-surface water interaction with evidence from environmental isotopes and hydrogeochemistry in Lake Babati Basin in northern Tanzania. Groundw. Sustain. Dev. 21, 1–21. https://doi.org/10.1016/j.gsd.2023.100940.

Onipe, T., Edokpayi, J.N., Odiyo, J.O., 2020. A review on the potential sources and health implications of fluoride in groundwater of sub-Saharan Africa. Journal of Environmental Science and Health - Part A Toxic/Hazardous Substances and Environmental Engineering 55 (9), 1078–1093.

Onipe, T., Edokpayi, J.N., Odiyo, J.O., 2021. Geochemical characterization and assessment of fluoride sources in groundwater of Siloam area, Limpopo Province. South Africa. Scientific Reports 11 (1), 1–19.

Ouedraogo, I., Defourny, P., Vanclooster, M., 2019. Application of random forest regression and comparison of its performance to multiple linear regression in modeling groundwater nitrate concentration at the African continent scale. Hydrogeol. J. 27 (3), 1081–1098. https://doi.org/10.1007/s10040-018-1900-5.

Parrone, D., Ghergo, S., Frollini, E., Rossi, D., Preziosi, E., 2020. Arsenic-fluoride co-contamination in groundwater: background and anomalies in a volcanic-sedimentary aquifer in Central Italy. J. Geochem. Explor. 217, 1–14.

Podgorski, J., Michael, B., 2022. Global analysis and prediction of fluoride in groundwater. Nat. Commun. 13 (1), 1–9. https://doi.org/10.1038/s41467-022-31940-x.

Prapanchan, V.N., Subramani, T., Karunanidhi, D., 2024. GIS and fuzzy analytical hierarchy process to delineate groundwater potential zones in southern parts of India. Groundw. Sustain. Dev. 25, 1–16. https://doi.org/10.1016/j.gsd.2024.101110.

Priyadarshini, I., Alkhayyat, A., Obaid, A.J., Sharma, R., 2022. Water pollution reduction for sustainable urban development using machine learning techniques. Cities 130, 1–15. https://doi.org/10.1016/j.cities.2022.103970.

Rajan, M., Karunanidhi, D., Subramani, T., Preethi, B., 2024. Evaluation of fluoride contamination in groundwater and its non-carcinogenic health hazards in a drought-prone river basin of South India. Phys. Chem. Earth 136, 1–14. https://doi.org/10.1016/j.pce.2024.103714.

Rango, T., Kravchenko, J., Atlaw, B., McCornick, P.G., Jeuland, M., Merola, B., Vengosh, A., 2012. Groundwater quality and its health impact: an assessment of dental fluorosis in rural inhabitants of the Main Ethiopian rift. Environ. Int. 43 (1), 37–47. https://doi.org/10.1016/j.envint.2012.03.002.

Rao, N.S., Dinakar, A., Kumari, B.K., 2021. Appraisal of vulnerable zones of non-cancer-causing health risks associated with exposure of nitrate and fluoride in groundwater from a rural part of India. Environ. Res. 202, 1–8. https://doi.org/10.1016/j.envres.2021.111674.

Ribeiro, M. H. D. M., da Silva, R. G., Moreno, S. R., Mariani, V. C., & Coelho, L. dos S. (2022). Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. Int. J. Electr. Power Energy Syst., 136, 1–23. doi:https://doi.org/10.1016/j.ijepes.2021.107712.

Sahoo, S., Russo, T.A., Elliott, J., Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. Water Resour. Res. 53 (5), 3878–3895.

Said, M., Komakech, H.C., Mjemah, I.C., Lufingo, M., Munishi, L.K., Kumar, S., 2022. Hydrogeochemical analysis of water quality dynamics under anthropic activities on the southern slopes of Mount Kilimanjaro. Tanzania. Chemistry Africa 5 (5), 1589–1610. https://doi.org/10.1007/s42250-022-00419-z.

Sakaa, B., Elbeltagi, A., Boudibi, S., Chaffaï, H., Islam, A.R.M.T., Kulimushi, L.C., Choudhari, P., Hani, A., Brouziyne, Y., Wong, Y.J., 2022. Water quality index modeling using random forest and improved SMO algorithm for support vector machine in Saf-Saf river basin. Environ. Sci. Pollut. Res. 29 (32), 48491–48508. https://doi.org/10.1007/s11356-022-18644-x.

Sarkar, S., Mukherjee, A., Chakraborty, M., Quamar, M.T., Duttagupta, S., Bhattacharya, A., 2023. Prediction of elevated groundwater fluoride across India using multi-model approach: insights on the influence of geologic and environmental factors. Environ. Sci. Pollut. Res. 30 (11), 31998–32013. https://doi.org/10.1007/s11356-022-24328-3.

Satheeshkumar, S., Thirukumaran, V., Karunanidhi, D., 2024. Modern River Science for Watershed Management GIS and Hydrogeological Application.

Sawangjang, B., Takizawa, S., 2020. Assessment of fluoride intake from rice consumption by using tap water containing fluoride for rice soaking water. Journal of Water and Environment Technology 18 (2), 117–131. https://doi.org/10.2965/JWET.19-084.

Schlüter, Thomas, 2006. *Geological Atlas of Africa with Notes on Stratigraphy, Tectonics, Economic Geology, Geohazards, Geosites and Geoscientific Education of each Country*, 2nd Fourth Coloured Revised Ed.

Shadbahr, T., Roberts, M., Stanczuk, J., Gilbey, J., Teare, P., Dittmer, S., Thorpe, M., Viñas Torné, R., Sala, E., Lió, P., Patel, M., Preller, J., Rudd, J.H.F., Mirtti, T., Rannikko, A.S., Aston, J.A.D., Tang, J., Schönlieb, C.-B., AIX-COVNET Collaboration, 2023. The impact of imputation quality on machine learning classifiers for datasets with missing values. Commun. Med. 3 (1), 1–15. https://doi.org/10.1038/s43856-023-00356-z.

Shaji, E., Sarath, K.V., Santosh, M., Krishnaprasad, P.K., Arya, B.K., Babu, M.S., 2024. Fluoride contamination in groundwater: A global review of the status, processes, challenges, and remedial measures. Geosci. Front. 15, 1–29. https://doi.org/10.1016/j.gsf.2023.101734.

Sharma, M.K., Kumar, M., 2020. Sulphate contamination in groundwater and its remediation: an overview. Environ. Monit. Assess. 192, 1–10. https://doi.org/10.1007/s10661-019-8051-6.

Shube, H., Kebede, S., Azagegn, T., Nedaw, D., Haji, M., Karuppannan, S., 2023. Estimating groundwater flow velocity in shallow volcanic aquifers of the Ethiopian highlands using a geospatial technique. Sustainability (Switzerland) 15 (19), 1–21. https://doi.org/10.3390/su151914490.

Singh, P., 2023. Prediction of groundwater contamination in an open Landll area using a novel hybrid clustering based AI model. Research Square 1–23. https://doi.org/10.21203/rs.3.rs-2803995/v1.

Smedley, P.L., Nkotagu, H., Pelig-Ba, K., Macdonald, A.M., Tyler-Whittle, R., Whitehead, E.J., Kinniburgh, D.G., 2002. *Fluoride in groundwater from high-fluoride areas of Ghana and Tanzania*. Minimising fluoride in drinking water in problem aquifers, R8033. Phase I final report, 1–74. www.thebgs.co.uk.

Subba Rao, N., Das, R., Sahoo, H.K., Gugulothu, S., 2024. Hydrochemical characterization and water quality perspectives for groundwater management for urban development. Groundw. Sustain. Dev. 24, 1–16. https://doi.org/10.1016/j.gsd.2023.101071.

Subba Rao, N., Dinakar, A., Sun, L., 2022. Estimation of groundwater pollution levels and specific ionic sources in the groundwater, using a comprehensive approach of geochemical ratios, pollution index of groundwater, unmix model and land use/land cover – A case study. J. Contam. Hydrol. 248, 1–18. https://doi.org/10.1016/j.jconhyd.2022.103990.

Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S.M., Wang, J., Lin, C., Cristea, N., Tong, D., Carande, W.H., Ma, X., Rao, Y., Bednar, J.A., Tan, A., Wang, J., Purushotham, S., Gill, T.E., Chastang, J., Howard, D., Holt, B., et al., 2022. A review of earth artificial intelligence. *Computers and geosciences*, A review of earth artificial intelligence. Comput. Geosci. 159, 1–16. https://doi.org/10.1016/j.cageo.2022.105034.

Sunkari, E.D., Abangba, T., Ewusi, A., Tetteh, S.E.K., Ofosu, E., 2023. Hydrogeochemical evolution and assessment of groundwater quality for drinking and irrigation purposes in the Gushegu municipality and some parts of east Mamprusi District. Ghana. Environmental Monitoring and Assessment 195 (1), 1–25.

Sunkari, E.D., Adams, S.J., Okyere, M.B., Bhattacharya, P., 2022. Groundwater fluoride contamination in Ghana and the associated human health risks: any sustainable mitigation measures to curtail the long term hazards? Groundw. Sustain. Dev. 16, 1–11.

Sunkari, E.D., Zango, M.S., Korboe, H.M., 2018. Comparative analysis of fluoride concentrations in Groundwaters in northern and southern Ghana: implications for the contaminant sources. Earth Syst. Environ. 2 (1), 103–117. https://doi.org/10.1007/s41748-018-0044-z.

Tan, Z., Yang, Q., Zheng, Y., 2020. Machine learning models of groundwater arsenic spatial distribution in Bangladesh: influence of Holocene sediment depositional history. Environ. Sci. Technol. 54 (15), 9454–9463. https://doi.org/10.1021/acs.est.0c03617.

Tay, C.K., Hayford, E.K., Hodgson, I.O.A., 2017. Application of multivariate statistical technique for hydrogeochemical assessment of groundwater within the lower Pra Basin. Ghana. Applied Water Science 7 (3), 1131–1150. https://doi.org/10.1007/s13201-017-0540-6.

Titus, Rian, Hans, B., Shafick, A., S. L., 2009. The Basement Aquifers of Southern Africa. Water Research Commission.

Tolera, M.B., Choi, H., Chang, S.W., Chung, I.M., 2020. Groundwater quality evaluation for different uses in the lower Ketar watershed. Ethiopia. Environmental Geochemistry and Health 42 (10), 3059–3078. https://doi.org/10.1007/s10653-019-00508-y.

Tshepo, K., 2017. Impact of Treated Wastewater on the Surface Water and Groundwater Quality: A Case Study in North East Gaborone (within Notwane Catchment). Botswana [MSc thesis].. University of Botswana http://hdl.handle.net/10311/1989.

Tuinhof, A., Foster, S., Van Steenbergen, F., Talbi, A., Wishart, M., 2011. Appropriate Groundwater Management Policy for Sub-Saharan Africa in face of demographic pressure and climatic variability. www.worldbank.org/gwmate.

Landtschoote, A. Van, 2016. Hydrogeological Investigation and Recharge Estimation of Gumera River Catchment in Lake Tana Basin. Ghent University, Northern Ethiopia [MSc thesis].

Vitor Portugal Gonçalves, M., Alves Santos, R., Alberto Machado Coutinho, C., Jerônimo Moreira Cruz, M., 2020. Fluoride levels in the groundwater and prevalence of dental fluorosis in the municipality of Santana, in region karstic of West Bahia. Brazil. Groundwater Hydrology 1–24. https://doi.org/10.5772/intechopen.85007.

Wadira, S.O., 2020. Hydrochemical Characteristics of Aquifers in Mwingi North-Kenya [MSc thesis].. University of Nairobi http://erepository.uonbi.ac.ke/handle/11295/154633.

World Health Organization (WHO), 2023. Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the Second Addendum. WHO, Geneva. https://washdata.org.

Xia, F., Jiang, D., Kong, L., Zhou, Y., Wei, J., Ding, D., Chen, Y., Wang, G., Deng, S., 2022. Prediction of Dichloroethene concentration in the groundwater of a contaminated site using XGBoost and LSTM. Int. J. Environ. Res. Public Health 19 (15), 1–24.

Yaghoubi, E., Yaghoubi, E., Khamees, A., Vakili, A.H., 2024. A systematic review and meta-analysis of artificial neural network, machine learning, deep learning, and ensemble learning approaches in field of geotechnical engineering. Neural Comput. & Applic. 36, 12655–12699. https://doi.org/10.1007/s00521-024-09893-7.

Yang, F.J., 2018. An implementation of naive bayes classifier. In: Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018, pp. 301–306. https://doi.org/10.1109/CSCI46756.2018.00065.

Yaseen, Z.M., 2023. A new benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. Knowledge-Based Engineering and Sciences 4 (3), 65–103. https://doi.org/10.51526/kbes.2023.4.3.65-103.

Zango, M.S., Pelig-Ba, K.B., Anim-Gyampo, M., Gibrilla, A., Sunkari, E.D., 2021. Hydrogeochemical and isotopic controls on the source of fluoride in groundwater within the Vea catchment, northeastern Ghana. Groundw. Sustain. Dev. 12, 1–13.

Zango, M.S., Sunkari, E.D., Abu, M., Lermi, A., 2019. Hydrogeochemical controls and human health risk assessment of groundwater fluoride and boron in the semi-arid north east region of Ghana. J. Geochem. Explor. 207, 1–21. https://doi.org/10.1016/j.gexplo.2019.106363.

Zhang, J., Zeng, L., Chen, C., Chen, D., Wu, L., 2015. Efficient Bayesian experimental design for contaminant source identification. Water Resour. Res. 51 (1), 576–598. https://doi.org/10.1002/2014WR015740.