DOI: https://dx.doi.org/10.4314/gjpas.v30i4.5

GLOBAL JOURNAL OF PURE AND APPLIED SCIENCES VOL. 30, 2024: 447-470
COPYRIGHT© BACHUDO SCIENCE CO. LTD PRINTED IN NIGERIA ISSN 1118 – 0579, e-ISSN: 2992 - 4464
www.globaljournalseries.com.ng, Email: globaljournalseries@gmail.com

A SYSTEMATIC REVIEW ON THE TRENDS, PROGRESSES, AND CHALLENGES IN THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN WATER QUALITY ASSESSMENT AND MONITORING IN NIGERIA

M. E. OMEKA, G. AMAH, M. I. MORPHY, B. O. OMANG, E. A. ASINYA, AND G.T. KAVE

Email: omeka.ekuru@unical.edu.ng; michaelekuru@gmail.com

ORCID: [a]0000-0003-0405-8616

(Received 3 June 2024; Revision Accepted 26 June 2024)

ABSTRACT

In recent decades, machine learning (ML) artificial intelligence has found wide application in water quality monitoring and prediction due to the increasing complexity of water quality data. This complexity has been attributed to the global upsurge in anthropogenic activities and climatic variations. It is therefore critical to identify the most accurate and suitable ML model for water quality prediction. In this study, a systematic literature review (SLR) was carried out to explore the trend and progress in the application of ML models in water quality monitoring and prediction in Nigeria from 2003-2024. A comprehensive review of the effectiveness of advanced ML models as well as the gaps in their application in the area of water quality assessment and monitoring was also carried out using the PRISMA-P meta-analysis technique. Forty publications were used to perform bibliographic analysis and visualization using the VOS viewer software. The study found that globally, the use of hybrid ML models in water quality prediction has not been well explored; a majority of the prediction has been based on the use of artificial neural networks (ANN). Among the ANN algorithms, the adaptive neuro-fuzzy inference system (ANFIS), and Wavelet-Adaptive Neural Fuzzy Interference System (W-ANFIS) hybrid models are the most accurate in prediction; with temperature, dissolved oxygen (DO), pH, electrical conductivity (EC), and total dissolved solids (TDS) among the most frequently predicted parameters. Nigeria is grossly lagging in the application of ML in water quality prediction. This limitation is largely attributed to inadequate data on environmental monitoring. It is critical therefore for future water quality monitoring and prediction studies in Nigeria to take advantage of the rapidly evolving field of machine learning; with more emphasis placed on the hybridized machine learning algorithms

KEYWORDS: Machine learning; Pollution; Artificial intelligence; Drinking water quality; Nigeria

INTRODUCTION

Approximately 75% of the earth's crust is made up of water. About 68.7 and 30.3% of the total water mass, are encased in ice caps or glaciers and ocean, respectively. Fewer than 1% is regarded as fresh (Al Aani et al. 2019).

The quality of water supplies has been declining dramatically in recent decades. This reflects both the increase in human activity and the variability of the global climate (Ayejoto et al. 2023; Abugu et al., 2024). Natural hydrological, hydrogeologic, and geogenic processes also influence water quality because they alter the water's chemistry

M. E. Omeka, Department of Geology, University of Calabar, P.M.B. 1115, Calabar, Cross River State, Nigeria
G. Amah, Department of Geology, University of Calabar, P.M.B. 1115, Calabar, Cross River State, Nigeria
M. I. Morphy, Department of Geology, University of Calabar, P.M.B. 1115, Calabar, Cross River State, Nigeria
B. O. Omang, Department of Geology, University of Calabar, P.M.B. 1115, Calabar, Cross River State, Nigeria
E. A. Asinya, Department of Geology, University of Calabar, P.M.B. 1115, Calabar, Cross River State, Nigeria
G.T. Kave, Department of Geology, University of Calabar, P.M.B. 1115, Calabar, Cross River State, Nigeria

(Xiao et al. 2021). The predominant pollution precursors that are unique to the location, either locally or regionally, will have a major influence on how quickly the water quality is altered, or polluted. Furthermore, the mechanisms and origins of declining water quality are typically numerous and intricate (Omeka, 2023). Examples of natural process variations in aquifer characteristics include water-rock interaction, chemical speciation, changes in stream dynamics in response to pollutant input, and variations in rainfall intensity. Both point (produce from mine and industrial effluents) and non-point (runoff from crop fields, effluents from town waste dumps seeping into groundwater, runoff from open defecation, etc.) sources can contribute to anthropogenic processes (Xiao et al. 2021; Okamkpa et al. 2022; Omeka et al., 2024a, b). Because of this intricacy, evaluating the quality of water usually involves difficult and thorough data gathering, laboratory investigation, and data analysis. Both conventional and non-traditional methods have been used to monitor and assess the quality of water in recent decades. The non-conventional approach uses multiple quantitative metrics, parametric and non-parametric models of statistics, health risk assessment mathematical models, and spatiotemporal models (by employing GIS) for quality checks, whereas traditional methods depend on the application of local and international benchmark standards. However, the disadvantage of using numerous numerical models is that it usually takes a lot of effort and expert knowledge to compute the sub-indices for model output. This may lead to inaccuracies in the output of the results and consequently affect the decision-making process, which can have a significant effect on the economy (Kouadri and Samir, 2021).

With a population density of more than 1.9 million, Nigeria has the largest human population of all the black African nations (National Population Commission, 2013). A startling 66.3 million people, however, lack access to safe drinking water (Ighalo et al., 2020; Okamkpa et al. 2022). Additionally, the accessible water sources are under a lot of stress because of the high population density and rising socioeconomic activity. It has been established that seasonal variations in temperature and rainfall intensity brought on by climate change significantly alter the quality of water resources (Kouadri and Samir, 2021; Omeka et al., 2024a). In this regard, there will be an indirect rise in the variability of climate change as a result of the growing anthropogenic activities (such as hydrocarbon exploration, industrialization, land usage, and population growth). The frequency of aquifer recharge/discharge rates will therefore be impacted by this (Omeka and Egbueri 2023; Ricolf et al., 2020; Abugu et al. 2024; Omang et al., 2023a, b). Therefore, efforts and focus should be directed towards developing a relationship between the various water quality precursors unique to an area in order to provide effective and sustainable water management and preservation. This will require the development of innovative, economical methods for comprehensive, reliable assessment and monitoring of water quality. Artificial Intelligence (AI) can be utilised to achieve this.

Globally, there has been some progress from the conventional water quality assessment and monitoring techniques towards artificial intelligence (AI) and Internet of Things (IoT) technologies (Geetha and Gouthami 2016). AI was first introduced in the early 1950s and has since gone through significant modifications and gained applications in various fields in data simulation and forecasting of events outcomes (Al Aani et al. 2019). The use of AI in water quality assessment and monitoring started in the early 1990s and has gained the attention of many researchers and environmental stakeholders since then (Hasan et al. 2011). As a result of the complexity, ubiquity, and nonlinearity usually associated with water quality datasets, a challenge may arise in the prediction of water quality through the use of conventional approaches. This is attributed to increasing sources of anthropogenic pollution sources and varying climatic changes. Therefore, developing a model that can efficiently predict and monitor water quality remains crucial, especially in the present Anthropocene age where there seems to be no sight of solution to the impending global pollution crisis.

As part of her industrial revolution plan, Nigeria projects to increase its industrial output from 4% to 10% of GDP by 2020 (NIRP 2014). This will mean that more pressure will be put on the environment and its valued components (such as water, air, and soil). The big question remains how the country will maintain a balance between its industrial revolution projections and the protection of valued environmental components. It is feared that with the continuous increase in population and impeding industrial revolution, Nigeria will be put under greater strain in terms of climatic variations. Additionally, the quantity and quality of water resources will experience a further decline and continuous shifts (Igwe and Omeka, 2022; Ayejoto et al. 2023; Aluma et al. 2024). It becomes imperative therefore for more conscious action plans to be put toward the assessment and prediction of the quality of water resources for informed decision-making and amelioration of contaminated sites through the use of cutting-edge approaches. As a response, in recent times, some studies have tilted towards the use of AI in water quality prediction in Nigeria, with some recorded progress. On a global scale, Ighalo et al. (2020) have carried out a comprehensive review of the application of different AI models in the assessment and monitoring of surface water in the last decade (2011-2020). Their study found that two AI models- Artificial Neural Networks (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) were the most applied models for water quality assessment and monitoring within the period.

The study also showed that Southeast Asia and Iran are among the countries with most studies on the use of these models in water quality prediction.

Studies in Nigeria on the application of AI in water quality prediction are very few in literature. Among these studies, most have focused on the use of standalone machine learning (ML) algorithms (Egbueri 2022; Omeka, 2023) while some have integrated ML and AHP in water quality prediction and prioritization (Ifediegwu 2022; Omeka et al. 2023b). For instance, Egbueri (2022) has integrated a multilayer perceptron artificial neural network (MLP-ANN) and multiple linear regression (MLR) with corrosion and scaling potential indices for the prediction of industrial water quality in southeastern Nigeria. This study concluded that the incorporation of ML with indexical models can provide robustness in industrial water quality assessment and management. In another study in the region, MLR and MLP-ANN were incorporated to predict groundwater quality using two numerical models- overall index of pollution (OIP) and water quality index (WQI) as predicted or output variables. The MLP-ANN showed higher efficiency of prediction as lower modeling errors were observed in their validation metrics (Egbueri and Agbasi, 2022). In an agricultural mine district in southeastern Nigeria, a composite irrigation suitability zonation map was generated through the integration of a GIS-based analytical hierarchy process (AHP) and ML algorithms (Omeka et al. 2023). The model successfully identified Mg, HCO_3, Cl, and SO_4 as the major contributors to the deterioration of the irrigation water quality in the region. In a similar study, simple kriging (SK) and inverse distance weighting (IDW) GIS-based spatiotemporal models have been used in combination with supervised ML algorithms for a composite assessment of the irrigation water quality for crop yield in a mining region in southeastern Nigeria (Omeka, 2023). This study revealed that soil is more susceptible to excessive sodium concentration, through the application of surface water for irrigation due to the higher pH of the surface water compared to groundwater.

Indeed, some efforts have been put toward the use of AI in water quality monitoring in Nigeria, especially, in the southeastern parts. It is however recommended that more research be channeled toward this area by embracing the application of hybridized (ensemble) ML algorithms. Due to the nonlinearity and randomness of water quality data, and the impending climate variability, the need for the use of improved ML models is imperative. The relationship between water quality and its controlling factors is known to vary over time (due to variations in anthropogenic factors and seasonal fluctuations), hence, the use of a predictive model that functions under the framework of historical data becomes inefficient over time (Khan et al. 2021; Omeka et al., 2024a). The use of hybridization in other parts of the world has shown very promising results. For instance, in Pakistan, the forward-rolling artificial intelligence algorithm has been used to generate a model that will show resiliency to variability in climate change in a hydrologic setting (Khan et al. 2021). Two ML algorithms – the recursive feature elimination and support vector machine learning (RFESVM) have been combined to monitor the variability in rainfall in Malaysia (Pour et al. 2020). In Penang, Malaysia, the support vector machines (SVM), backpropagation neural network (BP-NN), and radial basis function neural network (RB-NN) have been incorporated to monitor the quality of surface water in constructed wetlands (Mohammadpour et al. 2015). Nayak (2020) combined and compared the use of multiple linear regression (MLP) with correlation models in assessing the trend in the surface water quality of the Brahmani River. The MLR models were found to show better results in water quality trends. In the semi-enclosed Tolo Harbour water bay in Hon Kong, random forest machine (RF) and ANN machine learning models were combined to assess the spatial variations in eutrophication in the water. The ensemble model identified suspended solids, dissolved nutrients, and organic pollutants as significant contributors to water quality (Deng et al. 2022). In Iran, Sun et al. (2021), have combined two hybrid ML models- the multivariate adaptive regression spline (MARS) and intrinsic time-scale decomposition (ITD) with the multi-step supervised-based machine learning approach (MSMLEA) for the prediction of the water quality of the Tajan River in Iran. From the results, the hybrid ML models showed higher performance accuracy in water quality prediction.

Nigeria has seen serious incidences of cholera, diarrhoea, and other gastrointestinal ailments in the past few decades as a result of consuming polluted water (Martins et al. 2016). In the joint WHO/UNICEF (2012) report on monitoring programmes for sanitation and water supply, Nigeria comes in third place, right after China and India, among the nations without access to clean drinking water. Waterborne illnesses are thought to be the cause of approximately 130,000 child deaths in Nigeria each year (WHO/UNICEF 2012). If immediate action is not taken, it is anticipated that this number would rise by 2040 (WHO/UNICEF 2012). As per the 2016 report on cancer patterns in Nigeria by the Nigeria National System of Cancer Registries, around 80% of the country's cancer risks are attributable to pollution. Thirty percent of this are related to drinking polluted water (Morounke et al. 2017). This seems to support the 1996 Harvard report on cancer prevention, which found that of all the risk factors for cancer, the consumption of dissolved heavy metals (such as Ni, Cd, As, and Pb) in drinking water poses 2% of the world's cancer risks. Therefore, there is a need for an ongoing evaluation of innovative and state-of-the-art methods for comprehensive and sufficient water quality monitoring and prediction. In contrast to developing nations like Nigeria, data on seasonal variations and water quality is widely available in developed countries.

Several factors, such as a lack of government funding and inaccessibility to laboratory facilities in Nigeria, can be connected to this.

Consequently, a systematic literature review (SLR) covering the previous 20 years (2003–2024) on the use of machine learning models in water quality monitoring and assessment was conducted for this study. The trend in machine learning, a relatively recent field of study that has showed significant promise in the prediction of water quality, must be embraced, given the critical role clean water plays in both sustainable development and the preservation of human health. Furthermore, a comprehensive evaluation of the advancement from current well-observed developments to more advanced and hypothetical machine learning model simulation and comparative analysis has also been presented. The first section considers common factors that affect Nigeria's water quality in relation to the sources of contaminants and provides information on potential new contaminants. The second part considers the implications for prospects as well as the past and anticipated future trends in Nigeria's methods for managing and monitoring water quality. A priority for the frequency of input variables and the precision of validation measures used to show the findings has also been placed on the deployment of different ML models in water quality monitoring. Future research potential and investigations on water quality have also been recommended.

Study area

Location

Nigeria is bounded to the north by Niger, towards the east by Chad and Cameroon, towards the south by the Atlantic Ocean's Gulf of Guinea, and to the west by Benin Republic (Encyclopædia Britannica 2023) (Fig. 1).

Relief

Nigeria's terrain is characterized by lowlands in the north and south, which are punctuated by plateaus and mountains in the center of the nation. The Sokoto Plains are located in the country's northwest region, whereas the Borno Plains are located in the country's northeastern corner and extend to the Lake Chad basin. Porous, geologically young sedimentary strata underpin the Lake Chad basin and coastal regions, along with the Niger River delta and western sections of the Sokoto region in the extreme northwest (Encyclopædia Britannica, 2023; Ayejoto et al. 2023). The plateaus' distinctive features are high plains with vast, shallow valleys interspersed with several hills or solitary summits known as inselbergs; the base rocks are crystalline, with sandstones appearing in river basins. Some degraded features, including the Udi-Nsukka escarpment, rise suddenly above the plains at least 1,000 feet in height (300 meters) (Okamkpa et al. 2022).

Climate

The climate of Nigeria is tropical, with varying rainy and dry seasons depending on location. The southeast is hot and humid most of the year, but dry in the southwestern and further inland. The west and north have a savanna climate with distinct dry and wet seasons, whereas the extreme north has a steppe environment with limited precipitation.

The frequency of the rainy season generally reduces from south to north. The rainy season runs from March to November in the south, except for mid-May to September in the extreme north. The rains stop completely in the south during August, resulting in a brief dry season known as the "August break." Precipitation is greater in the south, particularly the southeastern, which experiences over 120 inches (3,000 mm) of rain every year, compared to around 70 inches (1,800 mm) in the southwest. Temperature and humidity stay largely consistent in the south throughout the year, but seasons fluctuate significantly in the north; and during the northern dry season, the average temperature range also increases.

Abbreviations

SAR	Sodium Absorption Ratio
RMSE	Root Mean Square Error
MLP-ANN	Multi-layer perceptron neural network
TP	Total Phosphorus
NN-CS	Neural Network trained by Cuckoo Search Algorithm
TS	Total Solids
TDS	Total Dissolved Solids
ANFIS	Adaptive Neuro-Fuzzy Inference System
BOD	Biochemical Oxygen Demand
TN	Total Nitrogen
TA	Total alkalinity
TH	Total Hardness
COD	Chemical Oxygen Demand
DO	Dissolved Oxygen
TSS	Total Suspended Solids
BPNN	Back Propagation Neural Network
R^2	Coefficient of Determination

PI Permanganate Index
SVM Support Vector Machine
W-BPNN Wavelet and back propagation neural network
W-ANFIS Wavelet and Adaptive Neuro-Fuzzy Inference System

Highlights

- Globally, the use of hybrid ML models in water quality prediction has not been well explored; a majority of the prediction has been based on the use of artificial neural networks (ANN).
- The adaptive neuro-fuzzy inference system (ANFIS), and Wavelet-Adaptive Neural Fuzzy Interference System (W-ANFIS) hybrid models are the most accurate in prediction.
- Nigeria is grossly lagging in the application of ML in water quality prediction. This limitation is largely attributed to inadequate data on environmental monitoring.

Fig. 1. Location Map of Nigeria showing the different geographic locations (source: Encyclopædia Britannica, 2023)

METHODOLOGY

Systematic Literature Review

The research method used in the present study was the systematic literature review. To carry out the research, the following research questions were raised:

(1) what are the major sources of water contamination in Nigeria?

(2) what are the most frequently used machine learning models in water quality analysis in the last two decades?

(3) what are the most accurate machine learning models based on their model predict metrics (such as coefficient of determination [R^2])?

(4) what are the most frequently monitored water quality parameters?

(5) what are the implications of the use of ML models in water quality monitoring?

The systematic literature review was comprised of published articles from Web of Science and Scopus databases. The search terms used for the search in the databases included the sources of contamination in the aquatic environment in Nigeria and the application of machine learning in water quality prediction.

In the Scopus database, each term was entered in the search directory including the abstract, title, and keywords. The filters used were the year of publication (within the date range of 2003-2024), the thematic area of environmental sciences, access type (open and none-open access articles), and the document type (article or review). In the Web of Science database, the search filter was based on Abstract, title, author keywords, and study within the date range of 2003-2024. Two thematic categories – machine learning in water quality analysis and environmental science were leveraged on for the search.

The articles were selected based on the following criteria:

(1) articles that give a detailed explanation of the major sources of water pollution

(2) articles that explain the use of hybridized or standalone machine learning algorithms in water quality monitoring

(3) articles that explain the peculiarities of the different validation metrics (e.g. coefficient of determination, standard error of estimate, the sum of square errors and residual error).

The exclusion was made on certain criteria such as:

(1) articles that appear outside the scope of the study

(2) articles that do not explicitly address the use of machine learning algorithms in water quality monitoring

(3) articles concerning other valued environmental components outside of water.

Creating a bibliographic map using the VOSviewer software

The VOSviewer was used in creating a bibliographic map for keywords in machine learning and water quality between the period of 2003-2024. Bibliographic data from Scopus and Web of Science databases were extracted with a filter in the area of environmental sciences. A total of 729 and 2,234 articles were respectively returned for the search terms for sources of contamination in the aquatic environment in Nigeria and the application of machine learning in water quality prediction in the Scopus database. For the Web of Science database, 521 and 1,672 articles respectively were returned. The articles were then exported from the database in CSV format to the VOSviewer. The bibliographic map was then created in the VOSviewer environment using keywords with a minimum of five occurrences selected (van Ech and Waltman 2020).

RESULTS AND DISCUSSION

Results from Databases

Table 1 shows the number of analyzed articles considered in the review, while Figure 2 shows the PRISMA 2000 flow diagram for systematic literature review (SLR) showing the results of the database searches and their registries. The variation in the number of articles using the two search items (sources of contamination in aquatic environment in Nigeria and application of machine learning in water quality prediction) in the two databases showed an upward increase in articles on the term application of machine learning in water quality prediction, compared to the former. The Scopus database showed a higher number of publications compared to the Web of Science database. This is because the Scopus database has a greater number of indexed journals compared to the Web of Science database.

The number of articles returned with the use of the search term sources of contamination in the aquatic environment in Nigeria and application of machine learning in water quality prediction were 729 and 2,234 respectively in the Scopus database. For the Web of Science database, 521 and 1,672 articles respectively were returned from the search (Table 1, Fig. 2).

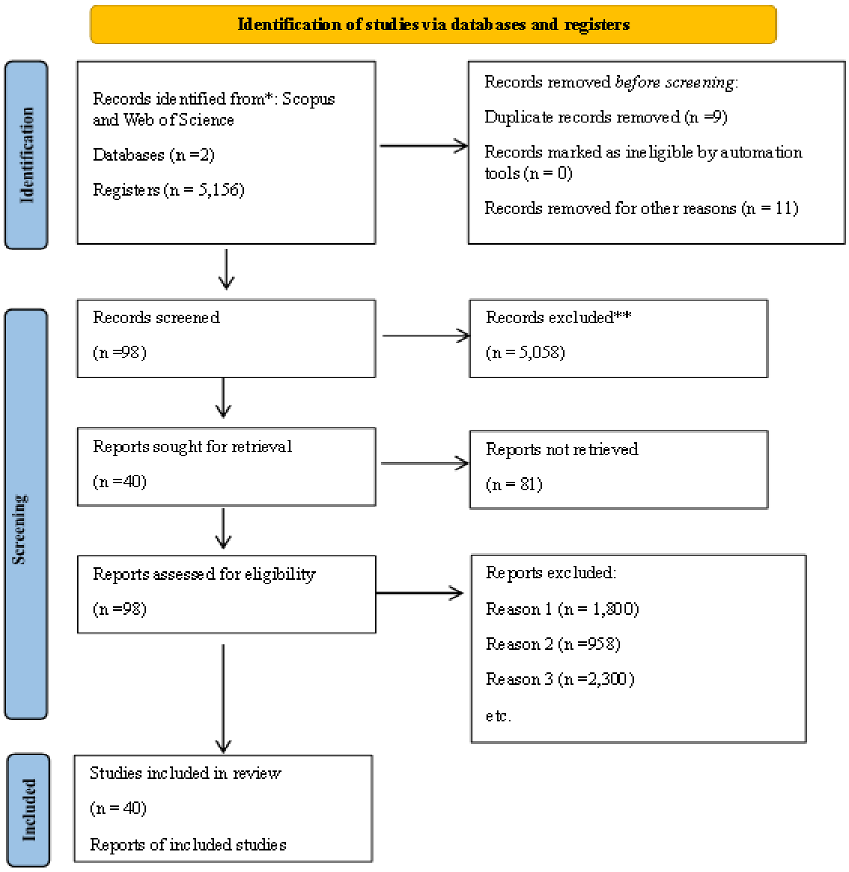

Fig. 2 PRISMA 2000 flow diagram for systematic literature review (SLR) showing the results of the database searches and their registries

Table 1 Number of articles analyzed and considered for the SLR

Search terms	Returned articles	Databases	Selected articles in the search
Sources of contamination in the aquatic environment in Nigeria	729	Scopus	35
Application of machine learning in water quality prediction	2,234	Scopus	20
Sources of contamination in the aquatic environment in Nigeria	521	Web of Science	25
Application of machine learning in water quality prediction	1,672	Web of Science	18
The sum of articles considered from abstracts and titles			98
The sum of articles considered after full-text analysis			40

The number of articles considered for review was 35 and 20 for the terms sources of contamination in the aquatic environment in Nigeria and application of machine learning in water quality prediction in the Scopus database. While 25 and 18 were selected for the Web of Science database (Table 1). Due to a large number of returned articles for the Scopus database, only the title and abstracts of the first 600 articles were read.

For inclusion and exclusion criteria, only articles that did not appear in the two databases were included for the review; as some articles appeared in both the Scopus and Web of Science databases. Therefore, a total of 98 articles were selected for reading of titles and abstracts while a total of 40 articles were selected for full-text analysis. Others were discarded. The discarded articles did not meet the established criteria. Based on the search term "application of machine learning in water quality prediction", out of the 257 publications that were considered, China occurred in 61, the United States in 57, India in 32, Malaysia in 15, south Korea in 14, Australia in 13, the United Kingdom and Canada in 10, Vietnam in 9, Iran in 8, Brazil and Iraq occurred in 7, Poland and Taiwan occurred in 6, Russian Federation, France, Germany, Italy, and Saudi Arabia occurred in 5; Morocco, Sweden, Bangladesh, Netherlands, and Hong Kong occurred in 4; Singapore, Spain, United Arabs Emirates, Norway, Palestine, Pakistan, and Greece occurred in 3; Croatia, Egypt, Pietro Florentine, Serbia, Belgium, Algeria, Nigeria, South Africa, Turkey, Columbia occurred in 2; Dakota, Philippines, Tunisia, Austria, Uganda, Romania, Chile, Denmark, Czech Republic, Luxembourg, Switzerland, Finland, Georgia, Peru, Sri Lanka, Portugal, State of Libya, Israel, and Mexico occurred in 1 had one publication each (Table 2). It can therefore be deduced frequency of the application of ML in these countries may be due availability of experts in the use of neural networks in engineering and environmental studies. Although this may be speculative, the author hopes that this observation will spur the interest of researchers in the field for more discussion and enlightenment.

Bibliographic mapping using VOSviewer

The bibliographic map was created using the VOSviewer software. Co-occurrence was chosen as the type of analysis, while All keywords were chosen as the unit of analysis, with fractional counting selected as the preferred counting method. For the application of machine learning in the water quality prediction search file, the software returned 377 keywords with 186 meeting the threshold and occurring at least six times across all analyzed articles. For each of the 186 keywords, the total strength of the co-occurrence links with other keywords was calculated. The keyword with the greatest link strength was then selected for generating the bibliographic map. Links between keywords are an indication of the relationship between them (the link strength) represented by a numerical value. Implying that the higher the link strength, the stronger the association where the two terms occur together (van Eck and Waltman 2020). Terms were then grouped into clusters of a different colour; with each colour representing the strength of the association.

For the application of machine learning in water quality prediction terms, analysis was made based on both the countries of study and the type of machine learning algorithm frequently used. The bibliographic map based on the overlay visualization (with the year of publication) and density visualization is shown in Fig. 3 based on the type of machine learning model. The graphical representation of the variation in the number of publications and the occurrence of the ML algorithms in water quality prediction is shown in Fig. 3. Artificial neural networks had the highest occurrence and strongest link strength in all analyzed publications, followed by regression analysis and random forest, while Bayes theorem and the convolutional neural network had the lowest occurrence and weakest link strength (Figs. 3 & 4). From the bibliographic map, high density and link strengths seem to occur mostly with Artificial neural networks, support vector machines, regression analysis, decision trees, and random forests. However, fuzzy systems, long short-term memory, convolutional neural networks, hybrid, and ensemble ML models had the weakest links and strength with the number of publications on the subject seeming to be increasing since 2021. This result implies that the use of hybrid ML models in water quality prediction has not been well explored globally; a majority of the prediction has been based on the use of artificial neural networks.

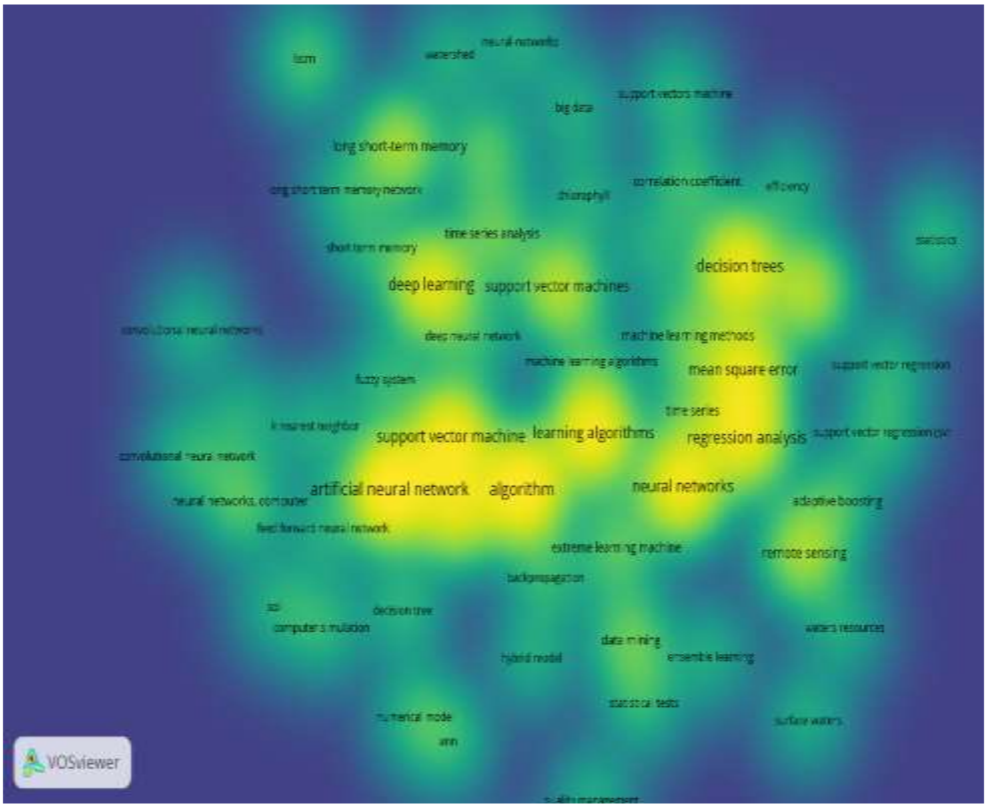

Fig. 3 Bibliographic coupling map showing results for frequently used machine learning in water quality prediction

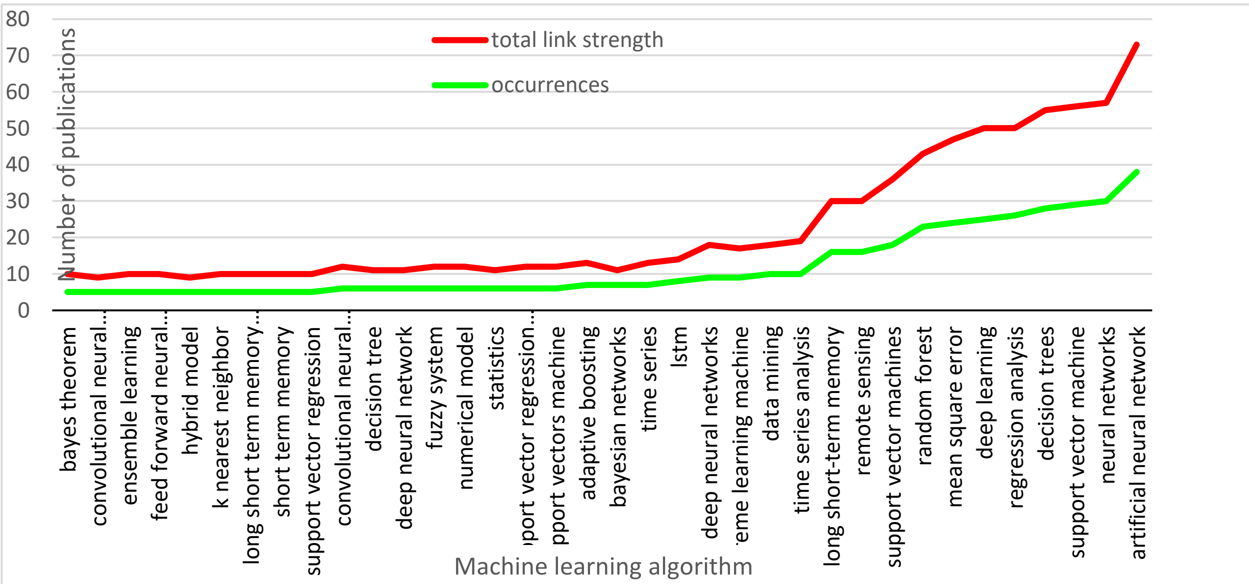

Fig. 4 Graphical representation of the frequently used machine learning model

Based on the countries of study, China, the United States and India had the highest number of publications and link strength concerning the term application of machine learning in water quality prediction (Fig. 5). The total number of publications involving the review developed by country was 61, 57, and 15 for China, the United States and India, respectively (Table 2, Fig. 5). This implies that countries like Croatia, Egypt, Pietro Fiorentini, Serbia, Belgium, Algeria, Nigeria, South Africa, Turkey, Columbia, Dakota, Philipines, Tunisia, Austria, Uganda, Romania, Chile, Denmark, Czech Republic, Luxembourg, Switzerland, Finland, Georgia, Peru, Sri Lanka, Portugal, State of Libya, Isareal and Mexico are grossly lagging behind in the use of ML in water quality monitoring and assessment.

Table 2 Number of articles per country occurring from the systematic literature review

Country	No. of Publication
China	61
United States	57
India	32
Malaysia	15
South Korea	14
Australia	13
United Kingdom	10
Canada	10
Vietnam	9
Iran	8
Brazil, Iraq	7
Poland, Taiwan, Poland	6
Russian Federation, France, Germany, Italy, Saudi Arabia	5
Morocco, Sweden, Bangladesh, Netherlands, Hong Kong	4
Singapore, Spain, United Arab Emirates, Norway, Palestine, Pakistan, Greece	3
Croatia, Egypt, Pietro Florentine, Serbia, Belgium, Algeria, Nigeria, South Africa, Turkey, Columbia	2
Dakota, Philippines, Tunisia, Austria, Uganda, Romania, Chile, Denmark, Czech Republic, Luxembourg, Switzerland, Finland, Georgia, Peru, Sri Lanka, Portugal, State of Libya, Israel, and Mexico.	1

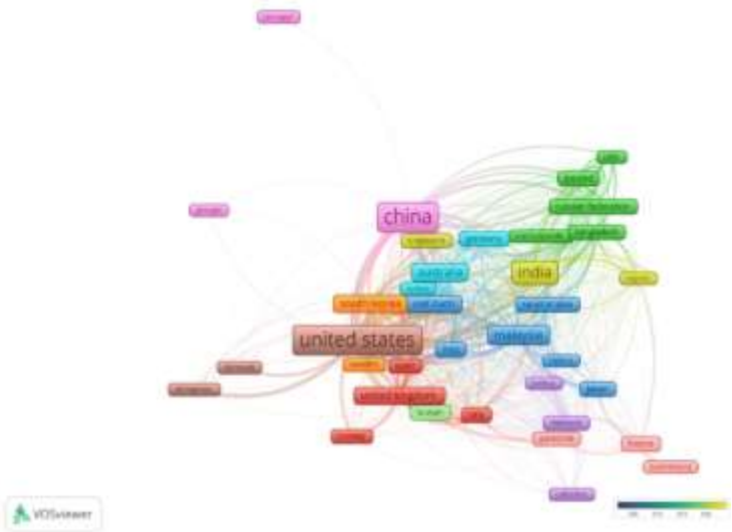

Fig. 5 Bibliographic coupling map of countries with the highest publication in the application machine learning in water quality prediction.

For the sources of contamination in the aquatic environment in Nigeria search term, analysis was made based on the relationship between anthropogenic and natural/geogenic sources of contamination. Based on the relationship between pollutant type, heavy metals, Lead, Cadmium, Copper, Chromium, and Zinc were reported in more publications and had the greatest link strength among others. Meanwhile, Antibiotic agents, aromatic compounds, bitumen, Cl, combustion, dumpsite, EC, mineralization, municipal solid wastes, organic carbon, phosphates, rainwater, saline intrusion, urbanization, and suspended particulate matter showed the lowest link strength and were reported in fewer publications (Fig. 6, Table 3).

Table 3 number of publications based on sources of aquatic contamination in Nigeria

Pollutant	Occurrences per publication	Total link strength
Heavy metal	67	62
Heavy metals	59	52
Lead	48	47
Cadmium	40	39
Copper	36	36
Chromium	35	35
Zinc	34	34
Escherichia coli	30	26
Geologic sediments	25	24
Polycyclic aromatic hydrocarbons	25	21
Manganese, Nickel	23	23
Arsenic, Iron	22	22
Nitrate	21	20
PAHs, pH	20	20
Electrical conductivity	19	16
Hydrocarbons	18	17
Coliform bacterium, sewage	14	13
Bacterium contamination, seasonal variation	13	13
Microbial contamination	12	10
Effluents, leachates, mercury, mining, petroleum, trace elements, trace metals, turbidity, weathering	11	11
Agriculture, Calcium, total dissolved solids	10	10
Aluminum, crude oil, feces, irrigation, sulfate, wastewater	9	9
Anti-infective agents, chlorine, fecal coliform, industrial waste, land use, NO_3, Na, population density, wastewater	8	6
Chlorine compounds, DO, geology, industries, Mg, oil spill, K	7	7
Co, F, landfill, leachates, leaching, organic matter, runoff, sulfur compounds, hardness, domestic wastes	6	6
Antibiotic agents, aromatic compounds, bitumen, Cl, combustion, dumpsite, EC, mineralization, municipal solid wastes, organic carbon, phosphates, rainwater, saline intrusion, urbanization, suspended particulate matter	5	4

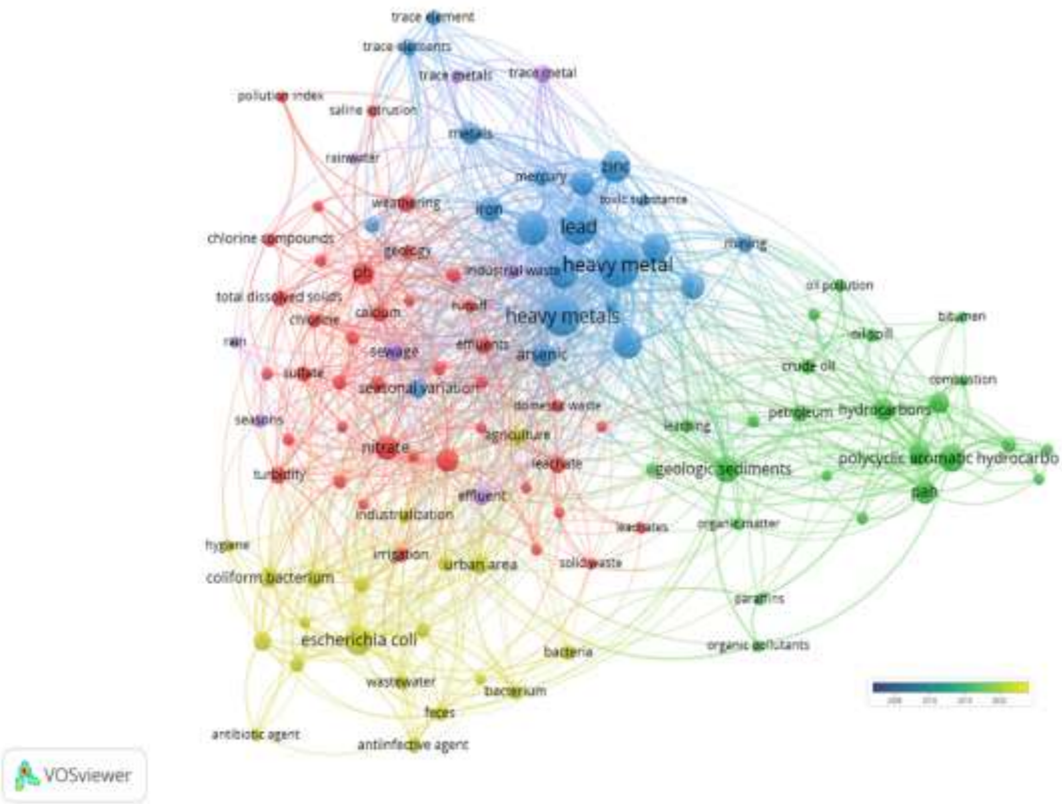

Fig. 6 Bibliographic coupling map for the sources of aquatic contamination in Nigeria

A relationship was established between anthropogenic and natural sources of contamination among all the reported articles on water quality contamination in Nigeria. From the results, it was observed that most studies attributed the occurrence of contamination in water to mining, followed by lithogenic and anthropogenic processes (Fig. 7).

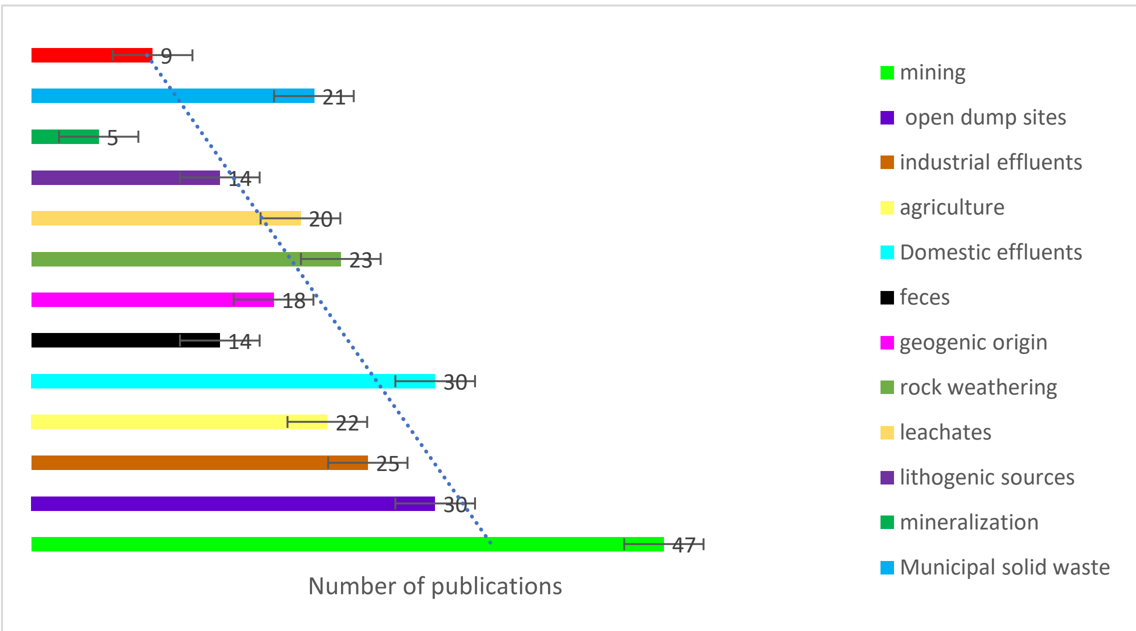

Fig. 7 Relationship of anthropogenic and natural sources of contamination in aquatic systems and the number of publications between 2003 and 2024

These results are in agreement with reports by Omeka & Egbueri (2023), that due to an upsurge in anthropogenic activities in the country, such as mining, industrialization, and socio-economic activities, high concentrations of heavy metal(loids) and potentially toxic elements have been reported in drinking water in the country. As reported by Obasi & Akudinobi (2020), there is an upward increase in the concentration of Cd, Pb, Cr, Zn, and cobalt in waters from the solid-mineral-rich southern Benue Trough in southeastern Nigeria. This has been attributed to the open-cast mining method that is prevalent among most mining operations in the country (Okolo et al. 2018; Omeka & Igwe 2021). In southeastern Nigeria, Ajala et al. (2022), provided a critical review of potential sources of heavy metals in water and fish as well as the human health risks from their consumption. High concentrations of Cd, As, Cu, and Zn were reported in both the drinking water and the shell of fishes. In Nsukka, southeastern Nigeria, there have been a reported high concentration of As, Cd, Hg, and Cr in water; attributed to anthropogenic influxes and run-off from domestic solid and sewage wastes (Nnaji et al. 2023).

Many studies in southwestern Nigeria have attributed the concentration of heavy metals in drinking water sources to industrialization. According to an extensive literature review carried out by Balogun et al. (2022) in the region, the long-standing sources of groundwater pollution cases have been attributed majorly to industrial effluents. In some parts of the Ibadan metropolis, southwest Nigeria, the concentration of heavy metals in shallow had-dug wells was observed to occur in the order of Zn > Fe > Pb > Cd > Mn; their occurrence is attributed to poor sanitation and industrialization (Ganiyu et al. 2021a). In the same region, the water quality of shallow aquifers was investigated for their metal content and bacterial load. The water was reportedly polluted due to pathogens and heavy metals; with the heavy metals occurring in the order of Cd > Pb > Zn > Fe > Mn, attributed to varying anthropogenic activities in the area (Ganiyu et al. 2021b). In the densely populated parts of the southwestern region, such as Lagos, water pollution has been associated with poor waste management practices, lowering of the water table (due to over-abstraction), and influx from industrial effluents (Ogundiran & Afolabi 2008). Some studies have also attributed the presence of heavy metals to both geogenic processes (such as weathering and leaching of subsurface geology) and anthropogenic sources (from industrial effluents). A study on the water quality assessment of Asa River in Ilorin, southwestern Nigeria revealed a high concentration of chromite ($FeCr_3O_4$) and pyrite (FeS) in river sediments, attributed to weathering of subsurface geology. Heavy metals such as Zn, Fe, Cr, and Mn were also reportedly found in elevated concentration due to the influx of industrial effluents in sediments (Adekola & Elleta 2007). Although some studies in the region have reported the presence of heavy metals in water due to vehicular emissions and related repair products, domestic sewage, and effluents (Tijani et al. 2004), most have been mostly due to industrial and agricultural effluents (Olaojo et al. 2016; Emenike et al. 2020).

In the northern and central parts of Nigeria, very few studies have reported water pollution due to industrial effluents. Most reports on heavy metal concentration in water have been attributed to mining activities. In the Anka gold mining area of northwestern Nigeria, there have been reported cases of Pb poisoning found in the hairs and nails of children within the vicinity of the mine site, (Adebwumi 2020). Although a study by Lar et al. (2015) reported the concentration of heavy metals (such as Zn, V, Pb, Cu, Co, Be Cr, As, Cd, Sb, and Se) in drinking water wells due to volcanic eruption in the Panyam volcanic province, in Jos, majority of the reports on metal concentration in the water have been attributed to mining and poor disposal of its effluents. Pb, Ni, and Hg poisoning have been reported in analyzed surface and groundwater samples from the Bagega gold mine province in Zamfara state, Northwestern Nigeria. The high elevated concentration of carcinogenic elements was attributed to artisanal gold mining (Nuhu et al. 2014).

In southeastern Nigeria, the sources of water pollution are highly ubiquitous – ranging from a wide range of anthropogenic sources (e.g., mining, poor waste disposal activities, poor hygiene, industries, population increase, land use, etc.) and geogenic sources (precipitation, chemical weathering, weathering, and rock-water interaction) (Omeka et al. 2023b; Aghamelu et al., 2022). Although most of the reports on contamination have been from both anthropogenic and geogenic activities (Nnorom et al. 2019; Edet et al. 2003), the majority have been from mining and poor waste disposal activities. Reports from mining have been majorly from the solid mineral-rich zones of the lower Benue Trough; with high concentrations of ore minerals occurring in association with Pb-Zn. (Omeka & Igwe 2023; Adamu et al. 2015). This has been mainly reported from Ebonyi, Enugu and Cross River states. A review carried out by Umeoguaju et al. (2022) for two decades (2000-2020) revealed that the concentration of heavy metals in most water sources in southeastern Nigeria is attributed majorly to anthropogenic influences from mining and oil exploration. This is in agreement with a study by Opuene & Agbozu (2008) on the heavy metal assessment in fish from Taylor Creek, southern Nigeria.

Application of neural networks in Water Quality Monitoring: implications for prospects

This section addresses the efficacy of artificial neural networks (ANNs) in water quality assessment, the most frequently used ANN models, the input or predictor variables, and the model accuracy based on the coefficient of determination (R^2).

To achieve this, a literature search was conducted on the application of neural networks in water quality modeling and prediction. The search was conducted on papers from the Scopus and Web of Science databases between 2003-2024, using the keywords "water quality and artificial intelligence" with emphasis on both surface and groundwater. The results from the search were compiled and presented in Table 4, and plots were generated based on the results in Table 4. As observed earlier from the bibliographic coupling map in Fig. 2, artificial neural networks appear to be the most frequently used machine learning model in water quality monitoring and assessment globally, followed only by support vector machines and decision trees. The wide usage of ANN in water quality studies has been attributed to its efficiency and versatility in predictions even in systems with poor computational strength (Egbueri et al. 2023; Omeka et al. 2024a). According to Ghavidel and Montazeri (2014), the unique ability of the ANN model to accurately match a broad range of nonlinear variables makes it a widely accepted ML model in most water quality studies. The ANN architecture is designed to mimic the human neural system; with the unique ability to quickly learn and send signals about a range of linear and nonlinear datasets through its interwoven parts known as "neurons" (Ozel et al. 2020). This makes it stand out among other water quality modeling techniques. Water quality data are usually nonstationary, random, nonlinear, and unpredictable (Beven, 2016). This means that the relationship between water quality and its controlling factors will tend to vary over time (as a result of varying anthropogenic and seasonal fluctuations), hence, the application of predictive models that work on historical data will become ineffective over time (Khan et al., 2021), hence more focus has been put on ANN due to its unique ability to process complex datasets.

Table 4 Dataset of selected articles from the Web of Science and Scopus databases used for SLR

Year	Location	ANN model	Input parameters	Highest R^2 (during the testing stage)	References
2023	1. Surface and groundwater, Ojoto suburb, southeastern Nigeria	MLP-ANN	TH, T, pH, TDS EC, Cl, Ca, SO_4, Pb, HCO_3, Zn Fe.	0.878	(Egbueri 2023)
	2. Groundwater samples from Osisioma, southeastern Nigeria	MLP-ANN	Cu, Pb, Fe As, Cr, Benzene, Ethylbenzene m-Xylene, Toluene, and o-Xylene	0.896	(Akakuru et al. 2023)
	3. Groundwater samples from Egbema, southeastern Nigeria	MLP-ANN	Fe, Zn, Ni, Cd, Cu and Pb	0.966	(Akakuru et al. 2023b)
	4. surface and groundwater from Okurumutet-Iyamitet mine province, southeastern Nigeria	MLP-ANN	HCO_3^-, SO_4^{2-}, NO_3^-, Cl^-, Mg^{2+}, Ca^{2+}, K^+ , and Na^+	0.861	(Omeka et al. 2023b)
2022	1. Bouregreg watershed, Morocco.	BPNN	pH and EC	0.87	(Bilali et al. 2022)
	2. Groundwater from Agartala municipality, India.	BPNN	Na^+, Mg^{2+}, Ca^{2+}, EC, HCO_3^-, and B	0.990	(Mallik et al 2022)
	3. groundwater from El Kharga Oasis, Western Desert of Egypt.	ANFIS	EC, pH, T°, TDS, Na^+, Mg^{2+}, K^+, Ca^{2+}, Cl^-, CO_3^{2-}, SO_4^{2-}, NO_3^- and HCO_3^-,	0.997	(Ibrahim 2022)
2021	(1) Groundwater wells from Illizi County, Algeria		pH, TH, EC, TDS, Ca, Mg, Na, K, HCO_3, Cl, SO_4, and NO_3	0.8957	(Kouadri et al. 2021)
	(2) groundwater from El Merk is an oil field in the SOUTHEAST of Algerian	BP-NN MLP-ANN	TH, NO_3, and NO_2	0.9967	(Kouadri & Samir 2021)

2020	1. The Elbe River, Germany	Wavelet and BPNN (W-BPNN)	Flow, pH, Fe, and DO	0.780	(Li et al. 2020)
	2. Groundwater from the Gaza Strip Palestine	MLP-NN	Abstraction average rate (AVR), relative humidity (RH), depth from the surface to well screen (DSWS), aquifer thickness (AT), recharge rate (RR), Initial chloride concentration (ICC), and groundwater level (GWL)	0.9770	(Kassem 2020)
	3. Yipin River, China	MLP-NN	pH, TP, Temperature, EC, PI, NH_3-N, COD and TN	0.717	(Zhu and Heddam 2020)
2019	Water samples from the Shivganga River basin, India	BP-NN	EC, pH, TH, TDS, Mg, Ca, K, Na, HCO_3, PO_4 NO_3, Cl, and SO_4	0.932	(Kadam et al. 2019)
	lakes, Tezpur University, India	MLP-ANN	BOD and TSS	0.783	(Ahamad et al. 2019)
2018	1. Gorganrood Basin, Iran	ANFIS	SAR	0.99	(Azad et al. 2018)
	2. Gorganrood Basin, Iran	ANFIS	EC	0.99	(Azad et al. 2018)

2017	1. Hooghly River, India	NN-CS	Chlorides, turbidity, pH, TA, and residual chlorine TH	-	(Chatterjee et al. 2017)
	2. Hooghly River, India	NN-GA	Chlorides, pH, residual chlorine TA TH, and turbidity	-	(Chatterjee et al. 2017)
	3. Saint John River, Canada	BPNN	TSS	0.976	(El Din and Zhang 2017)
	4. Langat River and Klang River, Malaysia	BPNN	BOD, DO, COD, Ph, NH3-N, and TSS	0.7267	(Hameed et al. 2017)
2016	1.Aji-Chay River, Iran	W-ANN	Salinity (EC)	0.9960	Barzegar et al. 2016
	2. Aji-Chay River, Iran	W-ANFIS	Salinity (EC)	0.9958	(Barzegar et al. 2016)
2015	1. Hilo Bay, Hawaii, USA	MLP	Salinity, DO, and Temperature	0.80	(Alizadeh and Kavianpour 2015)
	2. Hilo Bay, Hawaii, USA	W-ANN	Temperature, DO, and Salinity	0.967	(Alizadeh and Kavianpour 2015)
	3. Dahan River, Taiwan	BPNN	NH_3–N	0.979	(Chang et al. 2015)
	4. Nadong River, South Korea	W-ANN	Water level		(Seo et al. 2015)
2014	Karoon River, Iran	MLP-NN	DO	0.85	(Emamgholizadeh et al. 2014)
	Karoon River, Iran	MLP-NN	COD	0.74	(Emamgholizadeh et al. 2014)
2013	Johor River, Malaysia	MLP-NN	EC, TDS, and turbidity	0.799	(Najah et al. 2013)
	Jishan Lake, China	W-ANN	DO, Temperature, and pH	-	(Xu and Liu 2013)
	Jishan Lake, China	ENN	DO, Temperature, and pH	-	(Xu and Liu 2013)
2012	Kinta River, Malaysia	MLP-NN	36 parameters	0.765	(Gazzaz et al. 2012)
2011	Nile River, Egypt	MLP-NN	33 parameters	-	(Khalil et al. 2011)

Model utilization

It can be observed from Fig. 8 that the multilayer perceptron artificial neural network (MLP-NN) and back-propagated neural network (BP-NN) were the most frequently used neural network algorithms in water quality monitoring and prediction within the period of 2003-2024. This was closely followed by the adaptive neuro-fuzzy inference system (ANFIS), and other neural networks such as the Wavelet and BPNN (WNN) hybrid model, Neural Network trained by Cuckoo Search (NNCS), Neural Network trained by Genetic Algorithm (NN-GA), Elman Neural Network (ENN) and Wavelet-Adaptive Neural Fuzzy Interference System (WANFIS) in decreasing order. The high frequency of the MLP-NN and BP-NN in water quality monitoring and prediction is in line with a study by Rajaee et al. (2020). The wide acceptance of these algorithms has also been attributed to their ability to attain high modeling accuracy with fewer input parameters compared to other white-box models (Yetilmezsoy et al. 2011; Ighalo et al. 2020; Omeka 2023; Omeka et al., 2023b).

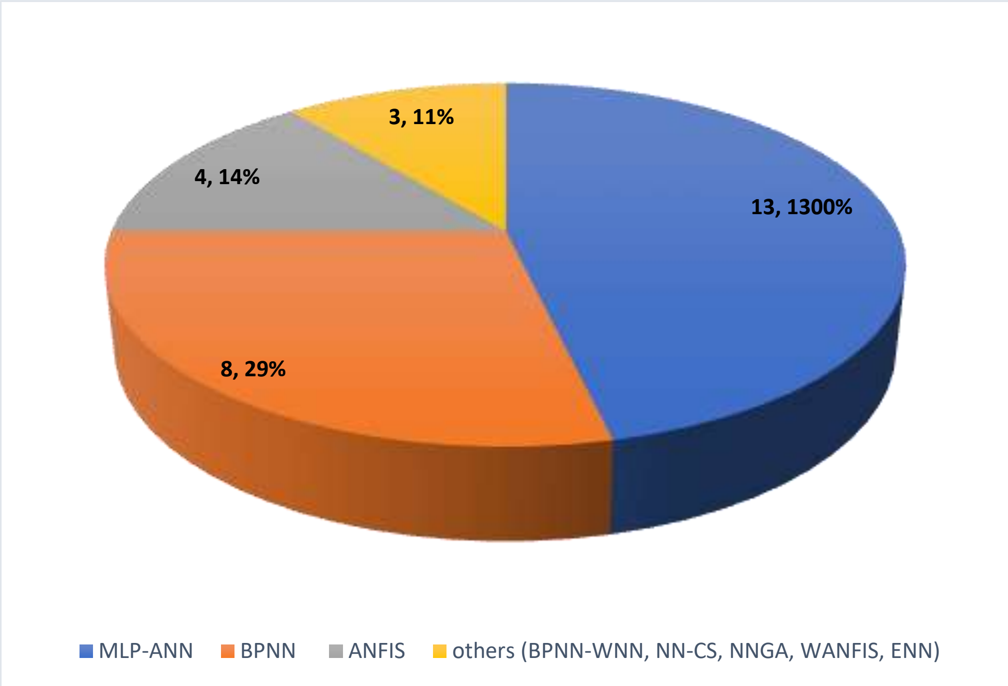

Fig. 8 Frequency in the application of ANN algorithms in water quality prediction

Model validation accuracy

In this assessment, only studies that used the coefficient of determination (R^2) as a validation metric were considered. The exclusion was made for the root mean square error (RMSE). The RMSE was excluded because of the difficulty of directly comparing the units to different parameters. This implies that its accuracy is dependent on the size (value) of the parameter being considered. This would mean that a parameter with a large numerical value will show a high RMSE value; resulting in inaccuracy (Ighalo et al. 2020). Unlike the RMSE, the R^2 is a measure of the extent of variation of a particular dataset, therefore giving better accuracy of parameters in a particular dataset (Adeniyi et al. 2019).

An observation of Table 4 shows that the adaptive neuro-fuzzy inference system (ANFIS), Wavelet-Adaptive Neural Fuzzy Interference System (W-ANFIS), and the Wavelet and BPNN (W-NN) hybrid models are the most accurate neural networks for both surface and groundwater monitoring and prediction. These findings are in agreement with the observation of Ighalo et al. (2020) on the frequently used neural networks for surface water prediction between 2011-2020. This study has revealed that although the MLP-NN and BP-NN seem to be the most popularly used algorithms in water quality prediction, however, they have low modeling accuracy compared to other hybrid models. The high accuracy of the W-ANFIS and W-NN hybrid models has been attributed to the in-cooperation of the wavelet decomposition of covariates (predictor variables) into detail and approximate components (Barzegar et al. 2016). On the other hand, the high accuracy of the ANFIS is because its architecture integrates the learning ability of both the Fuzzy Inference System and ANN (Emamgholizadeh et al. 2014).

Predicted water quality parameters

In this section, the most frequently analyzed water quality parameters for both surface and groundwater have been discussed.

An observation of Table 4 shows that temperature, dissolved oxygen (DO), pH, Electrical conductivity (EC; as a function of salinity), and total dissolved solids (TDS) are the most frequently investigated parameters in water quality prediction. A possible reason for their high prevalence is due to significance in the determination of the overall water quality in both surface and groundwater sources. Ighalo et al. (2020) have attributed the prevalence of DO, TDS, and pH to the high availability and low cost of measuring equipment. Moreover, these parameters are readily measured in situ before laboratory analysis because of their high susceptibility to surface environmental changes (Igwe and Omeka 2021).

CONCLUSION

With the aid of meta-analysis from Scopus and Web of Science databases and bibliometric coupling for data extraction and visualization, the present study has explored the recent trends to more hypothetical and advanced model simulation in the application of machine learning (ML) models in water quality monitoring and prediction in Nigeria for the last two decades (2003-2024). Emphasis has been made on the type of ML frequently used, the accuracy of the model in water quality prediction, and the frequently monitored water quality parameters. It was revealed that the use of hybridized ML models in water quality prediction has not been well explored globally; a majority of the prediction has been based on the use of artificial neural networks (ANN). The adaptive neuro-fuzzy inference system (ANFIS), and Wavelet-Adaptive Neural Fuzzy Interference System (W-ANFIS) hybrid models were observed to be the most accurate ANN algorithms, in prediction. Based on analysis of the literature, temperature, dissolved oxygen (DO), pH, conductivity (EC) and total dissolved solids (TDS) appear to be the most frequently predicted parameters. China and the United States have the highest number of publications concerning the application of ML in water quality prediction; while Nigeria appears among the countries grossly lagging behind.

Based on a literature analysis of the sources of contaminants in aquatic systems in Nigeria, heavy metals (Pb, Cd, Cu, Cr, and Zinc) were reported in more publications; linked majorly to mining and industrial effluents. This result implies that the increasing rate of anthropogenic activities coupled with climatic variation will continue to render water quality data complex; making the use of conventional assessment methods and standalone ML models inefficient. Therefore, future water quality monitoring and prediction studies in developing countries, especially in Nigeria – where data on environmental monitoring is lacking – should take advantage of the rapidly evolving field of machine learning; with more emphasis placed on the hybridized machine learning algorithms.

Although this study only considered the Scopus and Web of Science databases in its analysis, it is recommended that future studies should explore other databases such as Science Direct, PubMed, Cochrane Library, Lens, and Dimensions. Other bibliometric analytic techniques such as the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) should be explored as well.

Author contribution M. E. Omeka conceived and designed the project, conducted bibliometric modeling, and wrote the manuscript. M. I. Morphy, B. O. Omang1, E. A. Asinya1, & G.T. Kave carried out Manuscript review and editing. The first draft of the manuscript was written by M.E. Omeka., and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability Not applicable.

Declarations:

Consent for publication Not applicable.

Consent to participate Not applicable.

Competing interests The author declares no competing interests regarding this research

Funding No funding was received for this research

REFERENCES

Abugu, H. O., Egbueri, J. C., Agbasi, J. C., Ezugwu, A. L., Omeka, M. E., Ucheana, I. A., and Aralu, C. C., 2024. Hydrochemical Characterization of Ground and Surface Water for Irrigation Application in Nigeria: A Review of Progress. Chemistry Africa, 1-26.

Adamu, C. I., Nganje, T. N., and Edet, A., 2015. Heavy metal contamination and health risk assessment associated with abandoned barite mines in Cross River State, southeastern Nigeria. Environmental Nanotechnology, monitoring & management, 3, 10-21.

Adekola, F. A., and Eletta, O. A. A., 2007. A study of heavy metal pollution of Asa River, Ilorin. Nigeria; trace metal monitoring and geochemistry. Environmental Monitoring and Assessment, 125, 157-163.

Adeniyi AG, Ighalo JO, Marques G., 2020 Utilisation of machine learning algorithms for the prediction of syngas composition from biomass bio-oil steam reforming. Int J Sustain Energy. https://doi. org/10.1080/14786451.2020.1803862

Adewumi, A. J. "Contamination, sources and risk assessments of metals in media from Anka artisanal gold mining area, Northwest Nigeria." Science of the Total Environment 718, 2020: 137235.

Agbasi, J.C, Egbueri J.C, Ayejoto, D. A., Unigwe, C.O., Omeka, M.E., Nwazelibe V.E., Ighalo J.O., Pande C.B., Fakoya A.A., 2023. The impact of seasonal changes on the trends of physicochemical, heavy metal and microbial loads in water resources of southeastern Nigeria: A critical review. In: Egbueri J.C, Ighalo J.O, Pande C.B (eds) Climate change impacts on Nigeria. Springer, Cham. https://doi.org/10.1007/978-3-031-21007-5_25

Aghamelu, O. P., Omeka, M. E., and Unigwe, C. O., 2022. Modeling the vulnerability of groundwater to pollution in a fractured shale aquifer in SE Nigeria using information entropy theory, geospatial, and statistical modeling approaches. Modeling Earth Systems and Environment, 1-22.

Ahamad KU, Raj P, Barbhuiya NH, Deep A., 2019 Surface water

Ajala, O. A., Oke, M. R., Ajibade, T. F., Ajibade, F. O., Adelodun, B., Ighalo, J. O., ...and Silva, L. F., 2022. Concentrations, bioaccumulation, and health risk assessments of heavy metals in fishes from Nigeria's freshwater: a general overview. Environmental Science and Pollution Research, 29(55), 82660-82680.

Akakuru, O. C., Adakwa, C. B., Ikoro, D. O., Eyankware, M. O., Opara, A. I., Njoku, A. O., ... and Usman, A., 2023. Application of artificial neural network and multi-linear regression techniques in groundwater quality and health risk assessment around Egbema, Southeastern Nigeria. Environmental Earth Sciences, 82(3), 77.

Akakuru, O. C., Njoku, U. B., Obinna-Akakuru, A. U., Akudinobi, B. E., Obasi, P. N., Aigbadon, G. O., and Onyeanwuna, U. B., 2023. Non-carcinogenic health risk assessment and predicting of pollution indexing of groundwater around Osisioma, Nigeria, using artificial neural networks and multi-linear modeling principles. Stochastic Environmental Research and Risk Assessment, 1-31.

Al Aani S, Bonny T, Hasan SW, Hilal N., 2019. Can machine language and artificial intelligence revolutionize process automation for water treatment and desalination? Desalination 458:84–96

Alizadeh MJ, Kavianpour MR., 2015. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay Pacific Ocean. Mar Pollut Bull 98:171–178

Aluma, V. C., Igwe, O., Omeka, M. E., and Anyanwu, I. E., 2024. Combining multiple numerical and chemometric models for assessing the microbial and pollution level of groundwater resources in a shallow alluvial aquifer, Southeastern Nigeria. Arabian Journal of Geosciences, 17(7), 1-19.

Ayejoto, D.A, Egbueri, J.C, Agbasi, J.C, Omeka, M.E, Unigwe C.O, Nwazelibe V.E, Ighalo, J.O, Pande C.B., 2023. Influence of Seasonal Changes on the Quality of Water Resources in Southwestern Nigeria: A Review. In: Egbueri J.C, Ighalo J.O, Pande C.B (eds) Climate change impacts on Nigeria. Springer, Cham. https://doi.org/10.1007/978-3-031-21007-5_22

Azad A, Karami H, Farzin S, Saeedian A, Kashi H, Sayyahi F., 2018. Prediction of water quality parameters using ANFIS optimized by intelligence algorithms (case study: Gorganrood River). KSCE J Civ Eng 22:2206–2213

Barzegar R, Adamowski J, Moghaddam AA., 2016. Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. Stoch Environ Res Risk Assess 30:1797–1819

Beven K., 2016 Advice to a young hydrologist. Hydrol Process 30(20):3578–3582. https://doi.org/10.1002/hyp.10879

Chang F-J, Tsai Y-H, Chen P-A, Coynel A, Vachaud G., 2015. Modeling water quality in an urban river using hydrological factors– Data-driven approaches. J Environ Manage 151:87–96

Chatterjee S, Sarkar S, Dey N, Ashour AS, Sen S, Hassanien AE., 2017. Application of cuckoo search in water quality prediction using artificial neural network. Int J Comput Intell Stud 6:229–244

Deng T, Duan HF, Keramat A., 2022. Spatiotemporal characterization and forecasting of coastal water quality in the semi-enclosed Tolo Harbour based on machine learning and EKC analysis. Eng Appl Comput Fluid Mechanics 16(1):694–712. https://doi.org/10.1080/19942060.2022.2035257

Edet, A. E., Merkel, B. J., and Offiong, O. E., 2003. Trace element hydrochemical assessment of the Calabar Coastal Plain Aquifer, southeastern Nigeria using statistical methods. Environmental Geology, 44, 137-149.

Egbueri JC, Agbasi JC., 2022. Data-driven soft computing modeling of groundwater quality parameters in southeast Nigeria: comparing the performances of different algorithms. Environ Sci Pollut Res 29(25):38346–38373

Egbueri JC, Igwe O, Omeka ME, Agbasi JC., 2023. Development of MLR and variedly optimized ANN models for forecasting the detachability and liquefaction potential index of erodible soils. Geosystems and Geoenvironment 2(1):100104

Egbueri, J. C., 2022. Incorporation of information entropy theory, artificial neural network, and soft computing models in the development of integrated industrial water quality index. Environmental Monitoring and Assessment, 194(10), 693.

El Bilali, A., Taleb, A., Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment, Journal of the Saudi Society of Agricultural Sciences, 2020. doi:https://doi.org/10.1016/j.jssas.2020.08.001

El Din ES, Zhang Y., 2017. Improving the accuracy of extracting surface water quality levels (Swqls) using remote sensing and artificial neural network: a case study in the saint john river, Canada. International archives of the photogrammetry. Remote sensing & spatial information sciences, vol XLII-4/W4, 245–249

Emamgholizadeh S, Kashi H, Marofpoor I, Zalaghi E., 2014 Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. Int J Environ Sci Technol 11:645–656

Emenike, P. C., Tenebe, I. T., Neris, J. B., Omole, D. O., Afolayan, O., Okeke, C. U., and Emenike, I. K., 2020. An integrated assessment of land-use change impact, seasonal variation of pollution indices, and human health risk of selected toxic elements in sediments of River Atuwara, Nigeria. Environmental Pollution, 265, 114795.
Encyclopædia Britannica, Inc., 2023.

Ganiyu, S. A., Mabunmi, A. A., Olurin, O. T., Adeyemi, A. A., Jegede, O. A., and Okeh, A., 2021b. Assessment of microbial and heavy metal contamination in shallow hand-dug wells bordering Ona River, Southwest Nigeria. Environmental Monitoring and Assessment, 193, 1-22.

Ganiyu, S. A., Oyadeyi, A. T., and Adeyemi, A. A., 2021a. Assessment of heavy metals contamination and associated risks in shallow groundwater sources from three different residential areas within the Ibadan metropolis, southwest Nigeria. Applied Water Science, 11(5), 81.

Gazzaz NM, Yusof MK, Aris AZ, Juahir H, Ramli MF., 2012. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar Pollut Bull 64:2409–2420

Geetha S, Gouthami S., 2016. Internet of things enabled real-time water quality monitoring system Smart. Water 2:1. https://doi.org/10.1186/s40713-017-0005-y

Ghavidel SZZ, Montazeri M., 2014. Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. Stochastic Environ Res Risk Ass 28(8):2101– 2118. https://doi.org/10.1007/s00477-014-0899-y

Hameed M, Sharqi SS, Yaseen ZM, Afan HA, Hussain A, Elshafe A., 2017. Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in the tropical region, Malaysia. Neural Comput Appl 28:893–905

Hasan R, Raghav A, Mahmood S, Hasan MA Artificial Intelligence in River Quality Assessment. In: 2011 International Conference on Information Management, Innovation Management and Industrial Engineering, 2011. IEEE, pp 491-495

Ibrahim, H.; Yaseen, Z.M.; Scholz, M.; Ali, M.; Gad, M.; Elsayed, S.; Khadr, M.; Hussein, H.; Ibrahim, H.H.; Eid, M.H.; et al. Evaluation and Prediction of Groundwater Quality for Irrigation Using an Integrated Water Quality Indices, Machine Learning Models and GIS Approaches: A Representative Case Study. Water 2023, 15, 694. https://doi.org/10.3390/w15040694

Ifediegwu SI., 2022 Assessment of groundwater potential zones using GIS and AHP techniques: a case study of the Lafa district, Nasarawa State, Nigeria. Appl Water Sci 12:10. https://doi.org/ 10.1007/s13201-021-01556-5

Ighalo JO, Adeniyi AG, Marques G., 2020. Artificial intelligence for surface water quality monitoring and assessment A systematic literature analysis. Model Earth Syst Environ. https://doi.org/10. 1007/s40808-020-01041-z

Igwe, O., and Omeka, M. E., 2022. Hydrogeochemical and pollution assessment of water resources within a mining area, SE Nigeria, using an integrated approach. International Journal of Energy and Water Resources, 6(2), 161-182.

Kadam AK, Wagh VM, Muley AA, Umrikar BN, Sankhua RN., 2019. Prediction of water quality index using artificial neural network and multiple linear regression modeling approaches in Shivganga River basin. Model Earth Syst Environ, India. https://doi.org/10. 1007/s40808-019-00581-3

Khalil B, Ouarda T, St-Hilaire A., 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. J Hydrol 405:277–287

Khan N, Shahid S, Ismail TB, Behlil F., 2021. Prediction of heat waves over Pakistan using support vector machine algorithm in the context of climate change. Stochastic Environ Res Risk Assess. 1–19https://doi.org/10.1007/s00477-020-01963-1

Kouadri S, Elbeltagi A, Islam ARM, Kateb S., 2021 Performance of machine learning methods in predicting water quality index based on the irregular data set: application on Illizi region (Algerian southeast). Appl Water Sci 11(12):1–20

Kouadri S, Samir K., 2021. Hydro-chemical study with geospatial analysis of groundwater Quality Illizi Region, South-East of Algeria. Iran J Chem Chemical Eng (IJCCE) 40(4):1315–1333. https://doi. org/10.30492/ijcce.2020.39800

Kouadri S, Samir K., 2021. Hydro-chemical study with geospatial analysis of groundwater Quality Illizi Region, South-East of Algeria. Iran J Chem Chem Eng (IJCCE) 40(4):1315–1333. https://doi.org/10.30492/ijcce.2020.39800

Lar, U. A., and Gusikit, R. B., 2015. Environmental and health impact of potentially harmful elements distribution in the Panyam (Sura) volcanic province, Jos Plateau, Central Nigeria. Environmental earth sciences, 74, 1699-1710.

Mallik, S., Chakraborty, A., Mishra, U., and Paul, N., 2022. Prediction of irrigation water suitability using geospatial computing approach: a case study of Agartala city, India. Environmental Science and Pollution Research, 1-16.

Martins, P., Forstinus, N. O., and Ikechukwu, N. E., 2016. Water and Waterborne Diseases: A Review Water and Waterborne Diseases: A Review. on. JANUARY.

Mohammadpour R, Shaharuddin S, Zakaria N, Ghani A, Vakili M, Chan N., 2016. Prediction of water quality index in free surface constructed wetlands. Environ Earth Sci 75:1–12. https://doi. org/10.1007/s12665-015-4905-6

Morounke, S. G., Ayorinde, J. B., Benedict, A. O., Adedayo, F. F., Adewale, F. O., Oluwadamilare, I., Sokunle, S. S. and Benjamin, A., 2017. Epidemiology and Incidence of Common Cancers in Nigeria. J Cancer Biol Res, 5(3), 1105.

Najah A, El-Shafe A, Karim OA, El-Shafe AH., 2013. Application of artificial neural networks for water quality prediction. Neural Comput Appl 22:187–201

National Population Commission., 2013. Nigeria Demographic and Heath Survey 2013. National Population Commission, ICF International

Nayak S., 2020. Assessment of water quality of Brahmani River using correlation and regression analysis

Nigerian National System of Cancer Registries (2013) Federal Ministry of Health of Nigeria, Federal Secretariat Complex Shehu Shagari Way, Garki, Abuja P.M.B. 083, GarkiAbuja ISBN: 978-978-955-030-2

Nnaji, C. C., Chibueze, C. V., Mama, C. N., Alum, O. L., and Afangideh, C. B., 2023. Impact of anthropogenic and environmental conditions on surface run-off quality: a case study of Nsukka, Eastern Nigeria. International Journal of Environmental Science and Technology, 1-12.

Nnorom, I. C., Ewuzie, U., and Eze, S. O., 2019. Multivariate statistical approach and water quality assessment of natural springs and other drinking water sources in Southeastern Nigeria. Heliyon, 5(1), e01123.

Nuhu, A. A., Sallau, M. S., and Majiya, M. H., 2014. Heavy metal pollution: the environmental impact of artisanal gold mining on Bagega village of Zamfara state, Nigeria. Res J Pharm Biol Chem Sci, 5(6), 306-313.

Obasi, P. N., and Akudinobi, B. B., 2020. Potential health risk and levels of heavy metals in water resources of lead–zinc mining communities of Abakaliki, southeast Nigeria. Applied Water Science, 10, 184. https://doi.org/10.1007/s13201-020-01233-z

Ogundiran, O. O., and Afolabi, T. A., 2008. Assessment of the physicochemical parameters and heavy metals toxicity of leachates from the municipal solid waste open dumpsite. International Journal of Environmental Science & Technology, 5, 243-250.

Okamkpa, J.R. Omeka M.E., Igwe, O. and Iyiokwu M.U., 2022. An integrated geochemical and spatiotemporal assessment of groundwater resources within an industrial suburb, Southeastern Nigeria. International Journal of Energy and Water Resources, 1-20. https://doi.org/10.1007/s42108-022-00183-3

Okolo, C. C., Oyedotun, T. D. T., and Akamigbo, F. O. R., 2018. Open cast mining: a threat to water quality in the rural community of Enyigba in south-eastern Nigeria. Applied water science, 8, 1-11.

Olaojo, A. A., Oladunjoye, M. A., and Sanuade, O. A., 2018. Geoelectrical assessment of polluted zone by sewage effluent in University of Ibadan campus southwestern Nigeria. Environmental monitoring and assessment, 190, 1-12.

Omang, B. O., Vishiti, A., Kave, G. T., Enah, E. A., and Suh, C. E., 2023a. Microchemical characterization and stream sediment composition of alluvial gold particles from the Rafin Gora drainage system, Kushaka schist belt, North Western Nigeria. Communication in Physical Sciences, 9(3).

Omang, B. O., Omeka, M. E., Asinya, E. A., Oko, P. E., and Aluma, V. C., 2023b. Application of GIS and feedforward back-propagated ANN models for predicting the ecological and health risk of potentially toxic elements in soils in Northwestern Nigeria. Environmental Geochemistry and Health, 45(11), 8599-8631.

Omeka ME, Egbueri JC., 2023. Hydrogeochemical assessment and health-related risks due to toxic element ingestion and dermal contact within the Nnewi-Awka urban areas, Nigeria. Environ Geochem Health 1–29. https://doi.org/10.1007/s10653-022-01332-7

Omeka, M. E., and Igwe, O., 2023. Heavy metals concentration in soils and crop plants within the vicinity of abandoned mine sites in Nigeria: an integrated indexical and chemometric approach. International Journal of Environmental Analytical Chemistry, 103(16), 4111-4129.

Omeka, M. E., Ezugwu, A. L., Agbasi, J. C., Egbueri, J. C., Abugu, H. O., Aralu, C. C., Ucheana, I. A., 2024a. A review of the status, challenges, trends, and prospects of groundwater quality assessment in Nigeria: an evidence-based meta-analysis approach. Environmental Science and Pollution Research, 1-24.

Omeka, M. E., Igwe, O., Onwuka, O. S., Obajaja, H. A., Omang, B. O., Unigwe, C. O., and Aluma, V. C., 2024b. Modeling the vulnerability of water resources to pollution in a typical mining area, SE Nigeria using speciation, geospatial, and multi-path human health risk modeling approaches. Modeling Earth Systems and Environment, 1-30.

Omeka, M. E., Egbueri, J. C., and Unigwe, C. O., 2022a. Investigating the hydrogeochemistry, corrosivity, and scaling tendencies of groundwater in an agrarian area (Nigeria) using graphical, indexical, and statistical modeling. Arabian Journal of Geosciences, 15(13), 1-24. https://doi.org/10.1007/s12517-022-10514-7

Omeka, M.E., 2023. Evaluation and prediction of irrigation water quality of an agricultural district, SE Nigeria: an integrated heuristic GIS-based and machine learning approach. Environmental Science and Pollution Research, https://doi.org/10.1007/s11356-022-25119-6

Omeka, M. E., Egbueri, J. C., Unigwe, C. O., Agbasi, J. C., Ayejoto, D. A., and Nwazelibe, V. E., 2023a. A review on the influence of rainfall in the formation and expansion of gullies in Southeastern Nigeria. Climate change impacts on Nigeria: environment and sustainable development, 479-503.

Omeka, M. E., Igwe, O., Onwuka, O. S., Nwodo, O. M., Ugar, S. I., Undiandeye, P. A., and Anyanwu, I. E., 2023b. Efficacy of GIS-based AHP and data-driven intelligent machine learning algorithms for irrigation water quality prediction in an agricultural-mine district within the Lower Benue Trough, Nigeria. Environmental Science and Pollution Research, 1-30.

Opuene, K., and Agbozu, I. E., 2008. Relationships between heavy metals in shrimp (Macro brachium felicinum) and metal levels in the water column and sediments of Taylor Creek.

Ozel HU, Gemici BT, Gemici E, Ozel HB, Cetin M, Sevik H., 2020. Application of artificial neural networks to predict the heavy metal contamination in the Bartin River. Environ Sci Pollut Res. https:// doi.org/10.1007/s11356-020-10156-w

Pour SH, Wahab AKA, Shahid S., 2020. Physicalempirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia. Atmos Res 233:104720. https://doi.org/10.1016/j.atmosres.2019.104720

quality modeling by regression analysis and artificial neural network. In: Kalamdhad A, Singh J, Dhamodharan K (eds) Advances in waste management. Springer, Singapore. https://doi.org/10.1007/978-981-13-0215-2_15

Rajaee T, Khani S, Ravansalar M., 2020. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: a review. ChemometrIntell Lab Syst 200:103978

Ricolfi, L., Barbieri, M., Muteto, P. V., Nigro, A., Sappa, G., Vitale, S., 2020. Potential toxic elements in groundwater and their health risk assessment in drinking water of Limpopo National Park, Gaza Province, Southern Mozambique. Environmental Geochemistry and Health, 1–13.

Seo Y, Kim S, Kisi O, Singh VP., 2015. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. J Hydrol 520:224–243

Sun K, Rajabtabar M, Samadi S, Rezaie-Balf M, Ghaemi A, Band S, Mosavi A., 2021 An integrated machine learning, noise suppression, and population-based algorithm to improve total dissolved solids prediction. Eng Appl Comput Fluid Mech 15(1):251–271. https://doi.org/10.1080/19942060.2020.1861987

Tijani, Moshood N., Kennji Jinno, and Yoshinari Hiroshiro. "Environmental impact of heavy metals distribution in water and sediments of Ogunpa River, Ibadan area, southwestern Nigeria." Journal of Mining and Geology 40, no. 1, 2004: 73-83.

Umeoguaju, F. U., Akaninwor, J. O., Essien, E. B., and Amadi, B. A., 2022. Heavy metal profile of surface and groundwater samples from the Niger Delta region of Nigeria: a systematic review and meta-analysis. Environmental Monitoring and Assessment, 194, 1-37.

van Ech NJ, Waltman L., 2020. VOSviewer Manual. Universiteit Leiden, The Centre for Science and Technology Studies, Meaningful metrics. https://www.vosviewer.com/. Accessed February 2023

World Health Organisation/United Nations Children Fund. WHO/UNICEF 2010. Progress on drinking water and sanitation

Xiao Y, Liu K, Hao Q, Li J, Zhang Y, Cui W, Pei Q., 2021. Hydrogeochemical features and genesis of confined groundwater and health perspectives for sustainable development in urban Hengshui, North China Plain. J Chem

Xu L, Liu S., 2013. Study of short-term water quality prediction model based on wavelet neural network. Math Comput Model 58:807–813

Yetilmezsoy K, Ozkaya B, Cakmakci M., 2011. Artificial intelligence-based prediction models for environmental engineering. Neural Netw World 21:193

Zhu S, Heddam S., 2020 Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). Water Qual Res J 55:106–118