International Journal of Science Academic Research
Vol. 05, Issue 12, pp.8708-8721, December, 2024
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

GEOPHYSICAL AND MACHINE LEARNING INVESTIGATION OF SALTWATER INTRUSION INTO FRESHWATER AQUIFER IN SOME ESTUARY ENVIRONMENT IN NIGER DELTA

***Omajene Aghogho, EgbaiJames Chucks and Okolie Emmanuel Chukwuemeka**

Department of Physics, Delta State University, Abraka, Nigeria

Received 12th October 2024; **Accepted** 09th November 2024; **Published online** 16th December 2024

Abstract

Geophysical and physiochemical approach, combined with machine learning (ML) prediction and feature classifiers, were employed in a research study to investigate saltwater intrusion in freshwater aquifers in some Estuary environment in Niger Delta, Nigeria. Five ML classifiers and regressor models were employed to classify, predict, and optimize 17 resistivity features based on vertical electrical soundings (VES) data ranging from 0.4 to 769.9Ωm to predict saline intrusion into aquifer layers. The ML established that the Gradient Boost classifier and Random Forest (RF) regress or yielded cross-validation accuracy of 96% and 98% at R^2 ≤0.9642 and highlighted five significant predictors of saltwater intrusion in specified Estuary environment with F1 score ≤0.9300. The predictive performance of the selected ML confirmed the potential electrode was the most significant predictor of saltwater intrusion into freshwater aquifers, corresponding to an 85.3% normalized degree of importance. RF predicted saltwater intrusion in the freshwater aquifer based on optimum potential electrodes in the range of 0.5-9.24, layer depth of 0.5-4m, elevation (5-13), curve type (A), and resistivity (0.5-43.8Ωm)with a corresponding R^2≥0.8457. The findings from the groundwater occurrence and depth (GOD) index classified the study area into low and moderate vulnerability classes, with values ranging from 0.168 to 0.420. The hydraulic resistance values at 2.877m^{-1} to 27.28m^{-1}, determine the aquifer vulnerability index (AVI).Groundwater analysis indicated elevated levels of electrical conductivity, salinity, and total dissolved solids, exceeding WHO standards. The GOD index, AVI, and water quality index (WQI) from the coastal location were consistent with the ML prediction.

Keywords: Resistivity, saltwater intrusion, water analysis, Vertical electrical sounding, MLfeature classification and prediction.

INTRODUCTION

Groundwater is an important freshwater resource worldwide, particularly in coastal communities. It is a renewable and finite natural resource, vital for human life, social and economic development, and a valuable ecosystem component. Before drilling a borehole, knowledge of the subsurface hydrogeological properties is essential to ensure that pro-life aquifer repositories are present (Omajene, 2023). When aquifers are exposed to contaminants as a result of natural and anthropogenic activities, it can lead to groundwater quality degradation, including drinking water sources, and other consequences and will lead to serious health issues such as cancer, cholera, and typhoid (George, 2021; Ibuot *et al.,* 2022; Omejene et al.,2023). Saltwater intrusion can naturally occur in coastal aquifers, owing to the hydraulic connection between groundwater and seawater. Because saline water has a higher mineral content than freshwater, it is denser and has a higher water pressure. As a result, saltwater can push inland beneath the freshwater. Many studies were carried out on groundwater quality evaluation and hydrochemical characterization (Brindha *et al.,* 2014; Sajil *et al.,* 2014; Wu, 2014 & 2015; Bouzourra *et al.,* 2015; Vetrimurugan *et al.,* 2015; Li, 2016).

Groundwater vulnerability indicates contamination of the aquifer by applying geoelectric indices to ensure the protective nature of the aquifer (Omajene *et al.,* 2023). Groundwater quality generally encompasses the physical, chemical, biological, radiological, and morphological characteristics of the water (Ibuot *et al.,* 2019; Obiora *et al.,* 2020). The low resistivity zones seen in the area are most likely caused by geogenic processes rather than aquifer overstressing. This assertion is supported by evidence revealed by the groundwater flow and simulation of saltwater intrusion into the aquifer (Ohwoghene and Asuma *et al.,* 2021). Groundwater vulnerability indicates contamination of the aquifer by applying geoelectric indices to ensure the protective nature of the aquifer (Omajene *et al.,* 2023). Groundwater quality generally encompasses the physical, chemical, biological, radiological, and morphological characteristics of the water (Ibuot *et al.,* 2019; Obiora *et al.,* 2015).

A geophysical survey using the electrical resistivity method was applied to get background information on the distribution, formation, and type of rear subsurface aquifers as a means of delineating the areas that may be prone to groundwater contamination and determine the location and depth and appreciable and portable water supply could be achieved (Osegi, 2010; Egbai, 2011; Egbai, 2012; Okolie, 2013). The resistivity method has been used by various researchers to explore saltwater intrusion in the subsurface. It is preferred over other electrical techniques because it can provide a clearer understanding of the subsurface structure and is cost-effective (Egbai, 1998; Ayolabi, *et al.,* 2009). This method has been found suitable for determining freshwater and saltwater-bearing formations (Zolidy *et al.,* 1993; Reynolds, 1997; Kilner *et al.,* 2005; Sumsnorac, 2006; Ayolabi, 2005). The low resistivity zones seen in the area are most likely caused by geogenic processes rather than aquifer overstressing. This assertion is supported by evidence revealed from the groundwater flow and simulation of saltwater intrusion into the aquifer (Ohwoghene *et al.,* 2021).

Corresponding Author:** ***Omajene Aghogho,
Department of Physics, Delta State University, Abraka, Nigeria.

Machine learning (ML) modeling, classification, and optimization procedures involve transforming raw observations into desired features using machine-learning approaches (Ebere *et al.,* 2023; Singh *et al.,* 2013). The voting regressor, random forest classifier, baseline classifier, and gradient boost are ensemble learning methods that combine multiple machine learning models to create a stronger overall model (Raschka, 2018; Tahir *et al.,* 2023; Ovuoraye *et al.,* 2023). The VR and GB model's robustness is further enhanced by its ability to estimate error rates using out-of-bag samples, providing an unbiased estimate of its performance. By leveraging the strengths of each model, the current research seeks to optimize more accurate predictions and offer greater stability across different resistivity datasets or Vertical electrical sounding (VES) features.

The application of ML prediction, feature classification, and optimization has gained relevance in the various fields of geological field survey (Al-Fakih *et al.,* 2023; Lui *et al.,* 2020), seismic operations, petroleum exploration, and geotechnical and environmental health applications (Tran *et al.,* 2022; Yin *et al.,* 2024). VES, saltwater intrusion, and prediction of resistivity data about the salinity of water quality analysis with a given field are still limited in the literature (Al-Fakih *et al.,* 2023; Singh *et al.,* 2005; Nguyen *et al.,* 2021; Tran *et al.,* 2022). The previous research studied the hydrodynamic conditions and groundwater quality deterioration Mkilima, T. (2023); Eid *et al.,* (2024). Hodlur and Dhakate (2010) also investigated the correlation of VES and electrical borehole log data for groundwater exploration. Quaquah and Boateng (2023) reported the simultaneous optimization of VES and magnetotelluric data using a genetic algorithm. Roy *et al.,* (2024) reported an automatic model selection-based machine learning approach to predict seawater intrusion into coastal aquifers. Yin *et al.,* (2024) studied the uncertainty-based saltwater intrusion prediction using integrated Bayesian machine learning to model in a deep aquifer. Tran *et al.,* (2022) investigated the performances of different machine learning algorithms for predicting saltwater intrusion in the Vietnamese Mekong Delta using limited input data. Singh *et al.,* (2005) studied one-dimensional inversion of geo-electrical resistivity sounding data using artificial neural networks. Al-Fakih*et al.* (2023) also estimated electrical resistivity from logging data for oil wells using machine learning. Liu *et al*.,(2020) explored deep learning inversion of electrical resistivity data. Singh et al.,(2013) applied neural network modelling and prediction of resistivity structures using VES Schlumberger data over a geothermal area. There is still very limited research concerning the salinity intrusion into the many freshwater aquifers in the field that has combined resistivity, geochemical methods, and ML classifiers with the predictive model to investigate saltwater intrusion into aquifers in coastal regions across the globe (Singh *et al.,* 2013; Liu *et al.,* 2020; Quaquah and Boateng, 2023).

The current research investigated saltwater intrusion into freshwater aquifers using resistivity, geothermal and machine learning model prediction, and feature classifier. In this study, ML classifiers and regressor models including the random forest regressor, Gradient boost (GB) Baseline model, and voting regressor were employed to model, classify, and optimize the 16 VES features to predict saltwater intrusion in the study area. The hydraulic resistance was used to calculate the aquifer vulnerability index (AVI). The AVI delineated the study area into high and very-high vulnerability classes. The groundwater occurrence and Depth (GOD) index in the study area was classified into low and moderate vulnerability classes, while the water quality index (WQI), and pollution load index (PLI). The ML classifier's outputs and predictions will assist in experimental validation, reduce evaluation costs, and support environmental health, safety, and sustainability.

MATERIALS AND METHODS

The research methodology employed the electrical resistivity method using Schlumberger array, integrated physiochemical and machine learning to model approach to investigate saltwater intrusion into the freshwater aquifer in the Estuary environment of Niger Delta, Nigeria. The apparent resistivity of the subsurface and voltage generated by transmission of current between electrodes (Current and potential electrodes) placed on the surface of the various locations in the study area. The apparent electrical resistivity were then determined from the measured data and used to determine the geoelectric parameters. These geoelectric parameters are interpreted to determine subsurface resistivity anomalies, depths, and thicknesses following Telford *et al.,* 1990; Lowrie, 1997).

Vertical Electrical Sounding (VES) Data Collection

The 50 VES data were collected based on random sampling from different coastal areas across two major communities inBurutu and Ogulagha Local Government Areas of Delta State, Nigeria as shown in Fig 1. The coastal area under investigation lies within the Niger Delta region and has coordinates 5.3567° N latitude and 5.5073° E longitude. Burutu and Ogulagha are situated in the coastal plain of the Niger Delta, one of the world's largest deltaic plains known for its low-lying topography and network of creeks, rivers, and swamps. The area is intersected by major rivers, including the Niger River and its tributaries. The sedimentary rocks found in Burutu and Ogulagha consist mainly of shale, sandstone, and clay deposits, which are associated with the formation of oil and gas reservoirs. These geological formations have made the Niger Delta region a major oil-producing area in Nigeria and one of the largest oil-producing regions in Africa.

Collection of water samples

The water samples were collected from four different locations within the vicinity of the study area (Ogulagha, Youbebe Sea I, Youbebe Sea II, and Burutu Well II) without contamination. The freshwater sampling and preservation werecarried out in the Laboratory Unit, following the standard procedure of the Department of Chemistry, Delta State University, Abraka, Nigeria.

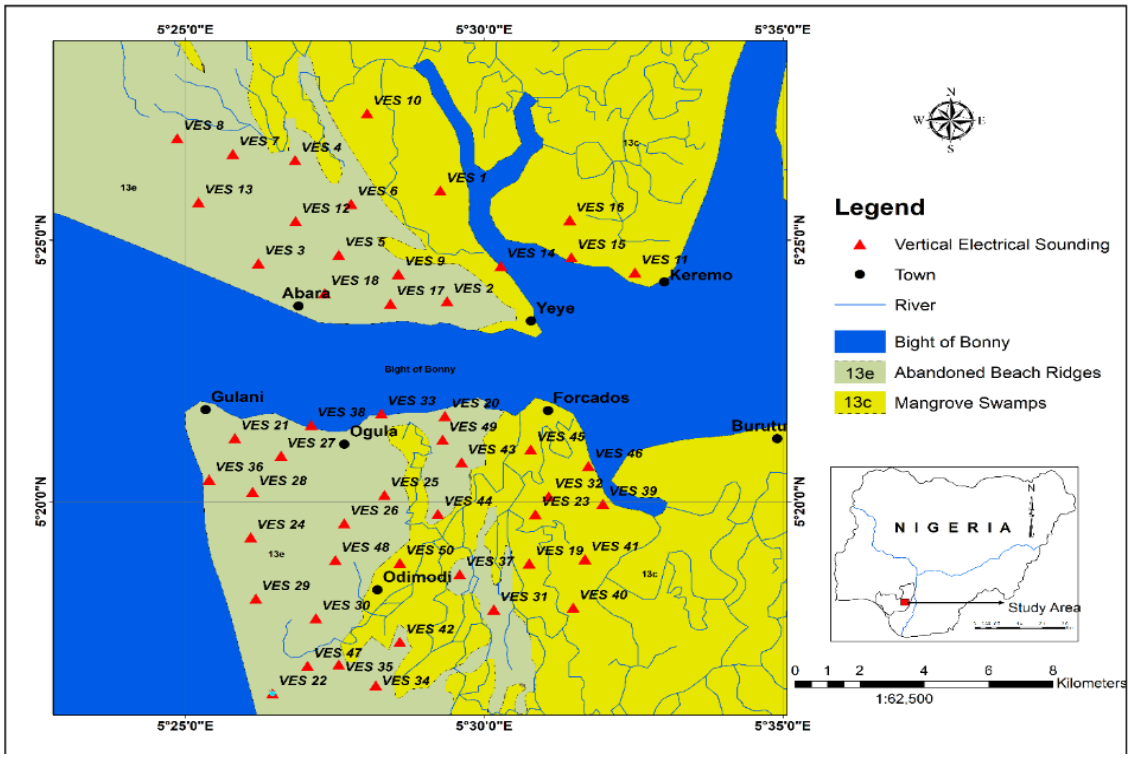

Fig. 1. The location of the Estuary Environment for collection of VES dataset

The water samples were split into two containers, one for examination of the anions, and the other for cations examination to determine their concentration in expressed in mg/L. The values of pH were measured using a multi-parameter analyzer. The values of electrical conductivity of the water samples were measured at the point of collection using a Wissenschaftlich-Technische Werkstätten LF91 (Ec) meter. The total dissolved solids (TDS) and dissolved oxygen (DO) were determined at the point of collection. The DO was measured with the aid of a dissolved oxygen meter and sensor. The values of the chemical oxygen demand (COD) and biological oxygen demand (BOD) were determined in the laboratory using standard procedures. These containers were initially washed with 0.05 M HCl filtered through membranes of 0.45 μm pores and then rinsed with ionized water. The water samples were acidified with concentrated nitric acid (HNO_3) to homogenize and prevent metallic ions from sticking to the walls. The analysis of the bicarbonates (${HCO_3}^-$) was carried out using a standard technique of titration to obtain their concentrations. The concentrations of the cations (K^+, ${Na^{2+}}$,${Zn^{2+}}{Ca^{2+}}$, ${Mg^{2+}}{Mn^{2+}}$, ${Pb^{2+}}$and ${Fe^{2+}}$) were determined using the Atomic Adsorption Spectrometer (AAS) model AA-7000 Shimadzu, Japan ROM version 1.01, while the anions (${SO_4}^{2-}$, Cl^-, ${HCO_3}^-$) were determined in the laboratory using the standard procedure of the titrimetric method.

Assessing drinking water quality and pollution level

This research employed different indices to assess the water quality of the study area. The indicators include the water quality index (WQI), contamination factor (CF), and pollution load index (PLI) by the World Health Organization Standard for Clean Water Assessment (WHO, 2020). The findings recorded from the PLI and WQI were employed to validate the predicted output recorded by the ML predictors and the feature classifier.

Water quality index (WQI): The Water Quality Index (WQI) is a numerical expression that summarizes the overall quality of water based on several physical, chemical, and biological parameters. It provides a simple way to communicate complex water quality information to the public and policymakers. The index typically combines multiple water quality parameters into a single value, allowing for an easy comparison of water quality over time or between different locations. This index was computed by employing the method of weighted arithmetic index. The sample concentration (C_i) in $\frac{mg}{L}$ of each is divided by each respective WHO standard (S_i) in $\frac{mg}{L}$ to obtain the quality rating scale (q_i). The ratio of C_i/S_i is then multiplied by a factor of 100 to give a mathematical expression in equation 1 (Verma *et al.,* 2020; Akakuru *et al.,* 2022).

$$q_i = \frac{C_i}{S_i} \times 100 \quad (1)$$

The inverse of the WHO standard corresponding to each of the analyzed parameters gives the relative weight (W_i) of each sample

$$W_i = \frac{1}{S_i} \quad (2)$$

The water quality index (WQI) is then expressed mathematically in equation 3 as the product of equations 1 and 2;

$$WQI = \sum q_i W_i \quad (3)$$

Determination of the GOD index: The GOD index, also known as the "GOD" vulnerability index, is a method used to assess groundwater vulnerability to pollution. This index considers geological and hydrogeological factors that influence the susceptibility of groundwater to contamination. The GOD index is determined by multiplying the effect of the three parameters, namely groundwater (G) (confined or unconfined aquifer), occurrence of lithological character of the vadose zone (O), and depth to the aquifer (D). The GOD index combines these factors to provide a comprehensive assessment of groundwater vulnerability to pollution. Areas with geologically permeable formations, shallow water tables, and thin or permeable overlying lithology are likely to have higher vulnerability scores, indicating a greater risk of contamination and vice versa. Table S1 **(**See supplementary material**)** gives the vulnerability ranges corresponding to the GOD parametric index. The findings based on the attribution of notes for GOD model parameters are presented in Table 1 and were used to validate the ML prediction and feature classifier output. The summary of the experimental design and workflow is presented in Fig 2**.**

$$GOD_{index} = G \times O \times D \quad (4)$$

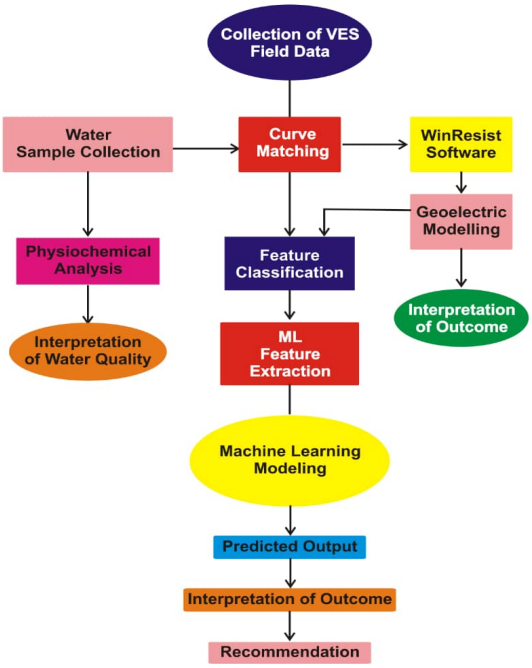

Fig. 2. Summary of the workflow of the experimental and design methodology

Table 1. Attribution of notes for GOD model Parameters (Khemiri *et al.*, 2013)

Aquifer type	Note	Lithology (Ω-m)	Note	Depth to aquifer (m)	Note
Non-aquifer	0	<60	0.4	<2	1
Artesian	0.1	60 - 100	0.5	2 – 5	0.9
Confined	0.2	100 - 300	0.7	5 – 10	0.8
Semi-confined	0.3 – 0.5	300 - 600	0.8	10 – 20	0.7

Machine learning methodology

The machine learning (ML) modeling, feature classification, prediction, and optimization procedure was performed using Scikit-learn with Python program interface (Abbas *et al.,* 2024). To select optimum VES features, three machine learning classifiers include the Dummy classifier (DC), Gradient Boost (GB), and Random Forest (RF). The prediction and optimization process involved three regressor models: the Baseline model, Votingregressor (VR), Gradient Boost (GB), and RF regressor. To classify the ML features, we used a feedback-forward selection approach (Tahir *et al.,* 2023). The approach involved employing classifier models based on 50 VES data collected from different aquifers in the Burutu and Ogulaha areas (refer to Fig 1, section 2.1). The ML feature classifiers provide predictive insights into which model predictors on locations based on VES data, resistivity values, Longitude, Latitude, elevation, location, curve type, depth, thickness, electrode potential, and current electrode are likely to generalize under specific salinity conditions, predicting saltwater intrusion of the locations. The classification models were first used to classify the VES variables in two locations, Ogulagha (Ogul) and Burutu (Buru) to obtain the top 10 features based on the normalized degree of importance of ML classifier models. The optimized VES features were used to predict the potential electrode as a response parameter using the ML regressor models. The output was then validated.

ML optimization and performance evaluation: This study employed used grid-search with cross-validation procedures to optimize the ML model outputs by automating and selecting the best hyper parameters or feature predictors to serve as a pivotal tool in unsupervised learning for exploring patterns and groupings within VES datasets based on environmental attributes (resistivity, location, depths, thickness, current potential, depths, curve type, longitude and latitude, and electrode potential) to determine saltwater intrusion into freshwater aquifer. To minimize prediction errors and avoid potential overfitting of the model,

the cumulative variance systematically modifies the tuning settings, following the method reported in the literature (Stackelbe *et al.,* 2021). The final predicted class label by choosing the class predictors with the highest normalized degree of importance ≤1 to ensure the model achieves maximum performance on the given VES dataset (Tahir *et al.,* 2023; Liang *et al.,* 2015).

This study used the following indicators to evaluate the performance of the ML feature classification model: accuracy score, cross-validation accuracy, precision, recall, and F1-score in model feature selection and prediction metrics. These performance indicators were used to validate the model performance and optimize the feature selection. True Positives (TP) outcome indicates the model correctly identified instances of saltwater intrusion. The False Negatives (FN) output indicates the instances model prediction misses any actual instances of saltwater intrusion (i.e., there were no incorrect rejections). The False Positives (FP): This output indicates the model incorrectly identified instances as saltwater intrusion when it was not (an incorrect alarm). The True Negatives (TN) metric measures the model correctly identified instances where there was no saltwater intrusion.

$$\text{Accuracy: } = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

This metric indicates the overall correctness of the model.

$$\text{Precision: } = \frac{TP}{TP+FP} \quad (6)$$

This measures the accuracy of the positive predictions.

$$\text{Recall (Sensitivity): } = \frac{TP}{TP+FN} \quad (7)$$

This measures the ability of the model to find all the relevant cases.

$$\text{F1-Score} = 2 \text{ x } \frac{Precision \text{ x } Recall}{Precision + Recall} \quad (8)$$

The F1-score metric measures the harmonic mean of precision and recall. This metric is useful to balance precision and recall, especially in cases with class imbalance (Tahir *et al.,* 2023). The high precision & low recall is cautious, leading to fewer false positives but possibly missing many true positives. However, a high recall & low precision indicate the model is aggressive in predicting the positive class, leading to more true positives but also more false positives (Tahir *et al.,* 2023).To evaluate the performance capacities of the ML regressor models, this study employed statistical evaluation and validation metrics to include mean absolute error (MAE), mean squared error (MSE), least squares coefficient (R^2), and root mean square error (RMSE) to validate model performance following procedures reported in the literature (Wager & Walther, 2015). The MAE values reflect the average error magnitude, easy to interpret, and less sensitive to outliers (Ovuoraye *et al.,* 2023). The RMSE offers a measure of error magnitude that is more sensitive to large mistakes, which is helpful in situations when significant deviations are a concern (Ugonabo *et al.,* 2022; Ovuoraye *et al.,* 2023). The value of the r-squared closest to a unit (R^2 ≤1.0) shows how well the model explains the variance in the target variable (Ovuoraye *et al.,* 2022; Enyoh *et al.,* 2024). A higher indicates a better fit. The summary of the ML feature classification and prediction methodology is presented in Fig 3**.**

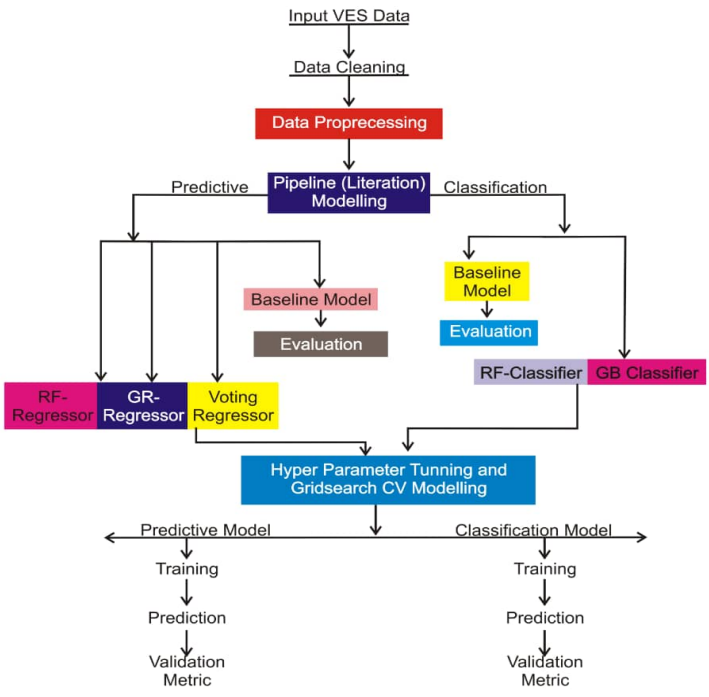

Fig 3: The summary of the Workflow for the Feature classification, prediction, and optimization

RESULTS AND DISCUSSION

Analysis of ML Classifiersmodel's Performance

Table 2 shows the confusion matrix for the interpretation of each model's performances and predictive capacities in handling the VES features important in predicting saltwater intrusion. The findings showed that the RF classifier selected ten features and achieved an accuracy score of 90%, indicating that it correctly classifies 90% of the instances on the test set, and cross-validation accuracy of 96% suggests that the model generalizes very well across different subsets of the data, which is higher than the single accuracy score, indicating robustness. Additionally, the 93% precision score of the RF classifier indicates that 93% of all instances projected to be positive are indeed positive. Furthermore, the recall of 90% implies that the model accurately detects 90% of all actual positive instances, and the F1-score of 91% balances precision and recall, demonstrating that the model performs well in both metrics.

The output of the GB-Classifier uses only 5 features toachieve the same degree of accuracy score of 90%, indicating that it is more efficient, requiring fewer features or model predictors to achieve similar performance.The cross-validation accuracy of 98% is the highest among the models, suggesting that the GB-Classifier has a better generalization performance on unseen data compared to the RF classifier.The Precision, recall, and F1-score are identical to those of the RF-Classifier, indicating that both model performed equally well in these metrics.However, the performance output of the Baseline Model (BM), with an accuracy of 50%, is essentially performing at the level of random guessing. (Enyoh *et al.,* 2024; Ovuoraye *et al.,* 2023). Comparatively, given that the GB-Classifier uses fewer features while maintaining performance, it could be considered more efficient and potentially less prone to overfitting.The GF-Classifier distinguishes itself from the RF-Classifier by using fewer features (5 vs. 10) and achieving higher cross-validation accuracy (98% vs. 96%) in the classification of the VES features, as demonstrated in Fig. 4. Both classifiers perform well overall, achieving identical accuracy scores and balanced precision, recall, and F1-scores,suggests that GB-Classifier is statistically the better choice of all three classification model.Although, the BM classifierwas outperformed by the other models, serving as a reference point that highlights the effectiveness of the trained classifiers; hence, both RF-Classifier and GB-Classifier are excellent choices, with GB-Classifier having a slight edge due to its higher cross-validation accuracy and use of fewer features.

The outline of Fig 4 confirmed the RF classifier model identified ten (10) predictors including potential electrode (MN/2) with a degree of importance score ≤ 0.35, p1, d1, and current electrode (AB/2) with corresponding importance scores ≤0.1. Other important predictors included p2, d2, h1, h2, p3, and latitude with degree of importance scores ≤0.05 as significant predictors to determining saltwater intrusion in the various locations based on the VES datasets. The GB classifier model selected potential electrodes with a degree of importance score >0.8, p1, d1, A, and elevation with a degree of importance score <0.1. The degree of importance of the potential electrode (MN/2) is significantly closer to a unit (1.0) suggesting it's the most significant VES feature for predicting saltwater intrusion into surface aquifers in the various locations under investigation using the geoelectrical sounding data. It can be observed from the outline of Fig. 4 that, in this case, the potential electrode (MN/2) appeared to be the most significantly important predictor of the saltwater intrusion following the classification of the VES features on the geoelectric datasets.

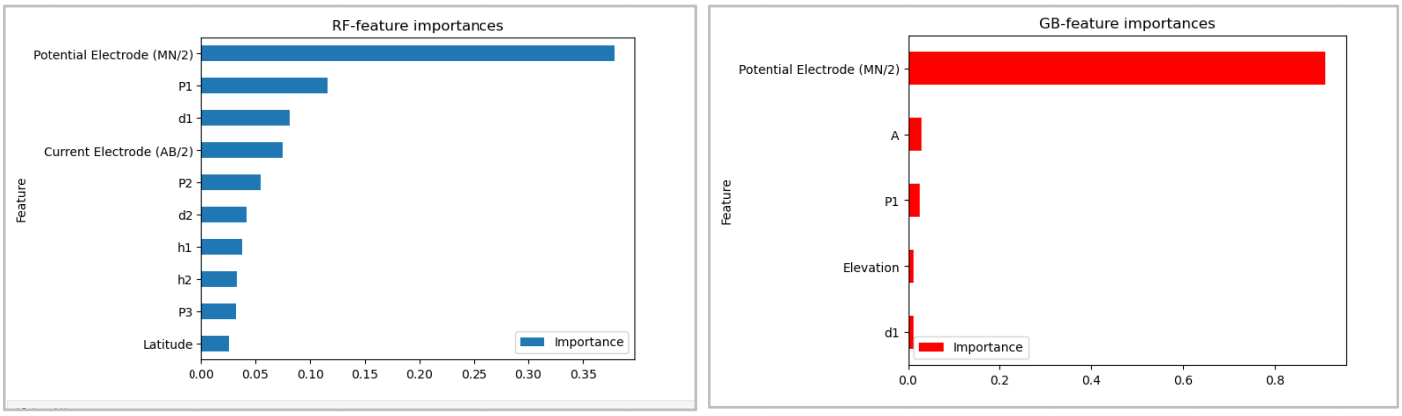

Fig. 4. ML classification Feature importance of the RF classifier and the GB classifier

Table 2. Confusion Matrix on the feature prediction and classification model

Models	No. of Feature	Accuracy score	Cross_validation Accuracy	Precision	Recall	F1-Score
RF-classifier	10	0.90	0.96	0.93	0.90	0.91
GB-classifier	5	0.90	0.98	0.93	0.90	0.91
Baseline model	NA	0.50	-	-	-	-

Additionally, the outputs from the confusion matrix of the selected classifier model is shown in Fig 5. The findings from the interpretation of the confusion matrices on the predictive performance of the GB classifier were consistent with the RF classifier model output. The value of True Positive (TP =2) obtained from the confusion matrix proved that both classifier models correctly identified 2 instances of saltwater intrusion in the target locations at Ogulagha and Burutu respectively. The False

Negatives(FN=0) indicate that both classifiers did not miss any actual instances of saltwater intrusion, such that there were no incorrect rejections. The value of False Positives **(**FP: 1) indicates that the selected classifier models incorrectly identified 1 instance as a saltwater intrusion when it was not (an incorrect alarm)which is crucial in environmental monitoring contexts.

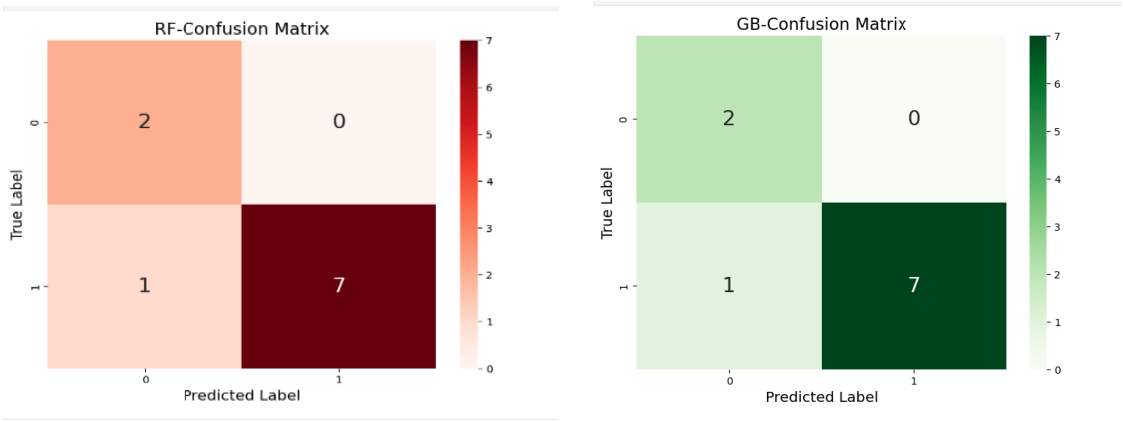

Fig. 5. Confusion matrix of the ML Feature Classification outputs of RF and GB classifier models

However, the comparative analysis of the class-specific metrics in **Table 3** established that for the target location in Burutu (Buru class), the Precision output of 0.67 indicates that 67% of predicted instances on the VES features for class Buru are predicted correctly. The recall score of 1 indicates that 100% of the actual class of the VES features instances Burutu location were correctly identified. While an F1-Score of 0.80 indicates a balanced measure of precision and recall. We concluded from the class-specific and evaluation metrics that the precision score of 100 indicates 100% of predicted instances for class Ogul (Ogulaha) are correct, with a corresponding recall value of 0.88 confirming that 88% of the actual classOgulaha instances were identified correctly. Additionally, the selected F1-Score of 0.93 confirmed a balanced measure of precision and recall) indicating that most of the predictions are correct.

Table 3. Summary of the ML feature classification evaluation statistics report

	RF Classifier				GB Classifier			
	Precision	**Recall**	**F1-score**	**Support**	**Precision**	**Recall**	**F1-score**	**Support**
Burutu	0.67	1.00	0.80	2	0.67	1.00	0.80	2
Ogulaha	1.00	0.88	0.93	8	1.00	0.88	0.93	8
Accuracy	-	-	0.90	10	-	-	0.90	10
Micro Ave	0.83	0.94	0.87	10	0.83	0.94	0.87	10
Weighted avg	0.93	0.90	0.91	10	0.93	0.90	0.91	10

Conclusively, the GB classifier probability and model assumption on the VES features proved to be more robust with 5 predictors, with the potential electrode voted as the most significant predictor of the saltwater intrusion into the aquifer. The moderate precision of 67% of the selected ML classifier implies that when the GB model predicts saltwater intrusion in the Burutu location, it is correct about two-thirds of the time. This means that although there is a risk of false positives, as indicated by the single false positive encountered. The recall score of 100% suggests that the model is excellent at detecting all instances of saltwater intrusion, which is crucial in environmental monitoring contexts (Ovuoraye *et al.,* 2022). Also, the lack of false negatives indicates that the model does not overlook any instances of intrusion, which is particularly important in cases where missing an intrusion could have serious consequences or environmental impact in the Burutu location. A higher precision of 88% confirmed an improved precision in the GB classifier model performance prediction of saltwater intrusion in the location Ogulagha class. The findings proved that RF-model assumptions are compromised due to false alarms. This outcome requires improving the precision of the RF model to minimize the number of false alarms. This is achieved by further tuning the model, adjusting hyperparameters, considering additional features, or using techniques like threshold adjustments for classification (Tahir *et al.,* 2023; Ovuoraye *et al.,* 2023).

Evaluation of the ML prediction performance and optimization

ML prediction and optimization was carried out taking the potential electrode as the target response on 5 predictors output from the GB model classifier. Three (3) ML regressor models were employed for the optimization, prediction, and interpretation of VES features and prediction of the classification outputs. The statistical evaluation metrics of the ML prediction and optimization output are shown in Table 4, and confirmed by the outlines of the actual versus predicted observation shown in Fig. 6(a-c). The ML optimization and statistical analysis result (Table 5) based on the RF-regressor yielded MAE (0.859), MSE:(1.9523), and RMSE (1.3973) respectively. Thisoutcome corresponded to apredicted R^2: 0.8457. The statistical outputs of theGB-regressortranslate to anMAE value of 0.8076, MSE value of 2.4376, and RMSE value of 1.5613 respectively. This outcome translates to a predicted-R^2 0.8073.Voting regressor (VR) model MAE =1.3567, MSE =3.8329, RMSE =1.9578, and predicted-R^2value of 0.7291 as indicated by the linearity of the actual and predicted observations in Fig 6(b).The Baseline model has MAE of 3.07999, MSE of 12.6524, and RMSEof3.65. The Baseline model (BM) serves as a reference point and shows significantly higher errors compared to the other models, indicating that the other models are performing much better.Its high MAE, MSE, and RMSE suggest that it is not a good predictor, with large deviations from the actual values. Comparatively,the predictive outputs of

the RF-regressor perform quite well with relatively low error values (MAE, MSE, and RMSE)with R^2 of 0.8457 indicating that the model explains about 84.57% of the variance in the most important VES feature (potential electrode), which is significantly strong.The GB-regressor has the least MAE (0.8076), suggesting the model has the smallest average error among other models (Enyor *et al.,* 2024; Ugonabo *et al.,* 2022).However, its MSE and RMSE outputs of the GB model are slightly higher than the RF-regressor, indicating some larger errors (Ovuoraye *et al.,* 2023; Sighn *et al.,* 2013).The predicted-R^2 of 0.8073, while slightly lower than the RF-regressor model, suggesting the model explains a good portion of the variance (80.73%). Overall, this model performs well but may struggle a bit more with larger errors compared to the RF-regressor. Furthermore, the VR model has the highest MAE, MSE, and RMSE among the three main models, indicating poorer performance (Ugonabo *et al.,* 2022; Lui *et al.,* 2020). The R^2 of 0.7291 suggests that it explains only 72.91% of the variance in the target variable, which is lower than the RF and GB Regressors.This model combines the other models in a way that unfortunately results in higher prediction errors. Therefore, it is less effective than using the individual models separately.The predicted optimal translates to potential electrode recorded from the RF model potential electrode of 0.5 corresponds to a depth (d_1) of 5,curve type A, elevation of 10 m, which translates to a resistivity (P_1) value of 0.5 in Ogulagha location. A higher magnitude of potential electrode (MN/2) value of 1.0 corresponded to the range of depth (d_1) in the range of 1.4- 4.8, curve type A, the elevation range of 5-13 m, which translates to a range of resistivity (P_1) values of 0.4 to 45.8 in Ogulagha location.

It can be concluded that the RF-regressor and GB-regressor are the top-performing predictive and optimization models. The predictive capacity of the RF-regressor slightly outperformedthe GB-regressor in terms of lower MSE, RMSE, and a higher R^2metric.This outcomesuggests that the RF-regressor is making more reliable and accurate predictionsdue to better generalization (Tahir *et al.,* 2023; Nadkarni *et al.,* 2023). The Voting Regressor (VR) did not improve upon the individual models and performed worse in terms of error metrics and R^2. This suggests that the combination of models in the VRmodel was not well-suited to this particular case study probably due to noise resistance (Ovuoraye *et al.,* 2023). Therefore, based on these evaluation metrics, the RF-regressor appears to be the best model for this task due to tunning ca, followed closely by the GB-regressor, while VR underperforms relative to these two. The predictive output of the baseline model served as a poor predictor, demonstrating that the other models add significant value to the training model. Consequently, the VR and BM model assumptions are compromised and will be ignored, while the RF regressor predictions and model probability will be adopted for further calibration and validation of the geoelectric analysis, GOD, geochemical, and WQI in the coastal regions under investigation.

Table 4. Performance evaluation metrics for ML regressormodels and predictions

Models	MAE	MSE	RMSE	R^2
RF-regressor	0.859	1.9523	1.3973	0.8457
GB-regressor	0.8076	2.4376	1.5613	0.8073
Voting-regressor	1.3567	3.8329	1.9578	0.7291
Baseline model	3.07999	12.6524	3.65	-

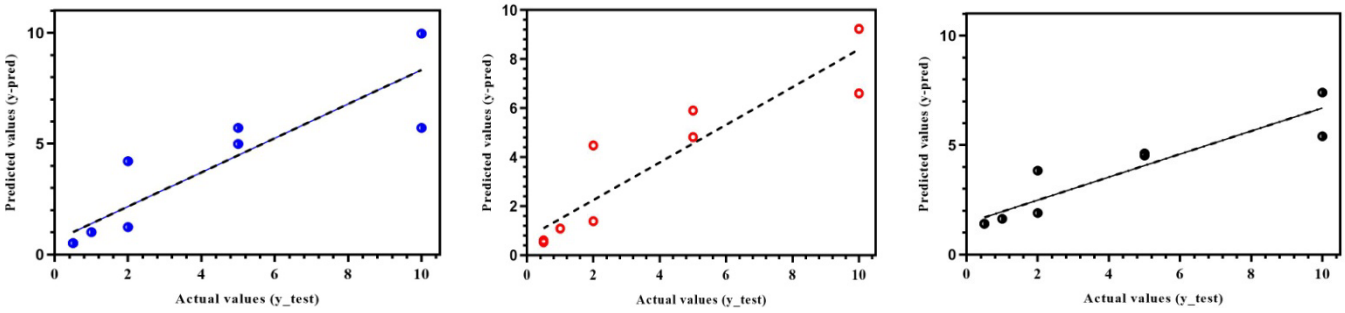

Fig 6. ML prediction and optimization output showing (a) actual versus predicted for RFregressor, (b) Actual versus predicted for GB, and (c) actual versus predicted for the VR model

Geoelectric Analysis

The geoelectric analysis and interpretation of the varying values of resistivity, thickness, and depths recorded from fifty (50) VES points shown in Table S2 (See supplementary material) reveal the heterogeneous nature of the subsurface. The frequency distribution of the curve types is shown in Fig 7**.** The outline showed that there are three to four geoelectric layers identified. The observed model curve types are dominated by H is 34 % of the total curve type, while other curve types are K, A, Q, AK, QH, AA, HA, KH, HK, and QQ. The topmost geoelectric layer has resistivity values vary from 0.1 0.5Ωm at VES 27 to 439.2Ωm at VES 32 and its thickness and depth range from 1.2 to 37.7 m and 4.0 to 38.3 m respectively. The resistivity and thickness of the aquifer layer (saturated layer) range from 0.4to769.9Ωm and 4.2 to 43.6 m was delineated as a low resistivity layer. This low resistivity may be attributed to saline water infiltration and high water content. Since the coastal areas often have intrusion of saline water from the sea into the aquifers. Saline water has lower resistivity compared to freshwater, so its presence can significantly decrease the overall resistivity of the aquifer layers. Comparatively, the RF regressor predicted resistivity output from the ML regressor (0.4-45.8 Ωm) is consistent with the geoelectric resistivity rangeof (0.4 to 769.9Ωm) layer, and the predicted thickness (5-13 m) of the aquifer is in the reasonable range (4.2 to 43.6 m) recorded from the geoelectric analysis. This low resistivity may be attributed to saline water infiltration and high water content. This output confirmed a saturated layer with low resistivity. The presence of salinity, indicated by a lower resistivity (≥0.4Ωm), explains the reduced overall resistivity of the

aquifer layers (≤45.8 Ωm) confirmed the presence of salinity which accounts for the reduced the overall resistivity of the aquifer layers.

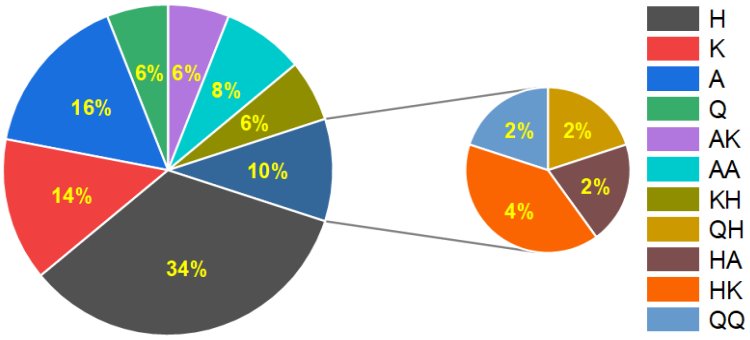

Fig. 7. Pie plot showing the frequency distribution of the curve type

Geochemical Approaches

Table 5 shows the water quality index (WQI) of the coastal regions under investigation. The geochemical output based on the WQI contents of 460.17 in Ogulaha exceeds the 440.85 benchmark in Burutu, indicating extremely poor water quality. The metallic ions/salts present in the water sample confirmed high concentrations of sodium (Na^+), magnesium (Mg^{2+}), and potassium (K^+) translates to ≥241 mg/L, with lower concentrations of heavy metal manganese (Mn^+), and zinc (Zn^{2+}) less than ≤0.98 mg/L respectively. This outcome indicates concentrations of alkaline metals are excessively high, suggesting high contamination contents, unpleasant taste, and likely presence of odor in the water (Ovuoraye *et al.,* 2022; Witkowska *et al.,* 2021). High sodium levels in drinking water can increase blood pressure, particularly in sodium-sensitive individuals, leading to hypertension, a risk factor for cardiovascular diseases (Chrysant, 2016; Jaques *et al.,* 2021). Additionally, water with high sodium content can taste salty, making it unpleasant to drink and potentially leading to reduced water consumption (Bigiani, 2020). Long-term accumulation of the toxic heavy in the body over time leads to long-term health problems for Indigenous people living in the coastal regions along Ogulaha communities.

Additionally, the high concentration of chloride (Cl^-), $HCO_3^-$and sulfate ions concentrations ≥ 26 mg/L suggests the water can irritate the eyes, skin, and respiratory system. At the same time, a high bicarbonate concentration can affect water quality and have other implications. The WQI is significantly greater than 440.85, and high bicarbonate levels can interfere with certain water treatment processes, such as coagulation and softening, making it more challenging to treat the water effectively. Comparatively, the WQI rating of both samples from Burutu and Ogulaha locations was significantly >300, suggesting that the water in these coastal regions is unsuitable for drinking (Akakuru *et al.,* 2021). The results revealed evidence of saltwater intrusion in freshwater aquifers. This outcome could be attributed to the thin layer or thickness observed across the study area. physiochemical parameters, such as in exchange, mineral dissolution, and precipitation, can affect the behavior of dissolved salts in groundwater, influencing the extent of saltwater intrusion. This outcome is in reasonable agreement with the findings recorded by the ML prediction and optimization output.

Table 5. Water Quality Index (WQI) from selected aquifer in predicted location

Sample	Concentrations (mg/L)									WQI
	SO_4^{2-}	Cl^-	HCO_3^-	Na^+	K^+	Ca^{2+}	Mg^{2+}	Mn^{2+}	Zn^{2+}	440.85
Ogunlaha well	26.00	210.00	74.00	849.00	241.00	14.00	314.00	0.98	0.73	460.17
Youbebe sea I	29.00	198.00	68.00	789.00	239.00	13.00	321.00	1.03	0.82	617.77
Youbebe sea II	40.30	214.00	81.00	910.00	256.00	11.37	298.00	1.43	0.91	486.99
Burutu well	28.10	194.00	91.00	841.00	291.00	14.63	428.00	1.08	0.76	440.85

Table 6. Summary of the range and averages of the physicochemical parameters of water from selected aquifer

S/N	Parameters	Minimum (mg/L)	Maximum (mg/L)	Average (mg/L)
1	pH	7.7	8.0	7.85
2	Electrical conductivity (µs/cm)	19600.00	27000.00	22935.00
3	Salinity(mg/L)	24.00	31.00	28.00
4	TDS (mg/L)	810.00	1100.00	925.00
5	DO (mg/L)	2.40	3.80	2.98
6	BOD (mg/L)	3.03	6.67	4.39
7	COD (mg/L)	2.00	3.07	2.39
8	SO_4^{2-}(mg/L)	26.00	40.30	30.85
9	Cl^-(mg/L)	194.00	214.00	204.00
10	HCO_3^-(mg/L)	68.00	91.00	78.50
11	Na^+(mg/L)	789.00	910.00	847.25
12	K^+(mg/L)	239.00	291.00	256.75
13	Ca^{2+}(mg/L)	11.37	14.63	13.25
14	Mg^{2+}(mg/L)	298.00	428.00	340.25
15	Mn^{2+}(mg/L)	0.98	1.43	1.13
16	Zn^{2+}(mg/L)	0.73	0.91	0.81

The physicochemical properties of the water (Table 6) confirmed high conductivity, with salinity values≥ 28, indicating the salinity of the region is significantly higher than freshwater (typically 0-1 mg/L) but lower than seawater (typically 35 mg/L). This result implies that the salinity levels in these areas could be negatively impacting certain organisms, resulting in their stress or exclusion. The water may require desalination or membrane treatment to reduce salinity levels and other salt concentrations for domestic usage (Olson *et al.,* 2022). The high TDS concentration recorded from the water analysis is consistent with the findings from the WQI report and confirmed ML predictions of a possibly high degree of saltwater intrusions in the coastal regions under investigation. However, the geochemical approach was also applied to investigate the percentage distribution of physical and oxygen-related water parameters contained in the water samples collected from four delineated layers across the coastal region. Refer to Figure 8 for the displayed results. The measurements of organic compositions (such as COD, DO, and BOD) indicate that the Burutu region has a higher degree of organic contamination compared to the Ogulaha districts. The water sample collected from the Ogulaha district shows relatively higher levels of salinity (25%) and conductivity (26%) compared to the Burutu district (21%). The significantly elevated conductivity of 26% confirmed saltwater intrusion into the freshwater aquifer in Ogulaha. The findings from the geochemical approach affirmed the significantly higher conductivity and salinity level in the ML predicted location (Ogulaha district) under investigation.

Table 7 shows the VES locations and their estimated values of the GOD index (Fig 8b) used for the description of the geological and hydrogeological conditions in the coastal area under investigation. The range of GOD index values, from 0.168 to 0.420, suggested that the study area can be classified into low and moderate vulnerability groups. However, regions with a low vulnerability rating indicate favorable geological and hydrogeological conditions in the area, and provide substantial protection against groundwater pollution. Regions of moderate vulnerability are indicated by vertical electrical sounding features at VES (6, 14, 16, 18, 24, 31, 33, 46, and 47), while the rest were delineated as low.

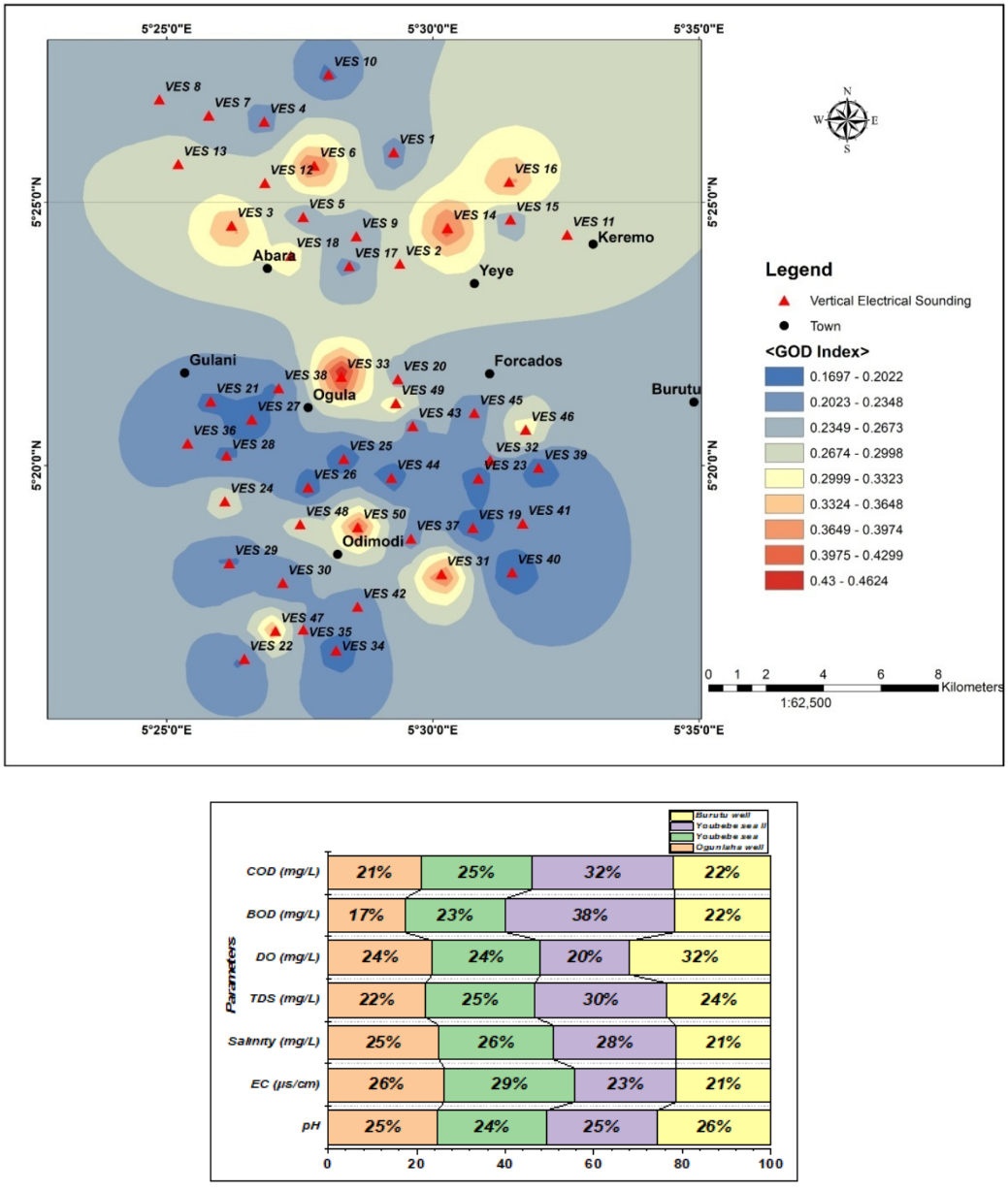

Fig 8. (a) Contour showing the variation and distribution of the GOD index(b)Confirmation of the percentage distribution of physical and oxygen-relatedwater parameters in the locations under investigation.

Table 7. Summary of aquifer vulnerability indices using GOD parametric model

VES points	Longitude (^{O}E)	Latitude (^{O}N)							GOD Index	Vulnerability class
			G	O (Ωm)	D	G	O	D		
1	5.4551	5.5080	Unconfined	13	3.7	0.6	0.4	0.9	0.216	Low
2	5.3558	5.5108	Unconfined	704.5	9.0	0.6	0.6	0.8	0.288	Low
3	5.3564	5.5122	Unconfined	133	3.2	0.6	0.7	0.9	0.378	Low
4	5.3586	5.5177	Unconfined	0.5	2.5	0.6	0.4	0.9	0.216	Low
5	5.3574	5.5164	Unconfined	5.2	0.6	0.6	0.4	1.0	0.240	Low
6	5.3567	5.5127	Unconfined	114.6	1.3	0.6	0.7	1.0	0.420	Moderate
7	5.3567	5.5143	Unconfined	19	0.5	0.6	0.4	1.0	0.240	Low
8	5.3522	5.5107	Unconfined	4.6	1.2	0.6	0.4	1.0	0.240	Low
9	5.3535	5.5118	Unconfined	8.3	0.6	0.6	0.4	1.0	0.240	Low
10	5.3549	5.5121	Unconfined	13.9	8	0.6	0.4	0.8	0.192	Low
11	5.3535	5.5060	Unconfined	1682.4	4.0	0.6	0.6	0.8	0.288	Low
12	5.3534	5.5060	Unconfined	73.4	3.1	0.6	0.5	0.9	0.270	Low
13	5.3513	5.5022	Unconfined	79.7	1.7	0.6	0.5	1.0	0.300	Low
14	5.3503	5.5060	Unconfined	257.3	1.2	0.6	0.7	1.0	0.420	Moderate
15	5.3506	5.5043	Unconfined	16.7	1.4	0.6	0.4	1.0	0.240	Low
16	5.3466	5.5019	Unconfined	122.2	3.6	0.6	0.7	0.9	0.378	Moderate
17	5.3510	5.5050	Unconfined	26.2	2.2	0.6	0.4	0.9	0.216	Low
18	5.3517	5.5065	Unconfined	743.4	2.1	0.6	0.6	0.9	0.324	Moderate
19	5.3564	5.3214	Unconfined	1.3	11.8	0.6	0.4	0.7	0.168	Low
20	5.3564	5.3214	Unconfined	0.5	1.3	0.6	0.4	1.0	0.240	Low
21	5.3564	5.3215	Unconfined	1.2	6.3	0.6	0.4	0.8	0.192	Low
22	5.3577	5.2215	Unconfined	0.3	4.4	0.6	0.4	0.8	0.192	Low
23	5.3584	5.3219	Unconfined	44.8	10.8	0.6	0.4	0.7	0.168	Low
24	5.3641	5.3214	Unconfined	69.5	1.6	0.6	0.5	1.0	0.300	Moderate
25	5.3642	5.3212	Unconfined	3.1	12.9	0.6	0.4	0.7	0.168	Low
26	5.3643	5.3211	Unconfined	1.2	10.8	0.6	0.4	0.7	0.168	Low
27	5.3654	5.3221	Unconfined	1.1	10.7	0.6	0.4	0.7	0.168	Low
28	5.3652	5.3222	Unconfined	0.5	5.0	0.6	0.4	0.8	0.192	Low
29	5.3652	5.3215	Unconfined	0.4	4.8	0.6	0.4	0.8	0.192	Low
30	5.3656	5.3230	Unconfined	0.7	4.5	0.6	0.4	0.8	0.192	Low
31	5.3620	5.3219	Unconfined	161.1	1.6	0.6	0.7	1.0	0.420	Moderate
32	5.3615	5.3294	Unconfined	45.8	1.4	0.6	0.4	1.0	0.240	Low
33	5.3598	5.3317	Unconfined	494.6	1.1	0.6	0.8	1.0	0.480	Moderate
34	5.3478	5.3230	Unconfined	1.6	10.8	0.6	0.4	0.7	0.168	Low
35	5.3479	5.3231	Unconfined	2.4	3.8	0.6	0.4	0.9	0.216	Low
36	5.3486	5.3231	Unconfined	1.4	3.9	0.6	0.4	0.9	0.216	Low
37	5.3492	5.3227	Unconfined	0.4	2.6	0.6	0.4	0.9	0.216	Low
38	5.3542	5.3212	Unconfined	0.6	7.1	0.6	0.4	0.8	0.192	Low
39	3.3556	5.3214	Unconfined	2.2	10.8	0.6	0.4	0.7	0.168	Low
40	5.3556	5.3210	Unconfined	2.6	11.8	0.6	0.4	0.7	0.168	Low
41	5.3556	5.3210	Unconfined	0.3	3.7	0.6	0.4	0.9	0.216	Low
42	5.3560	5.3220	Unconfined	0.7	2.3	0.6	0.4	0.9	0.216	Low
43	5.3560	5.3221	Unconfined	1.2	3.5	0.6	0.4	0.9	0.216	Low
44	5.3565	5.3226	Unconfined	5.7	11.2	0.6	0.4	0.7	0.168	Low
45	5.3601	5.3458	Unconfined	21.4	2.0	0.6	0.4	0.9	0.216	Low
46	5.3512	5.3447	Unconfined	239.9	5.9	0.6	0.7	0.8	0.336	Moderate
47	5.3505	5.3438	Unconfined	257.7	3.4	0.6	0.7	0.9	0.378	Moderate
48	5.3505	5.3430	Unconfined	150	10.4	0.6	0.7	0.7	0.294	Low
49	5.3516	5.3415	Unconfined	165.7	9.9	0.6	0.7	0.8	0.336	Moderate
50	5.3551	5.3382	Unconfined	142	1.5	0.6	0.7	1.0	0.420	Moderate

Comparative analysis of the ML model prediction with physiochemical resistivity methods

The analysis of the predicted outputs of the RF-regressor and GB classifier established that curve type (A), potential electrode (p = 0.5), elevation (10m), and depth (d =5) are significant predictors of saltwater intrusion corresponding to a resistivity value (p = 0.5Ωm), in Ogulahalocation. For a curve type (H), potential electrodes in the range of (0.53 - 9.24), elevation (5 - 13 m), and depth (1.4–4.3) are optimal predictors of saltwater intrusion in Ogulaha location in Delta State, and corresponding to the range of resistivity values (0.4 - 45.8Ωm). The predicted higher resistivity value of 45.8 is consistent with the GOD index lithology, and the resistivity parameter of < 60 indicates the coastal region under investigation is a non-confined aquifer (Khemiri *et al.,* 2013). Additionally, the predicted layer depths of 1.4–5 are in reasonable agreement with the attribution of notes for the GOD index in the range of layer depth of 2 – 5 m, suggesting that the aquifer type is likely artesian.

Additionally, the predicted curve type A is consistent with geoelectric analysis confirmed curve type A, which is controlled by 16% of location and second only to curve type H (see Fig1). The predicted thickness in the range of 1.4–4.8 delineates a saturated layer. The range of resistivity values of 1.4–45.8 can be attributed to saline water infiltration in non-permeable aquifers. The ML model's outputs demonstrated consistency with both the geochemical method and resistivity results. Based on the ratings obtained from the GOD index, PLI, and WQI, it can be concluded that the ML model classifier and predictive model have the potential to accurately forecast saltwater intrusion into the aquifer, with a rating between high and extreme (0.5 - 1.0). The significance of these findings lies in their potential to enhance local communities through improved access to safe drinking water, better irrigation facilities, and enhanced environmental health and safety measures for both aquatic ecosystems and humans.

Conclusion

The study investigated the feasibility of using a machine learning (ML) model to predict saltwater intrusion based on resistivity and geochemical validation. It utilized data from 50 VES datasets. The gradient boost (GB) classifier model was selected as the best ML classifier for analyzing VES features and predicting saltwater intrusion. It achieved a 100% recall and an 88% precision score. The ML classifier identified the Potential electrode as the most significant predictor of saltwater intrusion in two target locations, with an importance score of 0.85. The Random Forest regressor model predicted saltwater intrusion in the coastal region, based on five optimized features (potential electrodes, depth, curve type, and resistivity data). The RF-regressor produced reliable predictions with low statistical error metrics (MSE, RMSE, and MAE values less than 0.5) and a strong R^2 value of 0.8325, indicating consistency with actual observations. Analysis of water quality indexes confirmed saltwater intrusion into aquifers, primarily due to high levels of TDS salts (mostly chloride), heavy metal ions, stubborn organics (BOD < COD inclusive), and alkaline earth metals. The VES calibration and validation of the ML classifier and prediction model were consistent with geochemical and resistivity methods. The study recommends using a combination of the GB classifier and RF regressor model, utilizing 5 out of 17 VES features, for predicting saltwater intrusion into freshwater aquifers.

Declaration of Funding Declaration

No Funding was received for this project.

REFERENCES

1. Akakuru OC, Akudinobi B, Opara AI, Onyekuru SO, Akakuru OU (2021) Hydrogeochemicalfacies and pollution status of groundwater resources of Owerri and environs, Southeastern Nigeria. Environmental Monitoring Assess, 193, 623. https://doi.org/10.1007/s10661-021-09364-9
2. Akakuru OC, Eze CU, Okeke OC, Opara AI, Usman AU. Iheme O, Ibeneme SI, Iwuoha PO (2022). Hydrogeochemical evolution, water quality indices, irrigation suitability and pollution index of groundwater (PIG) around eastern Niger Delta
3. George NJ, Ibuot JC, and Obiora D. N (2015) Geoelectrohydraulic parameters of shallow sandy aquifer in Itu, AkwaIbom State (Nigeria) using geoelectric and hydrogeological measurements. *Journal of African Earth Sciences*, 110:52–63.
4. George, N. J. (2021) Geo-electrically and hydrogeologically derived vulnerability assessments of aquifer resources in the hinterland of parts of AkwaIbom State, Nigeria, Solid Earth Sciences, https://doi.org/10.1016/j.sesci.2021.04.002
5. Oghenero Ohwoghere-Asuma, I.FChinyem, K.E Aweto, (2020). The use of very low frequency electromagnetic survey in the mapping of groundwater condition of oporoza – gbamaratu area of Niger Delta, *Applied water science*. vol. 10:7 pg 1-4.
6. Omajene A, Ighrakpata C.F, Umayah.O.S (2023). Application of Geoelectric indices in Groundwater vulnerability mapping: A case study of Agbor, Delta State, Southern Region Nigeria Discovery 2023; 59:e51d1210.
7. Omajene. Aghogho. 2023. Investigation of Hydro geological Properties using Resistivity Data in parts of Delta State, Nigeria. *Nigerian institute of Physics Journal* vol. 23(1) 109-121
8. Omeje, E. T., Obiora, D. N., Okeke, F. N. Ibuot, J. C**.** and Omeje, V. D. (2023). Application of geoelectric technique and sensitivity analysis in assessment of aquifer vulnerability: a case study of Nsukka and Igbo-Etiti Area, Eastern Nigeria. *Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-023-03351-5
9. Thomas, J. E., George, N. J., Ekanem, A. M. &Nsikak, E. E. (2020). Electrostratigraphy and hydrogeochemistry of hyporheic zone and water-bearing caches in the littoral shorefront of AkwaIbom State University, Southern Nigeria. *Environmental Monitoring and Assessment,* 192:505. https://doi.org/10.1007/s10661-020-08436-6.
10. Verma P, Singh PK,Sinha RR, Tiwari AK (2020): Assessment of groundwater quality status by using water quality index (WQI) and geographic information system (GIS) approaches: a case study of the Bokaro district, India. Applied Water Science, 10:27, 7 https://doi.org/10.1007/s13201-019-1088-4
11. Sajil Kumar PJ, Elango L, James EJ (2014). Assessment of hydrochemistry and groundwater quality in the coastal area of South Chennai, India, *Arab J Geosci* 7:2641-2653. https://doi.org/10-1007/s12517-013-0940-3
12. Vetrimurugan E, Elango L. (2015). Groundwater chemistry and quality in an intensively cultivated River Delta. Water Qual Expo Heal 7:125-141. https://doi.org/10-1007/s12403-014-0133-7
13. Wu J, Li P, Qian H (2015). Hydrochemical characterization of drinking groundwater with special reference to fluoride in an arid area of China and the control of aquifer leakage on its concentrations. *Environ Earth sci.,* 73:8575-8588. htttps://doi.org/10-1007/s12665-015-4018-2
14. Ovuoraye, P. E., Ugonabo, V. I., Tahir, M. A., & Balogun, P. A. (2022). Kinetics-driven coagulation treatment of petroleum refinery effluent using land snail shells: An empirical approach to Environmental sustainability. *Cleaner Chemical Engineering*, *4*, 100084. https://doi.org/10.1016/j.clce.2022.100084
15. Ugonabo, V. I., Ovuoraye, P. E., Chowdhury, A., &Fetahi, E. (2022). Machine learning model for the optimization and kinetics of petroleum industry effluent treatment using aluminum sulfate. *Journal of Engineering and Applied Science*, *69*(1), 1-24. https://doi.org/10.1186/s44147-022-00164-7
16. Ovuoraye, P. E., Ugonabo, V. I., Fetahi, E., Chowdhury, A., Tahir, M. A., Igwegbe, C. A., &Dehghani, M. H. (2023). Machine learning algorithm and neural network architecture for optimization of pharmaceutical and drug manufacturing industrial effluent treatment using activated carbon derived from breadfruit (Treculiaafricana). *Journal of Engineering and Applied Science*, *70*(1), 1-21. https://doi.org/10.1186/s44147-023-00307-4
17. Wu J,Li P, Qian H et al.,(2014).Using correlation and multivariate statistical analysis to identify hydrogeochemical processes affecting the major ion chemistry of water: A case study in Laohebaphosphorite mine in Sichuan, China. *Arab J Geosci* 7:3973-3982. https://doi.org/10-1007/s12517-013-1057-4.

18. Bouzourra H, Bouhlila R, Elango L, et al., (2015). Characterization of mechanisms and processes of groundwater salinization in irrigated coastal area using statistics, GIS and hydrochemical investigations. *Environ scipollut Res.* 22:2643-2660. https://doi.org/10-1007/s11356-014-3428-0
19. Sheikh Mohammadi, A., Alamgholiloo, H., Golaki, M., Khakzad, P., Asgari, E., &Rahimlu, F. (2024). Cefixime removal via WO3/Co-ZIF nanocomposite using machine learning methods. *Scientific Reports*, *14*(1), 1-17. https://doi.org/10.1038/s41598-024-64790-2
20. Ebere E., Ovuoraye, P., Isiuku, O., & Igwegbe, C. (2023). Artificial Neural Network and Response Surface Design for Modeling the Competitive Biosorption of Pentachlorophenol and 2,4,6-Trichlorophenol to Canna indica L. in Aquaponia. *Analytical Methods in Environmental Chemistry Journal*, *6*(01), Pp.79-99 https://doi.org/10.24200/amecj.v6.i01.228
21. Singh, U. K., Tiwari, R., & Singh, S. (2013). Neural network modeling and prediction of resistivity structures using VES Schlumberger data over a geothermal area. *Computers & Geosciences*, *52*, 246-257. https://doi.org/10.1016/j.cageo.2012.09.018
22. Liu *et al.,* (2020). "Deep Learning Inversion of Electrical Resistivity Data," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5715-5728, https://doi.org/10.1109/TGRS.2020.2969040
23. Al-Fakih, A., Ibrahim, A.F., Elkatatny, S. *et al.,* (2023). Estimating electrical resistivity from logging data for oil wells using machine learning. *J Petrol Explor Prod Technol* 13, 1453–1461 (2023). https://doi.org/10.1007/s13202-023-01617-2
24. Singh, U., Tiwari, R., & Singh, S. (2005). One-dimensional inversion of geo-electrical resistivity sounding data using artificial neural networks A case study. *Computers & Geosciences*, *31*(1), 99-108. https://doi.org/10.1016/j.cageo.2004.09.014
25. Nguyen, T. G., Tran, N. A., Vu, P. L., Nguyen, Q., Nguyen, H. D., & Bui, Q. (2021). Salinity intrusion prediction using remote sensing and machine learning in data-limited regions: A case study in Vietnam's Mekong Delta. *Geoderma Regional*, *27*, e00424. https://doi.org/10.1016/j.geodrs.2021.e00424
26. Tran, T.T., Pham, N.H., Pham, Q.B. *et al.,* (2022). Performances of different machine learning Algorithms for predicting Saltwater intrusion in the Vietnamese Mekong Delta Using Limited Input Data: A Study from Ham Luong River. *Water Resour.* 49, 391–401. https://doi.org/10.1134/S0097807822030198
27. Yin, J., Huang, Y., Lu, C., & Liu, Z. (2024). Uncertainty-based saltwater intrusion prediction using integrated Bayesian machine learning modeling (IBMLM) in a deep aquifer. *Journal of Environmental Management*, *354*, 120252. https://doi.org/10.1016/j.jenvman.2024.120252
28. Roy, D.K., Paul, C.R., Munmun, T.H. *et al.,* (2024). An automatic model selection-based machine learning approach to predict seawater intrusion into coastal aquifers. *Environ Earth Sci* 83, 287 (2024). https://doi.org/10.1007/s12665-024-11589-z
29. Quaquah, M. A., & Boateng, E. (2023). Simultaneous Optimization of Vertical Electrical Sounding and Magnetotelluric Data using a Genetic Algorithm. Journal of Environment and Earth Science. Vol.13, No.7, ISSN 2224-3216. https://doi.org/10.7176/JEES/13-7-05
30. Hodlur, G. K., &Dhakate, R. (2010). Correlation of vertical electrical sounding and electrical borehole log data for groundwater exploration. *Geophysical Prospecting*, *58*(3), 485-503. https://doi.org/10.1111/j.1365-2478.2009.00831.x
31. Bai, H., Liu, P., Fu, X., Qiao, L., Liu, C., Xin, Y., & Ling, Z. (2023). Application of elastic net in quantitative analysis of major elements using Martian laser-induced breakdown spectroscopy datasets. *SpectrochimicaActa Part B: Atomic Spectroscopy*, *199*, 106587. https://doi.org/10.1016/j.sab.2022.106587
32. Aminul, M., & Jahan, N. (2017). Prediction of onset diabetes using machine learning techniques. International Journal of Computer Applications, 180(5), 7–11. https://doi.org/10.5120/ijca2017916020
33. Eid, M. H., Shebl, A., Eissa, M., Mohamed, E. A., Fahil, A. S., Ramadan, H. S., Abukhadra, M. R., El-Sherbeeny, A. M., Kovacs, A., & Szűcs, P. (2024). Comprehensive approach integrating remote sensing, machine learning, and physicochemical parameters to detect hydrodynamic conditions and groundwater quality deterioration in non-rechargeable aquifer systems. Heliyon, 10(12), e32992. https://doi.org/10.1016/j.heliyon.2024.e32992
34. Liang, M., Li, Z., Chen, T., &Zeng, J. (2015). Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 12(4), 928–937. https://doi.org/10.1109/tcbb.2014.2377729
35. RandomForestClassifier. (2024). Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
36. Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open-Source Software,* 3(24), 638. https://doi.org/10.21105/joss.00638
37. Stackelberg, P. E., Belitz, K., Brown, C. J., Erickson, M. L., Elliott, S. M., Kauffman, L. J., Ransom, K. M., & Reddy, J. E. (2021). Machine learning predictions of pH in the glacial aquifer system, northern USA. Ground Water, 59(3), 352–368. https://doi.org/10.1111/gwat.13063
38. Tahir, M. A., Farhat, Z., Khan, M. U. R., Gaber, J., Anwar, M., &Ovuoraye, P. E. (2023). Integration of machine learning algorithms classifiers and Sequential forward selection features in diabetes prediction. 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), 1531–1536.
39. Wager, S., & Walther, G. (2015). Adaptive concentration of regression trees, with application to random forests. In arXiv [math.ST]. http://arxiv.org/abs/1503.06388
40. Aminul, M., & Jahan, N. (2017). Prediction of onset diabetes using machine learning techniques. International Journal of Computer Applications, 180(5), 7–11. https://doi.org/10.5120/ijca2017916020
41. Liang, M., Li, Z., Chen, T., &Zeng, J. (2015). Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 12(4), 928–937. https://doi.org/10.1109/tcbb.2014.2377729

42. Random Forest Classifier. (2024). Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
43. Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. Journal of Open-Source Software, 3(24), 638. https://doi.org/10.21105/joss.00638
44. Scikitlearn.(2024).ScikitLearn.https://scikitlearn.org/stable/auto_examples/ensemble/plot_forest_hist_grad_boosting_comparison.html
45. Stackelberg, P. E., Belitz, K., Brown, C. J., Erickson, M. L., Elliott, S. M., Kauffman, L. J., Ransom, K. M., & Reddy, J. E. (2021). Machine learning predictions of pH in the glacial aquifer system, northern USA. Ground Water, 59(3), 352–368. https://doi.org/10.1111/gwat.13063
46. Witkowska D, Słowik J, Chilicka K. (2021). Heavy Metals and Human Health: Possible Exposure Pathways and the Competition for Protein Binding Sites. Molecules. 26(19):6060. https:/doi.org/10.3390/molecules26196060
47. Jaques D A, Wuerzner G, Ponte B. (2021). Sodium Intake as a Cardiovascular Risk Factor: A Narrative Review. Nutrients. 13(9):3177. https://doi.org/10.3390/nu13093177
48. Chrysant, S. G. (2016). Effects of High Salt Intake on Blood Pressure and Cardiovascular Disease: The Role of COX Inhibitors. *Clinical Cardiology*, *39*(4), 240-242. https://doi.org/10.1002/clc.22536
49. Bigiani, A. (2020). Salt Taste, Nutrition, and Health. *Nutrients*, *12*(5). https://doi.org/10.3390/nu12051537
50. Olson, S., Jansen, M. F., Abbot, D. S., Halevy, I., &Goldblatt, C. (2022). The Effect of Ocean Salinity on Climate and Its Implications for Earth's Habitability. *Geophysical Research Letters*, *49*(10). https://doi.org/10.1029/2021GL095748
51. Mkilima, T. (2023). Groundwater salinity and irrigation suitability in low-lying coastal areas. A case of Dar es Salaam, Tanzania. *Watershed Ecology and the Environment*, *5*, 173-185. https://doi.org/10.1016/j.wsee.2023.07.002
52. Nadkarni, S. B., Vijay, G. S., &Kamath, R. C. (2023). Comparative Study of Random Forest and Gradient Boosting Algorithms to Predict Airfoil Self-Noise. *Engineering Proceedings*, *59*(1), 24. https://doi.org/10.3390/engproc2023059024
53. Ovuoraye, P. E., Ugonabo, V. I., Fetahi, E., Chowdhury, A., Tahir, M. A., Igwegbe, C. A., &Dehghani, M. H. (2023). Machine learning algorithm and neural network architecture for optimization of pharmaceutical and drug manufacturing industrial effluent treatment using activated carbon derived from breadfruit (Treculiaafricana). *Journal of Engineering and Applied Science*, *70*(1), 1-21. https://doi.org/10.1186/s44147-023-00307-4
54. Abbas, U. L., Zhang, Y., Tapia, J., Md, S., Chen, J., Shi, J., & Shao, Q. (2024). Machine-Learning-Assisted Design of Deep Eutectic Solvents Based on Uncovered Hydrogen Bond Patterns. *Engineering*. https://doi.org/10.1016/j.eng.2023.10.020

****