Discover Quantum Science

RESEARCH Open Access

Hybrid quantum-classical machine learning with SABHQ for groundwater quality assessment

Jagadish Kumar Mogaraju[1*]

*Correspondence:
Jagadish Kumar Mogaraju
jagadishmogaraju@gmail.com
[1]IUCN, Climate Crisis Commission, New Delhi 110011, India

Abstract

Quantum Machine Learning is a growing field with the potential to unlock several opportunities in groundwater management. A novel approach integrating advanced quantum machine learning (QML) models with classical machine learning (ML) algorithms for robust groundwater quality assessment, using the SABHQ (Systematically Averaged Battery Hybrid Quantum) method, is proposed. We attempted to reduce noise and enhance generalization by stimulating composite sampling across both ML and QML pipelines. Accuracy, precision, recall, and F_1 scores were calculated across all models, using normalization and aggregation strategies to achieve relatively unbiased benchmarking. We observed that the dominant predictors, such as EC, Cl, HCO_3, SO_4, Na, TH, Latitude, and Longitude, reflect both spatial variability and anthropogenic influences on groundwater quality. The systematic averaging method yielded noticeable improvements for QML models, with QCNN achieving accuracy and F_1 scores of up to 0.95 and 0.98, respectively, outperforming classical machine learning models. We report that QCNN metrics improved due to aggregation, noise reduction, and possibly quantum entanglement effects, leading to better generalization, whereas VQC showed gradual but minor improvements. We believe our work can be extended to other environmental domains, not only to refine the method we proposed but also to enhance its adaptability.

Keywords Quantum machine learning, Classical machine learning models, QCNN, VQC, Groundwater quality

1 Introduction

Groundwater has been fetching attention in unexpected ways since the last few decades because of the sophisticated monitoring frameworks that revealed anthropogenic pressure on this precious source. A quantum-inspired recurrent neural network (QRNN) was developed and integrated with the convolutional neural network (CNN) and gradient-based optimizer (GBO) to enhance the capability of ML in predicting groundwater quality [1]. The Quantum Long Short-Term Memory (QLSTM) performed better than the classical LSTM models in predicting the groundwater levels [2].

A Quantum-Enhanced Gradient Boosting Neural Network (QGBNN) was used in the investigation of vital factors that can mitigate perfluorooctane sulfonic acid (PFOS) from aqueous solutions using nanofiltration membranes [3]. The quantum-behaved particle swarm optimization (QPSO) algorithm integrated with Levy flight (QPSO-LF) was used to improve the groundwater monitoring network [4]. The Quantum Particle Swarm Optimization (QPSO) and the Credal Decision Tree (CDT) were integrated to map the spring potential of groundwater [5]. Some subsurface hydrological inverse problems were solved using an improved quantum annealing performance [6]. Quantum Convolutional Neural Network (QCNN) was employed using the IBM Qiskit simulator and PennyLane simulator to predict the geothermal pumps-induced temperature variations in groundwater [6]. Bayesian artificial neural network (BANN), Bayesian support vector machine (BSVM), and Bayesian random forest (BRF) housed hybrid models were linked with quantum ML and deep learning models [7]. A quantum solver (quantum linear systems algorithms) for fracture systems was used to study the intricate hydrological linear systems [8]. The Adiabatic, fault-tolerant approach through the SSO algorithm (error-free) and the Variational linear solver (noise resilient) approach were used to accurately model vital fracture systems to simulate subsurface flow [9]. A Multi-objective quantum genetic algorithm combined with quantum state interference and coding of qubits was used to solve intricate problems in the management of agricultural water resources, i.e., utilization rate, deviation between area and yield, optimization effect, and convergence speed [10]. Absolute Quantum Gravimeter (AQG) and Differential Quantum Gravimeter (DQG) (FIQUgS project) were planned to precisely monitor groundwater and to detect subsurface voids, highlighting the potential of quantum gravity sensors [11]. Quantum Neural Networks (QNN) via hybrid quantum-classical training were used to assess the dissolution volume of carbon dioxide in saline aquifers through quantum circuit ansatz and Pauli Z operator measurement [12]. GA and a PSO algorithm were associated with quantum computing to investigate the contaminant sources of groundwater [13]. Quantum Group Method of Data Handling with Harmony Search (Q-GMDH-HS) and Quantum Group Method of Data Handling with Grey Wolf Optimizer (Q-GMDH-GWO) were used to estimate saturated hydraulic conductivity in semi-arid zones with the aid of superposition, collapse, rotation gates, and entanglement [14]. The hybrid quantum-classical distributionally robust optimization (DRO) was used to optimize water use efficiency and energy utilization [15]. Quantum variational crop-coil entanglement encoding, Quantum-guided agri-topological dynamics mapping, and Quantum federated learning for distributed farm intelligence were used as an integrated framework for climate-resilient farming that may also advocate groundwater governance [16]. SSA-BP, PSO-ELM, LSTM, BPNN, ELM, and CNN-LSTM models were compared with quantum neural networks in forecasting the river discharge [17]. Anaerobic digestion (wastewater treatment) exhibited enhanced performance when operational parameters were adjusted with inputs from quantum circuit learning (QCL), with a variational quantum circuit equipped with a classical optimizer, which was later proposed to work on a noisy intermediate-scale quantum (NISQ) device [18]. A hybrid model linking a quantum long short-term memory network (QLSTM), a random forest regression (RFR), and a temporal convolutional network to forecast precipitation with relatively higher accuracy [19]. A hybrid QLSTM Model (QGAPHnet) using Quantum Genetic Algorithm (QGA) and Particle Swarm Optimisation (PSO) was used to estimate the soil moisture

[20]. A quantum-behaved particle swarm optimization with a variational mode decomposition was used to forecast runoff [21]. A hybridizing Quantum Particle Swarm Optimization (QPSO) and the Credal Decision Tree (CDT) ensemble (QPSO-CDTreeEns model) was used to predict the flash floods and yielded susceptibility mapping in some zones [22]. RF, BRT, and SVM models were tested to identify areas that serve as potable and irrigation indicators [23]. Sequential Gaussian simulation (SGS) was used to study nitrates, As, and F levels in groundwater [24].We attempted to fill the research gaps from the previous works. There were limited comparative, metric-normalized benchmarks between ML and quantum circuits to study groundwater contamination. Previous work lacks a systematic study of how aggregation/averaging strategies affect the expression of quantum advantage. There was limited integration of explainability (ML-SHAP) with quantum models to associate performance gains with interpretation. We attempted to fill these gaps by providing a unified ML-QML benchmarking pipeline with normalized metrics. We propose the SABHQ averaging cohorts as a regulation framework to investigate quantum advantage and attempted to integrate ML-SHAP-derived dominant predictors with composite scores.

1.1 Novel contributions

In our opinion, the following are the key contributions we considered to present through this work.

1. We formulated the SABHQ (Systematically Averaged Battery augmented Hybridized Quantum machine learning) framework to investigate quantum advantage under different averaging cohorts, especially for groundwater investigations.
2. We attempted to integrate ML classifiers with QML models via a unified evaluation pipeline incorporating composite scores, SHAP-based interpretability, and integrated performance indices.
3. We introduced composite score equations that account for predictive metrics and agreement with dominant hydrogeochemical predictors.

2 Methodology

The rationale for the proposed work is to relatively reduce noise in hydrogeochemical measurements and to investigate whether quantum circuits extract patterns effectively from aggregated datasets than classical ML models. The proposed SABHQ framework integrates averaging cohorts and datasets into a unified benchmarking battery. This can enable controlled gradient analysis of aggregation effects on ML vs. QML performance. We used the QCNN convolutional structure that mimics classical CNNs but leverages entanglement pooling for hierarchical feature extraction from spatial hydrochemical patterns (https://pennylane.ai/qml/glossary/qcnn). We used VQC, which is the hardwar e-efficient variational baseline that can test architecture-independent quantum effects (https://pennylane.ai/qml/demos/tutorial_variational_classifier). The dataset needed for this study was obtained from the Central Ground Water Board, Ministry of Jal Shakti, Department of Water Resources, River Development and Ganga Rejuvenation, Govt. of India, through the https://cgwb.gov.in/en/ground-water-quality website. From the datas et (2023) procured from the source, we considered 16,776 observations with Longitude, Latitude, pH, EC, CO_3, HCO_3, Cl, F, SO_4, PO_4, Total Hardness, Ca, Mg, Na, K, Fe, As, U, and NO_3 (two levels: 'H' and 'L'). We considered 18 predictors and one target (NO_3)

for this study. We used an 80 (training data): 20 (test data) approach for this study as suggested in the previous studies [25–27]. We used CatBoost Classifier (catboost), Light Gradient Boosting Machine (lightgbm), Extreme Gradient Boosting (xgboost), SVM (svm), Logistic Regression (lr), Random Forest Classifier (rf), Extra Trees Classifier (et), Gradient Boosting Classifier (gbc), and Linear Discriminant Analysis (lda) models. We designed the methodology to integrate the machine learning (ML) models with quantum machine learning (QML) models, along with model interpretability (SHAP), to investigate groundwater contamination due to nitrate. We nurtured an approach to compare the ML models with variational and convolutional quantum circuits, with a data aggregation strategy (averaging) to understand if QML is actually generating a relative advantage over ML frameworks. Our workflow got initiated by preprocessing the dataset and standardizing it to feed into ML and QML frameworks (pipelines). The normalization of the dataset was done using the z-score method, and imputation was done using the mean (More information about the normalization and imputation can be obtained from the websites https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html and https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html). The imputation was done before the data split to avoid leakage. The categories were encoded using Labelencoder() that maps H/L → {0,1} → {+1,-1} for quantum ⟨Z⟩ matching. About 2% of outliers (removed) were detected due to extreme values of the signal from the dominating predictors. Averaging regimes (Systematically Averaged Battery) were constructed using subsets created by grouping samples, i.e., 3-average, 5-average, 7-average, and 9-average, and aggregating features within each group. We believe that this may stimulate composite sampling, allowing us to investigate quantum advantage with noise being reduced and potentially amplified class separation. We used a diverse ensemble of ML classifiers, i.e., Catboost, LightGBM, XGBoost, Random Forest, Extra Trees, Gradient Boosting Classifier, SVM (Support Vector Machine), Logistic Regression, and Linear Discriminant Analysis (more information on the models can be obtained from the https://scikit-learn.org/stable/supervised_learning.html website). We used evaluation metrics such as accuracy, precision, recall, and F1 score to compare models and determine the relatively better performing models. Model-specific feature importance was recorded, and the obtained metrics were averaged over multiple runs to maintain stability. We used quantum models, i.e., Variational Quantum Circuit (VQC) and Quantum Convolutional Neural Network (QCNN) (more information on the models can be obtained from the https://pennylane.ai/qml/demos/tutorial_variational_classifier/ and https://pennylane.ai/qml/glossary/qcnn websites). In order to confirm a relatively unbiased comparison, the model metrics are normalized to a 0 to 1 scale. We investigated the model agreements with the dominant variables while developing a composite score equation (CSE) to rank model performance and extracted the dominant predictors. We developed a framework combining Composite score (CS) and Dominant predictor agreement (DPA) to yield an Integrated score (IS). We combined SHAP-based agreement with integrated composite scores by extracting the SHAP value for the dominant predictors for each model and computing the dominant predictor agreement. We integrated SHAP values into the composite score equation (CSE), which resulted in the Integrated SHAP Composite Score Equation (ISCSE), which helped us comprehend how much the model predictions were influenced. We attempted to investigate whether systematic averaging consistently affects quantum ML performance metrics.

The dataset we considered has 18 continuous predictors and a target. The dataset was normalized using z-score scaling and fed into the circuit as a vector $x \in \mathbb{R}^{n_q}$, where $n_q = 4$ qubits. The encoding method involved Angle encoding via $R_Y(\theta)$ rotations. So, for each qubit $'i'$, $qml.RY(x[i], wires = i)$, and this maps each normalized feature value $x[i]$ to the Y- rotation angle $\theta = x[i]$ radians. Angle encoding can be efficient for continuous data, as it generally preserves feature magnitudes in the quantum state amplitudes. The proposed SABHQ averaging cohorts may further smooth inputs and aid encoding stability. The circuit implements a basic QCNN with two convolutional blocks on qubit pairs (0–1 and 2–3), followed by pooling that can mimic classical CNNs but leverages quantum entanglement. The initial encoding layer involves 4 parallel $R_Y(x[i])$ gates, one per qubit, while preparing the input state $|\psi\rangle = \otimes\_\{i = 0\} \wedge \{3\}\ R\_Y(x[i])|0\rangle$. In block 1 (qubits 0–1), the convolutional blocks involve $qml.CRY(weights, wires = [0,1])$. The controlled - RY gate involves $R_Y(\theta = weights)$ on target qubit 1, conditional on control qubit 0. This was followed by $qml.RY\left(weights[{}^{3}_{1}], wires = 0,\right.$ $qml.RY(weights, wires = 1)$. In block 2 (qubits 2–3), there was an identical structure, and CRY can enable data-dependent rotations. This can capture non-local correlations absent in classical convolutional layers, and the trainable RY layers may refine local features. More details can be obtained from https://pennylane.ai/qml/glossary/qcnn. The VQC (Variational Quantum Classifier) can serve as a baseline variational ansatz with a connected, layered architecture. It uses 4 qubits, i.e., $n_q = 4$ and n_{layers} that can be tuned for the ansatz depth. The z-score normalized 18 predictors reduced to $x \in R^4$ and encoding occurs within the circuit through angle embedding. The encoding method involves $R_Y(x[i])$ rotations on each qubit i and mapping features to the rotation angles $qml.RY(x[i], wires = i)\ \forall\, i \in \{0,1,2,3\}$. More details can be obtained from https://pennylane.ai/qml/demos/tutorial_variational_classifier. The methodology devised for this work is provided in Fig. 1.

3 Results

We used machine learning models and compared them using accuracy, recall, precision, and F_1 scores. The evaluation metrics for the ML models is provided in Table 1.

When we used the CatBoost classifier model, we observed that HCO_3, EC, and Cl were the dominant predictors (Fig. 2). Light Gradient Boosting Machine yielded HCO_3, Cl, and EC as dominant predictors (Fig. 3). The Extreme Gradient Boosting model identified EC as a dominant predictor (Fig. 4), whereas the SVM model identified longitude as a dominant predictor (Fig. 5). Using the Logistic regression model, we identified Cl and Na as the dominant predictors (Fig. 6), and the Random Forest classifier identified EC, Total hardness (TH), SO_4, and Cl as the dominant predictors (Fig. 7). Using the Extra Trees classifier, we identified TH, EC, and Mg as the dominant predictors (Fig. 8). The Gradient Boosting classifier model yielded EC and TH as the dominant predictors (Fig. 9), whereas the Linear discriminant analysis model resulted in Cl and Na as the dominant predictors (Fig. 10). More details about the interpretation may be obtained from https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html#A-simple-beeswarm-summary-plot.

We observed that the features HCO_3, Cl, and EC were most influential across the models, while the LightGBM, CatBoost, and SVM models provided higher total importance values to features, compared to XGBoost, RF, ET, and GBC models. We measured

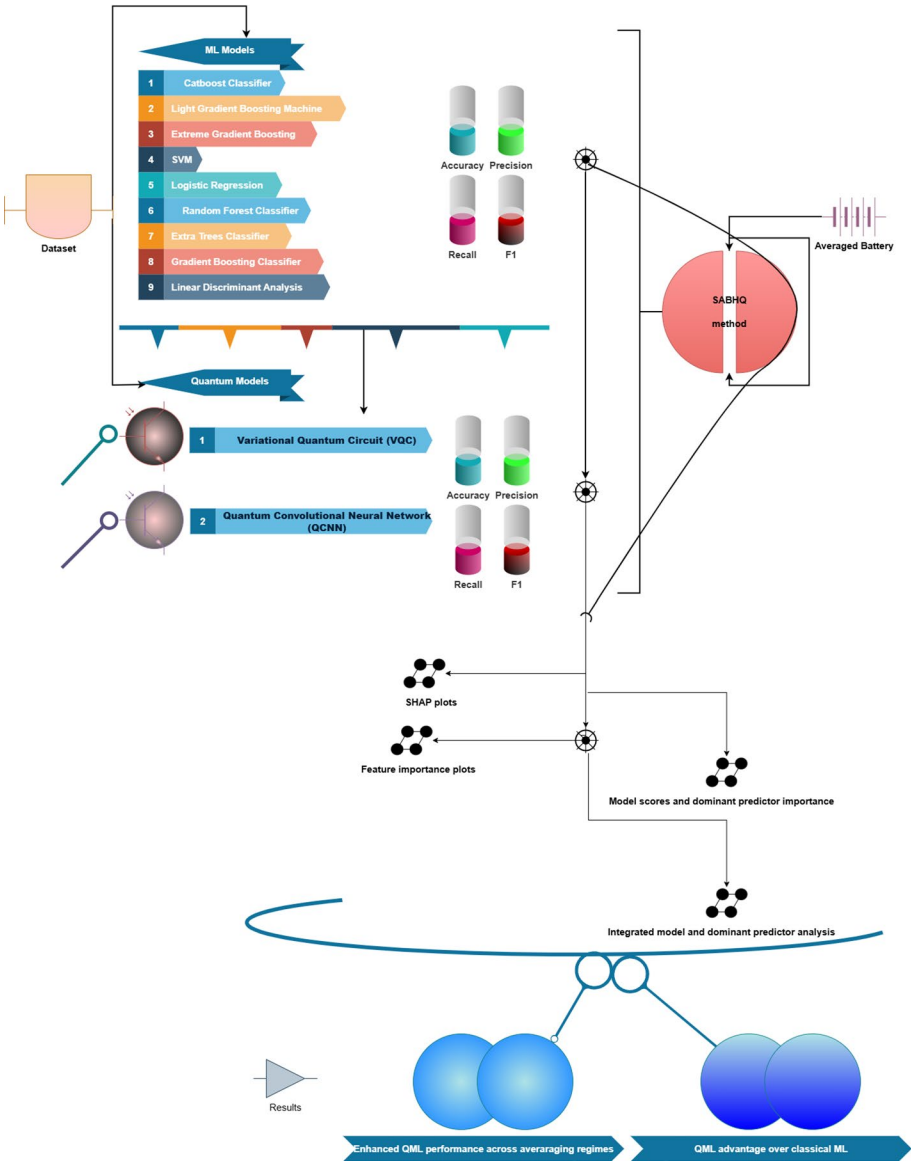

Fig. 1 Methodology

Table 1 Evaluation metrics (ML models)

	Model	Accuracy	Recall	Precision	F1
catboost	Catboost classifier	0.88	0.96	0.90	0.93
lightgbm	Light gradient boosting machine	0.88	0.96	0.90	0.93
xgboost	Extreme gradient boosting	0.88	0.95	0.91	0.93
svm	Support vector machine	0.88	0.96	0.89	0.93
lr	Logistic regression	0.87	0.97	0.88	0.92
rf	Random forest classifier	0.87	0.96	0.89	0.92
et	Extra trees classifier	0.87	0.96	0.89	0.92
gbc	Gradient boosting classifier	0.87	0.96	0.89	0.92
lda	Linear discriminant analysis	0.87	0.97	0.88	0.92

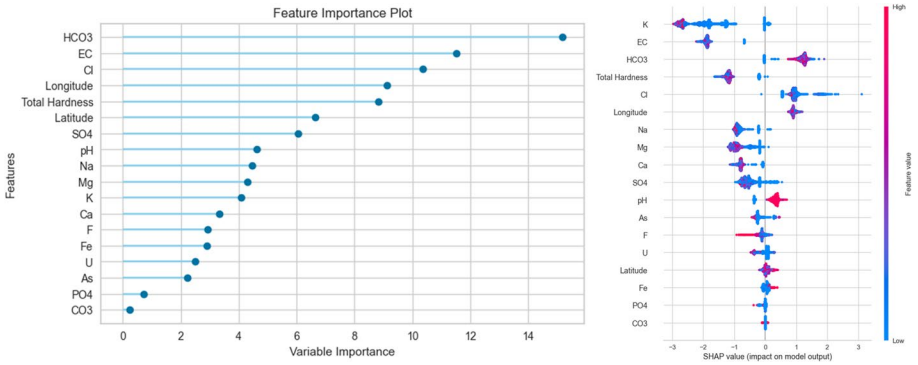

Fig. 2 CatBoost classifier (Feature importance and SHAP plots)

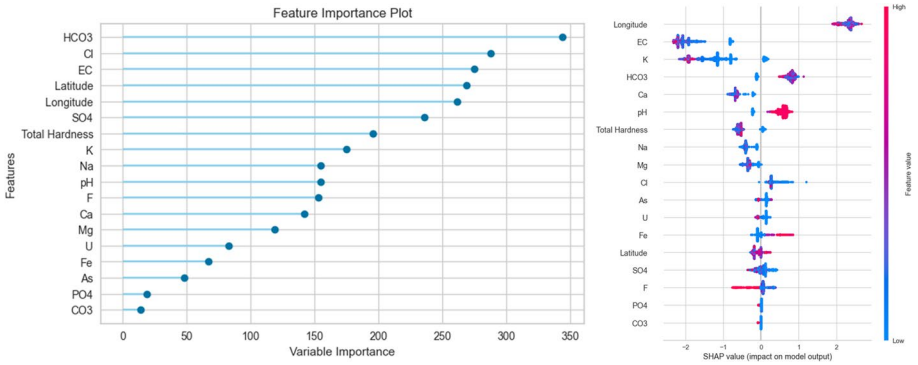

Fig. 3 Light gradient boosting machine (Feature importance and SHAP plots)

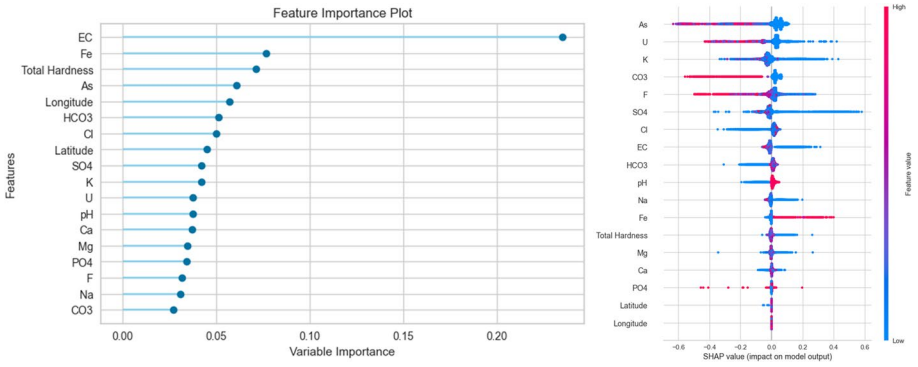

Fig. 4 Extreme gradient boosting (Feature importance and SHAP plots)

normalized variable importance and the normalized model importance to get the insights provided by ML (Table 2). The normalization (variable importance) was done using the equation

$$\text{Normalized Importance}_i = \frac{v_i}{S_v} \times 100$$

where, v_i is the importance of the variable i, and S_v is the total sum of mean importances for all variables.

The normalization (model importance) was done using the equation.

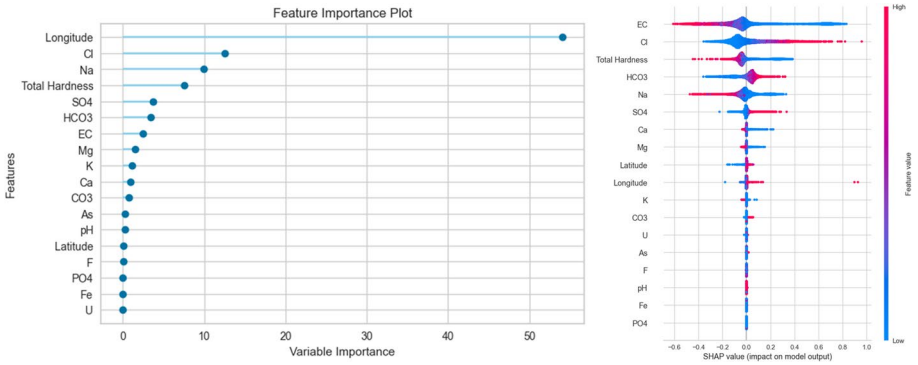

Fig. 5 SVM (Feature importance and SHAP plots)

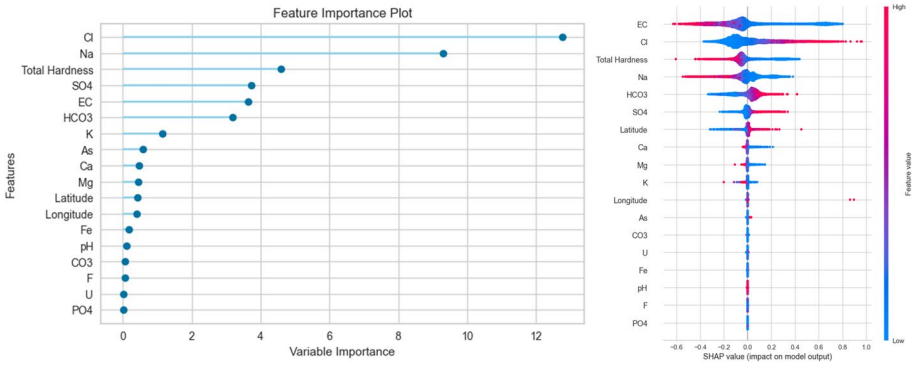

Fig. 6 Logistic regression (Feature importance and SHAP plots)

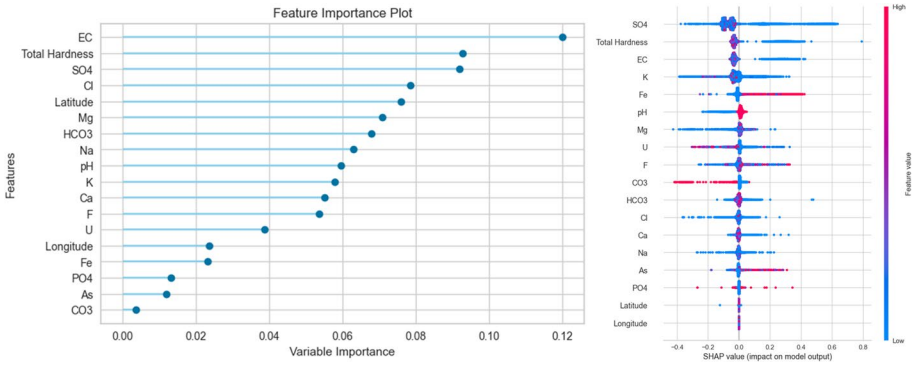

Fig. 7 Random forest classifier (Feature importance and SHAP plots)

$$\text{Normalized Importance}_k = \frac{m_k}{S_m} \times 100$$

where, m_k is the total importance assigned by the model and S_m is the total sum of importances for the models. SVM (87.98%) and LDA (87.29%) models show relatively better agreement with the dominant predictors (HCO_3, Cl, EC, Longitude, Latitude, SO_4, TH, and Na), while the LightGBM (64.64%) and ET (56.84%) models exhibited relatively lower agreement. Logistic regression (81.69%) and GBC (75.21%) models align with the key variables. The dominant variables, i.e., (HCO_3, Cl) - mineral dissolution and water-rock interactions, (EC) - salt concentration, (Longitude, Latitude) - geological and

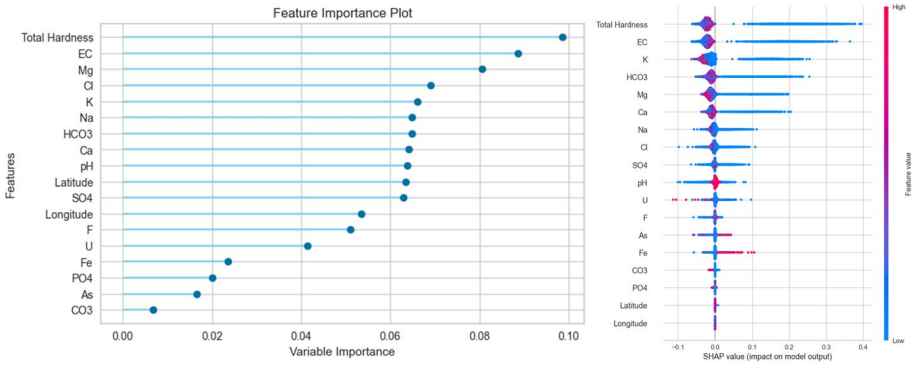

Fig. 8 Extra trees classifier (Feature importance and SHAP plots)

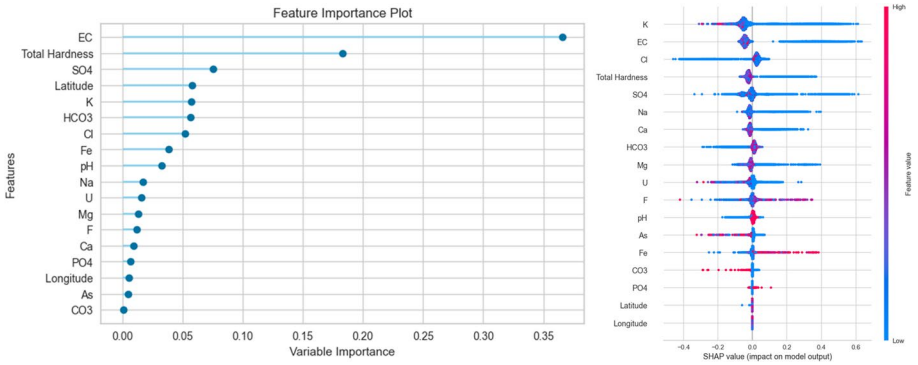

Fig. 9 Gradient boosting classifier (Feature importance and SHAP plots)

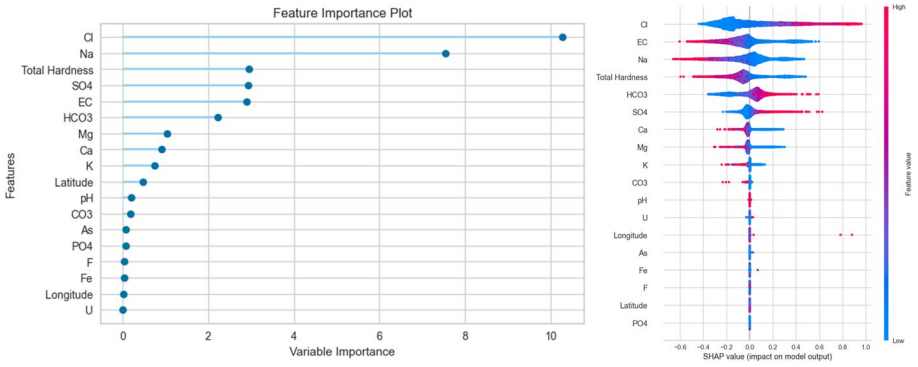

Fig. 10 Linear discriminant analysis (Feature importance and SHAP plots)

anthropogenic factors varying with spatial variation, SO_4, Total Hardness, and Na, may be treated as key indicators of groundwater quality, as they reflect natural geochemical and hydrochemical characteristics of groundwater. SVM and LDA models can provide reliable insights based on the established groundwater quality features. We considered Accuracy, Recall, Precision, and F1 scores to obtain the Composite Score Equation (CSE). Each metric was normalized using the equation.

$$X_{\text{norm}}(m_i) = \frac{X(m_i) - \min(X)}{\max(X) - \min(X)}$$

Table 2 Normalized variable and model importance

Variable	Total importance (%)	Model	Total importance (%)	Model	Agreement (%)
HCO3	13.44	LightGBM	89.84	XGBoost	68.87
Cl	13.32	CatBoost	3.30	LR	81.69
EC	12.14	SVM	3.27	GBC	75.21
Longitude	11.61	LR	1.33	LightGBM	64.64
Latitude	10.40	LDA	0.97	ET	56.84
SO4	9.34	GBC	0.03	CatBoost	71.00
Total Hardness	7.81	RF	0.03	LDA	87.29
Na	7.38	ET	0.03	RF	58.56
K	6.85	XGBoost	0.02	SVM	87.98
pH	6.22				
F	5.70				
Ca	5.43				
Mg	5.24				
U	4.51				
Fe	4.11				
As	2.98				
PO4	1.77				
CO3	1.21				

Table 3 Evaluation metrics with composite scores

Model	Accuracy	Recall	Prec.	F1	Composite score
XGBoost	0.8860	0.9550	0.9136	0.9339	0.5236
LightGBM	0.8862	0.9622	0.9083	0.9345	0.4816
CatBoost	0.8874	0.9647	0.9076	0.9353	0.4702
RF	0.8746	0.9649	0.8946	0.9284	0.4543
GBC	0.8726	0.9650	0.8926	0.9274	0.4524
ET	0.8735	0.9664	0.8924	0.9279	0.4473
SVM	0.8823	0.9691	0.8994	0.9328	0.4447
LDA	0.8709	0.9743	0.8844	0.9271	0.4185
LR	0.8759	0.9766	0.8876	0.9299	0.4131

$X(m_i)$ is the metric value for model m_i, max(X) and min(X) are the maximum and minimum values for the metric across the models. The composite score equation is as follows.

$$S(m_i) = \frac{1}{4}\left[\mathrm{Acc}_{\mathrm{norm}}(m_i) + \mathrm{Recall}_{\mathrm{norm}}(m_i) + \mathrm{Prec}_{\mathrm{norm}}(m_i) + \mathrm{F1}_{\mathrm{norm}}(m_i)\right]$$

$$S(m_i) = \frac{1}{4}\left[\begin{array}{l}\dfrac{\mathrm{Accuracy}(m_i) - \min(\mathrm{Accuracy})}{\max(\mathrm{Accuracy}) - \min(\mathrm{Accuracy})} + \dfrac{\mathrm{Recall}(m_i) - \min(\mathrm{Recall})}{\max(\mathrm{Recall}) - \min(\mathrm{Recall})} \\ + \dfrac{\mathrm{Prec}(m_i) - \min(\mathrm{Prec})}{\max(\mathrm{Prec}) - \min(\mathrm{Prec})} + \dfrac{\mathrm{F1}(m_i) - \min(\mathrm{F1})}{\max(\mathrm{F1}) - \min(\mathrm{F1})}\end{array}\right]$$

Note that each normalized metric gets equal weight, and the highest $S(m_i)$ specifies the best model. The XGBoost model performed better based on the composite scores (Table 3). We attempted to integrate dominant predictors with the composite model scores using the equation.

$$I_M = w_1 \cdot S_M + w_2 \cdot D_M$$

Table 4 Composite score, dominant predictor agreement, and integrated score

Model	Composite score	Dominant predictor agreement	Integrated score
LDA	0.4447	0.8729	0.6160
LightGBM	0.4816	0.8169	0.6157
LR	0.4131	0.8798	0.5998
XGBoost	0.5236	0.6887	0.5896
GBC	0.4702	0.7521	0.5830
ET	0.4473	0.7100	0.5524
CatBoost	0.4543	0.6464	0.5311
RF	0.4524	0.5684	0.4988
SVM	0.4185	0.5856	0.4853

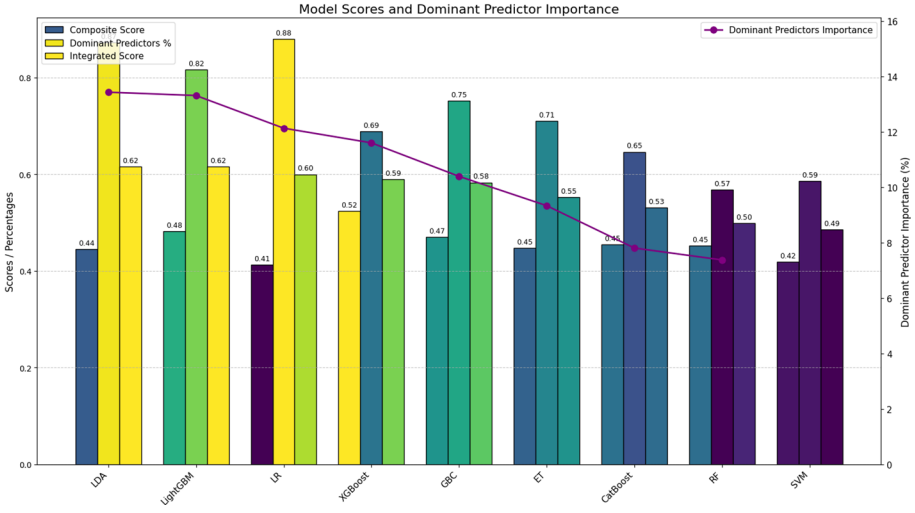

Fig. 11 Model scores with dominant predictor importance

w_1 and w_2 are the weights prioritizing predictive strength (0.6) and relevance (0.4), S_M is the composite model score, and D_M is the agreement percentage with the dominant predictors. Table 4 shows the composite score, dominant predictor agreement, and integrated scores. The integrated score balances model performance and alignment with key predictors. LDA and LightGBM achieved the highest integrated scores, combining both prediction and alignment to dominant attributes (Fig. 11). We considered SHAP values and observed that EC, Cl, TH, HCO3, Na, SO4, Latitude, and Longitude remained dominant predictors. We combined SHAP values and composite scores to yield better insights using the equation.

$$I(m_i) = w_1 \cdot S(m_i) + w_2 \cdot D_{SHAP}(m_i)$$

$S(m_i)$ is the normalized composite model score, $D_{SHAP}(m_i)$ is the dominant predictor agreement, and $w_1 \, and \, w_2$ are weights. The integrated scores are provided in Table 5. Based on the scores, XGBoost, LightGBM, CatBoost, GBC, RF, and ET models remained dominant models while EC, Cl, TH, HCO_3, Na, SO_4, Latitude, and Longitude emerged as dominant predictors (Fig. 12).

We expanded our work to investigate whether quantum ML models provide better metrics than ML models. VQC (Variational Quantum Classifier) and QCNN (Quantum Convolutional Neural Network) were considered in this study. SABHQ (Systematically Averaged Battery augmented Hybridized Quantum machine learning) method was

Table 5 Composite score, SHAP dominant agreement, and integrated score

Model	Composite score	SHAP dominant agreement	Integrated score
XGBoost	0.5236	0.70	0.589
LightGBM	0.4816	0.75	0.588
CatBoost	0.4543	0.72	0.559
RF	0.4524	0.67	0.539
ET	0.4473	0.68	0.541
LDA	0.4447	0.60	0.518
SVM	0.4185	0.61	0.507
LR	0.4131	0.65	0.510
GBC	0.4702	0.69	0.554

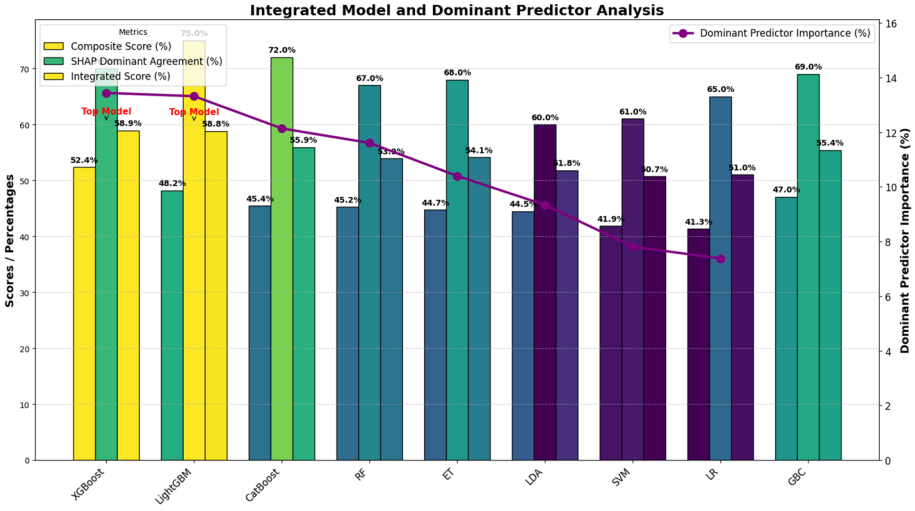

Fig. 12 Integrated model with dominant predictor analysis

Table 6 Quantum model metrics

Averaging	Model	Accuracy	Precision	Recall	F1
Total	QCNN	0.9088	0.9380	0.9666	0.9521
	VQC	0.8615	0.9470	0.9028	0.9244
By 3	QCNN	0.9449	0.9709	0.9724	0.9716
	VQC	0.8720	0.9700	0.8955	0.9313
By 5	QCNN	0.9313	0.9781	0.9511	0.9644
	VQC	0.8542	0.9831	0.8660	0.9208
By 7	QCNN	0.9517	0.9889	0.9621	0.9753
	VQC	0.9008	0.9911	0.9079	0.9477
By 9	QCNN	0.9517	0.9889	0.9621	0.9753
	VQC	0.9008	0.9911	0.9079	0.9477

developed. In the SABHQ method, we used systematically averaged cohorts (Battery) of averaged datasets (3, 5, 7, and 9), and the total dataset was also used for comparison (Table 6). With the QCNN model, the accuracy increases steadily from ~0.91 (all data) to ~0.95 (by 9), reflecting a consistent gain at each level. Precision increases from ~0.94 to ~0.99, suggesting the model is less likely to misclassify. Recall was maintained at high values, and the F1 score shifted from ~0.95 to ~0.98, reflecting a constant precision/recall balance. With the VQC model, the accuracy increases slowly, i.e., 0.86 to ~0.90, whereas the precision is stable with a slight increase, i.e., 0.95 to 0.99. Recall shifts from

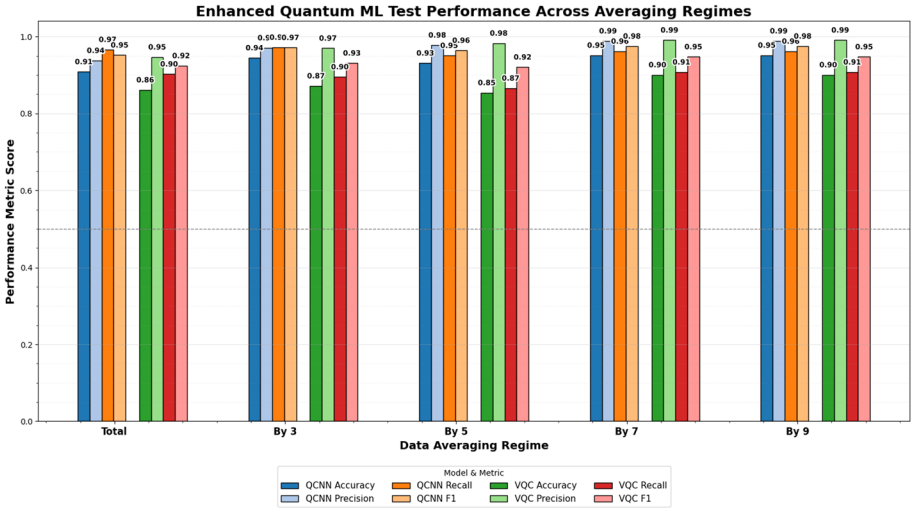

Fig. 13 Enhanced Quantum performance

Table 7 Deployment situations

Situation	Recommended model	Metric(s)	Averaging level	Reason(s)
Individual sample, no averaging	QCNN	Accuracy/F1	Total	Best overall; recall crucial for detection
Moderate noise, some batch	QCNN	Recall/F1	By 3/5	Robust to noise; high recall
Aggregate	QCNN or VQC	Precision	By 7/9	QCNN for reliability; VQC for stringent precision tasks

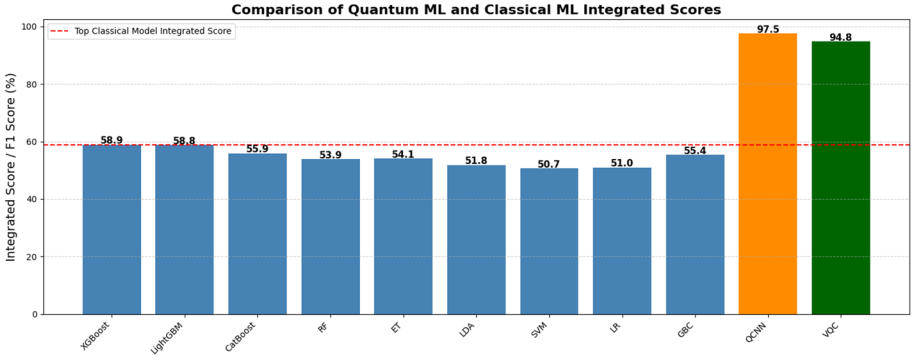

Fig. 14 Quantum ML and classical ML

~0.90 to ~0.91, reflecting less sensitivity to averaging than QCNN. F1 score slightly increased from ~0.92 to ~0.95. QCNN model metrics increase with higher averaging, peaking at 7/9. VQC model accuracy lags behind, but precision increases fastest while recall slowly ascends. QCNN provides F1 and recall advantage throughout the cohorts (Fig. 13). We observed that the metrics show reduced variance and higher means as averaging increases, especially for the QCNN model, which reflects that aggregation helps quantum circuits learn generalizable patterns (Table 7). Using moderate-to-large averaging (5–9) yields better accuracy, recall, and F1 for QCNN, signifying that this may be optimal for deployment (Fig. 14). The quantum ML models QCNN and VQC exceed the classical ML models (Fig. 14).

The evaluation metrics were compared (Fig. 15), and we observed that the QCNN accuracy starts high (~91%) and steadily improves to ~95% at higher averaging cohorts. VQC is lower (~86%), increasing to ~90%, reflecting that QCNN generalizes relatively better with the averaging method (SABHQ). In terms of accuracy, the classical ML best-reported integrated score was well below even the initial QCNN performance. The QCNN precision starts at ~94%, improving to ~99%, reflecting the ability to identify positive cases with a few false positive cases. VQC starts at ~95%, increases to ~99%, indicating that both quantum models improve in precision with the SABHQ method. This observation reinforced the quantum ML precision superiority. QCNN recall maintains high recall (~97%), with a dip at moderate averaging levels, but regains high recall at the highest averaging cohorts. VQC has a shifting recall (~90%) but improves slightly toward higher-averaging cohorts. The classical ML recall baseline is lower. The QCNN achieves F1 score above 95%, peaking around 98% with high averaging cohorts. VQC improves slowly from ~92% to ~95%. The classical ML baseline remains low, reflecting quantum ML superiority using the SABHQ method. We attempted to integrate QML and ML using the Composite Integrated Performance Score Equation (CIPSE).

$$IS\,(A) = w_{ML} \cdot\ S_{ML} + w_{QML} \cdot\ S_{QML}\,(A)$$

A {1,3,5,7,9} is the data averaging level (1 shows the 'Total' without averaging), S_{ML} is the classical ML integrated score, S_{QML} is the quantum ML model evaluation metric score at the averaging level A, and w_{ML} is the weighting factors reflecting priority.

The quantum advantage at the averaging level A can be calculated using

$$Q_{adv}\,(A) = S_{QML}\,(A)\ -\ S_{ML}$$

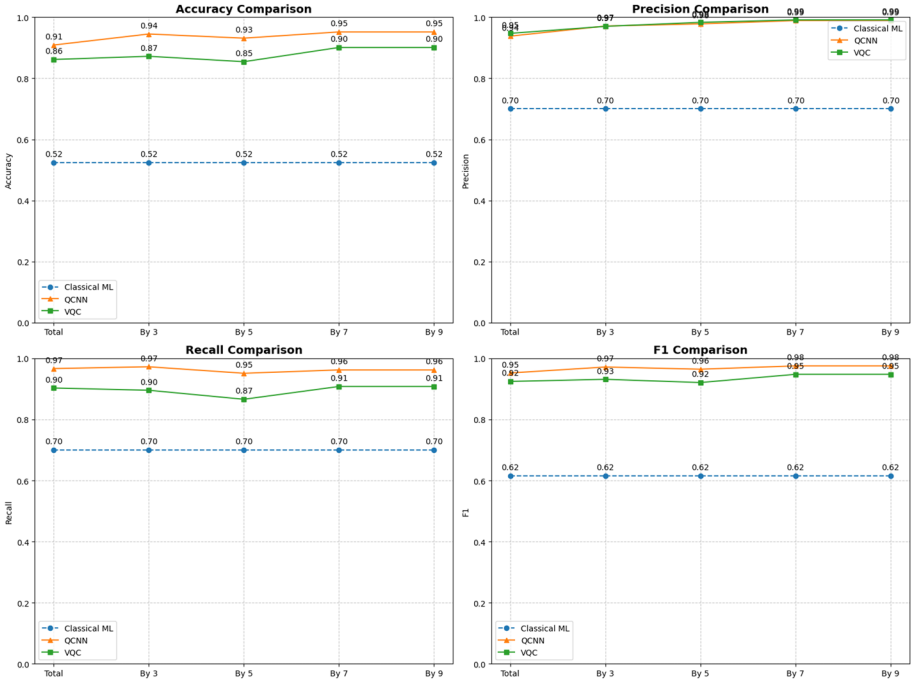

Fig. 15 Comparison of evaluation metrics (QML and Classical ML)

In case, $Q_{adv}(A)$ is greater than 0, then we can report that quantum ML outperforms classical at the averaging level A. If $Q_{adv}(A)$ rises with A, then it indicates that the SABHQ-treated dataset yields better QML advantage (Fig. 16).

4 Discussion

We systematically compared the XGBoost, LightGBM, CatBoost, SVM, and LDA with quantum models, i.e., Quantum Convolutional Neural Network (QCNN) and Variational Quantum Circuit(VQC), to predict nitrate contamination in groundwater. Our work resulted in a framework that harnesses metric normalization, SHAP-backed interpretability, and averaging regimes that enabled both performance benchmarking and task-driven interpretability. We observed that EC, Cl, HCO_3, SO_4, Na, TH, Latitude, and Longitude are dominant predictors, and our work aligns with [28], which highlighted the role of SHAP in interpretations. The quantum models exhibited advantages as sample averaging increased, with F1 scores rising at the highest aggregation levels, consistently exceeding the best classical models. We integrated normalized model scores and model interpretability, which can protect the prediction process from overfitting and black-box decision-making. Evaluation metrics (accuracy, recall, precision, F1) are augmented with SHAP-derived feature importance, yielding an integrated score that can rank models, which is not only for prediction capability but also throws light on validated predictors. SHAP analysis consistently supports the relevance of major geochemical features, making our work match the contemporary research works. We observed that averaging regimes (Battery) significantly affect quantum models in terms of boosted performance metrics through variance reduction and enhanced generalization. Hybrid classical-quantum methods are being considered for actual deployment, leveraging quantum ML algorithms for improved feature extraction and operational benefits, our work also supports the notion and may be used as one of the methods to use quantum ML for groundwater studies [29]. This empirical quantum advantage aligns with the theoretical expectation that quantum circuits surpass at extracting patterns from aggregated input regimes. We relied on a methodology that included stratified train-test splits and SHAP analysis, regulating overfitting and statistical leakage. We managed to ensure no overlap or data

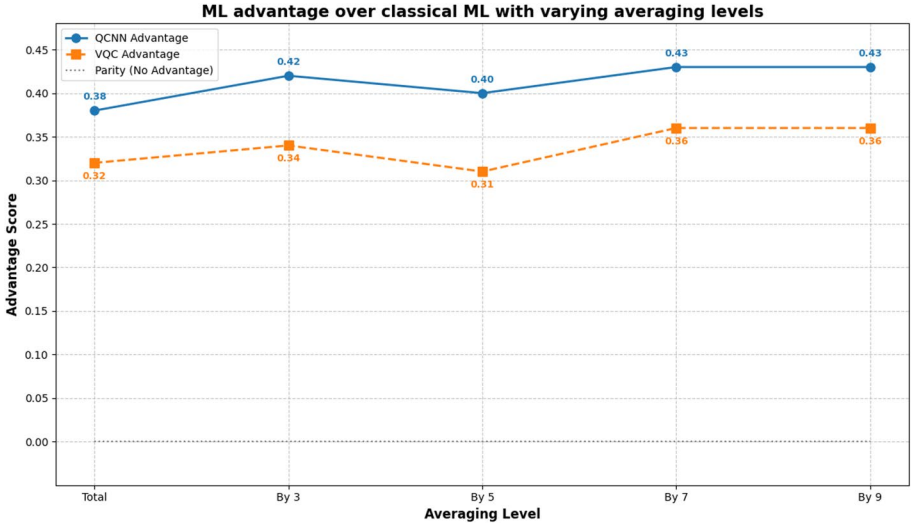

Fig. 16 QML advantage

Table 8 Leakage risk appraisal

Model type	Average train accuracy	Average test accuracy	Overfitting gap	Risk (leakage)
Classical ML (best)	~0.85–0.9	~0.8–0.85	Small to moderate	Low
Quantum ML (QCNN)	~0.94–0.96	~0.90–0.95	Small	Low
Quantum ML (VQC)	~0.87–0.88	~0.86–0.90	Small	Low

leakage between training and test sets using train-test splits and cross-validation. The trained classical ML models have balanced train/test metrics (with negligible to no overfit), so leakage is likely low. The Quantum ML exhibited robust generalization with test accuracies below train accuracies, which shows that there is no noticeable leakage. We observed stable or gradually improving metrics with averaging, which indicates noise reduction, but not leakage (Table 8). We believe that the effects of excessive averaging must be considered prudently, as over-aggregation may blur fine-grained hydrochemical variation critical for scientific assessment of groundwater quality. Developing explainable AI that integrates with QML could further increase reliability for deployment scenarios on a large scale. We observed the superiority of QCNN under optimal aggregation regimes (by 7 or by 9) that represent a significant leap in operational readiness for quantum ML in environmental monitoring. We believe that the proposed framework (ensemble ML, state-of-the-art QML, interpretability, and averaging battery) may be a new standard for scientific modeling pipelines, applicable well beyond groundwater [30, 31], and can extend to soil assessment, biodiversity, and spectral analysis.

In the case of qubit scalability, our study uses 4 - qubit circuits. The contemporary cloud quantum hardware supports this scale but has some inflections as to the barren plateaus in deeper ansätze and constrained connectivity, preventing complete pooling patterns. The proposed framework and the results we obtained may or may not yield the same results on actual quantum hardware, as there may be inclusion of gate errors and readout errors. We propose that the observed simulation advantages may yield promising directionality but may demand a 10–20% performance margin for hardware viability.

5 Conclusion

We initially worked towards finding whether classical machine learning models are being challenged by quantum machine learning models. Through this study, we developed a rigorous and interpretable framework for groundwater quality classification by integrating state-of-the-art classical machine learning (ML) models with advanced quantum machine learning (QML) models such as the Quantum Convolutional Neural Network (QCNN) and Variational Quantum Circuit (VQC). We employed a comprehensive multi-metric evaluation regime that included key evaluation metrics like accuracy, precision, recall, and F1-score, along with SHAP-based interpretability, and our methodology balances empirical predictive performance with scientific validity grounded in established hydrogeochemical information. We observed that the Quantum models, particularly QCNN, exhibited a significant performance advantage when the dataset was aggregated via averaging regimes (by 3, 5, 7, or 9). We find that the QCNN inherent architecture, leveraging quantum entanglement and non-classical correlations, permits noise suppression, variance reduction, and feature disentanglement that is critical/useful in highly noisy and aggregated environmental datasets. We observed that the VQC model also improved with averaging but lagged behind QCNN, indicating QCNN predominance for robustness and balanced performance in prediction tasks requiring

relatively high sensitivity. We noticed that the data averaging was not merely a preprocessing step but was an essential enabler of quantum advantage, which effectively amplified QML performance while preserving interpretability and environmental validity of predictions. We suggest that extreme aggregation risks over-smoothing, which may mask localized or transient groundwater anomalies vital for environmental monitoring. We operated on a balanced approach while working towards designing a framework, and our balanced framework supports informed model selection that may be tailored to specific needs. We are aware that the current QML simulations and devices constrain circuit depth and qubit count. We intend to further our work using quantum circuits and hybrid classical-quantum architectures (integrating PennyLane, Qiskit, Cirq) to comprehend complex hydrochemical interactions. We conclude that if advanced probabilistic quantum models or Bayesian uncertainty estimation are used along with our SABHQ framework, there will be enhanced risk assessment in contaminant detection and anomaly forecasting. We also propose a word of caution while using QML and ML frameworks through the following points. The sensitivity of a quantum circuit to input states should be considered (because noise, bias, or preprocessing errors risk can spread into model predictions). The cross-framework and model-type variations require caution due to the risk of overgeneralizing feature importance rankings. Actual quantum hardware is still emerging, and if available, noise, decoherence, and limited qubit counts restrict current QML deployment, making the simulation-derived results moderately reliable when compared to actual practical implementations. Integrating quantum algorithms into environmental monitoring systems involves heavy cost, infrastructure, and expertise, necessitating stringent measures before scale deployment.

Acknowledgements
The authors are thankful to the Central Groundwater Board (CGWB), Ministry of Jal Shakti, Department of Water Resources, River Development, and Ganga Rejuvenation, Government of India.

Author contributions
The corresponding author Jagadish Kumar Mogaraju has contributed to all sections of the manuscript.

Funding
The authors declare that no funding was received by anyone for this research work at any stage.

Data availability
The datasets used in this study can be obtained from the Central Groundwater Board (CGWB), Ministry of Jal Shakti, Department of Water Resources, River Development, and Ganga Rejuvenation, Government of India (https://cgwb.gov.in/en/ground-water-quality).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 15 November 2025 / Revised: 8 February 2026 / Accepted: 18 March 2026
Published online: 30 March 2026

References

1. Sharafi M, Gharehbaghi A, Mehdizadeh S. Development of explainable hybrid quantum-inspired recurrent neural networks for predicting groundwater quality: A case study at West Azerbaijan, Iran. J Water Process Eng. 2025;79:109013. https://doi.org/10.1016/j.jwpe.2025.109013.

2. Saha A, Rahman M, Wu F. (2024). Groundwater Level Prediction: Analyzing the Performance of LSTM and QLSTM Model. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 3755–3763). Presented at the 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA: IEEE. https://doi.org/10.1109/BigData62323.2024.10825269
3. Hu J, Xiao J, Yang W, Wang Y, Wang Y. Quantum-enhanced gradient boosting neural networks for optimizing PFOS removal with nanofiltration membranes. Desalination Water Treat. 2025;323:101266. https://doi.org/10.1016/j.dwt.2025.101266.
4. Gu Z, Bian J, Wang Y, Sun X, Ruan D, Li Y. Multiobjective Optimization of Groundwater Monitoring Network Based on Improved Quantum Particle Swarm Optimization Algorithm. J Water Resour Plan Manag. 2025;151(9):04025045. https://doi.org/10.1061/JWRMD5.WRENG-6867.
5. Msaddek MH, Ben Alaya M, Moumni Y, Ayari A, Chenini I. Enhanced machine learning model to estimate groundwater spring potential based on digital elevation model parameters. Geocarto Int. 2022;37(25):8815–41. https://doi.org/10.1080/10106049.2021.2007292.
6. Maheshwari D, Pelzer J, Schulte M. (2025). Predicting Heat Plume Temperature and Spatial Location Using Quantum Convolutional Neural Networks. In *2025 International Conference on Quantum Communications, Networking, and* Computing *(QCNC)* (pp. 623–627). Presented at the 2025 International Conference on Quantum Communications, Networking, and Computing (QCNC), Nara, Japan: IEEE. https://doi.org/10.1109/QCNC64685.2025.00103
7. Afrifa S, Zhang T, Zhao X, Appiahene P, Yaw MS. Climate change impact assessment on groundwater level changes: A study of hybrid model techniques. IET Signal Proc. 2023;17(6):e12227. https://doi.org/10.1049/sil2.12227.
8. Golden J, O'Malley D, Viswanathan H. Quantum computing and preconditioners for hydrological linear systems. Sci Rep. 2022;12(1):22285. https://doi.org/10.1038/s41598-022-25727-9.
9. Henderson JM, Podzorova M, Cerezo M, Golden JK, Gleyzer L, Viswanathan HS, O'Malley D. Quantum algorithms for geologic fracture networks. Sci Rep. 2023;13(1):2906. https://doi.org/10.1038/s41598-023-29643-4.
10. Yan Y. Comprehensive dispatch model of agricultural water resources based on multi-objective quantum genetic algorithm. Desalination Water Treat. 2021;239:192–201. https://doi.org/10.5004/dwt.2021.27807.
11. Sampietro D, Capponi M, Janvier C. (2025, March 15). FIQUgS Innovations in Quantum Gravity Sensing: Data Processing for an Archeological Case Study. https://doi.org/10.5194/egusphere-egu25-16898
12. Rao X, Luo C, He X, Hyung K. (2024). An Efficient Quantum Neural Network Model for Prediction of Carbon Dioxide CO2 Sequestration in Saline Aquifers. In *ADIPEC* (p. D021S061R005). Presented at the ADIPEC, Abu Dhabi, UAE: SPE. https://doi.org/10.2118/222257-MS
13. Zhao Y, Qu R, Xing Z, Lu W. Identifying groundwater contaminant sources based on a KELM surrogate model together with four heuristic optimization algorithms. Adv Water Resour. 2020;138:103540. https://doi.org/10.1016/j.advwatres.2020.103540.
14. Piri J, Kahkhamoghadam P, Chari MM. A comparison of quantum-inspired GMDH-HS and GMDH-GWO approaches to soil hydraulic conductivity prediction. J Hydrology: Reg Stud. 2025;61:102627. https://doi.org/10.1016/j.ejrh.2025.102627.
15. Qian L, Lin S, Muyeen SM, Zhang Z, Li D, Wang W. Integrated energy-water nexus optimization in rural microgrids: Leveraging quantum-classical robust optimization for sustainability. Prot Control Mod Power Syst. 2025;1–15. https://doi.org/10.23919/PCMP.2025.000039.
16. Khan AH, Saini DKJB, Khan TH, Rai BK, Pimpalkar A, Kumar G. A quantum-driven multi-stage framework integrating variational entanglement, reinforcement learning, and federated explainability for climate-resilient farming. Sci Rep. 2025;15(1):38363. https://doi.org/10.1038/s41598-025-22224-7.
17. Zhen L, Bărbulescu A. Quantum Neural Networks Approach for Water Discharge Forecast. Appl Sci. 2025;15(8):4119. https://doi.org/10.3390/app15084119.
18. Mohamed Y, Elghadban A, Lei HI, Shih AA, Lee P-H. Quantum machine learning regression optimisation for full-scale sewage sludge anaerobic digestion. npj Clean Water. 2025;8(1):17. https://doi.org/10.1038/s41545-025-00440-y.
19. Dong Y, Ding H. Bayesian optimization of hybrid quantum LSTM in a mixed model for precipitation forecasting. Mach Learning: Sci Technol. 2025;6(1):015065. https://doi.org/10.1088/2632-2153/adbbad.
20. Sai, S., Sen, A., Mallick, C., Mallik, A., Sen, U., Paul, M., … Roy, S. (2024). QGAPHnet:Quantum Genetic Algorithm Based Hybrid QLSTM Model for Soil Moisture Estimation. In*IGARSS 2024–2024 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5191–5194). Presented at the IGARSS 2024–2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece: IEEE. https://doi.org/10.1109/IGARSS53475.2024.10641651.
21. Feng Z, Niu W, Tang Z, Jiang Z, Xu Y, Liu Y, Zhang H. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. J Hydrol. 2020;583:124627. https://doi.org/10.1016/j.jhydrol.2020.124627.
22. Ngo, P.-T. T., Pham, T. D., Nhu, V.-H., Le, T. T., Tran, D. A., Phan, D. C., … Bui,D. T. (2021). A novel hybrid quantum-PSO and credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data. *Journal of Hydrology*, *596*, 125682. https://doi.org/10.1016/j.jhydrol.2020.125682.
23. Bordbar M, Busico G, Sirna M, Tedesco D, Mastrocicco M. A multi-step approach to evaluate the sustainable use of groundwater resources for human consumption and agriculture. J Environ Manage. 2023;347:119041. https://doi.org/10.1016/j.jenvman.2023.119041.
24. Busico G, Bordbar M, Rufino F, Sarracino A, Tedesco D. Assessment of 3 –, As, and F– background levels in groundwater bodies: A methodological review and case study utilizing sequential Gaussian simulation (SGS). Groundw Sustainable Dev. 2024;26:101211. https://doi.org/10.1016/j.gsd.2024.101211.
25. Khan I, Nizam S, Bamal A, Sajib AM, Mahammad Diganta MT, Shaida MA, Uddin MG. Optimized intelligent learning for groundwater quality prediction in diverse aquifers of arid and semi-arid regions of India. Clean Eng Technol. 2025;26:100984. https://doi.org/10.1016/j.clet.2025.100984.
26. Zhang T, Wu J, Chu H, Liu J, Wang G. Interpretable Machine Learning Based Quantification of the Impact of Water Quality Indicators on Groundwater Under Multiple Pollution Sources. Water. 2025;17(6):905. https://doi.org/10.3390/w17060905.
27. Shams MY, Elshewey AM, El-kenawy E-SM, Ibrahim A, Talaat FM, Tarek Z. Water quality prediction using machine learning models based on grid search method. Multimedia Tools Appl. 2023;83(12):35307–34. https://doi.org/10.1007/s11042-023-16737-4.
28. Lamane H, Mouhir L, Moussadek R, Baghdad B, Kisi O, Bilali E, A. Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration. Int J Sedim Res. 2025;40(1):91–107. https://doi.org/10.1016/j.ijsrc.2024.10.002.

29. Senthil Selvi A, Narendrakumar S, Samanvay GVMR, Vinayak A, Vembu CTA, S., Pandi S. S. (2024). From Classical to Quantum: Evaluating Machine Learning Enhancements. In *2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)* (pp. 778–784). Presented at the 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, India: IEEE. https://doi.org/10.1109/ICAICCIT64383.2024.10912309
30. Dai, Z., Zhan, C., Yin, H., Chen, J., Xu, L., Xia, Y., … Carroll, K. C. (2025). Incorporating Deep Learning Into Hydrogeological Modeling: Advancements, Challenges, and Future Directions. Journal of Geophysical Research: Machine Learning and Computation, 2(2), e2025JH000703. https://doi.org/10.1029/2025JH000703.
31. Maity R, Srivastava A, Sarkar S, Khan MI. Revolutionizing the future of hydrological science: Impact of machine learning and deep learning amidst emerging explainable AI and transfer learning. Appl Comput Geosci. 2024;24:100206. https://doi.org/10.1016/j.acags.2024.100206.

Publisher's Note