AVE Trends in Intelligent Computing Systems

Robust Learning Grey Wolf Optimiser-Based Feature Selection with ACBLSTM Classifier for Groundwater Quality Monitoring

M. Arunadevi Thirumalraj[1,*], V. Revathi[2], R. J. Anandhi[3], Prasanna Ranjith Christodoss[4]

[1]Department of Computer Science and Engineering, Karunya Institute of Technology and Science, Coimbatore, Tamil Nadu, India.
[1]Department of Computer Science and Business Management, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India.
[2]Department of Research and Development, New Horizon College of Engineering, Bengaluru, Karnataka, India.
[3]Department of Information Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India.
[4]Department of Computing, Mathematics and Physics, Messiah University, One University Ave, Mechanicsburg, Pennsylvania, United States of America.
aruna.devi96@gmail.com[1], revshank153@gmail.com[2], rjanandhi@hotmail.com[3], prchristodoss@messiah.edu[4]

*Corresponding author

Abstract: Improving the management of water resources globally, particularly in arid regions, requires evaluating the quality of the water. This study's objective is to evaluate and track groundwater quality using artificial intelligence (AI) techniques in conjunction with hydrochemical parameters. The scientific community has increasingly focused on groundwater quality monitoring over the past few decades. While the WQI is a useful instrument for evaluating groundwater quality, its classification accuracy isn't always optimal, particularly in large-scale databases. Consequently, this manuscript develops an ACBLSTM model for classifying groundwater quality. After applying Min-Max and Z-score normalisation methods to data collected from Indian and real-time water quality databases, the WQI calculation and data elimination are complete. The RLGWO method is employed to select the best features from the denoised data samples. Strong tolerance-based search direction adjustment and opposite-based learning reinforce this algorithm, which mimics the social hierarchy and hunting techniques of natural grey wolves. The suggested model has classified data from the Indian water quality database and real-time database with approximately 95% accuracy.

Keywords: Robust Learning Grey Wolf Optimisation (RLGWO); Water Quality Index (WQI); Convolutional Neural Network; Attention-based Convolutional Neural Network with Bidirectional Long-Short Term Memory (ACBLSTM); Groundwater Quality Monitoring.

Cite as: M. A. Thirumalraj, V. Revathi, R. J. Anandhi, and P. R. Christodoss, "Robust Learning Grey Wolf Optimiser-Based Feature Selection with ACBLSTM Classifier for Groundwater Quality Monitoring," *AVE Trends in Intelligent Computing Systems*, vol. 2, no. 3, pp. 142–154, 2025.

Journal Homepage: https://avepubs.com/user/journals/details/ATICS

Received on: 07/11/2024, **Revised on:** 14/02/2025, **Accepted on:** 13/04/2025, **Published on:** 09/09/2025

DOI: https://doi.org/10.64091/ATICS.2025.000198

1. Introduction

In many nations, groundwater serves as the primary source of water for farming, industry, and residential use. Groundwater is, in fact, the primary water source and, at the moment, the most precious natural resource for many different kinds of human endeavours [1]. For both direct consumption and food production, more than one-third of people on Earth are dependent on groundwater. Furthermore, many developing and underdeveloped nations have looked to groundwater as a source of clean drinking water, especially for people living in rural areas [2]. As a result, severe water shortages and poor water quality are widespread in many countries. For many years, there have been growing concerns about the health effects of drinking water quality [3]. Developing solutions to issues related to water resources, evaluating water quality, preventing flooding, understanding the natural world, and overseeing regional or local management of water resources can all be accomplished with the help of geographic information systems (GIS) [4]. Additionally, the GIS's spatial analysis extension enables interpolation of known values for groundwater-quality parameters at unknown locations, yielding a continuous surface that contributes to our understanding of the factors governing water quality and their distribution in the research region.

Furthermore, the overall water quality is evaluated using the WQI technique to rank the effects of each water quality parameter [5]. WQI helps provide targeted information to citizens and decision-makers about water quality. But assessing water quality presents several challenges, including obtaining a large sample size, conducting lab tests, and handling data. These are usually labour-intensive processes that cost a lot of money in terms of supplies, equipment, and labourers [6]. To guarantee that groundwater is safe for irrigation and human consumption, regular monitoring and quality assessment are essential [7]. It should be noted that the COVID-19 pandemic over the past three years has caused a shortage of chemical analysis reactors in laboratories across several nations [8]. Therefore, to accurately assess water quality and resolve the aforementioned issues, it is imperative to develop time- and cost-efficient methods. This can be done in many ways, including modelling, remote sensing, water quality sampling, and analysis. Recently, there has been an increased focus on artificial intelligence (AI) and its potential applications in managing and monitoring water quality [9]. AI has been applied to several tasks related to water quality, including data collection, analysis, and decision-making. Applications in water quality can leverage a variety of AI technologies. Machine learning is the most popular form of AI technology and can be applied to tasks such as prediction and data classification [10].

Artificial neural networks, rule-based systems, and evolutionary computation are other AI technologies that have been used to improve water quality. Artificial intelligence (AI) methods, such as support vector machines (SVMs), random forests (RFs), and artificial neural networks (ANNs), are widely used in water-related research worldwide [11]. In arid regions, evaluating and tracking groundwater quality has been the subject of numerous studies, particularly for managing irrigation water quality. To increase agricultural productivity in arid regions, groundwater accessibility and suitability for drinking and irrigation are essential. Based on specific input parameters, researchers have employed multiple artificial intelligence (AI) methods to forecast the irrigation water quality index (IWQI) [12]. Given the complexity of the factors affecting groundwater quality, a thorough evaluation approach is necessary. The WQI is determined by calculating several factors, including organic matter, pH, turbidity, temperature, and electrical conductivity. Consequently, WQI computation proved time-consuming and successful, but also entailed unintentional errors [13]. As a result, a large number of mathematical models are applied using both deep learning and machine learning techniques. This paper presents a novel groundwater quality classification model developed with artificial intelligence and advanced computing. The following are the contributions made by this work:

- Z-score and Min-Max normalisation techniques are used for the WQI calculation and data denoising after the databases on Indian water quality and real-time data are obtained. By scaling the range of the acquired data, the data denoising process improves its quality.
- Using the scaled data, the RLGWO technique is developed to select the best attributes/features. This technique can generate strong candidates for a workable solution while also successfully preventing search agents from becoming trapped in local optima. Furthermore, choosing the best features and attributes reduces the system's complexity and computation time.
- Lastly, the selected characteristics/features are fed into the ACBLSTM classifier to categorise groundwater quality into good, poor, very poor, and excellent. Evaluation metrics, including false discovery rate (FDR), sensitivity, accuracy, Matthews correlation coefficient (MCC), and specificity, are used to assess the efficiency of the proposed model.

2. Related Works

Raheja et al. [14] used 94 groundwater samples collected from tube wells across various locations in the Rohtak district of Haryana (India) during the pre- and post-monsoon seasons of 2022. Each sample's 14 hydrochemical parameters were measured and compared with the standard values recommended for drinking purposes by the Bureau of Indian Standards (BIS) 10,500:2015 and the World Health Organisation (WHO). The pre- and post-monsoon seasons showed variations in the Drinking Water Quality Index (DWQI), which was calculated from various hydrochemical parameters. These variations ranged from 95.02 to 448.92 and from 93.91 to 497.72, respectively. The western region of the study area had non-drinking-quality water values in both seasons, as indicated by spatial distribution maps of various hydrochemical parameters and the DWQI. The DWQI value was predicted using two machine learning algorithms: Support Vector Regression (SVR) with four kernel functions, and Gaussian Process Regression (GPR) with post-monsoon samples (test) and pre-monsoon samples (train). The findings show that when utilising both the SVR and GPR approaches, the SVR algorithm with the Radial basis kernel (SVR-RBF) outperforms other kernel functions. Sensitivity analysis indicates that three parameters are essential for DWQI prediction: TDS, $NO3^-$, and F^-. The government agencies will benefit from the suggested algorithms' results, which will offer alternative drinking water options in the affected areas.

Subbarayan et al. [15] evaluate groundwater's susceptibility to nitrate contamination using various classification models. Examining three and five classes further helps to understand the significance of the multi-class classification's use of classes. For the current study, three machine learning models with two classification schemes were developed: Extreme Gradient Boosting, Random Forest, and CART. The study's parameters, derived from the traditional DRASTIC approach, were combined with another parameter, land use. To choose the best model, evaluation metrics including Area Under the Curve of the Receiver Operating Characteristic (AUC-ROC), Positive Predictive Value, Negative Predictive Value, Accuracy, and Kappa were compared across all six models. Given the selected objective, the Random Forest model achieves an AUC of 0.95 and shows three-class classification, as evidenced by its consistent area distribution across the classes and model evaluation metrics. To evaluate groundwater vulnerability, this study emphasises the importance of data classification and the selection of the number of classes for machine learning models. The suggested method, leveraging cutting-edge machine learning techniques and Geographic Information Systems, provides insightful recommendations for improved groundwater management and contamination mitigation. Yazdi et al. [16] introduced the Fuzzy Groundwater Quality Index (FGWQI), which uses a fuzzy inference system to evaluate the groundwater quality of the Houmand-Absard aquifer, particularly for Iranian irrigation. Five quality indicators related to agriculture are combined into the FGWQI: Potential Salinity (PS), Magnesium Hazard Ratio (MHR), Sodium Percentage (Na%), and Kelly's Index (KI). Studies utilising water samples collected from 20 stations during the 2021 season were conducted to compare the FGWQI and the commonly used MHR.

The parameters of the samples, including magnesium, calcium, chloride, sodium, and sulphate, were determined by laboratory analyses. The FGWQI model performed better than conventional measures, most notably MHR, indicating more accurate evaluations. Even though MHR considered the water quality inappropriate in 47% of instances, the FGWQI was above 60, indicating acceptable outcomes (30–70), suggesting that more water resources could be allocated to agriculture if groundwater management were properly implemented. Farmers in the case study region will have more flexibility in using groundwater because FGWQI will replace MHR and prevent the decommissioning of water wells deemed suitable for irrigation. Kwak and Lee [17] employ a data-driven methodology to assess the effect of groundwater quality on road transport systems. In particular, an open data portal in Texas provided information on the state's road network and groundwater chemistry. The modelling stage for explainable artificial intelligence (XAI) and the multivariate analysis stage comprised the two stages of this investigation. Extreme Gradient Boosting (XGB) was used to develop a prediction model during the XAI modelling phase. The model used road transport attributes, such as speed limit, elevation, lane-miles, and annual average daily traffic, as input features, and groundwater chemistry parameters as output characteristics. Additionally, the accumulated local effect (ALE) and feature importance were used to analyse the relationships between groundwater chemistry and road transportation characteristics. The XAI models provided the groundwater chemistry parameters selected for the multivariate phase using principal component analysis (PCA) and Piper diagrams. The prediction model's results, with a mean percentage error as an absolute value of less than 0.20, demonstrated that the five groundwater chemistry parameters of pH, temperature, alkalinity, bicarbonate ($HCO3^-$), and aluminium (Al) were significantly impacted by road transport systems. Additionally, XAI models were created to understand the relationships between five specific parameters and road transport attributes.

The results as a whole showed that road transport systems within 50 metres and a well depth of 100 metres have a significant influence on the qualities of Texas groundwater. Using XAI techniques, this research provides a fresh perspective on point-source monitoring of groundwater pollution. Yang et al. [18] found that the prediction model's results, with a mean percentage error as an absolute value of less than 0.20, demonstrated that the five groundwater chemistry parameters of pH, bicarbonate ($HCO3^-$), temperature, alkalinity, and aluminium (Al) were significantly impacted by road transport systems. Additionally, XAI models were created to understand the relationships between five specific parameters and road transport attributes. Furthermore, the enhanced EWQI indicated that 1.41%, 63.38%, and 35.21% of the groundwater samples from Zhouzhi County had excellent, good, or moderate quality, respectively. Compared to other classification criteria, the enhanced EWQI method considers a greater number of water hydrochemical parameters and heavy metal elements, making it more appropriate and reliable for thoroughly assessing groundwater quality. Relangi et al. [19] have suggested that a manuscript aims to enhance a novel ensemble model for classifying groundwater quality. After the information is collected from the completion of the WQI computation, the real-time and Indian water quality databases, and the data denoising (using Z-score and Min-Max normalisation techniques). Using the enhanced whale optimisation algorithm (EWOA), the best features/attributes are selected from the denoised data sample. Given that exploring global solutions with traditional WOA is typically computationally intensive, a probability pro is often incorporated into WOA to improve classification accuracy and convergence speed. The ensemble model, which uses K-nearest neighbour (KNN) and AlexNet to classify different types of groundwater quality, is fed the selected optimal features/attributes. The real-time database and the database on Indian water quality have shown 99.88% and 99.98 (very close to100%) classification accuracy for the recently introduced ensemble-based EWOA model, respectively.

Zegaar et al. [20] developed economic models using machine learning and input counts that are most accurate for simulating the irrigation water quality index (IWQI). Eight classification algorithms were used: the K-Nearest Neighbours Algorithm, Support Vector Machines, Gradient Boosting classifiers, Random Forest, Extra Trees, Gradient Boosting classifiers, LightGBM classifier, and Support Vector Machines.

Two possible outcomes were examined: the first used conductivity, chloride, bicarbonate, sodium, calcium, and magnesium as inputs, and the second used total hardness (TH), chloride, and sulphate as inputs, as determined by the Mutual Information (MI) result. The models performed satisfactorily; the best models were the Extra Trees classifier, which achieved an 86.30% three-parameter F1 score, and the LightGBM classifier, which achieved a 91.08% F1 score with six inputs. Our work significantly advances the development of reliable, practical machine learning models for assessing water quality. Sajib et al. [21] present a novel strategy: the Groundwater Quality Index (GWQI) model, enhanced and assessed in the Savar subdistrict of Bangladesh. Ten water quality indicators, six of which include heavy metals, were gathered from 38 sampling sites within the study area and added to the GWQI model. The study increased the precision of water quality evaluation by utilising six tried-and-true machine learning (ML) techniques. The model's performance was evaluated using factors such as sensitivity, uncertainty, and reliability. The primary contribution of this study is the inclusion of heavy metals in the model's water quality index framework. This study, to the authors' knowledge, is the first to develop a GWQI model that includes heavy and trace elements. The study area's groundwater quality ranged from "good" to "fair" according to the GWQI assessment. This indicates that most indicators of water quality were within the established standard limits by the World Health Organisation and the Bangladeshi government. When it came to GWQI score prediction, artificial neural networks (ANN) outperformed the other machine learning models.

The superior effectiveness of ANN was demonstrated by performance metrics such as root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE) for testing (RMSE = 0.001; MSE = 0.00; MAE = 0.001), prediction evaluation statistics (PBIAS = 0.000), and training (RMSE = 0.361; MSE = 0.131; MAE = 0.262). Additionally, the GWQI demonstrated low uncertainty (<2%) and high sensitivity (R2 = 1.0) in grading water quality. These findings support the validity of a recently developed model for managing and monitoring groundwater quality, particularly regarding heavy metals. Gorgij et al. [22] conducted a study in the northwest of Iran utilising long short-term memory (LSTM) deep learning to forecast spatiotemporal groundwater suitability for irrigation. The sodium adsorptive ratio (SAR), an important irrigation water quality criterion, was used at 101 sampling points over an 18-year period (2002–2019) to train a deep learning model to predict irrigation water quality for 2020. Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and R were among the performance metrics used to assess the model's accuracy in spatiotemporal data forecasting. By computing the corresponding values of MAPE, RMSE, and R, 1.212, 0.312, and 0.89, these metrics validated the accuracy of the model. Conversely, the model's capacity was evaluated using generalisation ability (GA) and RBIAS.

The findings indicated that the LSTM model understated the targets, with an acceptable GA of 1.1832 and an RBIAS of roughly 1.539. Once the study area's map of irrigation water quality was completed, it became clear that roughly 78% of the water quality was acceptable to desirable for irrigation, while the remaining 22% had moderate to unacceptable quality. The residential area and the most unacceptable areas are contrasted to highlight how human activity affects groundwater quality by allowing pollutants and fertilisers to seep into groundwater supplies. Wang et al. [23] introduced an algorithm for extreme gradient boosting in machine learning, which was also utilised to calculate the aggregation function and to assign sub-index function weights carefully. It can save the time and computational work needed to determine the optimal parameters. The results showed that the study area's groundwater quality status was primarily maintained in the fair and good categories. Based on the BMA prediction, the WQI values in the research area ranged from 35.01 to 98.45. Seasonal variations in groundwater quality were observed in the study area between 2015 and 2020. The fair category exhibited the highest percentage; the marginal category, however, had the least. In terms of location, most sites were rated fair to good, with a few isolated areas receiving a marginal rating. This implies that the study area's groundwater quality has been suitably maintained. WQI-BMA, the study's model, has important ramifications for regional groundwater management because it is comparatively simple to use and understand.

3. Proposed Model

Four phases make up the proposed groundwater quality classification model: feature optimisation (RLGWO), data denoising (Z-score methods, Min-Max normalisation, and WQI calculation), data collection (Indian water-quality databases and real-time data), and groundwater quality classification.

3.1. Database Description

Using an Indian water-quality database and a real-time database, the effectiveness of the chosen groundwater-quality classification model was investigated [24]. Data from a lab monitoring water quality in Narsapuram, Andhra Pradesh, is included in the real-time database. Seven parameters make up the real-time database: pH, conductivity, temperature, nitrate + nitrite, total coliform, faecal coliform, and biological oxygen demand (BOD). Each groundwater sample's class labels are evaluated in this real-time database by computing the WQI. Furthermore, data from various Indian locations is entered into the Indian Water Quality Database from 2005 to 2014. One thousand six hundred seventy-nine samples total, collected from 666 different lakes and rivers, are included in this database. Seven parameters make up this database: BOD, pH, dissolved oxygen, temperature, faecal coliform, nitrate, and total coliform. The central government of India monitored the information to ensure the safety of drinking water.

3.2. Data Denoising

An essential component of the groundwater quality classification is the data denoising section, which enhances the quality of the gathered data. This section calculates the WQI using seven database parameters and then classifies the water samples based on the WQI values. Furthermore, to improve classification accuracy, Z-score and MinMax normalisation techniques are used. First, the WQI condenses the collected data into a single value, making water-quality information easier to understand. A weight function in WQI, Wi is allocated to each parameter according to its significance. Using seven quality parameters, the suggested model's efficacy is assessed on both databases, and the WQI is calculated using Eq. (1):

$$\mathrm{WQI} = \frac{\sum_{i=1}^{N} q_i \times W_i}{\sum_{i=1}^{N} W_i} \tag{1}$$

Where, q_i shows each parameter's quality estimation scale (QES) i, W_i Signifies each parameter's unit weight, and N lists all of the parameters. The term QES q_i is expressed mathematically in Eq. (2):

$$q_i = 100 \times \left(\frac{V_i - V_{Ideal}}{S_i - V_{Ideal}}\right) \tag{2}$$

Where, V_{Ideal} Reveals the ideal value, which is zero for all other parameters and a pH of 7.60 mg/L of dissolved oxygen, S_i indicates the standard value V_i Shows the measured value. The function of weight W_i It is calculated using Equation (3):

$$W_i = \frac{K}{S_i} \tag{3}$$

Wherein, K demonstrates the constant of proportionality and is calculated using Equation (4):

$$K = \frac{1}{\sum_{i=1}^{N} S_i} \tag{4}$$

Following WQI, the gathered data is rescaled using the lower and upper bounds of the Min-Max normalisation technique, which normally lie between 0 and 1 and -1 and 1. Furthermore, the Z-score method is used to normalise the collected data by computing the mean and standard deviation. The parametric values are scaled between 0 and 1 using the Z-score technique. Equations (5) and (6) provide the formulas for the Min-Max normalisation and z-score methods:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{5}$$

$$Z - \mathrm{Score} = \frac{(x - \mu)}{\sigma} \tag{6}$$

Where x shows the databases' tested samples, μ denotes the standard deviation value, and τ the mean value, $xmaxandxmin$ specifies the attribute's minimum and maximum values. RLGWO implementation also reduces the dimensionality of the rescaled data, thereby lowering the system's complexity and computational time.

Table 1: Allowable bounds for the seven parameters and their corresponding unit weights

Parameters	Permissible limits	Unit weight Wi
Dissolved-oxygen, mg/L	10	0.2213
BOD, mg/L	5	0.4426
Feal coliform/100 mL	100	0.0221
Nitrate, mg/L	45	0.0492
Ph	8.5	0.2604
Conductivity, $\mu S/cm$	1000	0.0022
Total coliform/100 mL	1000	0.0022

As a result, Table 1 shows the allowable bounds for the seven parameters along with their corresponding unit weights. Additionally, Table 2 shows the classification of water quality.

Table 2: Water quality classification

Range of water quality index	Classification
76 to 100	Very poor
51 to 75	Poor
26 to 50	Good
0 to 25	Excellent

3.3. Feature Selection: Robust Learning-Based GWO Algorithm (RLGWO)

Under the original GWO, despite circumstances in which the fittest solution (alpha) is trapped in a local optimum, every omega member of the hunting group updates their position by learning from the first three best leaders until the termination condition is met. This type of learning algorithm is ineffective for resolving issues in large, complex search spaces, but it can converge quickly and perform well in the exploitation phase. To preserve population diversity, some GWO techniques, such as those proposed by Walker et al. [26] and Gaidhane and Nigam [25], have been proposed. These techniques involve strategies to limit the omega learning mechanism. These tactics produce effective exploration outcomes. Finding a solution, though, takes longer and slows the algorithm's rate of convergence. This paper presents an algorithm (RLGWO) that balances discovery and use. In RLGWO, the adjustment of the searching direction is based on a robust tolerance mechanism. Omega can now narrow the search space by using this method to change the direction of their search, thereby preventing local optima. Furthermore, a candidate grey wolf strategy based on opposition learning is applied to generate potential leaders who can replace the hunting group's alpha, beta, and delta to operate in various regions of the search space. Next, a potential position update scheme is considered to ensure the algorithm's accuracy and efficiency, adjusting the candidate leader's potential to lead the grey wolf across various contexts.

3.3.1. Robust Tolerance-Based Adjust Searching Direction Mechanism (RTASDM)

The first GWO algorithm was known to have a high probability of becoming trapped in local optima when searching large, complex solution spaces. The hunting mechanism of a one-dimensional problem is shown in Fig. X, where the blue curve represents the objective function. $P_a$is the optimal solution, or alpha position, which leads the other participants. This makes it evident that each omega can follow the guiding direction of the alpha fitness value P_a, and after many iterations, the hunting party will be confined to a nearby maximum. Assume that $f(X_\omega^n)^k$represents the objective value of the n^{th}omega in the population at the k^{th}iteration. X_ω^nwill be updated using the basic GWO in the next iteration, producing a new fitness value denoted by $f(X_\omega^n)^{k+1}$. The total disparity between $f(X_\omega^n)^{k+1}$and $f(X_\omega^n)^k$is given in (14), where Ndenotes the population size:

$$\sum_{n=1}^{N}\left(f(X_\omega^n)^k\right) - \sum_{n=1}^{N}\left(f\left(f(X_\omega^n)^{k+1}\right)\right) = A(A \in \mathbb{R}) \quad (7)$$

The iterative process may ultimately lead to a solution convergent to the optimum (local or global). Given the circumstances, it seems more likely that A will be near 0. Establishing A in a range of values, researchers can ascertain when, in the case where φ is a small value near 0, the hunting group will converge at $[-\varphi, \varphi]$. Consequently, Formula (7) can be recast as:

$$\sum_{n=1}^{N}\left(f(X_\omega^n)^k\right) - \sum_{n=1}^{N}\left(\left(f(X_\omega^n)^{k+1}\right)\right) \in [-\varphi, \varphi] \quad (8)$$

To prevent hunting groups from becoming stuck in local optima and to ensure algorithmic efficiency, the omega's search direction can be modified whenever Equation (8) is satisfied. However, researchers cannot rely on the above conditions to alter omega's search direction in a vast, complex space. The other hunters in the group begin to look toward Pa as the global optimum, or alpha position, draws closer. This implies that, while meeting the requirements given in Equation (8), the generation k + 1 solution might not be enhanced by the alpha. Thus, over the next few iterations, a promising global optimum might be pointed out to the hunting group based on Pa's potential. Eventually, the grey wolf is said to be trapped around a local optimum when the number of satisfied conditions rises over many iterations, indicating that the current solution and earlier solutions are identical. The grey wolf must thus change the direction of its gaze. Consider the tolerance variable h that serves as a counter and is initially set to 0. Equation (9) can be used to update h as follows, provided the condition is satisfied:

$$h = h + 1 \quad (9)$$

The likelihood of getting trapped in a hunting group's local optima rises with increasing value of h. But in situations where P_a While conducting the above-described search for the global optimum, the omega ought to come next after the alpha direction rather than changing course. Thus, researchers present the probability. P_{adjust}, which permits the omega to modify the direction of its search. Equation (10), the current iteration is denoted by k, and MaxIt is the maximum number of iterations that can be made, can be used experimentally to achieve p adjust:

$$P_{adjust} = \frac{\exp(h)-1}{\exp\left(10+\frac{k\times 10}{MaxIt}\right)-1} \tag{10}$$

The P_{adjust} is not kept constant during iterations; instead, in line with h and k, its value is updated. When the quantity of P_{adjust} If the leader alpha (beta or gamma) exceeds an arbitrary value between 0 and 1, the leader will be replaced by a new candidate, who will then continue to lead the hunting party.

3.3.1.1. Algorithm 1

Below are the specifics of the methodology.

Algorithm 1: Robust Mechanism to Adjust the Searching Direction Based on Tolerance

At iteration kth;
Initialize $h = 0$
if $\sum_{n=1}^{N}\left(f(X_{\omega}^{n})^{k}\right) - \sum_{n=1}^{N}\left(\left(f(X_{\omega}^{n})^{k+1}\right)\right) \in [-\varphi, \varphi]$ then
$h = h + 1$
end if
generate a random number in the range of (0,1);
if $\left(\frac{\exp(h)-1}{\exp\left(10+\frac{k\times 10}{MaxIt}\right)-1} > rand()\right)$ then
Finding another candidate solution to replace alpha (beta or gamma)
End if

Algorithm 1 makes it evident that the P_{adjust} primarily on the tolerance value h, but also on k and h. When h is small in value, P_{adjust} has a high chance of decreasing to below the random number; in that case, the hunting group is still led by the top three leaders, and their leadership skills will be helpful in the following few iterations. Once it reaches the threshold, P_{adjust} rises sharply. This suggests that, over the next few iterations, the number of solutions has not increased and that the hunting party is more likely to become stuck in a local maximum. Thus, the significance of P_{adjust} It is most likely greater than the threshold, and a new leader will guide the omega search direction. On top of that, P_a increases the number of repetitions; especially when the number of iterations exceeds half, the likelihood of approaching the global optimum increases. Thus, to guarantee convergence, the value of h ought to be raised to eliminate the omega and alter the direction of the search.

3.3.2. Opposition-Based Learning for Candidate Generation Strategy

Recently, opposition-based learning (OBL) has been used to accelerate the convergence of different optimisation algorithms. By accounting for both the current population and its opposite, the OBL technique can produce viable candidate solutions. Compared with randomly generated candidate solutions, opposing candidate solutions are more likely to approach the global optimum, according to numerous studies. The opposite number and opposite point definitions, which highlight key components of OBL, are provided below:

Definition: Let $x \in R$ be a real number that falls within a specific range: $x \in [a, b]$. The opposite number x^{opp} is defined as follows:

$$x^{opp} = a + b - x \tag{11}$$

Definition: Let $P(x_1, x_2, \ldots, x_n)$ be a point in a candidate solution generated by n with $x_1, x_2, \ldots, x_n \in R$ and $x \in [a_i, b_i]$. The opposition point p^{opp} is completely defined by its coordinates $x_1^{opp}, x_2^{opp}, \ldots, x_n^{opp}$. Where:

$$x_1^{opp} = a_i + b_i - x_i, i = 1,2, \dots, n \tag{12}$$

It is straightforward to generate a random candidate solution in the suggested RLGWO search space after a potential solution has been identified to replace the leader. This will help the hunting group avoid the current local optimum. Nevertheless, there is no guarantee that the random candidate will improve the solution; it's likely to lead the hunting group into another local optimum, especially in large, complex spaces. Employing the OBL can guarantee a more efficient candidate generation process. The hunting group's investigation phase helps explain this situation. This suggests that, in an effort to identify potential locations for the global optimum, the grey wolves are searching the hold area. Since the alpha is the best option, it will have a big impact on the hunting group. Consequently, it makes sense to swap the alpha to its opposite position during this time frame to avoid a local optimum. The exploitation phase, which accounts for the other half of the iteration, is when the grey wolf narrows its scope and focuses on a particular region to identify the best solution. The alpha serves as the primary leader, and the beta (or gamma) may be substituted for its opposite position to prevent the omega from deviating from the global optimum and to ensure the algorithm's effectiveness. Because of their almost equal potential abilities, if the random number is more than 0.5, the beta is removed, and if it is less than this value, the gamma is added.

3.4. Classification Using Convolutional Attention Module

Following each convolutional layer are convolutional attention modules that adaptively capture important features. The two sub-modules that comprise the spectral and spatial attention modules form the convolutional attention module's structure. The output of each of the above convolutional layers is a 3D feature tensor $V \in R^{h_v \times w_v \times c_v}$, where h_v, w_v, and cv are the 2D feature maps' heights of V, the two variables that are being discussed are the number and width of two-dimensional feature maps. Researchers use V as the convolutional attention module's input. The spectral attention module is used to identify key spectral features for groundwater classification. When gathering spatial information, average pooling is commonly used, while maximum pooling is commonly used for extracting unique features. Consequently, by applying two pooling techniques as a spatial-wise average pooling and a spatial-wise maximum pooling, which are distinguished as:

$$C_{avg,i} = \frac{1}{h_v \times w_v} \sum_{h=1}^{h_v} \sum_{w=1}^{w_v} V_i(h, w), \quad i = 1,2,, \dots, c_v \tag{13}$$

$$C_{max,i} = \max(V_i), i = 1,2, \dots, c_v \tag{14}$$

Where $V_i \in R^{h_v \times w_v}$ represents the 2D feature map in the V channel's i-th channel, $C_{avg,i}$ represents the component of the spatial average representation in the i-th channel $C_{avg} \in R^{c_v}$, max(Z) gives back the biggest element in Z, and $C_{max,i}$ is the component located in the spatial maximum representation's i-th channel $C_{avg} \in R^{c_v}$. Next, researchers use two fully connected layers to implement spectral attention, with a sigmoid and a ReLU activation function. The definition of the sigmoid activation function is:

$$A_{spetral,avg} = W_2^S(\mathrm{Relu}(W_1^S(C_{avg}))) \tag{15}$$

$$A_{spectral,max} = W_2^S(\mathrm{Relu}(W_1^S(C_{max}))) \tag{16}$$

$$A_{spectral} = \mathrm{sigmoid}(A_{spetral,avg} \oplus A_{spectral,max}) \tag{17}$$

$$\mathrm{Relu}(x) = \max(x, 0) \tag{18}$$

$$\mathrm{sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{19}$$

Where W_1^S and W_2^S Parameters that can be learned $\oplus$ indicate the addition of elements, and $A_{spectral} \in R^{1 \times 1 \times c_v}$ The spectral focus. The components of $A_{spectral}$ indicate the importance of the matching 2D feature maps in the spectral domain. Following the creation of the spectral attention $A_{spectral}$Spectral attention module output is defined as:

$$V' = A_{spectral} \otimes V \tag{20}$$

Wherein V' indicates the 3D feature tensor that has been refined, and $\otimes$ stands for the element multiplication. To identify areas important for groundwater, the spatial attention module is utilised. To begin with, the spectral component of V' by means of maximum pooling and spectral-wise average is described as:

$$SPA_{avg,(h,w)} = \frac{1}{c_v}\sum_{c=1}^{c_v} S'_{h,w}(c), \quad h = 1,2,\dots,h_v; w = 1,2,\dots,w_v \tag{21}$$

$$SPA_{max,(h,w)} = \max\left(S'_{h,w}\right), \quad h = 1,2,\dots,h_v; w = 1,2,\dots,w_v \tag{22}$$

Where $S'_{h,w} \in R^{c_v}$ indicates the channel in the row and column h and w of V', $SPA_{avg,(h,w)}$ Represents the element found in the spectral average representation's h-th row and w-th column $SPA_{avg} \in R^{h_v \times w_v \times 1}$ and $SPA_{max}(h,w)$ is the element found in the spectral maximum representation's hth row and w-th column $SPA_{max} \in R^{h_v \times w_v \times 1}$. The spatial attention is implemented in the following using two activation functions: sigmoid and convolutional, which are described as:

$$SPA = Cat\left(SPA_{avg}, SPA_{max}\right) \tag{23}$$

$$A_{spatial} = Sigmoid(Conv(SPA)) \tag{24}$$

Where $Cat\left(SPA_{avg}, SPA_{max}\right)$ denotes the concatenation of SPA_{avg} and SPA_{max} along the spectral dimension, $Conv(SPA)$ symbolises the SPA convolutional layer, and $A_{spatial} \in R^{h_v \times w_v \times 1}$ is the focus on space? The components of $A_{spatial}$ Symbolise the significance of the relational spatial domain's regions. Therefore, one way to characterise the spatial attention module's output is:

$$V'' = A_{spatial} \otimes V' \tag{25}$$

Where $V'' \in R^{h_v \times w_v \times 1}$ represents the convolutional attention module's final 3D feature tensor output.

3.4.1. Classification Using Attention-Based Bidirectional LSTM

Every temporal slice $S_i \in R^{h \times w \times 2f}, i = 1,2,\dots,2T$, the attention-based CNN's final product is $P_i \in R^{150}$. Utilising an attention-based bidirectional long-term memory network, the model investigates the significance of distinct temporal slices, as variations among them contain temporal information relevant to groundwater recognition. A bidirectional LSTM connects two opposing-directional unidirectional LSTMs to the same output. A bidirectional LSTM retains information from the past and future, which improves its comprehension of context when compared to a unidirectional LSTM. The bidirectional LSTM used in this work consists of two 36-cell unidirectional LSTMs. The attention-based sequential output sequence is fed into the unidirectional LSTM for positive time direction, LSTMP CNN $P^P = (P_1, P_2, \dots, P_{2T})$ as the input sequence, whereas LSTMN uses the reverse sequence in the case of a negative time direction $P^N = (P_{2T}, P_{2T-1}, \dots, P_1)$ as the order of input. The unidirectional LSTMs' i-th node's outputs are $Y_i^P \in R^{36}$ and $Y_i^N \in R^{36}$, $i = 1,2,\dots,2T$, correspondingly. Next, researchers join together Y_i^P and Y_{2T+1-i}^N as the bidirectional LSTM's i-th node's output $Y_i \in R^{72}$. Unlike conventional methods, which utilise the final node's output for classification or other purposes, researchers utilise all of the bidirectional LSTM nodes' outputs $Y \in R^{2T \times 72}$ weighing and analysing the significance of various temporal slices while making use of the temporal attention mechanism. Two fully linked layers and an activation function called SoftMax, which has the following definition, and a ReLU activation function are used to implement the temporal attention mechanism:

$$Tem_i = W_2^T\left(Relu(W_1^T Y_i + b_1^T)\right) + b_2^T \tag{26}$$

$$A_{temporal} = softmax(Tem) \tag{27}$$

$$Softmax(x) = \frac{\exp(x)}{\sum \exp(x)} \tag{28}$$

Where W_1^T, W_2^T, b_1^T, and b_2^T are learnable parameters, Tem_i symbolises the i-th component of $Tem \in R^{2T \times 1}$ which undertakings $Y \in R^{2T \times 72}$ to a smaller scale, and $A_{temporal} \in R^{2T \times 1}$ is the focus of time. The components of $A_{temporal}$ Show how significant the matching temporal slices are. Afterwards, the sample's high-level representation X_n is characterised as:

$$L_n(e) = \sum A_{temporal} \otimes Y_e, e = 1,2,\dots,72 \tag{29}$$

Where $Y_e \in R^{2T \times 1}$ indicates the column e-th of $Y \in R^{2T \times 72}$ and $L_n(e)$ is the high-level representation's e-th element $L_n \in R^{72}$, which combines temporal, spectral, and spatial data of X_n.

3.4.2. Classifier

Considering the high-level depiction L_n of input data, the research predicts the groundwater sample's label using two layers: one fully connected layer and one with an activated SoftMax X_n, which has the following definition:

$$\mathrm{Pre} = \mathrm{softmax}(W^p L_n + b^p) \quad (30)$$

Where W^p, b^p are adjustable parameters and $\mathrm{Pre} \in R^C$ indicates the likelihood of X_n being a part of every C class. The RLGWO model optimally selects the proposed fine-tuning and also serves as a feature selection model.

4. Results and Discussion

The suggested model-based RLGWO is simulated for groundwater quality classification utilising an Intel Core i7 processor, 16GB of random-access memory, Windows 10, and a Python software environment. Sensitivity, FDR, specificity, accuracy, and MCC are performance measures that are used to assess how well the proposed model-based RLGWO predicts WQI and classifies groundwater quality. Next, Eqs. (31-35) present the mathematical definitions of the performance measures used, including sensitivity, FDR, specificity, accuracy, and MCC:

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \times 100 \quad (31)$$

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{FP+TP}} \times 100 \quad (32)$$

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN+FP}} \times 100 \quad (33)$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \times 100 \quad (34)$$

$$\mathrm{MCC} = \frac{\mathrm{TP\times TN - FP\times FN}}{\sqrt{(\mathrm{TP+FP})(\mathrm{TP+FN})(\mathrm{TN+FP})(\mathrm{TN+FN})}} \quad (35)$$

Where the values for true positive, true negative, false positive, and false negative are indicated by FP, FN, TP, and TN.

4.1. Quantitative Analysis Using a Live Database

The performance of the suggested model-based RLGWO is examined in this scenario using a real-time database, with metrics including MCC, specificity, accuracy, FDR, and sensitivity. This manuscript uses three-, five-, and seven-fold cross-validation to verify the efficacy of the proposed model-based RLGWO. Table 3 displays the experimental results. Compared with other cross-validation methods, 5-fold cross-validation (80:20% training: test) yielded good classification results. Cross-fold validation is included in this manuscript to reduce computational time and to reduce bias and variance in the developed RLGWO-based model.

Table 3: Validation analysis of the proposed model on a real-time dataset

Models	Accuracy	Precision	Recall	Specificity	MCC	KAPPA	F1 Score	AUC	FDR
AE	0.8070	0.8100	0.8100	0.8100	0.6000	0.5500	0.8100	0.8600	0.1900
DBN	0.7981	0.8000	0.8000	0.8000	0.0100	0.0100	0.8000	0.5100	0.2000
RNN	0.8031	0.8700	0.8700	0.8700	0.1100	0.0700	0.8700	0.5200	0.1300
CNN	0.8369	0.8800	0.8800	0.8800	0.0000	0.0000	0.8800	0.5000	0.1200
LSTM	0.8454	0.8700	0.8700	0.8700	0.1800	0.1600	0.8700	0.5600	0.1300
BLSTM	0.7047	0.7000	0.7000	0.7000	0.4100	0.4100	0.7000	0.7000	0.3000
ACBLSTM-RLGWO	0.8779	0.8800	0.8800	0.8800	0.0800	0.0100	0.8800	0.5000	0.1200

Table 3 presents the validation results of the proposed model on a real-time dataset, including performance metrics for various models. The accuracy ranges from 0.7047 to 0.8779, with the highest achieved by the ACBLSTM-RLGWO model. Precision values vary from 0.7 to 0.88, with the highest again achieved by ACBLSTM-RLGWO. Recall values are consistent with precision, ranging from 0.7 to 0.88 across different models. Specificity values are also uniform, ranging from 0.7 to 0.88. The Matthews correlation coefficient (MCC) varies from 0 to 0.41, with the highest achieved by BLSTM. Kappa values range from 0.01 to 0.41, again with the highest from BLSTM. F1 score values are consistent with precision and recall, ranging from 0.7 to

0.88. Area under the curve (AUC) values range from 0.5 to 0.86, with the highest achieved by AE. Finally, the false discovery rate (FDR) ranges from 0.12 to 0.3, with the lowest achieved by CNN. Overall, the ACBLSTM-RLGWO model appears to have the best overall performance across multiple metrics, followed closely by the LSTM and CNN models.

Table 4: Analysis of various models on the Kaggle dataset

Models	Accuracy	Precision	Recall	Specificity	MCC	KAPPA	F1 Score	AUC	FDR
AE	0.8896	0.8900	0.8900	0.8900	0.6300	0.6000	0.8900	0.7600	0.1100
DBN	0.8070	0.8100	0.8100	0.8100	0.6000	0.5500	0.8100	0.8600	0.1900
RNN	0.9186	0.9200	0.9200	0.9200	0.7600	0.7500	0.9200	0.8800	0.0800
CNN	0.8528	0.8500	0.8500	0.8500	0.6700	0.6400	0.8500	0.8900	0.1500
LSTM	0.8238	0.8200	0.8200	0.8200	0.3500	0.3200	0.8200	0.6300	0.1800
BLSTM	0.8266	0.8300	0.8300	0.8300	0.3500	0.2300	0.8300	0.5800	0.1700
ACBLSTM-RLGWO	0.9258	0.9300	0.9300	0.9300	0.7700	0.7700	0.9300	0.8800	0.0700

The analysis of various models from Table 4 on the Kaggle dataset reveals their performance across multiple metrics. The model accuracy ranges from 0.8070 to 0.9258, with ACBLSTM-RLGWO achieving the highest. Precision values range from 0.81 to 0.93, with ACBLSTM-RLGWO achieving the highest precision. Recall values are consistent across models, ranging from 0.81 to 0.93. Specificity values mirror recall, ranging from 0.81 to 0.93. Matthews correlation coefficient (MCC) values range from 0.35 to 0.77, with ACBLSTM-RLGWO having the highest MCC. Kappa values range from 0.23 to 0.77, with ACBLSTM-RLGWO again achieving the highest. F1 score values are consistent with precision and recall, ranging from 0.81 to 0.93. Area under the curve (AUC) values range from 0.58 to 0.89, with ACBLSTM-RLGWO having the highest AUC. Finally, the false discovery rate (FDR) ranges from 0.07 to 0.19, with the lowest achieved by ACBLSTM-RLGWO. Overall, ACBLSTM-RLGWO demonstrates superior performance across various metrics, followed by RNN and AE models.

5. Conclusion

A novel ACBLSTM-RLGWO model is used in this study to effectively classify groundwater quality. The first step in improving the quality of acquired data is to calculate the WQI and apply Min-Max normalisation and Z-score denoising techniques. Next, using the RLGWO technique, the most relevant characteristics for groundwater quality are selected. Finally, the ensemble classification model receives the dimensionally reduced features/attributes and uses them to categorise water quality into four categories: very poor, poor, good, and excellent. BiLSTM and attention-based CNN are employed to classify groundwater quality. Both the CNN module and the bidirectional LSTM module incorporate attention mechanisms. The CNN module handles the WQI's spatial and spectral information, and critical features are adaptively captured via spatial and spectral attention mechanisms. The bidirectional LSTM module extracts temporal dependencies from the CNN module's outputs, and the temporal attention mechanism investigates the significance of various temporal slices. Evaluation metrics are utilised to evaluate the efficacy of the proposed model, including sensitivity, specificity, accuracy, FDR, and MCC. As a result, the ensemble-based EWOA model achieved classification accuracies of 87% and 92% on the Indian water quality database and the real-time database, respectively. The experimental result is most closely associated with classical machine learning classifiers, such as CNNs, AEs, DBNs, RNNs, and LSTMs. Additionally, the system's complexity and computation time are greatly reduced by selecting the most relevant features and attributes. To further enhance groundwater quality classification, a new deep learning-based ensemble classification model is developed and verified using multimodal data.

Acknowledgement: The authors gratefully acknowledge the support and cooperation of Karunya Institute of Technology and Science, Saranathan College of Engineering, New Horizon College of Engineering, and Messiah University. Their valuable guidance and contributions were instrumental in the successful completion of this work.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Funding Statement: No Funding.

Conflicts of Interest Statement: The author declares no conflict of interest.

Ethics and Consent Statement: All authors have provided their full consent for this work to be published and made accessible for academic, educational, and research purposes.

References

1. E. F. Abdelaty, A. M. Abd-El-Hady, and S. F. Abouzahw, "Classification approaches to assess groundwater quality in Wadi El-Natron, Egypt," *Egyptian Journal of Soil Science*, vol. 62, no. 3, pp. 267–291, 2022.
2. H. Khairy and M. R. Janardhana, "Classification, hydrochemical characterization, and quality assessment of groundwater of coastal aquifer at Sari–Neka plain, Northern Iran," *Environmental Earth Sciences*, vol. 81, no. 12, pp. 1–25, 2022.
3. N. T. Giao, H. T. H. Nhien, and P. K. Anh, "Groundwater quality assessment using classification and multi-criteria methods: A case study of Can Tho City, Vietnam," *Environment and Natural Resources Journal*, vol. 20, no. 2, pp. 179–191, 2022.
4. R. Ayyandurai, S. Venkateswaran, and D. Karunanidhi, "Hydrogeochemical assessment of groundwater quality and suitability for irrigation in the coastal part of Cuddalore district, Tamil Nadu, India," *Marine Pollution Bulletin*, vol. 174, no. 1, p. 113258, 2022.
5. A. Peethambaran, M. A. Anso, T. S. Salumol, R. R. Krishnamurthy, and S. R. Mahapatra, "Classification and evaluation of groundwater in Cheyyar watershed, Thiruvannamalai district, Tamil Nadu," *Materials Today: Proceedings*, vol. 68, no. 6, pp. 669–678, 2022.
6. M. Najafzadeh, F. Homaei, and S. Mohamadi, "Reliability evaluation of groundwater quality index using data-driven models," *Environmental Science and Pollution Research*, vol. 29, no. 6, pp. 8174–8190, 2022.
7. M. O. Raimi and H. O. Sawyerr, "Preliminary study of groundwater quality using hierarchical classification approaches for contaminated sites in indigenous communities associated with crude oil exploration facilities in Rivers State, Nigeria," *Open Journal of Yangtze Oil and Gas*, vol. 7, no. 2, pp. 124–148, 2022.
8. D. J. Lapworth, T. B. Boving, D. K. Kreamer, S. Kebede, and P. L. Smedley, "Groundwater quality: Global threats, opportunities and realising the potential of groundwater," *Science of the Total Environment*, vol. 811, no. 3, p. 152471, 2022.
9. N. S. Rao, B. Sunitha, R. Das, and B. A. Kumar, "Monitoring the causes of pollution using groundwater quality and chemistry before and after the monsoon," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 128, no. 12, p. 103228, 2022.
10. X. Zhang, R. Zhao, X. Wu, and W. Mu, "Hydrogeochemistry, identification of hydrogeochemical evolution mechanisms, and assessment of groundwater quality in the southwestern Ordos Basin, China," *Environmental Science and Pollution Research*, vol. 29, no. 1, pp. 901–921, 2022.
11. X. Qu, L. Shi, and J. Han, "Spatial evaluation of groundwater quality based on toxicological indexes and their effects on ecology and human health," *Journal of Cleaner Production*, vol. 377, no. 12, p. 134255, 2022.
12. V. Sunitha and B. M. Reddy, "Geochemical characterization, deciphering groundwater quality using pollution index of groundwater (PIG), water quality index (WQI) and GIS in hard rock aquifer, South India," *Applied Water Science*, vol. 12, no. 3, p. 41, 2022.
13. G. Hinge, B. Bharali, A. Baruah, and A. Sharma, "Integrated groundwater quality analysis using water quality index, GIS and multivariate technique: A case study of Guwahati City," *Environmental Earth Sciences*, vol. 81, no. 16, p. 412, 2022.
14. H. Raheja, A. Goel, and M. Pal, "Evaluation of groundwater quality for drinking purposes based on machine learning algorithms and GIS," *Sustainable Water Resources Management*, vol. 10, no. 1, p. 11, 2024.
15. S. Subbarayan, S. Thiyagarajan, S. Karuppannan, and B. Panneerselvam, "Enhancing groundwater vulnerability assessment: Comparative study of three machine learning models and five classification schemes for Cuddalore district," *Environmental Research*, vol. 242, no. 2, p. 117769, 2024.
16. S. H. Yazdi, M. Robati, S. Samani, and F. Z. Hargalani, "Assessing the sustainability of groundwater quality for irrigation purposes using a fuzzy logic approach," *Environmental and Sustainability Indicators*, vol. 22, no. 6, p. 100342, 2024.
17. K. Kwak and E. H. Lee, "Impact of road transport system on groundwater quality inferred from explainable artificial intelligence (XAI)," *Science of the Total Environment*, vol. 917, no. 3, p. 170388, 2024.
18. Y. Yang, P. Li, V. Elumalai, J. Ning, F. Xu, and D. Mu, "Groundwater quality assessment using EWQI with updated water quality classification criteria: A case study in and around Zhouzhi County, Guanzhong Basin (China)," *Exposure and Health*, vol. 15, no. 4, pp. 825–840, 2023.
19. N. D. S. S. K. Relangi, A. Chaparala, and R. Sajja, "Effective groundwater quality classification using enhanced whale optimization algorithm with ensemble classifier," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 1, pp. 214–223, 2023.
20. A. Zegaar, S. Ounoki, and A. Telli, "Machine learning for groundwater quality classification: A step towards economic and sustainable groundwater quality assessment process," *Water Resources Management*, vol. 38, no. 2, pp. 621–637, 2024.

21. A. M. Sajib, M. T. M. Diganta, A. Rahman, T. Dabrowski, A. I. Olbert, and M. G. Uddin, “Developing a novel tool for assessing the groundwater incorporating water quality index and machine learning approach,” *Groundwater for Sustainable Development*, vol. 23, no. 11, p. 101049, 2023.
22. A. D. Gorgij, G. Askari, A. A. Taghipour, M. Jami, and M. Mirfardi, “Spatiotemporal forecasting of groundwater quality for irrigation purposes using deep learning method: Long short-term memory (LSTM),” *Agricultural Water Management*, vol. 277, no. 3, p. 108088, 2023.
23. X. Wang, Y. Tian, and C. Liu, “Assessment of groundwater quality in a highly urbanized coastal city using water quality index model and Bayesian model averaging,” *Frontiers in Environmental Science*, vol. 11, no. 3, p. 1086300, 2023.
24. W. Long, J. Jiao, X. Liang, and M. Tang, “An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization,” *Engineering Applications of Artificial Intelligence*, vol. 68, no. 2, pp. 63–80, 2018.
25. P. J. Gaidhane and M. J. Nigam, “A hybrid grey wolf optimizer and artificial bee colony algorithm for enhancing the performance of complex systems,” *Journal of Computational Science*, vol. 27, no. 7, pp. 284–302, 2018.
26. D. Walker, M. Smigaj, and N. Jovanovic, “Ephemeral sand river flow detection using satellite optical remote sensing,” *Journal of Arid Environments*, vol. 168, no. 9, pp. 17–25, 2019.