Physics and Chemistry of the Earth 144 (2026) 104657

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

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

A hybrid expert–PCA–ANN framework for hydrogeochemical optimization of the water quality index in a semi-arid aquifer

Mostafa Dougha [a,*], Tahar Selmane [b], Ashraf A. Ahmed [c,**], Ahmed A. Arafat [d], Sherif S.M. Ghoneim [e], Johnbosco C. Egbueri [f,g], Lakhdar Seraiche [a], Enas E. Hussein [h,***]

[a] Department of Hydraulics, Faculty of Technology, University of M'sila, M'sila, 28000, Algeria
[b] Department of Hydraulics and Civil Engineering, Faculty of Technology, University of Ghardaia, Ghardaia, 47000, Algeria
[c] Department of Civil and Environmental Engineering, Brunel University of London, Kingston Lane, Uxbridge, UB83PH, UK
[d] Department of Civil Engineering, College of Engineering, Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia
[e] Department of Electrical Engineering, College of Engineering, Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia
[f] Department of Geology, Chukwuemeka Odumegwu Ojukwu University, Uli, 431124, Nigeria
[g] Research Management Office (RMO), Chukwuemeka Odumegwu Ojukwu University, Igbariam, Nigeria
[h] Water Management Research Institute, National Water Research Center, Shubra El-Kheima, 13411, Egypt

ARTICLE INFO

Keywords:
Groundwater quality
Hydrogeochemistry
Machine learning
Principal component analysis
Weighting factor optimization

ABSTRACT

Groundwater quality in semi-arid regions is controlled by both natural hydrogeochemical processes and human activities. This study examines and compares four approaches for constructing a groundwater quality index in the southern Chott Hodna Basin (Algeria), combining expert judgment, principal component analysis (PCA), and artificial neural network (ANN) optimization. Eleven physicochemical parameters related to salinity, mineralization, and agricultural contamination were analyzed from 33 groundwater samples. Water quality index values range from 52.67 to 303.44, with a mean of 147.79, which indicates generally poor groundwater quality and marked spatial variability. Higher values are mainly linked to salinity and nitrate enrichment. All WQI weighting approaches show strong agreement ($r > 0.97$), suggesting consistent representation of the aquifer's hydrochemical structure. The PCA-based approach emphasizes salinity-related variables such as electrical conductivity, total dissolved solids, chloride, and sulfate, which reflect evaporite dissolution and water–rock interaction. In contrast, the ANN-optimized model increases the influence of nitrate, giving prominence to the role of agricultural contamination. The hybrid PCA–ANN model provides the lowest prediction error (around a 12–18% reduction compared to non-optimized models) and better captures complex relationships among variables. Spatial interpolation using ordinary kriging identifies degraded water quality zones mainly in the northern and northeastern parts of the aquifer. In general, combining statistical and machine learning approaches improves the objectivity and reliability of groundwater quality assessment in semi-arid regions.

1. Introduction

Groundwater is a major freshwater resource used for domestic supply, irrigation, and industrial activities, particularly in Semi-arid regions. However, groundwater quality is increasingly affected by both natural hydrogeochemical processes and human activities. Factors

Abbreviations: WQI_Expert, traditional expert-based allocation; WQI_Expert_ANN, expert weights optimized through artificial neural networks; WQI_PCA, principal component analysis–derived weights; WQI_PCA_ANN, PCA weights refined using ANN optimization; WQI, water quality index; ANN, artificial neural networks; PCA, principal component analysis; EC, electrical conductivity; TDS, total dissolved solids; Cl^-, chlorides; SO_4^{2-}, sulfates; NO_3^-, nitrates; HCO_3^-, bicarbonates; Ca^{2+}, Calcium; Mg^{2+}, Magnesium; Na^+, Sodium; K^+, Potassium; pH, Potential hydrogen; R^2, coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; OK, Ordinary Kriging.

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: mostafa.dougha@univ-msila.dz (M. Dougha), selmane.tahar@univ-ghardaia.edu.dz (T. Selmane), Ashraf.ahmed@brunel.ac.uk (A.A. Ahmed), aaarafat@tu.edu.sa (A.A. Arafat), s.ghoneim@tu.edu.sa (S.S.M. Ghoneim), johnboscoegbueri@gmail.com (J.C. Egbueri), enas_el-sayed@nwrc.gov.eg (E.E. Hussein).

https://doi.org/10.1016/j.pce.2026.104657
Received 15 March 2026; Received in revised form 9 June 2026; Accepted 15 July 2026
Available online 17 July 2026

such as mineral weathering, evaporite dissolution, and ion exchange strongly influence groundwater chemistry (Shaibur et al., 2019b; Abdullahi et al., 2025; Nigam and Indu, 2025). Vulnerable aquifers are affected by several human activities, including intensive irrigation, fast urban development, and wastewater infiltration. (Egbueri and Agbasi, 2022; Shaibur et al., 2023; Seraiche et al., 2025a; Seraiche et al., 2025b; Nigam and Indu, 2025). All of these pressure factors show how important it is to have reliable methods for analyzing and managing groundwater over the long term.

The water quality index (WQI) is widely used to simplify complex hydrochemical data into a single numerical value (Hassan et al., 2021; Chakraborty et al., 2023). The WQI can help determine whether water is safe for drinking, suitable for irrigation, and safe for the environment by combining multiple physicochemical parameters. Several studies have applied WQI methods in Algeria, Bangladesh, India, Africa, and China to examine groundwater affected by mineralization, agricultural contamination, and hydrochemical variability (Zhang et al., 2020; Hassan et al., 2021; Selmane et al., 2022; Chakraborty et al., 2023; Maliki et al., 2024; Sen et al., 2024; Clemente et al., 2026). In Bangladesh, physico-chemical and hydrochemical analyses have identified contamination of vulnerable aquifers linked to salt intrusion, agricultural practices, and geogenic influences (Shaibur et al., 2019a, 2021). Although the conventional water quality index (WQI) method is practical and widely used. It is based on weighting methods defined by specialists that may result in subjectivity and reduced reproducibility of results in different hydrogeological contexts.

Groundwater chemistry is generally governed by a few hydrochemical variables such as electrical conductivity (EC), total dissolved solids (TDS), chlorides (Cl^-), sulfates (SO_4^{2-}), nitrates (NO_3^-), and bicarbonates (HCO_3^-), as demonstrated by multiple studies. These variables reflect processes such as evaporation, mineral dissolution, and fertilizer leaching (Shaibur et al., 2019b; Shaibur, 2022; Mahamadou et al., 2024; Guenouche et al., 2024; Marfo et al., 2024). Similar hydrochemical conditions have been reported in other aquifers across Asia and Africa, where natural mineralization and anthropogenic activities together affect groundwater evolution (Abidi et al., 2024). This indicates that improving the objectivity and reliability of parameter weighting remains a major challenge in WQI development.

To reduce constraints associated with subjective weighting, multivariate statistical techniques, including principal component analysis (PCA), are widely employed in groundwater research. The PCA reduces data dimensionality by transforming correlated variables into a smaller number of independent components that explain most of the dataset variance (Khoi et al., 2022; Zhang et al., 2024; Guenouche et al., 2024; Abdullahi et al., 2025). This method helps identify the most important hydrogeochemical processes and assigns weights to the resulting parameters. Previous studies from India, Bangladesh, and North Africa indicated that PCA-based approaches improve interpretability and reproducibility in groundwater assessment (Shaibur et al., 2019b; Tripathi and Singal, 2019; Shaibur et al., 2021; Guenouche et al., 2024; Nishat et al., 2025). However, PCA mainly describes linear relationships and may not fully represent complex interactions among hydrochemical variables.

In recent years, artificial neural networks (ANN) have proven good at simulating nonlinear interactions in hydrogeochemical datasets. ANN models can learn complex interactions among variables and improve prediction accuracy compared with conventional statistical approaches (Hassan et al., 2021; Mengistu et al., 2026). ANN-based WQI models have shown promising results in groundwater studies from India, West Africa, Saudi Arabia, and Iraq (Fatah et al., 2020; Egbueri and Agbasi, 2022; Ibrahim et al., 2023; Nayak et al., 2023; Mahamadou et al., 2024; Ismail et al., 2024; Al-Zubaidi et al., 2025). In hydrochemical applications, better optimization techniques like Levenberg–Marquardt and Bayesian regularization have also made model convergence and prediction stability better (de Jesús Rubio, 2020). However, ANN methods may suffer from limited interpretability and possible overfitting, especially when applied to relatively small hydrochemical datasets.

Recent researches are looking more and more at hybrid approaches that combine hydrogeochemical knowledge, multivariate statistics, and machine-learning techniques. This is because each weighting method has its own strengths and weaknesses. Expert-based methods preserve hydrogeochemical interpretation but remain partly subjective. PCA-based approaches improve statistical objectivity but mainly describe linear variance structures. ANN models better capture nonlinear interactions but require careful validation to maintain reliability. Combining these methodologies that work well together may improve WQI performance while preserving hydrogeochemical consistency (Taşan, 2023; Ismail et al., 2024; Abdullahi et al., 2025; Acquah et al., 2025).

Although recent advancements have occurred, comparative studies assessing expert-based, PCA-derived, and ANN-optimized weighting methods within a unified framework are still few, especially in North African aquifers. The Chott Hodna basin (Maadher Plain, Algeria) is a hydrogeologically vulnerable environment characterized by low rainfall, a lot of evaporation, evaporitic geological formations, and more and more groundwater exploitation (Selmane et al., 2022; Seraiche et al., 2025a). In this region, groundwater quality is affected by both natural mineralization and contamination linked to irrigation return flow and wastewater infiltration. Local communities are strongly dependent on groundwater for domestic and agricultural uses; reliable, regionally adapted assessment methods are needed for long-term water resource management.

This study, therefore, aims to: (i) examine the spatial variability of major hydrochemical parameters and identify the dominant geogenic and anthropogenic processes controlling groundwater chemistry; (ii) develop and compare WQI models based on expert weighting, PCA-derived weighting, and ANN optimization; and (iii) evaluate the relative performance, interpretability, and environmental relevance of these approaches in a semi-arid aquifer system. This study proposes a hybrid framework—by combining hydrogeochemical analysis, multivariate statistics, and machine-learning techniques—to improve the reliability and applicability of groundwater quality assessment in semi-arid regions.

Recent advancements in hydrogeological artificial intelligence include approaches such as self-organizing maps, probabilistic learning models, and generative algorithms. (Karimi-Rizvandi et al., 2021; Gholami and Sahour, 2024). However, the assessment of groundwater quality still requires analytical frameworks that are computationally efficient and hydrogeochemically relevant. In this context, hybrid methods combining statistical analysis and machine learning optimization remain suitable for regional assessment of the water quality index and for the environmental management of semi-arid aquifer systems. Ensuring the reproducibility of these hybrid methods is also a key objective.

2. Materials and methods

2.1. Study area

The region is located south of the Chott Hodna basin (Fig. 1); The Maadher plain extends around the coordinates 35°17′13.39″ N and 4°16′05.81″ E, at an average altitude of approximately 450 m. The climate there is semi-arid, characterized by strong seasonal contrasts: cool, wet winters and hot, dry summers. The average annual rainfall is about 225 mm, however it changes a lot from year to year because the climate in the area is so variable. (Selmane et al., 2022; Salim et al., 2023; Lakhdar et al., 2026). The average temperature is about 21.9°C. Since the 1970s, tremendous growth in agricultural production and heavy groundwater extraction have made farmers more reliant on irrigation for their crops. Consequently, groundwater levels have fallen sharply, while the proliferation of poorly constructed wells has

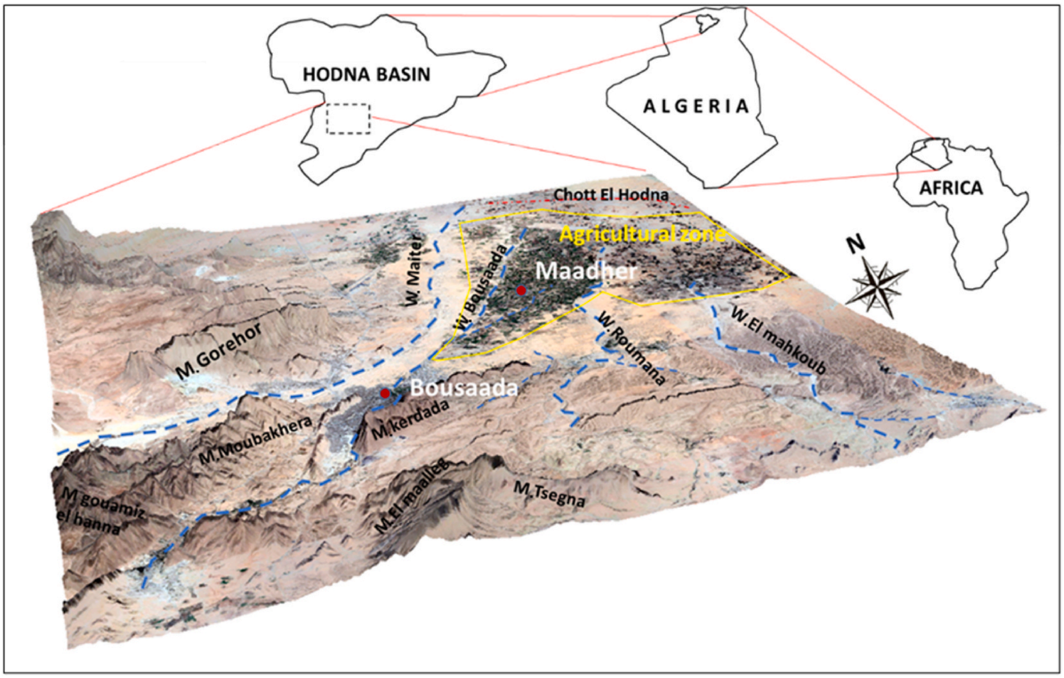

Fig. 1. A 3D map showing the topography of the study area (Maadher), and its location in relation to Algeria and the Hodna Basin.

accelerated the salinization of aquifers and increased their vulnerability to contamination (Ahmed et al., 2022; Seraiche et al., 2025a). These events pose risks to groundwater quality and regional environmental health, necessitating sustainable practices for aquifer protection.

2.2. Geology and hydrogeology of region

The landscape of the study area is characterized by sand dunes, recent alluvial deposits, and isolated rocky reliefs, including Mount Meharga and Mount Kerdada. The geological history of the Hodna Basin records successive marine transgressive–regressive cycles from the Late

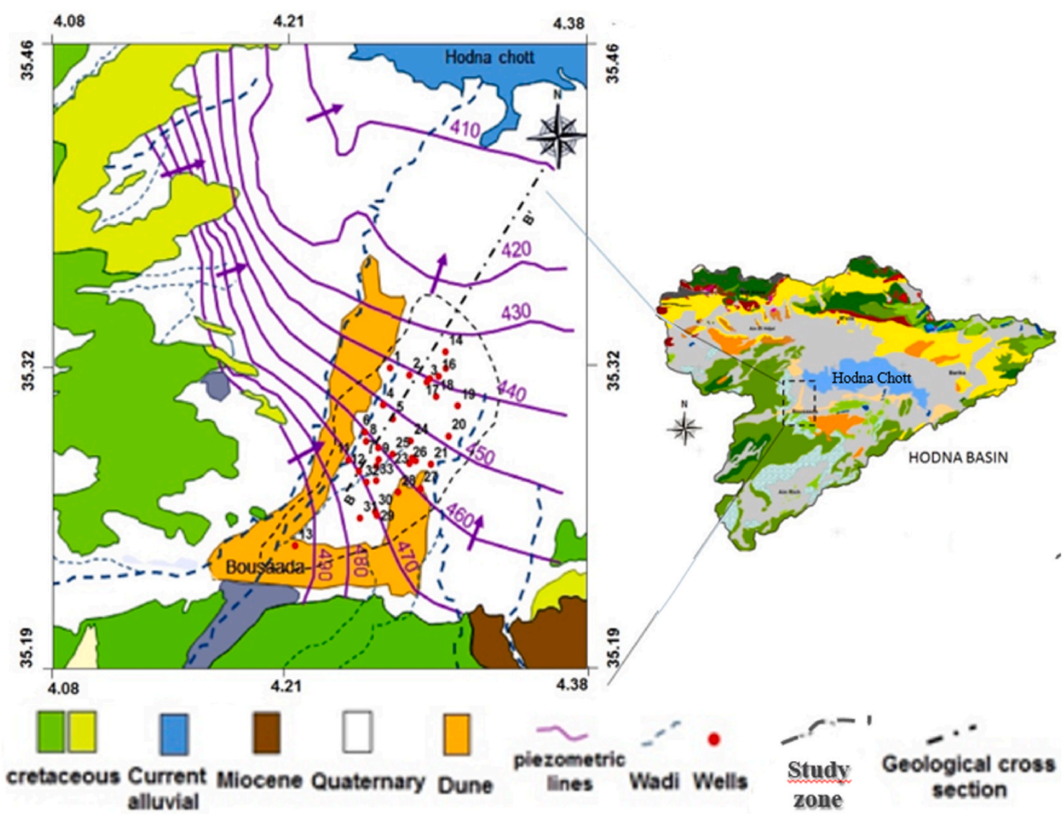

Fig. 2. Geological map of the Maadher region showing boundaries of studied zone, piezometric lines and sampling locations.

Triassic to the Late Cretaceous (Guiraud, 1973; Seraiche et al., 2025b). The stratigraphic succession includes Triassic formations, Cretaceous limestones, Miocene sandstones, and Lower Eocene gypsiferous marls. These are all on top of Quaternary deposits that have not yet solidified (Fig. 2). This geological context exerts a strong influence on groundwater flow regimes and mineralization processes.

The Mio-Plio-Quaternary aquifer is the principal source of groundwater. It is more than 250 m thick. It is the main aquifer that people use in the area. (Guiraud, 1973; Dougha and Hasbaia, 2019). Hydraulic connectivity with the deeper Albian aquifer occurs through vertical exchanges, influencing the chemistry and salinity of the groundwater. The recharge comes mainly from the infiltration of episodic floods of wadis and return water from irrigation, while the low natural recharge in semi-arid conditions reduces the system's dilution capacity. The interactions between evaporite formations, carbonate rocks, and anthropogenic inputs collectively shape the hydrogeochemical evolution of groundwater, which contributes to salinization and nitrate enrichment.

2.3. Data collection and water quality parameters

We took samples of groundwater from the Maadher plain in Algeria, which is south of Chott Hodna, between November 17 and 20, 2021. We got 33 samples from wells. We select these wells carefully to show how different the hydrogeochemical conditions are on the plain, taking into account both natural geochemical processes and how people could affect the quality of groundwater.

Eleven physico-chemical parameters commonly used to assess groundwater quality were examined: pH, EC, TDS, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, SO_4^{2-}, HCO_3^{-} and NO_3^{-}. The selected variables are particularly relevant in semi-arid hydrogeological contexts (Selmane et al., 2022), where processes such as evaporite dissolution, carbonate alteration and ion exchange significantly influence groundwater chemistry.

All chemical analyses were carried out in certified laboratories in accordance with World Health Organization guidelines (WHO, 2017; Dougha et al., 2018). The ionic balance error remained within ±6%, indicating satisfactory analytical accuracy. Descriptive statistical analysis, together with standardized (z-score) evaluation, showed relatively stable pH values throughout the study area. On the other hand, the parameters related to salinity (EC, TDS, Cl^{-}, SO_4^{2-}), the main cations and nitrates showed marked spatial variability with several localized outliers. These trends suggest heterogeneous mineralization processes and the presence of areas of high contamination that are probably associated with both geogenic sources and agricultural runoff (Selmane et al., 2022).

2.4. Methodological framework

The overall methodological approach (Fig. 3) combines conventional hydrogeochemical assessment, multivariate statistical techniques and machine learning methods to improve the calculation of the WQI index. Three complementary weighting approaches were applied: (i) Expert-based weighting (conventional WQI); (ii) Principal Component Analysis (PCA)-based weighting; (iii) Artificial Neural Network (ANN)-based weight optimization.

Model performance was assessed using standard statistical indicators, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), along with residual and sensitivity analyses. Finally, the spatial distribution of WQI index values was generated using the ordinary kriging technique to delineate contamination zones and assess spatial trends in the study area.

2.5. Water quality index (WQI)

The water quality index (WQI) is a comprehensive tool that combines numerous physicochemical parameters into a single numerical value indicative of overall water quality (Hassan et al., 2021; Selmane et al., 2023; Maliki et al., 2024). Converting complex hydrochemical information into an index facilitates interpretation and allows for comparison between sampling sites, thus contributing to the informed decision-making in water resource management. (Sen et al., 2024; Seraiche et al., 2025b).

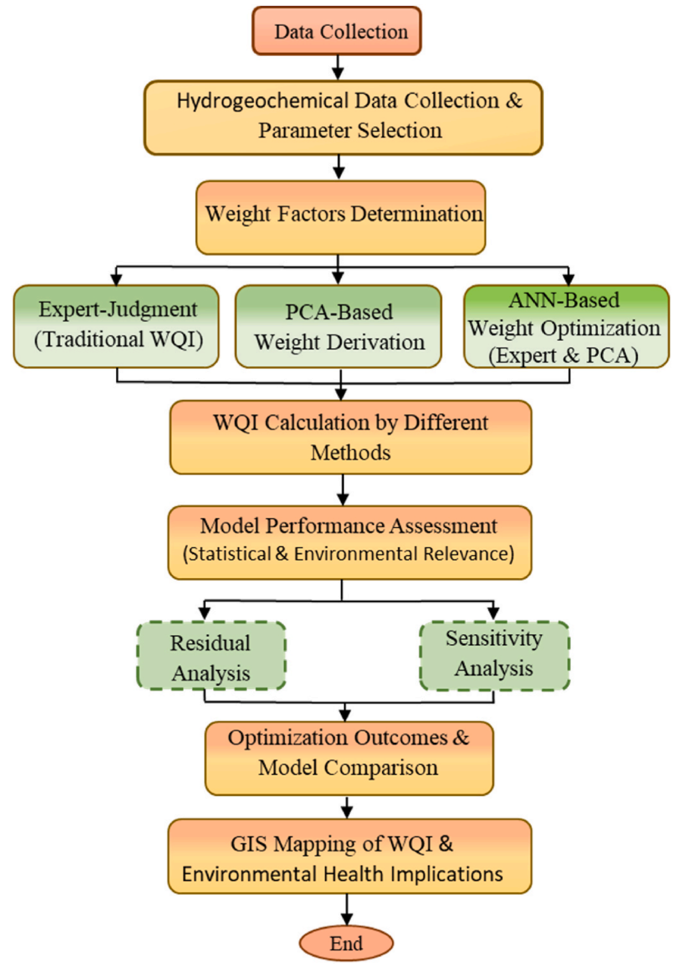

Fig. 3. Flowchart of methodology following, integrate the hydrogeochemical assessment framework using expert weighting, Principal Component Analysis (PCA), and Artificial Neural Networks (ANN) to evaluate groundwater quality index (WQI) and contamination risk.

The general formulation of the WQI is expressed as:

$$WQI = \sum_{i=1}^{n} W_i \times P_i \tag{1}$$

W_i is the relative weight assigned to the parameter i. P_i is the normalized value of the parameter i, calculated according to the water quality thresholds. n is the total number of parameters considered.

Standardization of parameters is usually carried out according to threshold values recommended by the WHO or other environmental bodies:

$$P_i = \left(\frac{C_i}{S_i}\right) \times 100 \tag{2}$$

C_i is the measured concentration of the parameter i. S_i is the recommended standard or limit value.

WQI values were classified into five quality categories (Appelo and Postma, 2005): excellent (<50), good (50–100), medium (100–150), poor (150–200), and very poor (>200).

2.6. Hybrid optimization framework

2.6.1. Expert-based weights

In the classical approach, importance coefficients (I_i) were assigned based on hydrogeochemical relevance and importance to regional environmental health. The relative weightings were calculated as follows:

$$W_i = \frac{I_i}{\sum_{j=1}^{n} I_j} \tag{3}$$

This weighting structure reflects the main regional hydrochemical processes, including salinization (EC, TDS, Cl^-, SO_4^{2-}), carbonate buffering capacity (HCO_3^-), and anthropogenic inputs such as nitrate (NO_3^-) contamination (Mahamadou et al., 2024; Sen et al., 2024; Abdullahi et al., 2025; Nigam and Indu, 2025). Although this method is widely adopted and benefits from expertise in the field, it inherently contains a degree of subjectivity. Therefore, its performance (Table 1) was evaluated against data-driven alternatives.

2.6.2. PCA-based weighting scheme

Principal Component Analysis (PCA) was applied to identify the main hydrogeochemical processes controlling groundwater chemistry and to derive statistically grounded parameter weights (Tripathi and Singal, 2019; Zhang et al., 2024; Abdullahi et al., 2025).

Prior to analysis, all variables were standardized as:

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \tag{4}$$

Z_{ij} is the standardized value for the parameter j in sample i. X_{ij} is the observed value. μ_j is mean of parameter j. σ_j: standard deviation of parameter j.

The correlation matrix was computed and eigenvalue decomposition performed to obtain eigenvalues (λ_k) and component loadings (L_{ik}). Components with eigenvalues greater than one were retained following the Kaiser criterion. Relative weights were then determined using a multi-component contribution approach:

$$W_i = \frac{\sum_{k=1}^{m} (|L_{ik}| \times V_k)}{\sum_{j=1}^{n} \sum_{k=1}^{m} (|L_{jk}| \times V_k)} \tag{5}$$

V_k is the percentage of the total variance that component k explains. m is the number of components that are kept (for example, eigenvalues >1). n is the total number of parameters.

This strategy minimizes subjective biases by leveraging the intrinsic statistical structure of the dataset, thus offering a more objective alternative to weighting (Singhal et al., 2020; Maliki et al., 2024).

The component retention was based on the Kaiser criterion, scree plot, and cumulative explained variance. The first three principal components (>80%) were retained. Unrotated components were used for weighting based on absolute loadings and explained variance, ensuring robustness to sign indeterminacy. Variables with cross-loadings were retained due to the cumulative nature of the weighting scheme.

2.6.3. ANN-based weight optimization

Artificial Neural Networks (ANN), specifically a multilayer perceptron (MLP), were employed to model complex relationships among hydrochemical variables and further refine WQI weighting (Ibrahim et al., 2023; Aldrees et al., 2024; Guenouche et al., 2024; Al-Zubaidi et al., 2025).

2.6.3.1. Data preparation. The eleven physicochemical parameters were used as input variables and normalized using Min–Max scaling. The target output corresponded to the reference WQI derived from either expert-based or PCA-based weighting.

2.6.3.2. Network architecture. The ANN consisted of: (i) An input layer with 11 neurons; (ii) One hidden layer (optimized experimentally); (iii) An output layer with a single neuron representing predicted WQI, see Fig. 4.

Nonlinear activation functions (ReLU or tanh) were used in the hidden layer, but a linear activation function was used in the output layer (Nayak et al., 2023; Ngwenya et al., 2025).

2.6.3.3. Training procedure. The dataset was randomly divided into training (≈70%), validation (≈15%), and testing (≈15%) subsets. Model performance was optimized by minimizing the Mean Squared Error (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(WQI_{pred,i} - WQI_{actual,i}\right)^2 \tag{6}$$

N: total number of samples, $WQI_{pred,i}$: predicted WQI_ANN, $WQI_{actual,i}$:

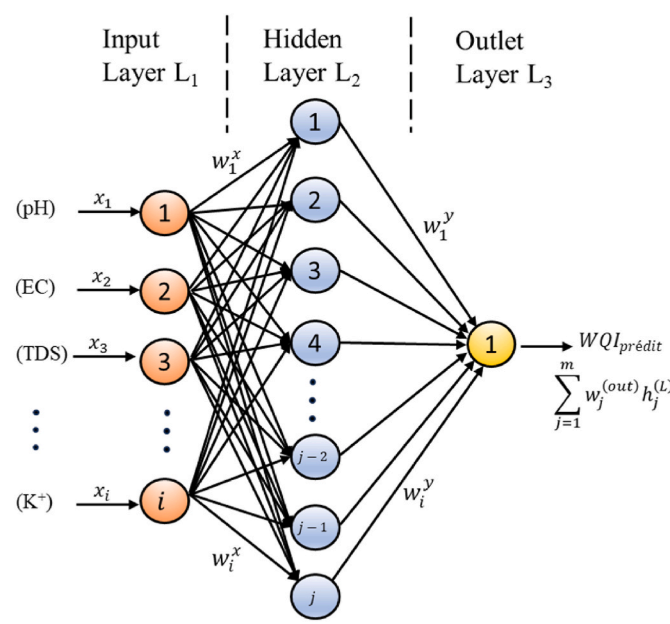

Fig. 4. Flowchart represent the architecture of Artificial Neural Network (ANN) and Training and optimization procedure for Water Quality Index (WQI) modeling. X_i is the input (variable) to the neuron. W_i^x is the relative weight assigned to parameter i in the layer x. $h_j^{(L)}$ denotes the activation (output) of neuron j in layer L. m is the neuron number in the last layer. WQI_{pred}: predicted WQI value.

Table 1
Expert-based weights and relative weights (W_i) of physicochemical parameters in Water Quality Index (WQI) assessment of the south Chott Hodna region. Potential of Hydrogen (pH), Electrical conductivity (EC), total dissolved solids (TDS), Calcium (Ca^{2+}), Magnesium (Mg^{2+}), bicarbonates (HCO_3^-), chlorides (Cl^-), sulfates (SO_4^{2-}), nitrates (NO_3^-), Sodium (Na^+), and Potassium (K^+).

Parameter	pH	EC	TDS	Ca^{2+}	Mg^{2+}	HCO_3^-	Cl^-	SO_4^{2-}	NO_3^-	Na^+	K^+	$\sum()$
Importance weight (I_i)	2	5	6	3	3	7	6	7	5	2	1	47
Expert (W_i)	0.04	0.11	0.13	0.06	0.06	0.15	0.13	0.15	0.11	0.04	0.02	1.00

actual/reference WQI

Optimization algorithms such as Adam or Levenberg–Marquardt were implemented. Early stopping and validation monitoring were applied to prevent overfitting and to improve generalization capacity.

2.6.3.4. Extraction of optimized weights. Relative parameter contributions were quantified using the connection weight approach (Olden and Jackson, 2002; Kulisz et al., 2021):

$$W_i = \frac{\sum_{l=1}^{L}(|w_{il} \times w_{l0}|)}{\sum_{j=1}^{n}\sum_{l=1}^{L}(|w_{jl} \times w_{l0}|)} \tag{7}$$

W_i: relative weight of parameter i, w_{il}: weight from input i to the hidden neuron l, w_{l0}: weight from hidden neuron l to output, L: number of hidden neurons, n: number of inputs.

These W_i values are then normalized to sum to one, forming optimized relative weights. These values ensure that only positive influences are taken into account.

2.6.3.5. Model evaluation. The model's performance was evaluated using the coefficient of determination (R^2), mean squared error (MSE), mean absolute error (MAE) and residual diagnostics (Sekarlangit et al., 2024). In addition, a sensitivity analysis was conducted to assess the relative influence of each parameter on the prediction of the WQI.

The integration of expert judgment, PCA, and ANN establishes a complementary and robust modeling framework. Expertise-based weighting ensures hydrogeochemical relevance, PCA enhances statistical objectivity, and ANN captures complex relationships between parameters. By combining these strengths, the proposed hybrid methodology improves the predictive reliability of the reproducibility and practical applicability of WQI modeling in semi-arid aquifers.

The combination of expert judgment, PCA and ANN creates a complementary and robust modeling framework. Expertise-based weighting ensures hydrogeochemical relevance, PCA improves statistical objectivity, and ANN captures complex non-linear relationships between parameters. The proposed hybrid methodology enhances the reliability of predictions, reproducibility and practical applicability of WQI modeling for semi-arid aquifers by integrating these strengths. (Taşan, 2023).

Because the dataset was relatively small, model validation relied on a hold-out approach with repeated random initialization and early stopping to limit overfitting. Although this method provided stable model performance, it may not fully capture data variability. Future studies could apply k-fold cross-validation to improve model robustness and reliability (Jung et al., 2020). In addition, the ANN model was guided by hydrogeochemically meaningful indices, including Expert-WQI and PCA-WQI, which helped reduce physically unrealistic predictions.

3. Results

3.1. Statistical analysis of groundwater quality

The evaluation of groundwater quality from 33 samples in the El Maadher area reveals significant non-compliance with WHO (2017) drinking water standards (Table 2), particularly for key physicochemical parameters. All samples' pH was within the allowed range (6.5-8.5), but more than 50% exceeded electrical conductivity and TDS limitations, indicating hydrogeochemical-related salinity and mineralization. All samples had elevated calcium and sulfate contents, with averages of 314.12 and 666.15 mg/L, respectively. These values exceed the WHO recommendations, which range from 75 to 250 mg/L. These anomalies probably originate from geogenic sources, including the dissolution of limestone and gypsum. (Selmane et al., 2023). Nitrate contamination is another major concern, with more than 90% of samples exceeding safe levels. An average nitrate level of 173.47 mg/L was calculated across the study area, representing a serious environmental health risk. Concentrations of Mg^{2+}, Cl^-, and Na^+ are also high (93.9%, 66.7%, and 30.3%, respectively), reflecting the combined effects of salinization, anthropogenic pollution, and agricultural inputs on the aquifer system. These high ionic concentrations are consistent with the geological composition of the region, whereby the Triassic evaporite formations and the Eocene gypsum marls contribute to a natural enrichment in Na^+, Cl^-, and Mg^{2+} by dissolution of evaporites and ion exchanges. The hydraulic connection between the Mio-Plio-Quaternary aquifers and the deeper Albian aquifers could promote the mineralization of groundwater. On the other hand, the concentrations of K^+ and HCO_3^- remain within acceptable limits, which indicates limited dissolution of carbonates and a low agricultural influence on these elements. The hydrogeochemical profile indicates the influence of geological formations and hydrogeological connections on salinity and quality of groundwater. She also emphasizes the importance of continuous monitoring. Furthermore, it demonstrates the potential of advanced multivariate statistical techniques and machine learning to refine the estimation of WQI and spatial interpretation in complex semi-arid aquifers. The asymmetry study revealed that pH values were approximately symmetrical, while EC, TDS, Ca^{2+}, Cl^-, and NO_3^- showed slight positive asymmetry, suggesting localized enrichment. Conversely, Mg^{2+}, SO_4^{2-}, Na^+, and especially K^+ showed a significant positive asymmetry, reflecting the influence of extreme values linked to mineral dissolution and anthropogenic inputs (Auwal et al., 2026). These results confirm the significant variability in groundwater quality in the studied region and its consequences for sustainable water management and public health.

3.2. Derivation of weight factors for WQI

3.2.1. Expert-based weight

In the normalized formulation (Eq. (8)), the effect of each parameter on the final WQI score is determined by its weight, which indicates how

Table 2
Descriptive statistical analysis of groundwater physicochemical parameters in the study area and comparison with WHO (2017) guideline values for drinking water quality.

Parameter	Standards WHO (2017)	Min	Max	Mean ± SD	Skewness	% Sample exceeding the WHO standard
pH	6.5-8.5	7.15	7.89	7.61 ± 0.16	−0.59	0%
EC	2500	1035.81	4901.78	2648.33 ± 871.21	+0.92	51.5%
TDS	1000	463.14	1825.51	1031.39 ± 307.01	+0.92	57.6%
Ca^{2+}	75	108.22	601.20	314.12 ± 116.2	+0.85	**100%**
Mg^{2+}	50	36.46	255.20	97.57 ± 45.55	+2.03	**93.9%**
HCO_3^-	500	149.45	244.00	178.84 ± 27.64	+0.3	0.0%
Cl^-	250	88.75	923.00	425.89 ± 220.68	+0.9	**66.7%**
SO_4^{2-}	250	250.00	1300.00	666.15 ± 232.38	+1.3	**100%**
NO_3^-	50	12.00	407.00	173.47 ± 104.41	+1.0	**90.9%**
Na^+	200	42.78	379.07	127.52 ± 75.73	+1.5	30.3%
K^+	12	5.10	16.00	7.35 ± 1.9	>+3.0	6.1%

All parameter units are in mg/l, EC (μS/cm) and pH without unit. SD is the Standard Deviation. WHO: World Health Organization.

much it can affect human health and groundwater use. Parameters such as SO_4^{2-}, HCO_3^-, TDS, and Cl^- were given higher relative weighting due to their importance for salinity and mineralization. In contrast, K^+ and Na^+ were given lower weighting because they are less toxic and have less effect at normal levels.

$$WQI = 0.04 \times P_{pH} + 0.11 \times P_{EC} + 0.13 \times P_{TDS} + 0.06 \times P_{Ca^{2+}} + 0.06 \times P_{Mg^{2+}} + 0.15 \times P_{HCO_3^-} + 0.13 \times P_{Cl^-} + 0.15 \times P_{SO_4^{2-}} + 0.11 \times P_{NO^{3-}} + 0.04 \times P_{Na^+} + 0.02 \times P_{K^+} \quad (8)$$

The weighted-aggregation method developed in this study gives a more accurate and understandable assessment of groundwater quality especially in semi-arid regions where salinity and hardness problems are widespread. To support efficient resource management, the resulting WQI is used to classify water quality into categories such as excellent, good, poor, or unfit for consumption.

3.2.2. PCA-based weight

Principal Component Analysis (PCA) was applied to the standardized dataset (11 parameters) to derive objective, data-driven weights. The correlation matrix revealed strong multicollinearity among EC, TDS, Ca^{2+}, and Cl^-, confirming their shared hydrogeochemical origin related to mineral dissolution.

Three principal components with eigenvalues greater than 1 explained 80.13% of the total variance (PC1 = 55.27%, PC2 = 14.15%, PC3 = 10.71%) (Table 3). PC1 was primarily associated with EC, TDS, Ca^{2+}, SO_4^{2-}, Cl^-, and Na^+, representing salinity-driven processes. PC2 and PC3 were linked with nutrient inputs and buffering processes, especially NO_3^-, HCO_3^-, K^+ and pH. According to variance-scaled absolute loadings, the biggest weights were attributed to Cl^-, EC, TDS, SO_4^{2-} and Ca^{2+} which validated that groundwater variability is predominantly dominated by salinity processes. NO_3^- and HCO_3^- contributed less to the total statistical variance yet nitrate still remains relevant from an environmental perspective due to health consequences and is of importance from a public health perspective. These results are consistent with previous research in semi-arid regions. Hamma et al. (2024) shown that the integration of PCA and classification approaches may effectively identify hydrogeochemical parameters and their related environmental hazards. In the same way, a study carried out in the F'kirina plain (Algeria) (Salima and Belgacem, 2017) proved that the PCA makes it possible to separate natural geochemical fingerprints from anthropogenic pollution sources in the modeling of the WQI.

Column W_i (PCA) shows the extent to which each parameter contributes to explaining overall changes in groundwater quality. The highest weights are associated with EC (0.106), TDS (0.106), Cl^- (0.107), and SO_4^{2-} (0.106). These indicate that these parameters are the main drivers of variability within the dataset. In contrast, NO_3^- (0.062) and HCO_3^- (0.067) have the lowest weights, suggesting a relatively minor influence. This ranking can be applied to sensitivity analysis to identify the most influential indicators, to prioritize monitoring strategies, or to refine the model.

The values of PC1 and PC2 shown in Fig. 5 are variable loadings, which indicate each variable's contribution to the dimensions. W_i shows how much each of them contributes to the PCA space. Variables far from the origin along an axis strongly influence that principal component; those near each other are positively correlated, those opposites are negatively correlated, and those near the origin have little influence on PC_1 and PC_2, with their variation explained more by other components like PC_3. The analysis provides a clear hydrogeochemical evaluation of the primary parameters that control variability in groundwater quality in the semi-arid aquifer system. (Singhal et al., 2020; Maliki et al., 2024). In the standardized formulation (Eq. (9)), the final WQI score is based on its weightings.

$$WQI = 0.092 \times P_{pH} + 0.106 \times P_{EC} + 0.106 \times P_{TDS} + 0.104 \times P_{Ca^{2+}} + 0.074 \times P_{Mg^{2+}} + 0.067 \times P_{HCO_3^-} + 0.107 \times P_{Cl^-} + 0.106 \times P_{SO_4^{2-}} + 0.062 \times P_{NO^{3-}} + 0.089 \times P_{Na^+} + 0.086 \times P_{K^+} \quad (9)$$

Exploratory Varimax rotation produced a similar factor structure and comparable variable rankings. The Water Quality Index (WQI) classification patterns also remained consistent, with no major differences between the rotation methods (Acal et al., 2020). This shows that the interpretation of the results and the final groundwater quality categorization were stable regardless of the chosen rotation strategy.

3.2.3. ANN-based weight (optimization)

Deriving weights for the WQI using ANN is a sophisticated optimization method that aims to determine the relative contribution (W_i) of each groundwater quality parameter to the overall index (Olden and Jackson, 2002; Kulisz et al., 2021. It uses two distinct benchmarks: the expert-assigned WQI (WQI_expert) and the PCA-derived WQI (WQI_PCA). A multilayer perceptron neural network (ANN) was used in this study to optimize these weights. The main objective was to reduce the discrepancy between the WQI values predicted by ANN and the chosen target index, thereby improving the accuracy of groundwater quality assessment and its implications for environmental health risk assessment.

As shown in Table 4, the optimal weights of Expert_ANN and PCA_ANN models are similar, although with some subtle differences. This is because the importance of each parameter varies depending on the target definition. In the Expert_ANN configuration, nitrate (NO_3^-) has the highest weight (0.18), showing that the effect is most significant on water quality in the semi-arid study area. This disparity can be explained by agricultural runoff and wastewater infiltration, which are some of the primary factors of anthropogenic pollution. Subsequently, the moderate weights are magnesium Mg^{2+} (0.12), TDS (0.10), electrical conductivity EC (0.09), SO_4^{2-} (0.09), and pH (0.09). In the PCA_ANN setting, the importance of nitrate is slightly reduced (0.15), while electrical conductivity (0.11), Ca^{2+} (0.10) and SO_4^{2-} (0.10) become more important.

Table 3
Principal component analysis results of groundwater physicochemical variables, including eigenvalues, explained and cumulative variance percentages, factor loadings (PC1, PC3, PCA) derived parameter weights ($W_{i,\text{PCA}}$) for hydrogeochemical interpretation and water quality assessment.

N°	Principal component	Explained variance (%)	Cumulative variance (%)	Parameter	PC1	PC2	PC3	W_{i_PCA}
1	**PC1**	**55.27**	**55.27**	pH	−0.168	0.279	−0.695	**0.092**
2	**PC2**	**14.15**	**69.42**	EC	0.396	0.045	−0.131	**0.106**
3	**PC3**	**10.71**	**80.13**	TDS	0.396	0.045	−0.131	**0.106**
4	PC4	6.87	87.01	Ca^{2+}	0.384	0.113	−0.059	**0.104**
5	PC5	6.14	93.14	Mg^{2+}	0.298	−0.051	0.001	**0.076**
6	PC6	3.47	96.61	HCO_3^-	−0.11	−0.613	−0.025	**0.067**
7	PC7	2.06	98.68	Cl^-	0.371	0.109	−0.206	**0.107**
8	PC8	1.08	99.76	SO_4^{2-}	0.363	−0.249	−0.029	**0.106**
9	PC9	0.14	99.9	NO_3^-	0.029	0.647	0.295	**0.062**
10	PC10	0.1	100	Na^+	0.286	−0.179	−0.169	**0.089**
11	PC11	0	100	K^+	0.237	−0.015	0.5637	**0.086**

Wi: was estimated from the three principal components (3 PCs). Wi_PCA is weight of the i-th parameter derived from Principal Component Analysis

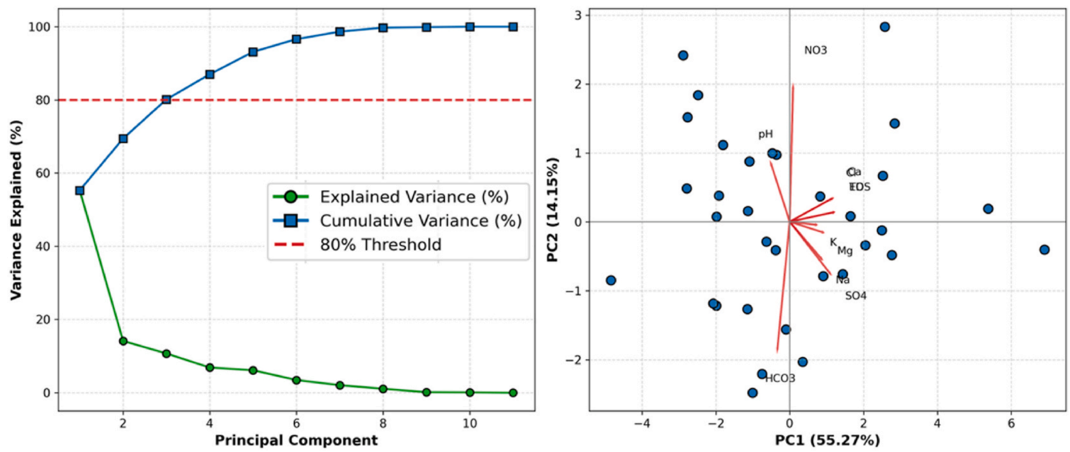

Fig. 5. Projection of hydrogeochemical variables on the PC_1/PC_2 factorial plane, illustrating parameter correlations and relative contributions, accompanied by a scree plot showing the variance explained by principal components Analysis (PCA) for groundwater quality assessment.

Table 4
Relative weights assigned to physicochemical parameters for Water Quality Index (WQI) calculation using the Expert–ANN and PCA–ANN models.

Parameter	pH	CE	TDS	Ca^{2+}	Mg^{2+}	HCO_3^-	Cl^-	SO_4^{2-}	NO_3^-	Na^+	K^+	$\sum W_i$
$W_{i\_Expert\ \_ANN}$	0.09	0.09	0.1	0.07	0.12	0.06	0.08	0.09	0.18	0.06	0.06	1.00
$W_{i\_PCA\_ANN}$	0.09	0.11	0.09	0.1	0.12	0.05	0.09	0.1	0.15	0.04	0.07	1.00

Wi_Expert_ANN is weight of parameter i derived from expert knowledge combined with Artificial Neural Network (ANN), Wi_PCA_ANN is weight of parameter i derived from Principal Component Analysis (PCA) combined with Artificial Neural Network (ANN).

The result indicates that the objective set by PCA focuses on parameters that affect the overall variance of hydrogeochemical processes, rather than solely on the observed pattern of water quality degradation identified by experts. The weightings of Na+, K+ and HCO_3^- are very low in both models, suggesting that these parameters contribute limitedly to the worsening of water quality in this dataset.

Table 5 provides the performance indicators of the model. Both ANN configurations achieved high accuracy with R^2 values greater than 0.987 during training. The Expert_ANN model converged after 66 epochs with a mean absolute error (MAE) of 0.0354 and a mean squared error (RMSE) of 0.0396. In contrast, the WQI_PCA_ANN converged in only 20 epochs, but its error rate was slightly higher (RMSE = 0.0454; MAE = 0.04035). This suggests that the PCA based objective leads to faster network learning at the cost of accuracy.

In Fig. 6, the scatter plots clearly demonstrate that the ANN-enhanced models closely reproduce the observed WQI values, confirming the strong effectiveness of ANN optimization in improving the weighting strategies. In the first graph (WQI_Expert vs. WQI_Expert_ANN), the data points cluster closely around the 1:1 line, indicating that the ANN successfully captures complex multivariate relationships that the model based solely on expertise does not account for. The second graph (WQI_PCA vs. WQI_PCA_ANN) shows an even more compact distribution with minimal deviation from the regression line; it demonstrates that the PCA-derived weighting scheme benefits more from ANN refinement. Overall, these visual results confirm that the ANN significantly improves the predictive accuracy of both expertise-based and PCA-based WQI formulations, and also that the PCA-ANN model proves to be the most stable and reliable for representing real-world groundwater quality conditions. The training loss fluctuated slightly during the optimization of the different parameters, the validation loss remained low and stable throughout the training. The model's validation means squared error (MSE) remained stable between approximately 0.05 and 0.18, and its coefficient of determination (R^2) on the test set remained consistently high (0.92–0.98). These indicators show that the artificial neural network (ANN) has demonstrated good predictive agreement despite slight fluctuations during training.

In terms of variability, the PCA_ANN model exhibits the largest interquartile range, indicating the greatest dispersion among the four approaches. The Expert_ANN model also shows noticeable variability, although it remains more stable than PCA_ANN. The classical Expert method presents a moderate spread, whereas the PCA approach demonstrates slightly lower variability compared to its ANN-enhanced counterpart.

With respect to extreme values, all methods consistently identify high outliers associated with heavily polluted samples. However, the ANN-based models tend to amplify higher WQI values, resulting in greater dispersion at the upper end of the distribution. In contrast, low WQI values are classified consistently across all approaches, suggesting a stable and reliable identification of good-quality groundwater.

Although the ANN models showed good predictive performance, the relatively small sample size may reduce the reliability of the connection-weight estimates and increase the risk of overfitting (D'souza et al., 2020). As a result, ANN-derived parameter weights should be interpreted as relative indicators of hydrochemical influence rather than precise measures of sensitivity. Several studies suggest that small datasets can increase uncertainty in model-based importance rankings, which highlights the need for caution when interpreting results. To improve robustness, future research could use bootstrap resampling or Monte Carlo cross-validation or rely on larger regional datasets. These

Table 5
Performance evaluation of ANN-based Water Quality Index (WQI) prediction models optimized using different algorithms during the training phase based on RMSE, MAE, and R.[2].

WQI	Phase	epoch	RMSE	MAE	R^2
*WQI*_Expert_ANN	Training	66	0.039587	0.035354	0.98929
*WQI*_PCA_ANN	Training	20	0.045428	0.04035	0.987729

WQI_Expert_ANN: Water Quality Index model based on Expert weighting combined with an Artificial Neural Network (ANN). WQI_PCA_ANN: Water Quality Index model using PCA-based weighting combined with an Artificial Neural Network (ANN). RMSE: Root Mean Square Error. MAE: Mean Absolute Error. R^2: the coefficient of determination.

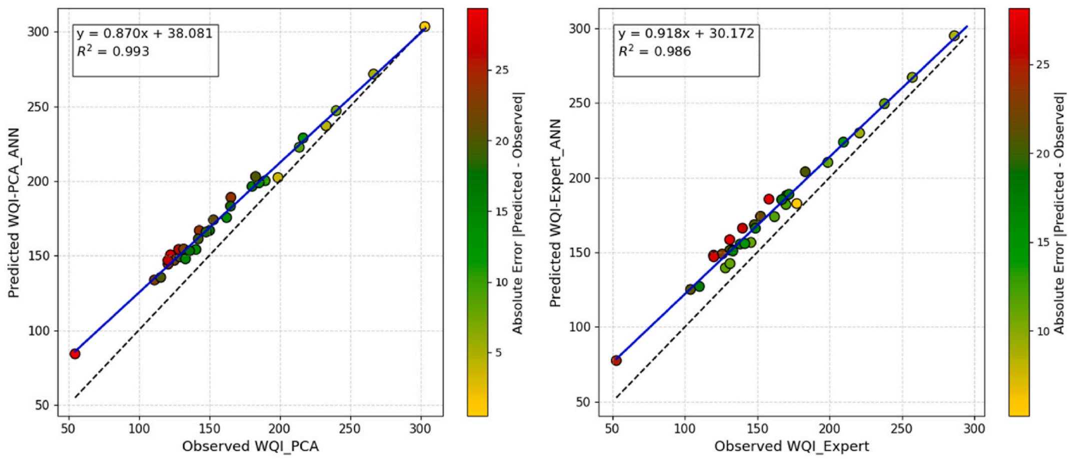

Fig. 6. Comparison of observed and ANN-Predicted WQI Values under Expert-based and PCA-based weighting strategies for model performance evaluation. R^2 is the coefficient of determination.

approaches would help reduce uncertainty and provide more reliable estimates of variable importance in groundwater quality modelling.

3.3. WQI estimation by different approaches

Fig. 7 shows how four different methods measure WQI values in 33 groundwater samples. The majority of the samples fall within the medium to poor quality range, but a minority is classified as very poor. ANN-optimized methods tend to produce slightly higher WQI values, while PCA-based WQI generally produce lower estimates. This study shows that both the quality of the data and the method used can affect the results.

The mean WQI values indicate overall poor groundwater quality in the study area. WQI_PCA (161.79) and WQI_Expert (157.71) fall within the "poor" category, while the ANN-based models produced slightly higher averages (178.87 and 175.03), suggesting greater sensitivity to degradation. Overall, all four approaches show consistent trends, with ANN models slightly amplifying WQI values compared to conventional methods.

Table 6 shows that the majority of samples were classified as "medium" quality in the PCA and expert-weighted models (48.5% and 51.5%, respectively), while the ANN-enhanced models shifted toward "poor" and "very poor" classifications, reflecting their sensitivity to complex, nonlinear parameter interactions. The percentage of wells rated as "very poor" increased from 15.2% (WQI_Expert) and 18.2% (WQI_PCA) to 27.3% (WQI_PCA_ANN) and 21.2% (WQI_Expert_ANN). Across all methods, several wells—F5, F19, F22, F30, and F31—were repeatedly marked as "very poor," highlighting risks of groundwater contamination and potential environmental health issues. This index highlights areas of concern and potential sources of anthropogenic pollution. The difference between artificial neural network models improves diagnostics by capturing hydrogeochemical processes that linear water quality assessment frameworks may miss. Overall, these results underscore the importance of combining data-driven (PCA) and intelligent (ANN) methods in hydrogeochemical assessment and groundwater quality evaluation for semi-arid aquifers.

3.4. GIS-based spatial mapping of WQI

The Ordinary Kriging (OK) method was used to model the spatial distribution of the WQI in the Maadher plain (Fig. 8). The OK method was selected because it provides unbiased spatial estimates and allows evaluation of prediction uncertainty, which is important for studying vulnerable aquifers in arid and semi-arid regions. Although the sampling density was moderate, strong spatial autocorrelation and relatively consistent hydrochemical gradients supported the use of this method. Similar approaches have been reported in previous groundwater studies under comparable hydrogeological conditions. Increasing the number of sampling points in future work may further improve interpolation reliability and reduce spatial uncertainty. A total of 33 groundwater samples were analyzed in the study, with WQI values ranging from 52.67 to 303.44. The observed readings exhibit moderate to high variability (standard deviation: 42.8–50.3), indicating distinct differences in the hydrochemistry of this groundwater system. The Pearson correlation coefficients for the four WQI formulations (PCA, PCA_ANN, Expert, and Expert_ANN) were all above 0.97. The observed consistency suggests that the major hydrochemical gradients are preserved across weighting formulations, despite differences in parameter prioritization and nonlinear optimization. The significant autocorrelation and robust sample density indicated that kriging was an effective method (Ahmad Saleh et al., 2021; Redjem et al., 2021; Kumar et al., 2022). Kriging uses variogram modeling to account for the spatial structure inherent in hydrochemical information. This sets it apart from deterministic methods like inverse distance weighting (IDW). An experimental variogram was generated from all pairwise sample comparisons (two-point statistics) to quantify how WQI values change with distance. The data were fitted to a spherical model, a standard approach in groundwater geostatistical analysis.

Spatially, across all four models, very poor water quality (WQI >200) is consistently observed in the north and northeast, especially near Chott Hodna's recharge fringes, where salinity and nitrate contamination are likely to increase due to evaporative concentration and anthropogenic sources such as return flows from agriculture (Dougha & Hasbaia, 2019; Selmane et al., 2023). The central basin is dominated by medium to poor water (WQI: 100–200), while good quality water (WQI: 50–100) is restricted to small areas in the southwest, where seepage zones and upstream wadi segments intersect.

ANN optimization in both PCA-based and expert-based formulations resulted in sharper class boundaries and a slightly expanded "good water" zone in the southwest. This demonstrates that ANN can detect subtle complex hydrogeochemical relationships in the dataset (Nishat et al., 2025). However, the persistence of the northern "very poor water" belt indicates a high environmental health risk, regardless of the weighting method. This geostatistical mapping method, which uses a spherical variogram and an ordinary kriging framework, generates a spatial model that is both quantitative and sensitive to errors. Such a degree of precision is essential for effective risk-based groundwater management. This is particularly applicable in the Maadher Plain, where overexploitation, salinization, and health risks from nitrates necessitate targeted mitigation strategies in the most vulnerable areas.

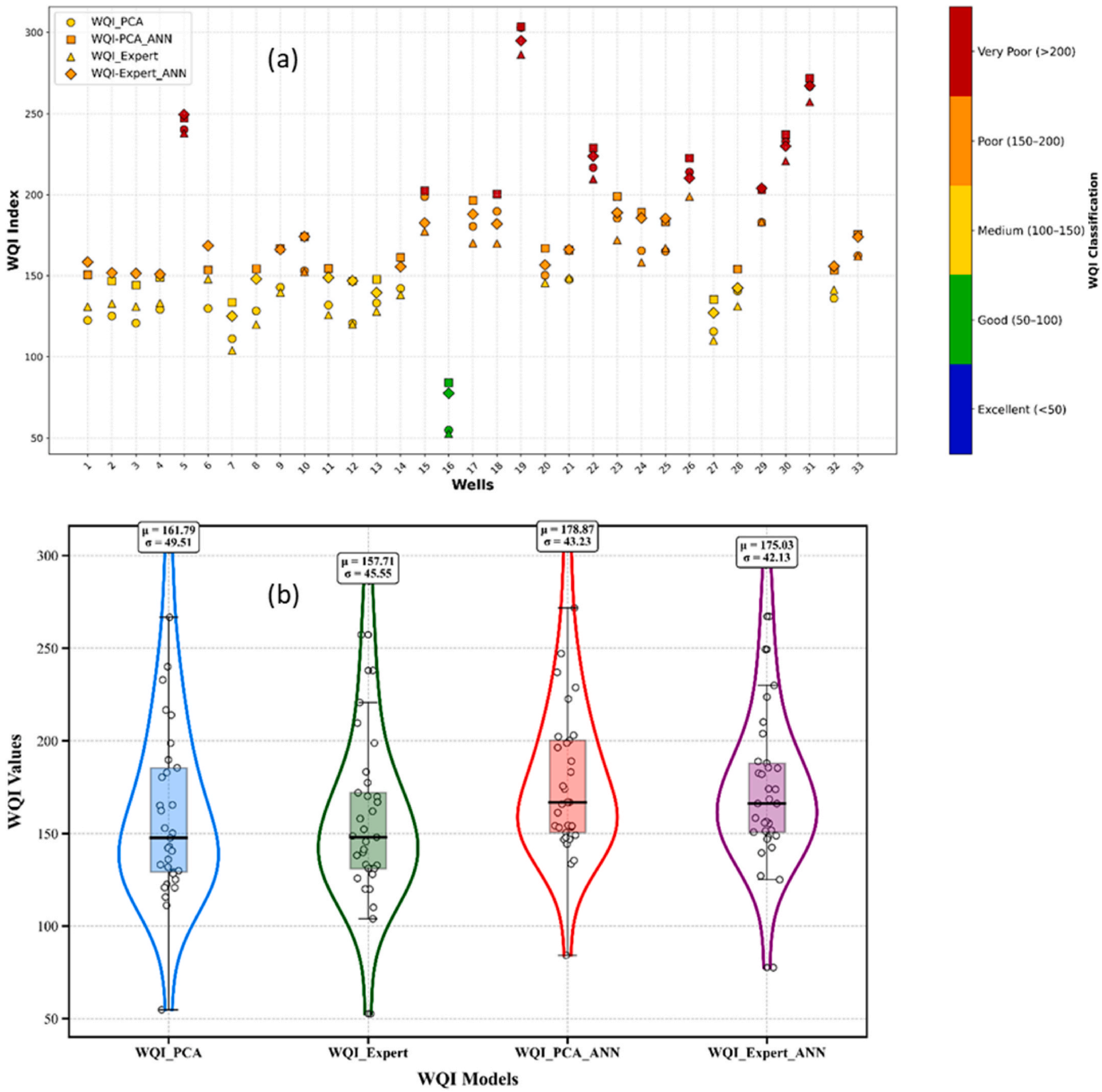

Fig. 7. Comparative analysis of Water quality index (WQI) values obtained from different conventional (Expert, Principal Component Analysis (PCA)) and ANN-based models. a) Comparison of WQI values for all investigated wells derived from the different models, including groundwater quality classification categories. b) Violin and boxplot distributions illustrating the statistical variability, density distribution, mean (μ), and standard deviation (σ) of WQI values for each model. WQI_Expert: Water Quality Index model using Expert-based weighting, WQI_Expert_ANN: Water Quality Index model based on Expert weighting combined with an Artificial Neural Network (ANN), WQI_PCA: Water Quality Index model using PCA-based weighting, WQI_PCA_ANN: Water Quality Index model using PCA-based weighting combined with an ANN.

Although ordinary kriging provides unbiased spatial estimation and accounts for spatial autocorrelation, this study does not explicitly quantify prediction uncertainty using kriging variance maps (Redjem et al., 2021). Also, incorporating uncertainty layers would strengthen the spatial analysis. Evaluating sensitivity to variogram model selection could further improve the reliability of the results. These steps are important for improving spatial decision support for groundwater management, especially in arid and semi-arid regions where sampling density is limited.

4. Discussion

4.1. Residual analysis of WQI methods

Fig. 9a illustrates the disparities among the three models in comparison to the WQI_PCA reference. The WQI_Expert model (blue) has residuals predominantly near zero, with fluctuations ranging from approximately −5 to +22. This pattern indicates that the prediction errors are somewhat balanced and that the WQI_PCA baseline is nearer

Table 6
Groundwater quality classification and drinking suitability assessment using the conventional and ANN-based WQI models in the Maadher region.

WQI Range (n = 33)	Water Quality Class	WQI_Expert (Nbr of wells)	WQI_Expert (%)	WQI_Expert_ANN (Nbr of wells)	WQI_Expert_ANN (%)	Difference (%)	Reclassified Wells (Expert → ANN)
<50	Excellent water	0	0.00	0	0.00	0.00	–
50–100	Good water	1	3.03	1	3.03	0.00	–
100–150	Moderately polluted water	17	51.52	7	21.21	−30.31	F1, F2, F3, F4, F6, F9, F14, F32, F33 → Poor
150–200	Poor water	10	30.30	18	54.55	+24.25	F1, F2, F3, F4, F6, F9, F14, F32, F33 ← Medium
>200	Highly polluted water	5	15.15	7	21.21	+6.06	F26, F29: Poor → Very Poor
WQI Range (n = 33)	**Water Quality Class**	**WQI_PCA (Nbr of wells)**	**WQI_PCA (%)**	**WQI_PCA_ANN (Nbr of wells)**	**WQI_PCA_ANN (%)**	**Difference (%)**	**Reclassified Wells (PCA → PCA_ANN)**
<50	Excellent water	0	0.00	0	0.00	0.00	–
50–100	Good water	1	3.03	1	3.03	0.00	–
100–150	Moderately polluted water	16	48.48	7	21.21	−27.27	F1, F6, F8, F9, F11, F14, F28, F32, F33 → Poor
150–200	Poor water	10	30.30	16	48.48	+18.18	F1, F6, F8, F9, F11, F14, F28, F32, F33 ← Medium
>200	Highly polluted water	6	18.18	9	27.27	+9.09	F15, F18, F29: Poor → Very Poor
Total	—	33	100	33	100	–	–

WQI_Expert: Water Quality Index model using Expert-based weighting, WQI_Expert_ANN: Water Quality Index model based on Expert weighting combined with an Artificial Neural Network (ANN), WQI_PCA: Water Quality Index model using PCA-based weighting, WQI_PCA_ANN: Water Quality Index model using PCA-based weighting combined with an ANN.

to the actual values. The WQI_Expert_ANN (orange) and WQI_PCA_ANN (green) models predominantly exhibit negative residuals, ranging from −38 to +18 and −35 to +8, respectively. This persistent negative bias indicates that individuals tend to overrate WQI relative to WQI_PCA, leading to lower residual values. The ANN-based models have larger residuals, with WQI_Expert_ANN displaying significant outliers, particularly for sample 5. This may reflect local overfitting or model sensitivity to extreme hydrochemical conditions in some samples. The results demonstrate that the ANN models exhibit consistent patterns, whereas the classic WQI_Expert produces predictions that are more closely aligned with the PCA-derived reference. Furthermore, the ANN methodologies may require recalibration or improved feature selection to mitigate systematic bias.

4.2. Sensitivity analysis of weighting factors

Fig. 9b illustrates variations in parameter weights (W_i) across the four weighting strategies. In the Expert model, the largest weight was given to NO_3^- (0.18), followed by Mg^{2+} (0.12) and TDS (0.10), suggesting the joint influence of agricultural contamination and salinity. The relatively high weights assigned to HCO_3^- and SO_4^{2-} reflect expert recognition of buffering processes and evaporite dissolution.

The Expert_ANN configuration maintains nitrate as the dominant parameter but increases the relative importance of Ca^{2+} and SO_4^{2-} while reducing the weight of HCO_3^-. This redistribution highlights the capacity of machine-learning models to refine parameter influence by detecting complex dependencies within the dataset.

PCA-based approaches, in contrast, distribute weights more evenly, particularly among EC, TDS, and Cl^-, consistent with their strong contribution to total variance. Adjustments introduced by the PCA_ANN model remain moderate compared with the expert-based ANN refinement.

Three main observations emerge: (i) All approaches consistently identify EC, TDS, and NO_3^- as key indicators of groundwater degradation. (ii) Expert-based weighting emphasizes parameters with recognized geochemical and anthropogenic significance. (iii) ANN optimization shifts emphasis toward variables with stronger predictive power under complex conditions.

Methodological divergence is more strongly emphasized by Spearman correlation analysis (Fig. 10). Expert weights are weakly correlated to the other strategies (~0.23-0.26), indicating a different priority structure. In contrast, the weights of Expert_ANN and PCA_ANN are strongly correlated (0.84), indicating internal consistency among ANN-derived solutions. PCA-only weights correlate weakly with ANN-based techniques, demonstrating that variance-based statistical weighting is conceptually quite different.

Together, these findings support the value of combining complementary methodologies to balance interpretability with predictive strength.

4.3. Optimization results and model performance evaluation

Table 7 shows the performance indicators where we find significant differences between weighting schemes. The W_i_Expert model (R^2 = 0.78) offers strong interpretability but limited predictive performance. Incorporating ANN optimization (W_i_Expert_ANN) substantially improves explanatory power (R^2 = 0.92), confirming the benefit of modeling complex multivariate relationships. Similarly, the W_i_PCA model (R^2 = 0.80) provides an objective and statistically based approach, but is less flexible to accommodate site-specific hydrochemical complexity. The hybrid W_i_PCA_ANN model showed the best agreement with the reference WQI formulations, with R^2 = 0.94, RMSE = 1.92, and MAE = 1.48. This suggests that combining PCA-based statistical structure with ANN learning improves the nonlinear representation of WQI behavior in the available dataset. However, these results mainly reflect internal model optimization rather than confirming the absolute accuracy of groundwater quality assessment.

Table 8 shows a consistent hydrogeochemical signal across the four weighting strategies (PCA, PCA-ANN, Expert, and Expert-ANN). In all cases, nitrate (NO_3^-; Average weight = 0.124), sulfate (SO_4^{2-}; 0.111), total dissolved solids (TDS; 0.108), electrical conductivity (EC; 0.102), and chloride (Cl^-; 0.101) appear as the principal factors controlling groundwater quality in the semi-arid South Chott Hodna aquifer. The ANN-optimized models further accentuate the role of NO_3^- (reaching 0.178) and Mg^{2+} (0.118), underscoring the combined effects of fertilizer leaching, agricultural return flows, and ongoing mineral dissolution.

Salinity and nitrate contamination are the main causes of groundwater deterioration in semi-arid aquifers. These are also common challenges that lead to declines in water quality in regions with limited water supply (Egbueri and Agbasi, 2022; Auwal et al., 2026). In the South

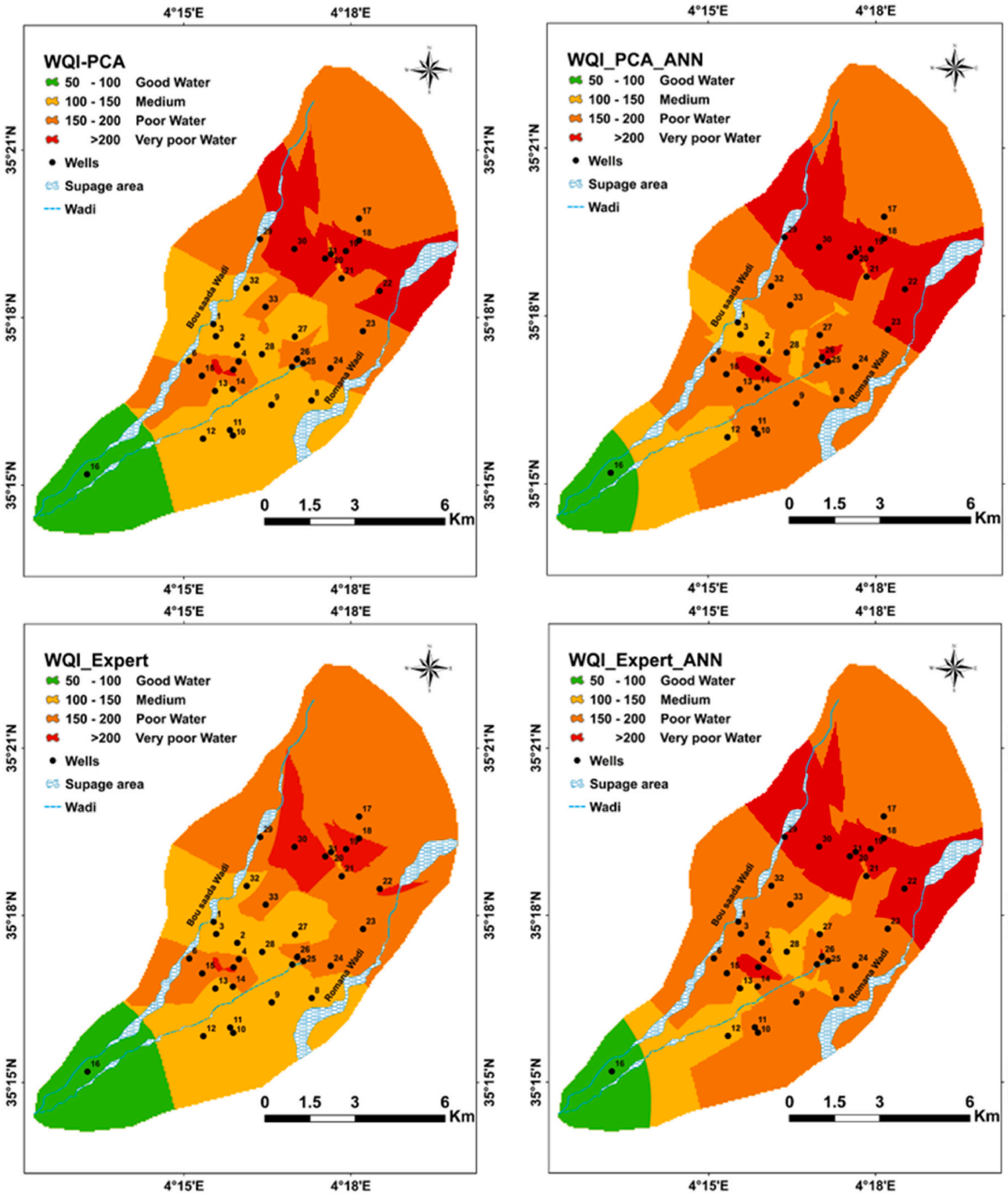

Fig. 8. Spatial distribution maps of the Water Quality Index (WQI) derived from Principal Component Analysis (PCA), Expert, and ANN-based weighting approaches in the Maadher zone, illustrating hydrogeochemical variability and identifying potential groundwater contamination zones. WQI_Expert: Water Quality Index model using Expert-based weighting, WQI_Expert_ANN: Water Quality Index model based on Expert weighting combined with an Artificial Neural Network (ANN), WQI_PCA: Water Quality Index model using PCA-based weighting, WQI_PCA_ANN: Water Quality Index model using PCA-based weighting combined with an ANN.

Chott Hodna aquifer, NO_3^- (0.124), SO_4^{2-} (0.111), TDS (0.108), EC (0.102), and Cl^- (0.101) were the main variables affecting groundwater quality. This pattern indicates a synergistic effect of anthropogenic activities and natural mineral-dissolving processes. Similar hydrochemical conditions have been documented throughout North Africa, particularly in Algeria, Morocco, and Tunisia, where EC, TDS, Cl^-, and SO_4^{2-} are extensively employed to detect evaporite dissolution, carbonate weathering, evaporation, and irrigation return flow in semi-arid aquifers (Bouderbala et al., 2016; Bradai et al., 2022; Guenouche et al., 2024; Maliki et al., 2024).

Similar circumstances have been noted in India and Bangladesh. In these agricultural regions, EC, TDS, NO_3^-, and Mg^{2+} are strongly linked to fertilizer leaching, groundwater overuse, and progressive salt accumulation (Shaibur et al., 2019a; Chakraborty et al., 2023; Sen et al., 2024). Long-term agricultural expansion in China has similarly elevated nitrate enrichment and salinity-related variables in intensively irrigated aquifers subjected to environmental stress and increasing population demand. Similar trends were reported in Middle Eastern countries such as Iran, Iraq, Saudi Arabia, and Egypt, where SO_4^{2-}, Cl^-, EC, and NO_3^- are associated with evaporite weathering, fast urban growth, and irrigation return flows (Bahrami et al., 2022; Fatah et al., 2020; Alrowais et al., 2023; Asmoay et al., 2025). Similar

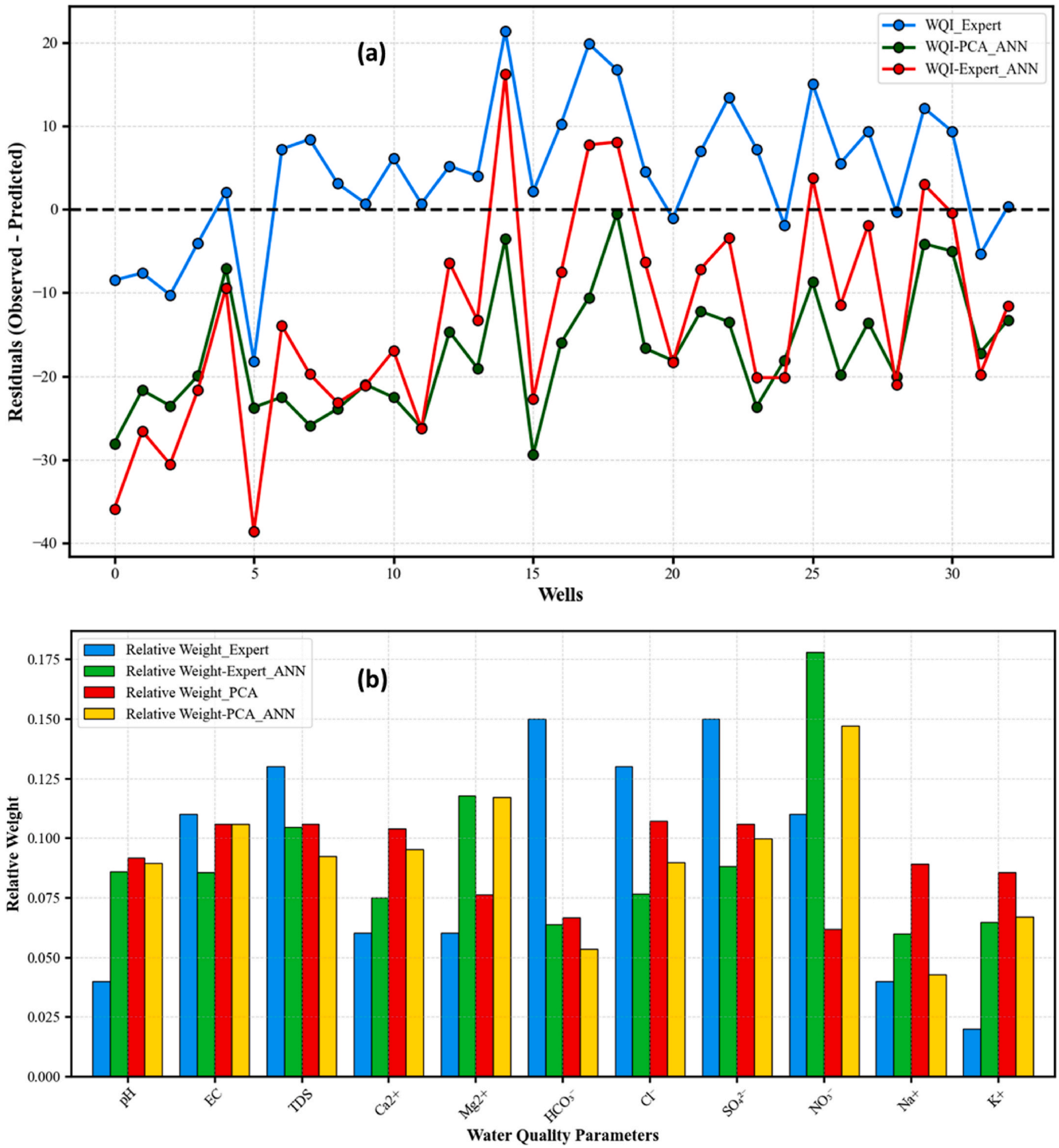

Fig. 9. a) Residual comparison of Water Quality Index (WQI) models against the PCA-derived reference. b) Comparison of the relative weights assigned to each hydrogeochemical parameter to the four WQI weighting approaches (bar charts). WQI_Expert: Water Quality Index model using Expert-based weighting, WQI_Expert_ANN: Water Quality Index model based on Expert weighting combined with an Artificial Neural Network (ANN), WQI_PCA_ANN: Water Quality Index model using PCA-based weighting combined with an ANN.

groundwater quality patterns were seen in arid agricultural basins of the southwestern United States, where elevated TDS and nitrate concentrations are associated with extensive irrigation and restricted natural recharge.

In contrast, pH, Na^{+}, and K^{+} showed relatively low weights in the present study. This behavior is typical of carbonate-buffered aquifers, where these variables usually stay the same unless salinity or sodification gets worse. (Appelo and Postma, 2005). The ANN-based models also increased the relative importance of NO_3^{-} and Mg^{2+}, indicating that agricultural runoff, wastewater infiltration, and mineral dissolution strongly affect groundwater chemistry. Overall, the results agree with previous international studies and show that groundwater degradation

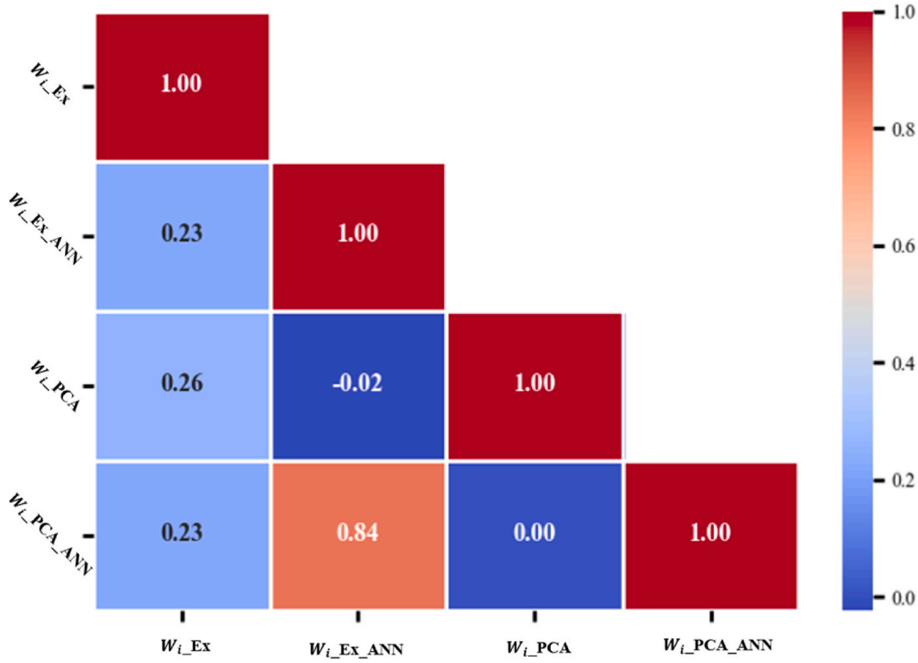

Fig. 10. Spearman correlation between weighting methods (Expert, PCA, Expert-ANN, PCA-ANN) used for the construction of the Water Quality Index (WQI), illustrating the consistency of weighting strategies in the assessment of groundwater quality. W_i_Ex is the weight of parameter i derived from expert knowledge, W_i_Ex_ANN is the weight of parameter i derived from expert knowledge combined with Artificial Neural Network (ANN), W_i_PCA is weight of parameter i derived from Principal Component Analysis (PCA), W_i_PCA_ANN is weight of parameter i derived from (PCA) combined with Artificial Neural Network (ANN).

Table 7
Performance comparison of WQI models (W_{i_Expert}, $W_{i\_Expert\_ANN}$, W_{i_PCA}, and $W_{i\_PCA\_ANN}$) using statistical indicators (R^2, RMSE, MAE) for hydrogeochemical assessment of groundwater quality in the Maadher Plain.

Model	R^2	RMSE	MAE
W_{i_Expert}	0.78	4.21	3.15
$W_{i\_Expert\_ANN}$	0.92	2.15	1.63
W_{i_PCA}	0.80	3.98	2.94
$W_{i\_PCA\_ANN}$	**0.94**	**1.92**	**1.48**

Wi_Expert is weight of parameter i derived from expert knowledge, Wi_Expert_ANN is weight of parameter i derived from expert knowledge combined with Artificial Neural Network (ANN), Wi_PCA is weight of parameter i derived from Principal Component Analysis (PCA), Wi_PCA_ANN is weight of parameter i derived from (PCA) combined with Artificial Neural Network (ANN).

in semi-arid regions is controlled by the interaction between climate-related salinity and human-induced nutrient enrichment. The ANN framework also improved the prediction of WQI values and helped represent the combined hydrogeochemical processes affecting groundwater quality.

5. Methodological and management implications

The close agreement between the hybrid PCA–ANN results and findings reported in regional and international studies indicates that groundwater degradation in semi-arid aquifers is mainly associated with two interacting processes: nitrate enrichment linked to human activities and climate-related salinity increase. This suggests that agricultural pressure and hydroclimatic conditions together strongly influence groundwater chemistry. In this study, the ANN-enhanced framework improves conventional WQI methods while remaining consistent with established hydrogeochemical interpretations of groundwater evolution. The model combines PCA and ANN to use both statistical analysis and nonlinear prediction. This makes it easier to interpret and more

Table 8
Comparative analysis of the relative weights (Wi) assigned to groundwater quality parameters using PCA-, Expert-, and ANN-based weighting approaches, including statistical indicators for evaluating parameter importance and dominance in groundwater quality assessment.

Parameter	$W_{i\text{-PCA}}$	$W_{i\text{-PCA\_ANN}}$	$W_{i\text{-Expert}}$	$W_{i\text{-Expert\_ANN}}$	Mean ± SD	CV (%)	Range	Rank	Category
pH	0.092	0.089	0.040	0.086	0.077 ± 0.022	28.6	0.052	8	Minor
EC	0.106	0.106	0.110	0.086	0.102 ± 0.010	9.8	0.024	4	Moderate
TDS	0.106	0.092	0.130	0.105	0.108 ± 0.014	13.0	0.038	3	Moderate
Ca^{2+}	0.104	0.095	0.060	0.075	0.084 ± 0.018	21.4	0.044	6	Moderate
Mg^{2+}	0.076	0.117	0.060	0.118	0.093 ± 0.026	28.0	0.058	5	Moderate
HCO_3^-	0.067	0.054	0.150	0.064	0.084 ± 0.040	47.6	0.096	7	Moderate
Cl^-	0.107	0.090	0.130	0.077	0.101 ± 0.020	19.8	0.053	5	Moderate
SO_4^{2-}	0.106	0.100	0.150	0.088	0.111 ± 0.024	21.6	0.062	2	Dominant
NO_3^-	0.062	0.147	0.110	0.178	0.124 ± 0.044	35.5	0.116	1	Dominant
Na^+	0.089	0.043	0.040	0.060	0.058 ± 0.023	39.7	0.049	11	Minor
K^+	0.086	0.067	0.020	0.065	0.060 ± 0.025	41.7	0.066	10	Minor

Category criteria: Dominant (Mean weight >0.11); Moderate (Mean weight between 0.08 and 0.11); Minor (Mean weight <0.08). Wi_Expert is weight of parameter i derived from expert knowledge, Wi_Expert_ANN is weight of parameter i derived from expert knowledge combined with Artificial Neural Network (ANN), Wi_PCA is weight of parameter i derived from Principal Component Analysis (PCA), Wi_PCA_ANN is weight of parameter i derived from (PCA) combined with Artificial Neural Network (ANN).

reliable for assessing the quality of groundwater.

The proposed PCA–ANN framework is less reliant on expert judgment and established weighting schemes than conventional multi-criteria decision-making methods such as the Analytic Hierarchy Process (AHP), TOPSIS, and fuzzy inference systems. Instead, it learns from the structure of the hydrochemical dataset itself, which reduces methodological subjectivity. The PCA component identifies the fundamental hydrogeochemical processes that affect how groundwater changes, while the ANN component finds nonlinear correlations between physicochemical variables that linear statistical methods could miss. As a result, the model becomes more sensitive to contamination indicators, particularly nitrate pollution and salinity-related variables such as EC, TDS, SO_4^{2-}, and Cl^-.

From a management point of view, the hybrid WQI framework could help with groundwater monitoring and figuring out environmental risks in semi-arid regions where hydrochemical data and monitoring facilities are generally limited. The results indicate that nitrate contamination and salinity are the main factors affecting groundwater quality in the Maadher aquifer. This indicates that practical control measures need to be implemented, such as improved fertilizer use, stricter rules for agricultural return flows, reduced wastewater infiltration, and protection of key pumping zones. GIS-based spatial maps also help identify areas with higher contamination risk, which can support monitoring priorities and remediation planning.

Overall, the proposed framework provides a flexible decision-support tool for groundwater quality assessment under data-scarce conditions. Although this study focuses on a specific aquifer system, the methodological approach may also be applied to other semi-arid aquifers and combined with multi-temporal monitoring data in future studies. Some limitations should nevertheless be considered, especially the relatively small sample size and the lack of independent external validation datasets. Future work should include larger spatial and temporal datasets, seasonal monitoring programs, and cross-aquifer validation to further examine the robustness and long-term applicability of the hybrid PCA–ANN framework.

6. Conclusions

Groundwater quality deterioration is becoming a serious issue in arid and semi-arid regions. This study developed a hybrid framework that combines expert knowledge, principal component analysis (PCA), and artificial neural networks (ANN) to improve Water Quality Index (WQI) assessment in the semi-arid aquifer of the South Chott Hodna Basin. The approach combined hydrogeochemical interpretation with statistical and machine-learning methods, which reduced subjectivity in groundwater quality evaluation.

The four tested models (WQI_Expert, WQI_Expert_ANN, WQI_PCA, and WQI_PCA_ANN) showed clear differences in performance. PCA improved the statistical consistency of parameter weighting, whereas ANN better represented nonlinear relationships between hydrochemical variables. Among the tested approaches, WQI_PCA_ANN produced the best results ($R^2 = 0.94$, RMSE = 1.92, and MAE = 1.48). This indicates that combining PCA and ANN improves the reliability of WQI prediction.

Hydrogeochemical analysis showed that groundwater degradation is mainly related to salinity and nitrate contamination linked to human activities. Higher weights assigned to EC, TDS, Cl^-, SO_4^{2-}, and NO_3^- suggest strong effects from evaporite dissolution, irrigation return flow, agricultural activities, and wastewater infiltration. Spatial analysis also showed that poor to very poor groundwater quality is mainly concentrated in the northern and northeastern parts of the basin. This suggests that both natural conditions and land-use practices affect groundwater chemistry.

Overall, the proposed framework provides a useful tool for groundwater quality assessment in semi-arid aquifers. However, future studies should include larger datasets, external validation, and uncertainty analysis to improve the robustness and wider application of the approach.

Authors ethical responsibilities

All authors confirm that they have read and understood the ethical responsibilities specified in the Instructions for Authors of the journal and that they have adhered to these requirements during manuscript preparation.

Funding

This work is funded and supported by the Deanship of Graduate Studies and Scientific Research, Taif University.

CRediT authorship contribution statement

Mostafa Dougha: Conceptualization, Investigation, Methodology, Software, Visualization, Writing – original draft. **Tahar Selmane:** Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Ashraf A. Ahmed:** Data curation, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Ahmed A. Arafat:** Data curation, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Sherif S.M. Ghoneim:** Data curation, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Johnbosco C. Egbueri:** Conceptualization, Investigation, Methodology, Software, Visualization, Writing – original draft. **Lakhdar Seraiche:** Conceptualization, Investigation, Methodology, Software, Visualization, Writing – original draft. **Enas E. Hussein:** Data curation, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University for funding this work.

Data availability

Data will be made available on request.

References

Abdullahi, I.M., Abubakar, L., Saidu, G., Yusuf, H., 2025. Hydrogeochemical characterization of groundwater using water quality index and multivariate statistical analysis in Binji town and environs, Sokoto basin, Northwestern Nigeria. Appl. Water Sci. 15 (2), 33. https://doi.org/10.1007/s13201-025-02358-9.

Acal, C., Aguilera, A.M., Escabias, M., 2020. New modeling approaches based on varimax rotation of functional principal components. Mathematics 8 (11), 2085. https://doi.org/10.3390/math8112085.

Acquah, T., Appiah-Brempong, M., Anornu, G.K., 2025. Groundwater quality and associated health risks in the Eastern Region of Ghana. Heliyon 11 (2), e041910. https://doi.org/10.1016/j.heliyon.2025.e41910.

Ahmad Saleh, A.Y., Balakrishnan, P., Al-Ghouti, M.A., 2021. Comparison of GIS-based interpolation methods for mapping groundwater quality in the State of Qatar. Groundw. Sustain. Dev. 13, 100573. https://doi.org/10.1016/j.gsd.2021.100573.

Ahmed, F., El Houda, B.N., Mostafa, D., Tahar, S., Abdessamed, O.M., Ali, H., 2022. Spatio-temporal assessment of groundwater quality: a case study of M'sila province (Algeria). Arabian J. Geosci. 15 (24), 1775. https://doi.org/10.1007/s12517-022-11044-y.

Al-Zubaidi, H.A., Kadhim, E.J., Jasim, H.K., Naje, A.S., Al-Jassani, F.D., Chelliapan, S., 2025. Machine learning analysis for the Hilla River water quality index, Iraq. Glob. Nest J. 27 (3). https://doi.org/10.30955/gnj.06980.
Aldrees, A., Khan, M., Taha, A.T.B., Ali, M., 2024. Evaluation of water quality indexes with novel machine learning and SHapley Additive ExPlanation (SHAP) approaches. J. Water Process Eng. 58, 104789. https://doi.org/10.1016/j.jwpe.2024.104789.
Alrowais, R., Abdel Daiem, M.M., Li, R., Maklad, M.A., Helmi, A.M., Nasef, B.M., Said, N., 2023. Groundwater quality assessment for drinking and irrigation purposes at Al-Jouf area in KSA using artificial neural network, GIS, and multivariate statistical techniques. Water 15 (16), 2982. https://doi.org/10.3390/w15162982.
Appelo, C.A.J., Postma, D., 2005. Geochemistry, Groundwater and Pollution, second ed. A.A. Balkema Publishers.
Asmoay, A.A., Shams, E.M., Galal, W.F., et al., 2025. Geochemical characterization and health risk assessment of groundwater in Wadi Ranyah, Saudi Arabia, using statistical and GIS-based models. Environ. Geochem. Health 47, 208. https://doi.org/10.1007/s10653-025-02517-6.
Auwal, A.M., Obidiegwu, M.U., Kwaya, M.Y., Ben-Owope, O.A., Korie, J.I., Muhammad, A.Y., et al., 2026. Integrated hydrogeochemical and multivariate statistical assessment of groundwater in Gashua, Northeastern Nigeria: implications for arid zone aquifer management. Model. Earth Syst. Environ. https://doi.org/10.1007/s40808-026-02727-6.
Bahrami, M., Khaksar, E., Bahrami, A., 2022. Groundwater quality evaluation for potable and irrigation uses in the semi-arid region of southern Iran. Irrig. Drain. 71 (1), 143–158. https://doi.org/10.1002/ird.2671.
Bouderbala, A., Remini, B., Hamoudi, A.S., Pulido-Bosch, A., 2016. Assessment of groundwater vulnerability and quality in coastal aquifers: a case study (Tipaza, North Algeria). Arabian J. Geosci. 9 (3), 181. https://doi.org/10.1007/s12517-015-2151-6.
Bradai, A., Yahiaoui, I., Douaoui, A., Abdennour, M.A., Gulakhmadov, A., Chen, X., 2022. Combined modeling of multivariate analysis and geostatistics in assessing groundwater irrigation sustenance in the Middle Cheliff Plain (North Africa). Water 14 (6), 924. https://doi.org/10.3390/w14060924.
Chakraborty, T.K., Islam, M.S., Ghosh, G.C., Ghosh, P., Zaman, S., Habib, A., et al., 2023. Human health risk and hydro-geochemical appraisal of groundwater in the southwest part of Bangladesh using GIS, water quality indices, and multivariate statistical approaches. Toxin Rev. 42 (1), 285–299. https://doi.org/10.1080/15569543.2022.2134572.
Clemente, M.H., Amwele, H.R., Fohrer, N., Smit, E., Jacobs, F., Muacahila, A.N.S., Ajayi, O.G., 2026. Suitability assessment of water for human consumption using water quality index along the Cubango–Kavango River, in Angola and Namibia. Phys. Chem. Earth, 104320. https://doi.org/10.1016/j.pce.2026.104320.
De Jesús Rubio, J., 2020. Stability analysis of the modified Levenberg–Marquardt algorithm for the artificial neural network training. IEEE Transact. Neural Networks Learn. Syst. 32 (8), 3510–3524. https://doi.org/10.1109/TNNLS.2020.3015200.
Dougha, M., Hasbaia, M., Girou, A., Redjem, A., 2018. Numerical simulation of hydrodynamics in swimming pools. Euro-Mediterr. J. Environ. Integr. 3 (1), 33. https://doi.org/10.1007/s41207-018-0076-7.
Dougha, M., Hasbaia, M., 2019. Contribution of multivariate analysis to understanding groundwater quality of mixed aquifer in the M'sila plain (Algeria). Int. J. Hydrol. Sci. Technol. 9 (2), 154–172. https://doi.org/10.1504/IJHST.2019.098160.
D'souza, R.N., Huang, P.Y., Yeh, F.C., 2020. Structural analysis and optimization of convolutional neural networks with a small sample size. Sci. Rep. 10, 834. https://doi.org/10.1038/s41598-020-57866-2.
Egbueri, J.C., Agbasi, J.C., 2022. Performances of MLR, RBF-NN, and MLP-NN in the evaluation and prediction of water resources quality for irrigation purposes under two modeling scenarios. Geocarto Int. 37 (26), 14399–14431. https://doi.org/10.1080/10106049.2022.2087758.
Fatah, K.K., Hamed, M., Saeed, M.H., Dara, R., 2020. Evaluation of groundwater quality using GIS and water quality index in Bardarash area, Northern Iraq. Iraqi Geol. J. 53 (2c), 87–104. https://doi.org/10.46717/igj.53.2c.7Rs-2020-09.07.
Gholami, V., Sahour, H., 2024. A hybrid approach of supervised self-organizing maps and genetic algorithms for predictive mapping of arsenic pollution in groundwater resources. Expo. Health 16, 775–790. https://doi.org/10.1007/s12403-023-00592-7.
Guenouche, F.Z., Mesbahi-Salhi, A., Zegait, R., Chouia, S., Kimour, M.T., Bouslama, Z., 2024. Assessing water quality in North-East Algeria using WQI and PCA. Water Pract. Technol. 19 (4), 1232–1248. https://doi.org/10.2166/wpt.2024.073.
Guiraud, R., 1973. Hydrogeology of the Hodna Chott Basin, Northern Algeria. Publ. Serv. Géol. 39, 159–170.
Hamma, B., Alodah, A., Bouaicha, F., Bekkouche, M.F., Barkat, A., Hussein, E.E., 2024. Hydrochemical assessment of groundwater using multivariate statistical methods and water quality indices. Appl. Water Sci. 14 (2), 33. https://doi.org/10.1007/s13201-023-01999-8.
Hassan, M.M., Hassan, M.M., Akter, L., et al., 2021. Efficient prediction of water quality index (WQI) using machine learning algorithms. Hum. Cent. Intell. Syst. 1, 86–97. https://doi.org/10.2991/hcis.k.211203.001.
Ibrahim, H., Yaseen, Z.M., Scholz, M., Ali, M., Gad, M., Elsayed, S., Khalifa, M.M., 2023. Evaluation and prediction of groundwater quality for irrigation using water quality indices, machine learning and GIS. Water 15 (4), 694. https://doi.org/10.3390/w15040694.
Ismail, R., Rawashdeh, A., Al-Mattarneh, H., Hatamleh, R., Telfah, D.B., Jaradat, A., 2024. ANN-based determination of river water quality index. Civ. Eng. J. 10 (7), 2261–2274. https://doi.org/10.28991/CEJ-2024-010-07-012.
Jung, K., Bae, D.-H., Um, M.-J., Kim, S., Jeon, S., Park, D., 2020. Evaluation of nitrate load estimations using neural networks and canonical correlation analysis with K-fold cross-validation. Sustainability 12 (1), 400. https://doi.org/10.3390/su12010400.
Karimi-Rizvandi, S., Goodarzi, H.V., Afkoueieh, J.H., Chung, I.-M., Kisi, O., Kim, S., Linh, N.T.T., 2021. Groundwater-potential mapping using a self-learning Bayesian network model: a comparison among metaheuristic algorithms. Water 13 (5), 658. https://doi.org/10.3390/w13050658.
Khoi, D.N., Quan, N.T., Linh, D.Q., Nhi, P.T.T., Thuy, N.T.D., 2022. Predicting water quality index in the La Buong River using machine learning. Water 14 (10), 1552. https://doi.org/10.3390/w14101552.
Kulisz, M., Kujawska, J., Przysucha, B., Cel, W., 2021. Forecasting groundwater WQI using artificial neural networks. Energies 14 (18), 5875. https://doi.org/10.3390/en14185875.
Kumar, M., et al., 2022. Geospatial water quality index modeling in arid aquifers using kriging versus inverse distance weighting. J. Hydrol., 128112 https://doi.org/10.1016/j.jhydrol.2022.128112.
Lakhdar, S., Messaoud, G., Tahar, S., Mostafa, D., 2026. Hydrochemical assessment and groundwater pollution in the southern Hodna Chott region (Algeria). Bull. Serb. Geogr. Soc. 106 (1), 21–36.
Mahamadou, K., Adama, T., Wendpagnagda, C., Eulodie, I.E., Ibrahim, B.O.Y.L., Hassan, A.M., 2024. Groundwater quality assessment and modeling in Bam Province, Burkina Faso. Int. J. Environ. Clim. Change 14 (12), 740–756. https://doi.org/10.9734/ijecc/2024/v14i124659.
Maliki, A., Radouane, E.M., Chahlaoui, A., Mourad, D.S.J., Ouballouk, Y., Abderrazzaq, B., Khaffou, M., Taouraout, A., 2024. Groundwater quality in the Zagora region using GIS and WQI. Geol. Ecol. Landsc. https://doi.org/10.1080/24749508.2024.2438402.
Marfo, J.N., Quaye-Ballard, J.A., Kwakye, S.O., Obeng, K., Arko-Adjei, A., Quaye-Ballard, N.L., 2024. Groundwater quality assessment in the Ashanti Region, Ghana. Heliyon 10 (6), e27545. https://doi.org/10.1016/j.heliyon.2024.e27545.
Mengistu, T.D., Chung, I.M., Chang, S.W., 2026. Machine learning for water quality prediction and uncertainty assessment. Phys. Chem. Earth, 104319. https://doi.org/10.1016/j.pce.2026.104319.
Nayak, J.G., Patil, L.G., Patki, V.K., 2023. Artificial neural network based water quality index (WQI) for river Godavari (India). Mater. Today Proc. 81, 212–220. https://doi.org/10.1016/j.matpr.2021.03.100.
Ngwenya, B., Paepae, T., Bokoro, P.N., 2025. Monitoring ambient water quality using machine learning and IoT: a review and recommendations for advancing SDG indicator 6.3.2. J. Water Process Eng. 73, 107664. https://doi.org/10.1016/j.jwpe.2025.107664.
Nigam, N., Indu, S., 2025. Groundwater quality assessment around industrial area of Gautam Budh Nagar, India. Int. J. Sci. Res. Eng. Trends. https://doi.org/10.22214/ijraset.2025.66578.
Nishat, M.H., Khan, M.H.R.B., Ahmed, T., et al., 2025. Comparative analysis of machine learning models for predicting water quality index in Dhaka's rivers of Bangladesh. Environ. Sci. Eur. 37, 31. https://doi.org/10.1186/s12302-025-01078-w.
Olden, J.D., Jackson, D.A., 2002. Illuminating the "black box": variable contributions in artificial neural networks. Ecol. Modell. 154 (1–2), 135–150. https://doi.org/10.1016/S0304-3800(02)00064-9.
Redjem, A., Benyahia, A.Z., Dougha, M., Nouibat, B., Hasbaia, M., Ozer, A., 2021. Combining AHP with GIS for landfill site selection. Rev. Roum. Géogr 65 (2).
Salim, D., Doudja, S.G., Ahmed, F., Omar, D., Mostafa, D., Oussama, B., Mahmoud, H., 2023. Comparative study of different discrete wavelet based neural network models for long term drought forecasting. Water Resour. Manag. 37 (3), 1401–1420. https://doi.org/10.1007/s11269-023-03432-0.
Salima, D., Belgacem, H., 2017. Hydrochemical characterization of the upper plains aquifers in F'kirina Ain-Beïda, Northeastern Algeria. Desalination Water Treat. 92, 90–97. https://doi.org/10.5004/dwt.2017.21507.
Sekarlangit, S., Widodo, C.E., Tarno, T., 2024. Artificial neural network versus long short-term memory for river water quality prediction. Int. J. Curr. Sci. Res. Rev. 7 (4), 2000–2005. https://doi.org/10.47191/ijcsrr/v7-i4-02.
Selmane, T., Dougha, M., Djerbouai, S., Djemiat, D., Lemouari, N., 2023. Groundwater quality evaluation using water quality index and GIS in Maadher plain, Algeria. Environ. Sci. Pollut. Res. 30 (11), 30087–30106. https://doi.org/10.1007/s11356-022-24338-1.
Selmane, T., Dougha, M., Hasbaia, M., Ferhati, A., Redjem, A., 2022. Hydrogeochemical processes and multivariate analysis of groundwater in arid Maadher region, Algeria. Acta Geochim. 41 (5), 893–909. https://doi.org/10.1007/s11631-022-00553-y.
Sen, P., Saxena, A., Mehta, P., Mehta, R., 2024. Analysis of groundwater quality in semi-arid region of Eastern-South Rajasthan on the basis of water quality index (WQI). J. Geomatics 18 (2), 15–23. https://doi.org/10.58825/jog.2024.18.2.112.
Seraiche, L., Dougha, M., Ghodbane, M., Selmane, T., Ferhati, A., Djemiat, D.A., 2025a. Groundwater vulnerability assessment using GIS-based DRASTIC and fuzzy AHP. Desalination Water Treat., 101380 https://doi.org/10.1016/j.dwt.2025.101380.
Seraiche, L., Ghodbane, M., Selmane, T., Dougha, M., 2025b. Advanced groundwater quality assessment using geochemical analysis and water quality indices with fuzzy logic. J. Eng. Sustain. Dev. 29 (6), 700–711. https://doi.org/10.31272/jeasd.3013.
Shaibur, M.R., 2022. Salinity levels in pond, deep tube well, and pond sand filter water in two unions of southwestern coastal district Satkhira, Bangladesh. In: Naddeo, V., Choo, K.H., Ksibi, M. (Eds.), Water-Energy-Nexus in the Ecological Transition. Adv. Sci. Technol. Innov. Springer, Cham, pp. 193–197. https://doi.org/10.1007/978-3-031-00808-5_45.
Shaibur, M.R., Ahmmed, I., Rumpa, S.S., Parvin, S., Hossain, M.S., Rahman, M.M., Sarwar, S., 2019a. Physico-chemical parameters of groundwater at Jashore University of Science and Technology campus and its surrounding villages. J. Jessore Univ. Sci. Technol. 4 (1), 34–45.

Shaibur, M.R., Hossain, M.S., Sony, S.J., 2019b. Drinking water quality of hand tube well water at sub-urban areas of Jashore Municipality, Bangladesh. J. Jessore Univ. Sci. Technol. 4 (1), 11–22.
Shaibur, M.R., Parvin, S., Ahmmed, I., Rahaman, M.H., Das, T.K., Sarwar, S., 2021. Gradients of salinity in water sources of Batiaghata, Dacope and Koyra Upazila of coastal Khulna district, Bangladesh. Environ. Chall. 4, 100152. https://doi.org/10.1016/j.envc.2021.100152.
Singhal, A., Gupta, R., Singh, A., Shrinivas, A., 2020. Groundwater quality monitoring in a semi-arid region. Groundw. Sustain. Dev. 11, 100381. https://doi.org/10.1016/j.gsd.2020.100381.
Taşan, S., 2023. Estimation of groundwater quality using water quality index, artificial intelligence and GIS in Turkey. Appl. Water Sci. 13 (1), 15.
Tripathi, M., Singal, S.K., 2019. PCA-based parameter selection for water quality index: river Ganga, India. Ecol. Indic. 96, 430–436. https://doi.org/10.1016/j.ecolind.2018.09.025.
World Health Organization, 2017. Guidelines for drinking-water Quality, fourth ed. World Health Organization, Geneva.
Zhang, B., Hu, X., Li, B., Wu, P., Cai, X., Luo, Y., Jiang, M., 2024. A groundwater quality assessment model for WQI using PCA, entropy weight and CV methods. Water Environ. Res. 96 (12), e11155. https://doi.org/10.1002/wer.11155.
Zhang, Q., Xu, P., Qian, H., 2020. Groundwater quality assessment using improved water quality index (WQI) and human health risk (HHR) evaluation in a semi-arid region of Northwest China. Expo. Health 12, 487–500. https://doi.org/10.1007/s12403-020-00345-w.