



Interpreting optimised data-driven solution with explainable artificial intelligence (XAI) for water quality assessment for better decision-making in pollution management

Javed Mallick¹ · Saeed Alqadhi¹ · Hoang Thi Hang² · Majed Alsubih¹

Received: 5 December 2023 / Accepted: 3 June 2024 / Published online: 17 June 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

In Saudi Arabia, water pollution and drinking water scarcity pose a major challenge and jeopardise the achievement of sustainable development goals. The urgent need for rapid and accurate monitoring and assessment of water quality requires sophisticated, data-driven solutions for better decision-making in water management. This study aims to develop optimised data-driven models for comprehensive water quality assessment to enable informed decisions that are critical for sustainable water resources management. We used an entropy-weighted arithmetic technique to calculate the Water Quality Index (WQI), which integrates the World Health Organization (WHO) standards for various water quality parameters. Our methodology incorporated advanced machine learning (ML) models, including decision trees, random forests (RF) and correlation analyses to select features essential for identifying critical water quality parameters. We developed and optimised data-driven models such as gradient boosting machines (GBM), deep neural networks (DNN) and RF within the H2O API framework to ensure efficient data processing and handling. Interpretation of these models was achieved through a three-pronged explainable artificial intelligence (XAI) approach: model diagnosis with residual analysis, model parts with permutation-based feature importance and model profiling with partial dependence plots (PDP), accumulated local effects (ALE) plots and individual conditional expectation (ICE) plots. The quantitative results revealed insightful findings: fluoride and residual chlorine had the highest and lowest entropy weights, respectively, indicating their differential effects on water quality. Over 35% of the water samples were categorised as ‘unsuitable’ for consumption, highlighting the urgency of taking action to improve water quality. Amongst the optimised models, the Random Forest (model 79) and the Deep Neural Network (model 81) proved to be the most effective and showed robust predictive abilities with R^2 values of 0.96 and 0.97 respectively for testing dataset. Model profiling as XAI highlighted the significant influence of key parameters such as nitrate, total hardness and pH on WQI predictions. These findings enable targeted water quality improvement measures that are in line with sustainable water management goals. Therefore, our study demonstrates the potential of advanced, data-driven methods to revolutionise water quality assessment in Saudi Arabia. By providing a more nuanced understanding of water quality dynamics and enabling effective decision-making, these models contribute significantly to the sustainable management of valuable water resources.

Keywords Water quality assessment · Machine learning · Sustainable water management · Explainable artificial intelligence (XAI) · Entropy-weighted WQI

Responsible Editor: Xianliang Yi

✉ Javed Mallick
jmallick@kku.edu.sa

Saeed Alqadhi
sdalqadi@kku.edu.sa

Hoang Thi Hang
hlhangstac@gmail.com

Majed Alsubih
malsubih@kku.edu.sa

¹ Department of Civil Engineering, College of Engineering, King Khalid University, P.O. Box: 394, Abha 61411, Kingdom of Saudi Arabia

² Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi 110025, India

Introduction

In the Kingdom of Saudi Arabia (KSA), the challenge of ensuring water quality is particularly acute due to the unique geographical and climatic conditions (Al-Omran et al. 2016). As a country largely devoid of natural freshwater bodies like rivers and lakes, KSA predominantly relies primarily on groundwater and desalination of seawater to meet its water needs for drinking, irrigation and industrial purposes. This dependence is exacerbated by the country's rapid economic development and population growth (Al-Omran et al. 2016). The extensive use of groundwater, particularly in agriculture, has led to considerable challenges. The use of artificial fertilisers, which is necessary to meet the increased demand for food, often leads to excessive nutrient extraction from the soil (Fallatah 2020). This excess fertiliser, which is rich in nitrates and phosphates, then seeps into the groundwater and contaminates it. Furthermore, the transport of these pollutants through the groundwater system contributes to the broader problem of water pollution, which affects both human health and the environment (Saud and Abdullah 2009; Alghamdi et al. 2020).

Saudi Arabia's geographical landscape, characterised by thick Mesozoic and Cenozoic sedimentary rocks, forms productive aquifers that are central to the country's groundwater resources (Alharbi and Zaidi 2018). However, over-reliance on these aquifers in an arid environment is leading to declining groundwater levels and deteriorating quality, posing a major environmental challenge (Alharbi 2018). Rapid urbanisation and the expansion of agricultural activities across the country have further increased the demand for freshwater resources. As a result, the task of assessing and managing water quality is becoming increasingly complex and requires more efficient and reliable methods (Khanfar 2008; Al-Hammad and Abd El-Salam 2016). The traditional approach to water quality assessment is the use of water quality indices (WQIs) (Khuan et al. 2002; Asadollah et al. 2021). WQIs are an important tool for water resources management as they represent a single measure that encompasses various physical and chemical parameters of water quality (Zali et al. 2011; Hameed et al. 2017). However, the calculation of these indices is often associated with challenges, such as being time-consuming, complex and prone to inconsistencies due to the use of different equations and methods (Kouadri et al. 2021). This complexity is exacerbated by the lack of a universal WQI method, leading to different interpretations and assessments of water quality (Leong et al. 2021).

In response to these challenges, artificial intelligence (AI)-based WQI prediction emerges as a promising

solution (Aldhyani et al. 2020; Hmoud Al-Adhaileh and Waselallah Alsaade 2021; Sajib et al. 2024). AI-based models offer a transformative solution by eliminating the need for tedious sub-index calculations, enabling fast and efficient water quality assessment (Sarafaraz et al. 2024; Tiyyasha et al. 2021). These models are characterised by their non-linear structures that can handle large data sets with different scales and are resistant to missing data (Elbeltagi et al. 2022). The strength of AI algorithms in predicting complex phenomena lies in their ability to analyse data and recognise patterns (Irwan et al. 2023; Sidek et al. 2024). In this process, algorithms are constructed using a subset of the data set (training data) and the prediction performance is validated using a separate subset (test set) (Irwan et al. 2023). Notable AI algorithms that have been successfully applied in water quality prediction include adaptive boosting (Adaboost), gradient boosting (GBM), extreme gradient boosting (XGBoost), decision trees (DT), extra trees (ExT), random forest (RF), multilayer perceptron (MLP), radial basis function (RBF), deep feed-forward neural network (DFNN) and convolutional neural network (CNN) (Kim et al. 2022; Khoi et al. 2022; Aldrees et al. 2022; Nayan et al. 2020; Talukdar et al. 2023a; Al-Sulttani et al. 2021; Sinha 2023; Yusri et al. 2022; Ho et al. 2019; Sakaa et al. 2022; Sheikh Khozani et al. 2022; El-Shebli et al. 2023; Kogekar et al. 2021; Mei et al. 2022; Sidek et al. 2024). These AI models have been used in various contexts e.g. in the prediction of manganese removal prediction (Bhagat et al. 2020; Erickson et al. 2021), flood susceptibility studies (Talukdar et al. 2020a; 2023a; Islam et al. 2021; Saha et al. 2021; Ahmed et al. 2022; Mahato et al. 2021), the identification of pollution sources (Mia et al. 2023) and the general prediction of water quality (Talukdar et al. 2023b; Sinha 2023; Yusri et al. 2022; Khoi et al. 2022), with varying degrees of accuracy. The application of AI in water quality assessment represents a significant advance as it offers a more streamlined, accurate and efficient approach compared to conventional methods (Lap et al. 2023).

Grid search optimisation in machine learning is a methodological approach to improve the performance of models by fine-tuning their hyperparameters (Wu et al. 2019; Kim and Seo 2024). In contrast to general model parameters derived from training data, hyperparameters are predefined settings that control the learning process. For example, while the coefficient of a logistic regression (LR) model is determined during training, the number of decision trees in an RF model is a hyperparameter that is set before training (Talukdar et al. 2020b). The importance of hyperparameters in machine learning cannot be overstated, as they directly influence the accuracy, speed and reliability of the models. Grid search involves exhaustively exploring a given range of hyperparameters, iteratively running through all possible

combinations and evaluating their performance to select the optimal set (Dodangeh et al. 2020). This process ensures that the chosen hyperparameters maximise the effectiveness of the model (Talukdar et al. 2024). However, this task can be computationally intensive and time-consuming, particularly for complex models with numerous hyperparameters (Raheja et al. 2024). However, the meticulousness of grid search makes it an indispensable tool in machine learning, especially for applications where precision and model performance are critical (Fang et al. 2021).

Although advanced AI models, including hybrid and deep learning systems, can provide highly accurate predictions, they often operate as "black boxes" that offer little insight into how they arrive at these predictions (Alshehri and Rahman 2023). This ambiguity limits their applicability, particularly in areas where understanding the motivations behind decisions is critical (Park et al. 2022a). Explainable artificial intelligence (XAI) addresses this fundamental challenge in the field of AI and machine learning for decision-making processes (Talukdar et al. 2023a, 2024). XAI aims to demystify these complex models and provide clarity on the internal mechanisms and decision-making processes (Talukdar et al. 2023b; Park et al. 2022a). Techniques such as SHapley Additive exPlanations (SHAP), partial dependence plot (PDP), permutation-based feature importance, accumulated local effect (ALE) have gained prominence in this field (Ahmed et al. 2023, 2024, Mia et al. 2023). SHAP, an additive feature assignment method, and PDP decompose the prediction of a model into the contribution of each feature, providing insights into the interaction and relative importance of the different variables to the model's results (Ahmed et al. 2024). This interpretive approach not only increases the transparency of AI models, but also builds trust with users and stakeholders and ensures that AI-driven decisions are understandable and justifiable (Talukdar et al. 2023b). The application of XAI is critical in areas such as geohazard prediction and environmental monitoring, where understanding the basis of predictive modelling is as important as the predictions themselves (Ahmed et al. 2024).

This study addresses critical gaps in water quality assessment in Saudi Arabia, where conventional methods are inadequate given the dynamic environmental changes and the region's dependence on groundwater and seawater desalination, exacerbated by agricultural and industrial pressures. We use advanced machine learning (ML) models to conduct real-time data-driven analyses with an entropy-weighted WQI and integrate XAI to identify key water quality parameters. This enables targeted policy measures and improves the understanding of water quality interactions. In addition, the H2O API in R programming, which is central to our methodology, facilitates both grid search optimisation and XAI integration, simplifying the exploration of multiple parameter combinations to optimise model configuration

and interpretability (Talukdar et al. 2024; Šandera and Štych 2024). This integration of grid search optimisation with XAI within the H2O framework ensures that model performance is not only improved, but that the results are transparent, trustworthy, and meet the requirements of complex data-driven decision-making processes. This approach represents a significant innovation in environmental monitoring as it improves both the performance and interpretability of ML models, which is crucial for effective environmental management and policy formulation. These innovations hold great potential for global sustainable water management and could influence future academic and policy directions.

Materials and methodology

Study area

The Asir Province in Saudi Arabia is located within the coordinates 18°12'029.355"N, 18° 12'051.436"N latitude and 42° 29'05.157"E, 42° 29'019.795"E longitude, covering an area of 84,250 square kilometres (as shown in Fig. 1b). Asir experiences various climatic conditions, including hot desert, cold desert, cold semi-arid and hot semi-arid zones. It receives an annual rainfall of 350 mm, making it a significant region for agriculture. The geological composition of Asir consists of aquifers such as quaternary alluvium, quartz sandstone and conglomerates, with secondary aquifers primarily composed of calcareous deposits undergoing lateral diagenetic modifications. These aquifers exhibit greater porosity and karstification (Mallick et al. 2018). Figure 1B provides a geological map of the Asir region. One crucial source of groundwater in the region is the unconfined quaternary alluvial aquifers, which are replenished by runoff from the Asir highlands. These shallow aquifers have an estimated annual recharge of 1196×106 m³ and exhibit varying water quality, ranging from poor to good, as noted by Dabbagh and Abderrahman in 1997. The high-quality groundwater found in Wadi-al-Dawassir is attributed to a 100-m-thick layer of alluvial fill.

The Asir region is rapidly advancing its rainwater harvesting efforts through the construction of check dams. These dams enable the collection of sufficient water to cultivate 15,000 hectares of agricultural land. Presently, if just a quarter of the runoff water currently lost could be effectively harvested, it would fulfil all of Saudi Arabia's existing agricultural water requirements. The majority of Saudi Arabia's runoff occurs along the escarpment in the Asir region, where wadis flow towards the coastal area, contributing approximately 60% of the nation's total runoff. Most wadi structures are filled with sand and gravel, and after a short distance, the runoff seeps into subsurface water bodies (wadi) and forms a sub-flow, recharging the groundwater. Storm runoff can

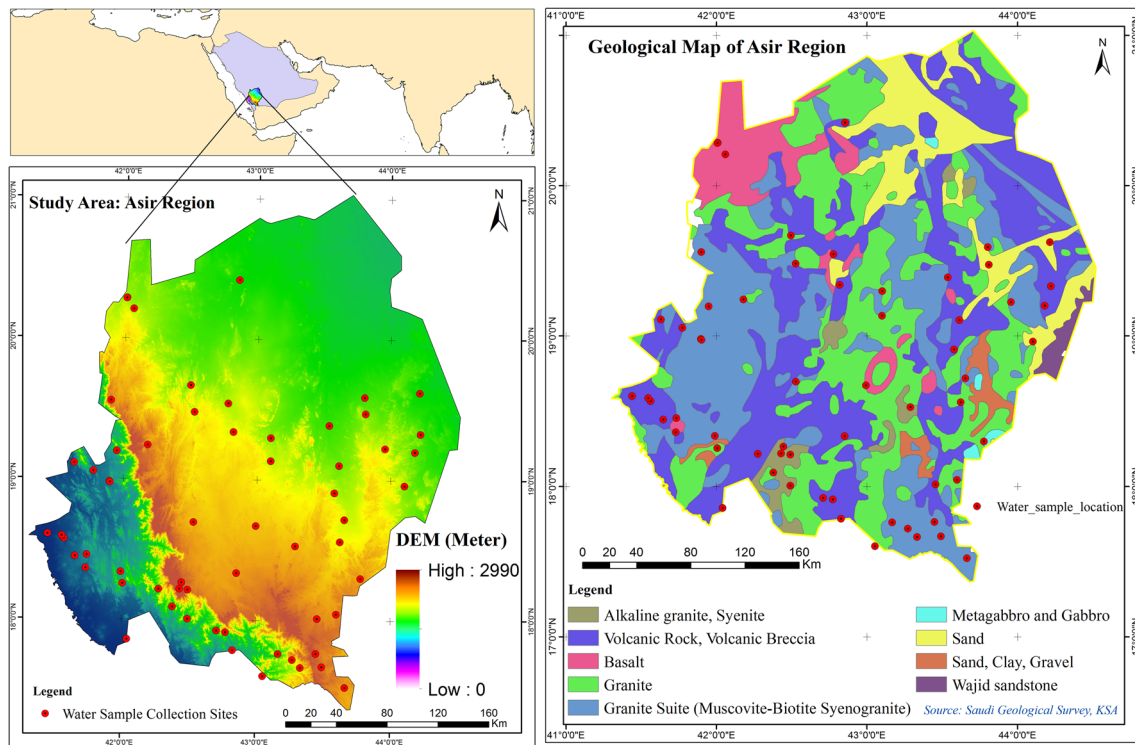


Fig. 1 Study area and geological map

occur in the Asir region at any time of the year. Saudi Arabia has a longstanding tradition of constructing dams, particularly in the Hijaz and Asir regions. As of 2018, according to the Saudi Arabia Ministry of Environment, Water and Agriculture, there are 509 dams throughout the kingdom, with 117 of them located in the Asir region. The primary purpose behind constructing these dams is to capture runoff and replenish the groundwater network, although some dams also serve as sources of drinking water and direct irrigation for agriculture.

Sampling and laboratory analysis

A total of 62 groundwater samples were taken from various wells within the study region, as shown in Fig. 1b. The locations of these wells were randomly selected to ensure a representative distribution across the different geological and hydrological conditions in the area. This random selection was intended to provide a comprehensive overview of the water quality in the entire study region. During sampling, several parameters, including electrical conductivity (EC), pH and total dissolved solids (TDS), were measured to assess water quality. Handheld sensors manufactured by HANNA were employed for these measurements. Prior to each sampling session, the sensor was calibrated daily using standard solutions with pH values of 4.0, 7.0 and 10.0, as well as EC standards at 84 $\mu\text{S}/\text{cm}$, 1413 $\mu\text{S}/\text{cm}$ and 12.8

mS/cm . For each sampling site, two groundwater samples were collected in high-density polyethylene (HDPE) bottles. One of these samples was acidified in the field using a 1:1 nitric acid solution and was subsequently utilised for cation analysis. The second sample was left unaltered and was designated for laboratory-based anion estimation. To determine the concentrations of cations such as calcium (Ca^{2+}), sodium (Na^+), magnesium (Mg^{2+}), potassium (K^+) and iron (Fe), an atomic absorption spectrophotometer from Thermo Scientific's M series was employed. On the other hand, the analysis of anions, including chloride (Cl^-), fluoride (F^-), nitrate (NO_3^-) and sulphate (SO_4^{2-}), was carried out using an ion chromatograph (Dionex) in gradient mode. Additionally, bicarbonate (HCO_3^-) was determined using a titrimetric method, while total alkalinity and hardness were assessed following the standard procedures outlined in APHA 1995. All the reagents, standards and chemicals utilised in these analyses were of analytical grade and sourced from Merck.

Water quality index estimation using entropy weighted arithmetic method

The theoretical method for estimating the WQI using the entropy-weighted arithmetic method revolves around the application of entropy theory to objectively determine the weighting of each water quality parameter, thereby eliminating the subjective biases typically associated with expert

opinion (Singh et al. 2019). In this method, entropy, a concept derived from information theory, is used to quantify the degree of disorder or uncertainty associated with each water quality parameter (Verma et al. 2022). The greater the variability of a parameter across different water samples, the higher its entropy value, indicating a more significant role in the overall water quality assessment. This process begins with the collection and normalisation of water quality data, followed by the calculation of entropy values for each parameter. These entropy values are then used to derive objective weights that reflect the relative importance of each parameter to overall water quality. These weights are applied to the normalised values of each parameter and the entropy-weighted arithmetic mean of these values is calculated to obtain the final WQI. This entropy-based weighting approach ensures that the WQI is a more accurate and objective representation of water quality, as it minimises the influence of human bias and subjectivity that is inevitable in methods that rely heavily on expert opinion. In this way, it provides a more reliable and scientifically sound basis for water quality assessment, which is crucial for effective environmental monitoring and management.

Methods for machine learning–based feature selection techniques

In the theoretical framework of machine learning-based feature selection techniques, in particular the decision tree, random forest and correlation methods, each technique fulfils a specific role in identifying the most important features for predicting the WQI. The decision tree method works by creating a tree-like model of decisions where the importance of features is determined by how effectively they contribute to the partitioning of the data, indicating their influence on the outcome variable, in this case the WQI. The random forest approach, an ensemble of decision trees, further refines this process by constructing multiple trees and aggregating their results, thereby improving the reliability and generalisability of the feature importance scores. This method is particularly effective in dealing with overfitting and provides a more comprehensive understanding of feature relevance. Finally, the correlation method involves statistical analysis to assess the strength and direction of the relationship between each feature and the WQI. By assessing the correlation coefficients, this method helps to identify traits that have a significant linear relationship with the WQI. Together, these three methods provide a robust framework for feature selection, ensuring that the most predictive and relevant features are identified for use in WQI prediction models. This multi-layered approach utilises both statistical and machine learning techniques to improve the accuracy and efficiency of water quality assessments.

Selection and optimisation of ML models in H2O API for assessing water quality

The selection of the GBM, DNN and RF models in the H2O API for this water quality assessment study is based on their different capabilities in dealing with complex, non-linear data patterns commonly found in environmental datasets. These models are favoured due to their robustness, their ability to handle a large number of input features and their resistance to overfitting, which makes them particularly suitable for water quality analysis. Grid search optimisation is chosen for these models to systematically explore a wide range of hyperparameter combinations to determine the optimal model configuration. This approach is crucial for improving model performance and prediction accuracy. The H2O API is used in this study due to its user-friendly interface, scalability and efficient handling of large datasets for optimising and deploying ML models. It provides a comprehensive environment that simplifies the implementation of complex models and the grid search optimisation process, making it an ideal choice for this application.

GBM

The GBM model in H2O is an ensemble learning method that sequentially builds a series of decision trees, where each tree is designed to correct the errors of its predecessor (Talukdar et al. 2023a). This method combines the predictions from multiple trees to produce a final, more accurate prediction (Ahmed et al. 2024) (see Eq. 1). GBM is particularly effective in assessing water quality due to its ability to model complex interactions between parameters and its high prediction accuracy. The iterative nature of GBM allows it to focus on difficult-to-predict instances, making it highly adaptable to varying water quality data patterns.

$$GBM(x) = \sum_{t=1}^T \gamma_t h_t(x) \quad (1)$$

where $h_t(x)$ represents the decision trees and γ_t are the weights for each tree.

DNN

DNNs in the H2O framework consist of multiple layers of interconnected nodes or neurons, each designed to progressively extract and refine features from the input data. The transformation in each layer of a DNN can be mathematically represented as:

$$a_{i+1} = f(w_i \cdot a_i + b_i) \quad (2)$$

where a_i are the activations from the previous layer, w_i and b_i represent the weight matrix and bias vector of the current

layer, respectively, and σ denotes a non-linear activation function, such as ReLU or sigmoid. This structure allows DNNs to capture intricate relationships within the data, modelling complex patterns that are not readily apparent (Ahmed et al. 2024). The depth of the network (number of layers) and the number of nodes in each layer can be adjusted to suit the complexity of the dataset, with deeper networks generally being more capable of learning nuanced features of the data (Talukdar et al. 2023a).

RF

RF, as implemented in H2O, operates as an ensemble of decision trees, each constructed from a randomly selected subset of the training data and features, according to the bagging approach. The final prediction of the RF model is typically the average of the predictions from all the individual trees, which can be represented as:

$$RF(x) = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (3)$$

where $h_t(x)$ are the individual trees. Each tree is built independently, and the random selection of features and samples for each tree reduces variance and avoids overfitting, making RF particularly robust against model bias and variance issues (Palkar et al. 2022). This methodology not only enhances the stability and accuracy of the predictions but also makes RF an excellent tool for handling datasets with a high dimensional feature space.

Assessment of the performance of ML models

When evaluating the performance of machine learning models for water quality prediction, a comprehensive set of metrics is used, each providing a unique perspective on the accuracy and reliability of the model. Mean square error (MSE) and root mean square error (RMSE) are used to quantify the mean square difference and square root of this difference between the predicted and actual values, respectively, providing insight into the overall prediction error of the model. The mean absolute error (MAE) measures the average magnitude of the errors in a set of predictions without considering their direction. The Root Mean Squared Logarithmic Error (RMSLE) is particularly useful when dealing with exponential growth as it evaluates the logarithmic difference between the predicted and actual values, making it less sensitive to large errors in predicting higher values. Similar to the MSE, the mean residual deviance is a measure of the variance of the prediction errors and indicates the deviation of the model from the observed data. The coefficient of determination (R^2) indicates the proportion of the variance in the dependent variable that can be predicted by

the independent variables and is therefore a measure of the explanatory power of the model. In addition, the Taylor diagram is used as a visual tool to compare the statistical summary of the model's performance, including its correlation, standard deviation and RMSE, with the observed data. This comprehensive approach to performance evaluation ensures a thorough assessment of the accuracy, reliability and suitability of the model for predicting water quality.

Interpreting optimised ML models through XAI

Interpreting optimised machine learning models in the H2O API, in particular GBM, DNN and RF, through XAI is critical to understanding the decision-making process of these models. XAI provides transparency and insight into the complex workings of these advanced models and makes them more interpretable for users and stakeholders (Talukdar et al. 2023b; Ahmed et al. 2023, 2024; Mia et al. 2023). To this end, the DALEX package in R is used, which provides a set of tools and methods to explain and understand the behaviour and predictions of machine learning models. Using the features of DALEX, we can analyse the models to gain a comprehensive understanding of how they work, to understand how different features influence the predictions, and to identify possible biases or inconsistencies.

Model diagnostic

Model diagnosis in XAI involves evaluating the performance and reliability of the optimised ML models. This is primarily done by residual analysis, which analyses the difference between the observed values and the predictions of the model (residuals). Residual analysis helps to identify patterns or anomalies in the predictions, such as systematic biases, overfitting or underfitting. By analysing these residuals, we gain insights into the accuracy of the model in different data segments and can diagnose problems that could affect the performance of the model. This diagnostic process is important to validate the effectiveness of the model and ensure that it generalises well to new, unseen data.

Model parts

The 'model parts' aspect of XAI focuses on understanding the contribution of each feature to the model's predictions. This is achieved through permutation-based feature importance, a technique that assesses the impact of reshuffling each feature on the accuracy of the model. By randomly permuting the values of each feature and measuring the change in the model's performance, we can determine the importance of each feature to the predictions. This method provides a ranking of the features based on their importance

and helps to identify the most influential variables in the model and understand their role in the prediction process.

Model profile

Model profiling in XAI includes techniques such as Partial Dependence Plots (PDP), Accumulated Local Effects (ALE) and Individual Conditional Expectation (ICE) plots. These methods provide a deeper insight into the relationship between characteristics and the predicted outcome. PDPs show the average impact of a feature on the model's predictions and allow us to identify patterns and trends in how feature values influence the outcome. ALE charts provide a similar view, but focus on the local effects of features and provide a more accurate representation in the presence of correlated features. ICE graphs, on the other hand, show how predictions for individual instances change as a feature is varied, providing a detailed insight into the behaviour of the model. Together, these profiling techniques provide a comprehensive picture of how different features and their interactions affect the model's predictions, improving the interpretability of complex ML models.

Results

Assessment of water quality condition

Water quality was assessed by measuring physicochemical parameters such as TDS, conductivity, pH and concentrations of ions such as nitrates, sulphates, chlorides, iron and magnesium. The descriptive statistics show considerable variability. For example, the TDS value showed a high standard deviation of 479.47, which indicates a different mineral content of the samples (supplementary Section 1). The confidence intervals for pH (7.61–7.92) indicate consistent acidity, while wider intervals for TDS (583.89–822.59) and conductivity (1045.30–1397.61) reflect greater variability in these readings. Such discrepancies indicate different water sources and possible contaminants. Probability distribution plots further illustrate this variability: ammonia and nitrite levels are predominantly low, as indicated by right-skewed distributions, while broader distributions for TDS and conductivity indicate a range of dissolved concentrations (Supplementary Fig. 1). Parameters such as pH show narrow peaks, indicating uniformity in the samples, while flatter distributions for TDS and conductivity indicate greater variation. In terms of overall water quality impairment, parameters with a broad distribution are of particular concern, especially if they include ranges that exceed environmental or health standards. They may indicate the need for targeted water treatment procedures or further investigation of possible sources of pollution. The shape and distribution

of these distributions can inform water quality management decisions, such as whether to focus on general treatment methods or specific pollutants.

The descriptive statistics and distributions emphasise the challenges of predicting overall water quality from isolated parameters alone. Therefore, we calculated the WQI using the entropy weighting method, which is based on WHO standards. The WHO sets acceptable limits such as 0.50 mg/L for ammonia (weight 0.05), 0.10 mg/L for nitrite (weight 0.09) and 10 mg/L for nitrate (weight 0.06) to ensure water safety. TDS has a higher allowable limit of 600 mg/L (weight 0.02), indicating the ubiquitous presence of TDS in water, while the limits for chloride and sulphate are both set at 250 mg/L (weight 0.03). Total hardness has a limit of 500 mg/L (weight 0.02), and the limit for total calcium is 75 mg/L with a lower weight of 0.02. Magnesium and iron have limits of 30 mg/L (weight 0.05) and 0.30 mg/L (weight 0.05), respectively, reflecting their moderate impact on the WQI. Fluoride has the highest weighting of 0.26, with a limit of 1.50 mg/L, as it has a significant impact on health at varying concentrations. The standard for alkalinity is 80 mg/L (weighting 0.03), while conductivity is set at 1000 $\mu\text{S}/\text{cm}$ (weighting 0.01). The standard for pH is 8.50 (weight 0.00), which means that it has less direct impact. Turbidity and residual chlorine are weighted at 0.17 and 0.10, with standards of 5.00 NTU and 0.50 mg/L respectively, emphasising their importance in determining water clarity and microbial safety. These parameter weights, reflecting the degree of impact of each parameter on overall water quality, are used to calculate the WQI, which has a mean of 188.14 and a high standard deviation of 794.92, indicating considerable variation in water quality between samples. The WQI was then categorised into five categories—excellent, good, poor, unsuitable and very poor—as in Alam et al. (2021). The distribution analysis showed that over 35% of the samples fell into the unsuitable category, indicating the need for treatment, while less than 10% were very poor, indicating heavy contamination. Conversely, around 40% of samples were categorised as excellent or good and 15% as poor, illustrating the different water quality requirements in the different regions (Fig. 2).

Analysis of feature selection

Feature selection is a crucial step in the modelling process in both machine learning and deep learning, as it helps to improve the performance of the model by reducing complexity, preventing overfitting and increasing computational efficiency. By identifying and retaining only the most informative features, models can achieve higher accuracy with simpler, more interpretable results. Therefore, we used three ML models in this study, such as correlation of all variables with WQI, decision tree

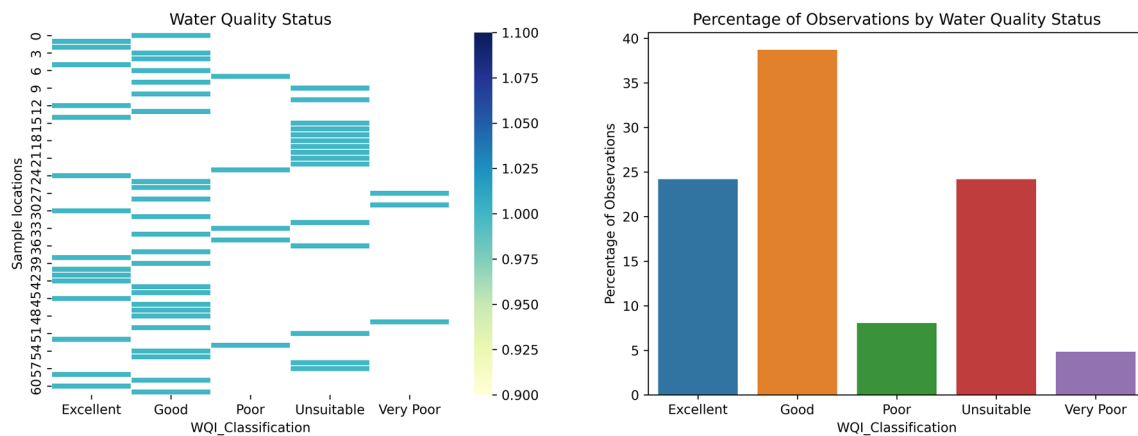


Fig. 2 A composite assessment of water quality across various sample locations, utilizing the WQI as a metric. On the left, the horizontal bar chart displays the WQI classification for individual sam-

ple locations, with varying lengths of bars representing the index's numerical value mapped to a colour gradient

and random forest models as feature selection (Fig. 3). In the given analysis for the model of correlation with WQI (panel a), the five most influential parameters are residual chlorine, nitrate, TDS, sulphate and total hardness, which all show strong positive correlations, suggesting that they are significant predictors of water quality. The least influential parameters with the lowest correlations include ammonia, fluoride, alkalinity, total calcium and iron, indicating a weaker linear relationship with WQI. For the decision tree model (panel b), the parameters with the highest values for importance are nitrite, turbidity, sulphate, magnesium and TDS. These parameters are considered to be the most critical in determining the splits in the decision tree and therefore have a large influence on the predictions of the model. Conversely, the least important parameters that contribute least to decision making are iron, alkalinity, total calcium, fluoride and ammonia. Finally, for the Random Forest model (panel c), the left graph shows that the top five parameters that increase the mean square error the most when omitted (i.e. the most important) are residual chlorine, fluoride, pH, nitrate and TDS. This indicates that omitting these features significantly degrades the performance of the model. The diagram on the right shows IncNodePurity, where the most important factors for node purity are residual chlorine, nitrate, TDS, pH and conductivity. Based on these quantitative assessments, ammonia was found to have minimal impact on model performance and was therefore removed from the dataset for further WQI assessment with DL models. The low importance of the features and the minimal impact on model accuracy and node purity justify the exclusion of ammonia, allowing the models to focus on parameters with stronger predictive relationships to WQI.

Implementation of ML models in H2O API for assessing water quality

GBM, DNN and RF algorithms were used in the implementation of ML models within the H2O API framework to assess the WQI. The H2O API facilitates the streamlined application of these complex algorithms and enables efficient optimisation and assessment of WQI, which is crucial for the development of a data-driven system for fast and accurate water quality monitoring. This approach not only expedites the processing of large data sets but also provides a scientific interface for in-depth analyses that support more informed decision making. By utilising the computing power and user-friendly features of the H2O API, the application of these water quality assessment models becomes more accessible, enabling continuous improvements in environmental management.

Optimization of ML models using grid search algorithm

The versatility of the H2O API in grid search algorithms enables optimal identification of hyperparameters for robust ML models, which are essential for WQI assessments. For the GBM model, a grid search across 36 hyperparameter combinations (e.g. balance_classes, col_sample_rate, max_depth) identified the best model, `gbm_grid1_model_38`, based on the lowest RMSE. This model, shown in detail in Supplementary Fig. 2, consists of 658 trees with a depth of 3 to 6 and optimises complexity and fit to the training data. The DNNs underwent a more extensive optimization testing 97 hyperparameter variations, including activation, epochs, hidden layers and L1 and L2 regularisation (Supplementary Fig. 3). The

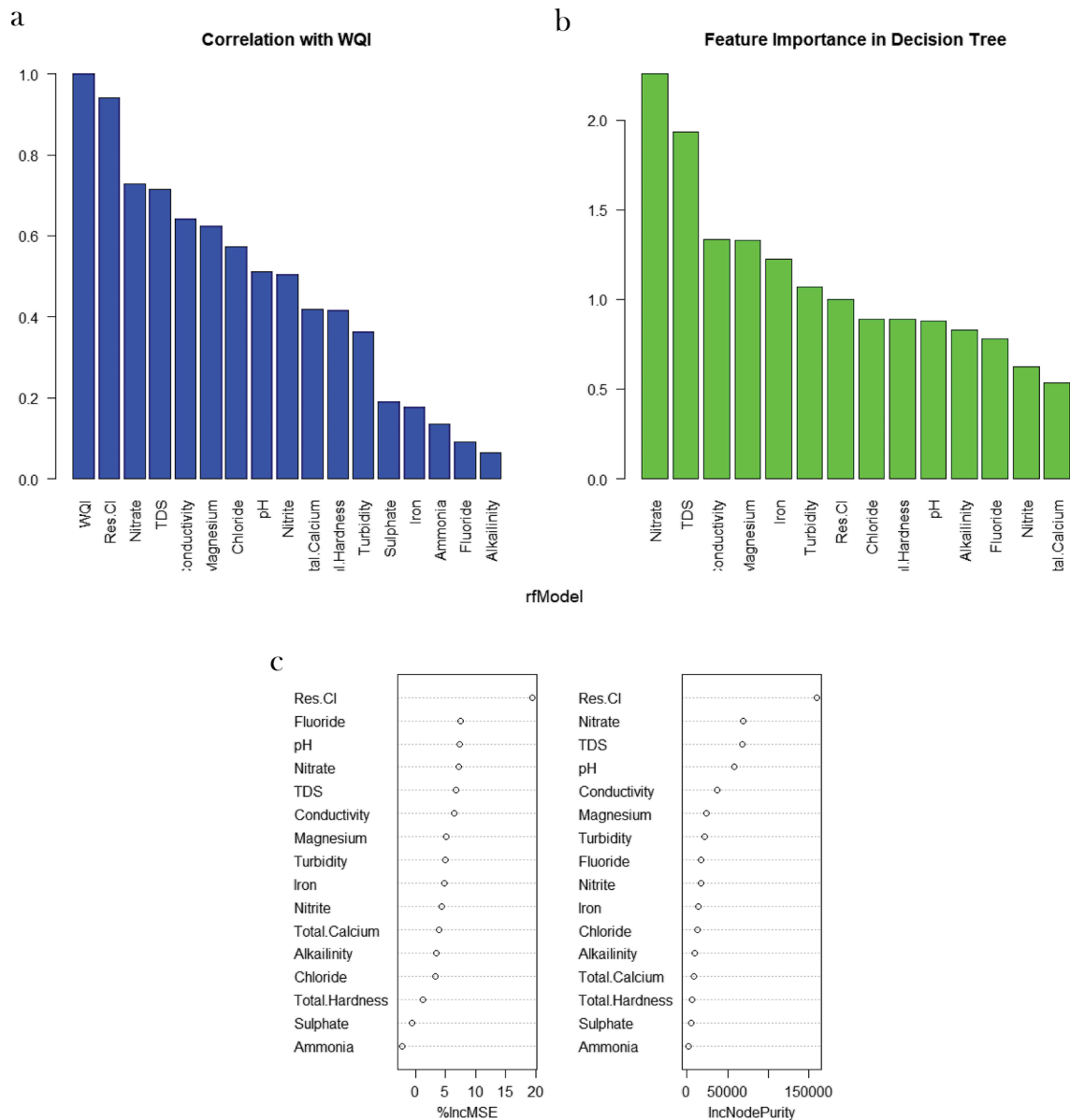


Fig. 3 A three-part feature selection analysis for predictive modeling of WQI with ML algorithms. Panel **a** shows the Pearson correlation coefficients between individual water quality parameters and the WQI, highlighting the parameters that are most strongly linearly related to the WQI. Panel **b** shows the feature importance values derived from a decision tree model, indicating the relative predic-

tive value of each parameter within the model. Panel **c** contrasts two metrics from a random forest model: The percentage increase in mean squared error (%IncMSE) when a feature is excluded and the increase in node purity (%IncNodePurity), both of which quantify the impact of each parameter on the accuracy of the model and the decision process

optimal DNN model, DNNs_model_81, was characterized by an architecture with 206 weights/biases, 15 input neurons and multiple hidden layers of 5 neurons each, using a Tanh activation function. This configuration emphasises the depth of the model and the tailored complexity for accurate WQI prediction. Meanwhile, the robustness of the RF model was tested in 389 trials, optimising parameters such as max_depth, mtries, ntrees and sample_rate, without a single error occurring (Supplementary Fig. 4). The

best RF model, RF_grid1_model_79, configured with 50 trees and a maximum depth of 18, shows its ability to recognise complex, non-linear patterns in the dataset. These optimization efforts for the GBM, DNN and RF models ensure a data-driven, science-based approach to rapid WQI assessment. The optimal hyperparameters, facilitated by the H2O API grid search, improve model performance and promote accurate water quality monitoring that supports informed environmental decision making.

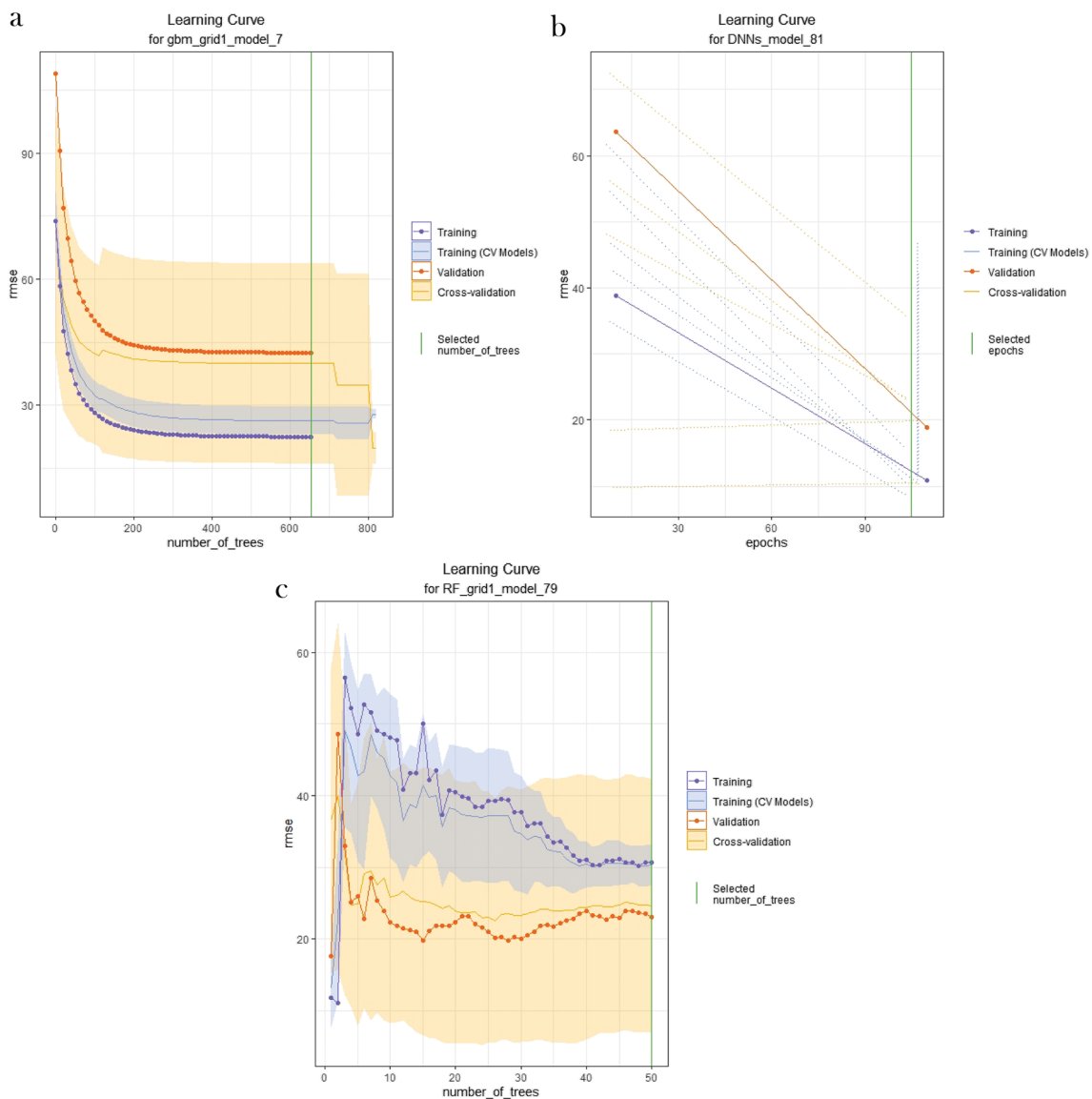


Fig. 4 The learning curves of three optimised ML models for WQI prediction: Panel **a** shows the learning curve of the GBM model *gbm_grid1_model_38*. Panel **b** shows the learning progress of the

DNN model *DNNs_model_81*. Panel **c** shows the learning curve of the RF model *RF_grid1_model_79*

To validate the model optimization, the learning curves for the three best models—*gbm_grid1_model_38*, *DNNs_model_81* and *RF_grid1_model_79*—were analysed (Fig. 4). For *gbm_grid1_model_38*, the curve stabilises at 600 trees, with training and cross-validation error rates close to 30%, indicating an optimal number of trees without overfitting. *DNNs_model_81* shows a decrease in error rates and reaches a plateau at 80 epochs, indicating effective learning. *RF_grid1_model_79* stabilises the error reduction at around 40 trees, confirming the adequate complexity and generalizability of the model. These curves show that the models are well tuned and capture the necessary data patterns without being too specific to the training set.

Assessment of optimised ML models

The performance assessment of machine learning models is a crucial step to ensure their reliability and effectiveness in predictive tasks. In the context of water quality analysis, evaluating the accuracy and generalisability of models such as GBM, DNN and RF is essential to determine their practical utility in predicting WQI from various water parameters (Table 1).

The GBM model (*gbm_grid1_model_38*) was evaluated using various regression metrics. The performance of the model on the training data shows an MSE of 501.84, RMSE of 22.40, MAE of 8.89, RMSLE of 0.16 and mean residual

Table 1 Statistical analysis of model performance in predicting WQI using GBM, DNN and RF models

Model	GBM (gbm_grid1_ model_38)	DNN (DNNs_ model_81)	RF (RF_grid1_ model_79)
MSE (training)	501.84	118.91	943.01
RMSE (training)	22.4	10.9	30.71
MAE (training)	8.89	4.99	10.39
RMSLE (training)	0.16	0.12	0.27
MSE (validation)	1868.66	354.97	532.33
RMSE (validation)	43.23	18.84	23.07
MAE (validation)	19.33	9.34	12.8
RMSLE (validation)	0.26	0.13	0.22
MSE (cross-validation)	1723.09	601.77	980.1
RMSE (cross-validation)	41.51	24.53	31.31
MAE (cross-validation)	18.11	11.69	11.63
RMSLE (cross-validation)		0.31	0.27

deviance of 501.84. These results indicate that the model fits the training data well, with a relatively low error rate. However, when evaluating the validation data, the error metrics are higher, with an MSE of 1868.66, RMSE of 43.23, MAE of 19.33, RMSLE of 0.26 and mean residual deviance of 1868.66. The increase in error rates in the validation set compared to the training set indicates that the model may not generalise as effectively to new, unseen data, which could be a sign of overfitting. Cross-validation, a more robust metric, reports an MSE of 1723.09, RMSE of 41.51 and MAE of 18.11. The summary of the cross-validation metric shows variation in the model's performance across different folds, with an RMSE ranging from 19.92 to 65.39 and an R^2 (coefficient of determination) metric ranging from 0.56 to 0.87, highlighting some variability in the model's predictive accuracy. The average MAE value of 18.73 and the RMSE value of 38.52 from the cross-validation are higher than the values from the training data, but lower than the values from the validation data. This suggests that while the model may be slightly overfitting, it still has a reasonable degree of predictive power that could be applicable to new data, especially when considering the mean R^2 value of 0.64, which suggests that a good proportion of the variance is explained by the model. The slight overfitting observed does not preclude the use of the model for new data, but indicates that the predictions of the model should be considered with an awareness of its limitations and potential for error.

The DNN model (DNNs_model_81) shows different performance levels in the training, validation and cross-validation data sets. For the training data, the model has an

MSE of 118.91, an RMSE of 10.90, an MAE of 4.99, an RMSLE of 0.12 and a mean residual deviation of 118.91. These metrics indicate strong performance in the training set with relatively low error rates, suggesting that the model has learnt to fit the training data effectively. When evaluating the validation data, the error increases with an MSE of 354.97, an RMSE of 18.84, an MAE of 9.34, an RMSLE of 0.13 and a mean residual error of 354.97. Although the error is not too high, the predictive ability of the DNN model for unseen data is still very good. The cross-validation results, which allow a more robust assessment by training the model on multiple folds of the data, show further increased error rates: an MSE of 601.77, RMSE of 24.53, MAE of 11.69 and RMSLE of 0.31, with a mean residual deviance of 601.77. The summary of cross-validation metrics illustrates the variability in the model's performance across different folds, with RMSE ranging from 9.84 to 35.55 and R^2 values ranging from 0.77 to 0.90. The mean cross-validation RMSE of 22.02 and MAE of 11.32 are higher than the training metrics, but not excessively so, indicating that the model maintains its predictive ability across different subsets of the data.

The RF model (RF_grid1_model_79), which was evaluated for water quality prediction, shows different performance metrics for training, validation and cross-validation data. On the training data, the RF model reports an MSE of 943.01, an RMSE of 30.71, an MAE of 10.39, an RMSLE of 0.27 and a mean residual error of 943.01. These figures suggest that when applied to the training data on which it was trained (particularly out-of-bag samples), the model has a moderate level of error, which is to be expected with a diverse dataset. Moving to the validation data, the performance of the model improves with a lower MSE of 532.33, RMSE of 23.07, MAE of 12.80, RMSLE of 0.22 and mean residual error of 532.33. The reduction in MSE and RMSE in the validation set compared to the training set is atypical, as models typically perform better on the training data. This could indicate that the model is very robust and does not over-fit as it maintains its performance when exposed to unseen data. The cross-validation metrics provide a comprehensive assessment of the generalisability of the model. In a fivefold cross-validation of the training data, the model shows an MSE of 980.10, an RMSE of 31.31, an MAE of 11.63 and an RMSLE of 0.27, with a mean residual deviance of 980.10. The summary of the cross-validation metrics shows some variability in the model's performance, with an average RMSE of 24.62 and a standard deviation of 17.65, indicating that the model's prediction error can fluctuate but generally maintains a consistent level of performance across different data subsets. The R^2 values, which indicate the proportion of variance explained by the model, range from 0.75 to 0.97 and average 0.86, indicating that the model captures a substantial proportion of the variance in the data. The RF model RF_grid1_model_79 thus shows robust performance

with consistency across training and cross-validation datasets, with a slight increase in error rates in cross-validation. The ability of the model to maintain a relatively stable error rate across different subsets of the data, without a significant increase in error rate during cross-validation, suggests that the model is not overfitting and is likely to perform well on new data. The high R^2 values also speak in favour of the model’s ability to reliably predict water quality.

Comparisons of the performance of ML models

The performance of the statistical comparison of the ML models’ performance in predicting water quality indices using the scatter heat map (Fig. 5) and the Taylor diagram

(Fig. 6). For the GBM model (gbm_grid1_model_38), an R^2 of 0.92 for training indicates that 92% of the variability in the training dataset is captured by the model, dropping slightly to 0.90 in the testing phase. This drop means that although the GBM model is robust, there may be some overfitting as the model is slightly less effective at predicting unseen data. The DNN model (DNNs_model_81) has R^2 values of 0.98 for training and 0.97 for the test phase, indicating exceptional performance and generalisation from training to unseen data. The minimal decrease in the R^2 value from training to testing indicates that the DNN model captured the underlying patterns in the data very well without overfitting. The RF model (RF_grid1_model_79) has the highest R^2 of 0.99 during training,

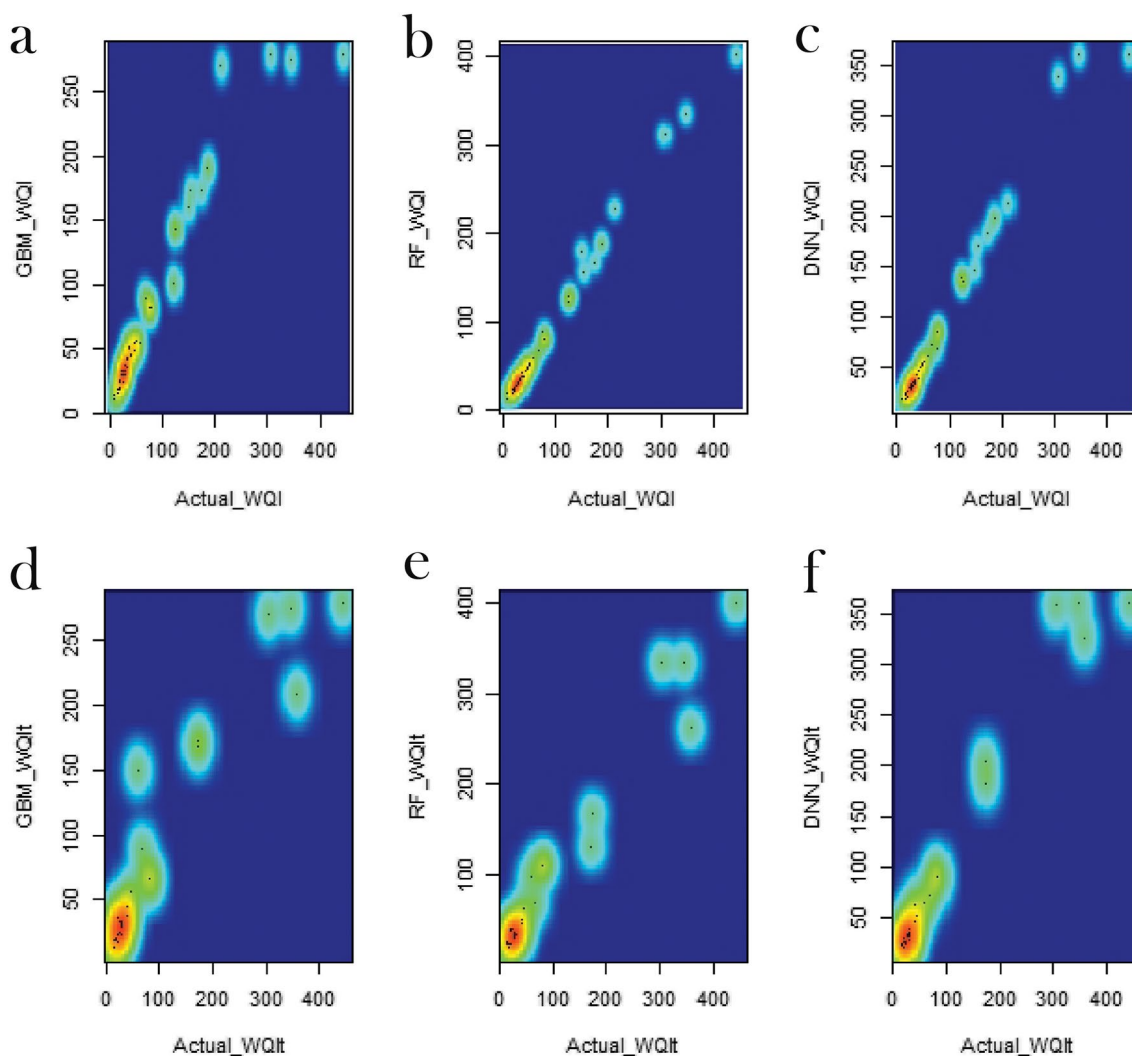


Fig. 5 The predictive performance of three machine learning models—GBM, RF and DNN—in estimating the WQI for training and testing phase. Each plot contrasts the predicted WQI values (y-axis) against the actual WQI values (x-axis) for both training (top row) and testing (bottom row) datasets. Areas with higher colour intensity

indicate a higher concentration of data points, with the diagonal line representing perfect prediction. The closeness of data points to this diagonal reflects the accuracy of each model, with the RF and DNN models showing tighter clusters around the diagonal line

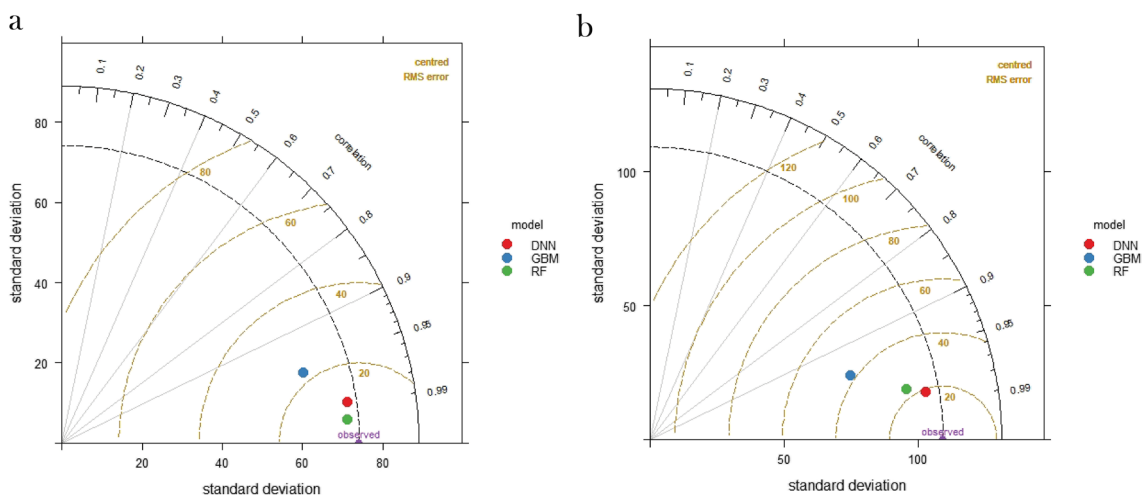


Fig. 6 Comparative assessment of ML model performance in WQI prediction. Panel a presents the performance of three machine learning models—DNN, GBM and RF—during the training phase. Panel b shows the same models' performance in the testing phase

which indicating almost perfect predictability. During testing, however, the R^2 drops to 0.96, which is a slight decrease, but still indicates a highly predictive model with strong generalisation capabilities.

In a Taylor diagram, which provides a visual summary of several aspects of model performance, the correlation coefficient (radial distance from the origin), centred RMSE (contours) and standard deviation (distance along the x-axis) of the model predictions relative to the observed values are displayed simultaneously (Fig. 6). The DNN and RF models with high R^2 values and a smaller drop between training and testing are closer to the 'observed' point in the diagram, indicating better performance. The GBM model still performs well, but is slightly further away due to its lower R^2 values.

Overall, although the RF model shows a slight decrease in R^2 value from training to test, its high R^2 values indicate that it performs best and is good at capturing and predicting the variance in the water quality data. The DNN model also shows excellent performance and is comparable to the RF model, but the slightly higher R^2 of the RF model in training gives it an advantage. Despite its good fit, the GBM model is slightly outperformed by the other models, as R^2 decreases more strongly from training to testing, which can be crucial in model selection for predictive tasks where the highest accuracy is required.

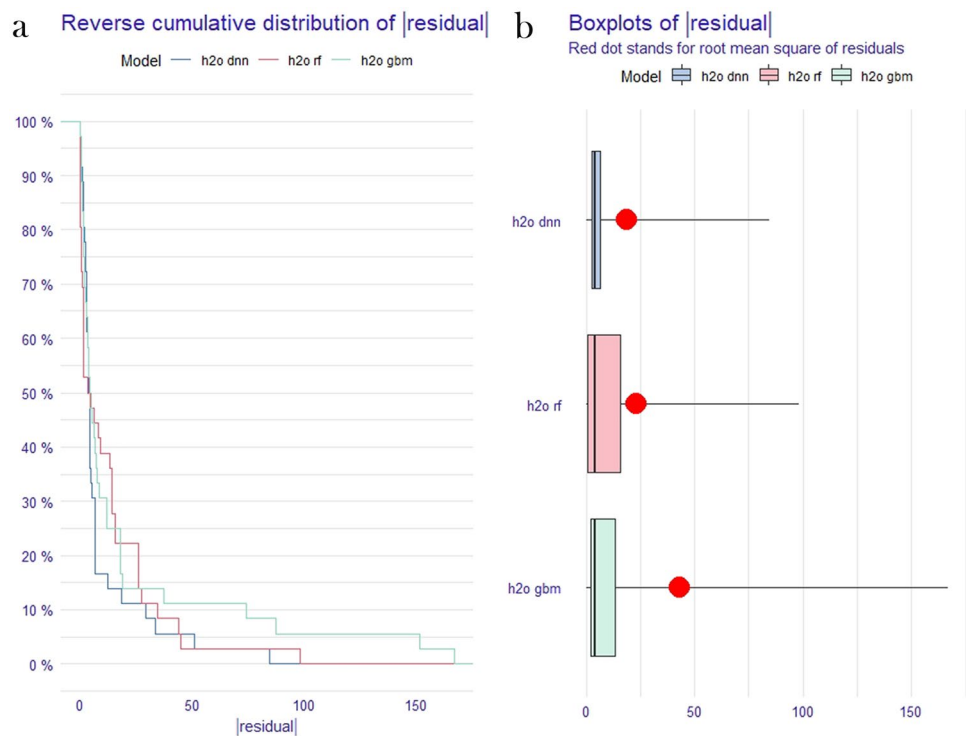
Interpreting optimised WQI-ML models using XAI for better decision making in water pollution management

The interpretation of optimised WQI-ML models using XAI is crucial for informed decision making in water pollution management. XAI facilitates understanding and confidence in ML models by providing insights into their decision-making processes.

Model diagnostic

This step involves assessing the overall statistical health and robustness of the model. Technique such as residual analysis of model assumptions are used to ensure that the model's predictions are reliable and consistent. The model diagnostic plots for three optimised machine learning models—h2o dnn, h2o rf and h2o gbm—used to predicting the WQI show a detailed scientific analysis that takes into account the inverse cumulative distribution and boxplot of the residuals for each model (Fig. 7). In the inverse cumulative distribution graph, the h2o rf model has a higher percentage of lower residuals compared to the other two models over the entire range of residuals, suggesting that the predictions of the h2o rf model are more consistently close to the true values. A higher percentage of smaller residuals indicates that the model better represents the underlying pattern in the data without overfitting. This is also supported by the boxplot, in which the h2o rf model has a median closer to zero and a smaller interquartile range, suggesting that most of its predictions are closer to the true values and vary less. The root mean square of the residuals, indicated by the red dot, is also lower for the h2o rf model than for the h2o dnn and h2o gbm models, supporting the assumption that the h2o rf model has smaller prediction errors on average. Conversely, the h2o gbm model appears to have a wider spread of residuals, as indicated by the step-down pattern in the inverted cumulative distribution plot and the larger interquartile range in the boxplot. Although it is not the worst model, it shows a higher variability in its predictions. Although the h2o dnn model has a relatively lower median of the residuals compared to the h2o gbm model, it still shows greater variability and a higher root mean square of the residuals than the h2o rf model. This indicates that although the h2o

Fig. 7 Model diagnostic analysis comparing ML algorithms for WQI prediction. Panel **a** shows the reverse cumulative distribution of absolute residuals for the optimised models (h2o dnn, h2o rf, h2o gbm). Panel **b** depicts boxplots of absolute residuals for each model, where red dots signify the root mean square of residuals, offering an at-a-glance assessment of each model's precision and consistency



dnn model makes several accurate predictions, it is on average less accurate than the h2o rf model. Overall, the h2o rf model stands out as the most consistent and accurate model for predicting WQI amongst the three models evaluated, as evidenced by both the distribution of its residuals and the root mean square error.

Model parts

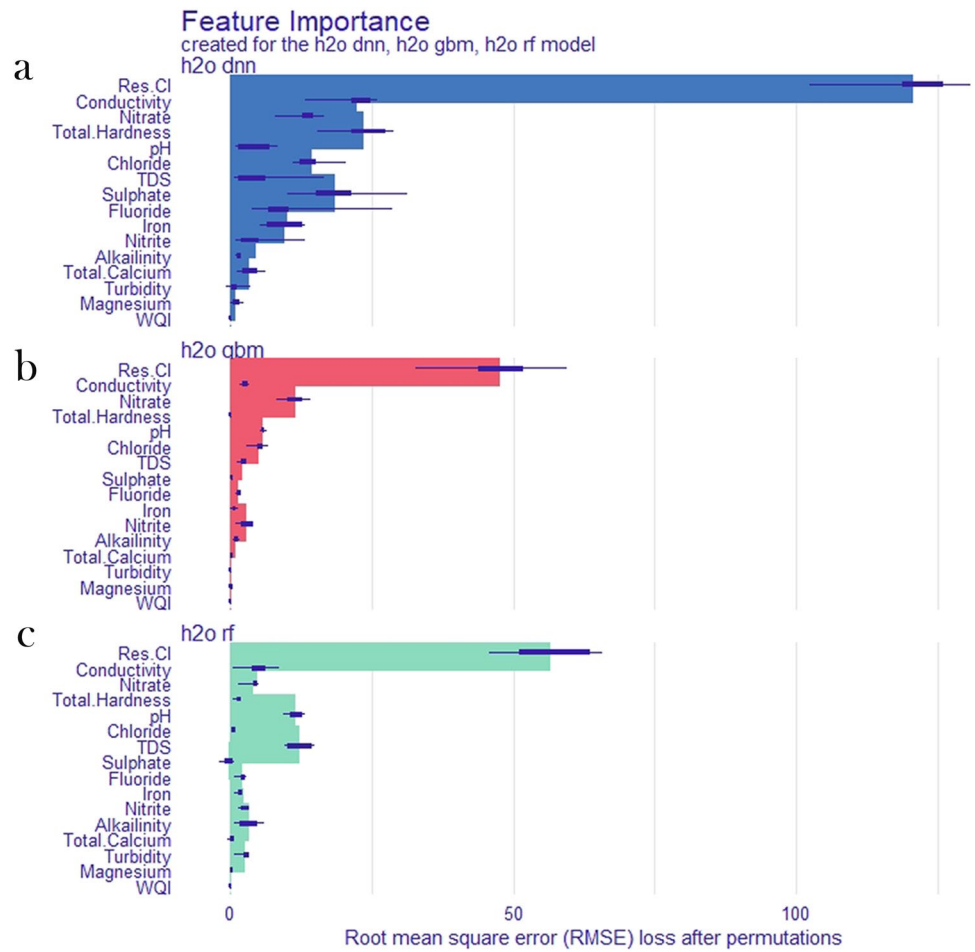
In this stage, the model is broken down into its individual components to assess the contribution of each feature to the prediction. Technique such as permutation-based feature importance ranking can be used to help understand which parameters are most influential in determining water quality and should therefore be prioritised in management strategies. Analysing the feature importance plots for the machine learning models, like h2o dnn, h2o gbm and h2o rf, we can identify the most and least influential water quality parameters for predicting the WQI (Fig. 8). For the h2o dnn model, the most important influencing factors are residual chlorine (Res.Cl), conductivity, nitrate and total hardness, with residual chlorine being the most important. In contrast, magnesium, total alkalinity and nitrite are the least influential. The h2o gbm model categorises conductivity as the most critical parameter, followed by nitrates, residual chlorine and pH, while magnesium, total alkalinity and turbidity have the least influence. The h2o rf model shows a similar pattern for the most influential parameters, with conductivity in first place, followed by total hardness, chloride and

sulphate. The least influential characteristics for this model are also magnesium, total alkalinity and iron. The overall assessment of all three models shows that conductivity and nitrate are consistently amongst the most influential parameters, indicating their critical role in water quality and their potential as primary indicators of WQI. The least influential parameters, such as magnesium, may be less variable or have less direct impact on water quality in the context of these models. This analysis is of great importance for water management decisions. It shows that monitoring and control of conductivity and nitrate should be prioritised in order to maintain or improve the WQI. This prioritisation can help to design more effective strategies for water treatment and management. This will ensure that resources are allocated to the most important factors affecting water quality, thereby improving environmental outcomes and public health.

Model profile

Profiling the model is about understanding how changes in input characteristics affect the output predictions of the model. This can be achieved through methods such as PDP, ALE and ICE. The result shows the PDPs for three ML models (h2o dnn, h2o gbm, h2o rf) used to predicting the WQI (Fig. 9). The PDPs illustrate the relationship between a set of values for a particular feature and the average prediction result of the model, holding all other features constant. In this way, we can understand the impact of a single feature on the predicted outcome. We can derive several relationships

Fig. 8 Feature importance analysis for **a** h2o dnn, **b** h2o gbm and **c** h2o rf models in predicting WQI



Partial Dependence profile

Created for the h2o gbm, h2o rf, h2o dnn model

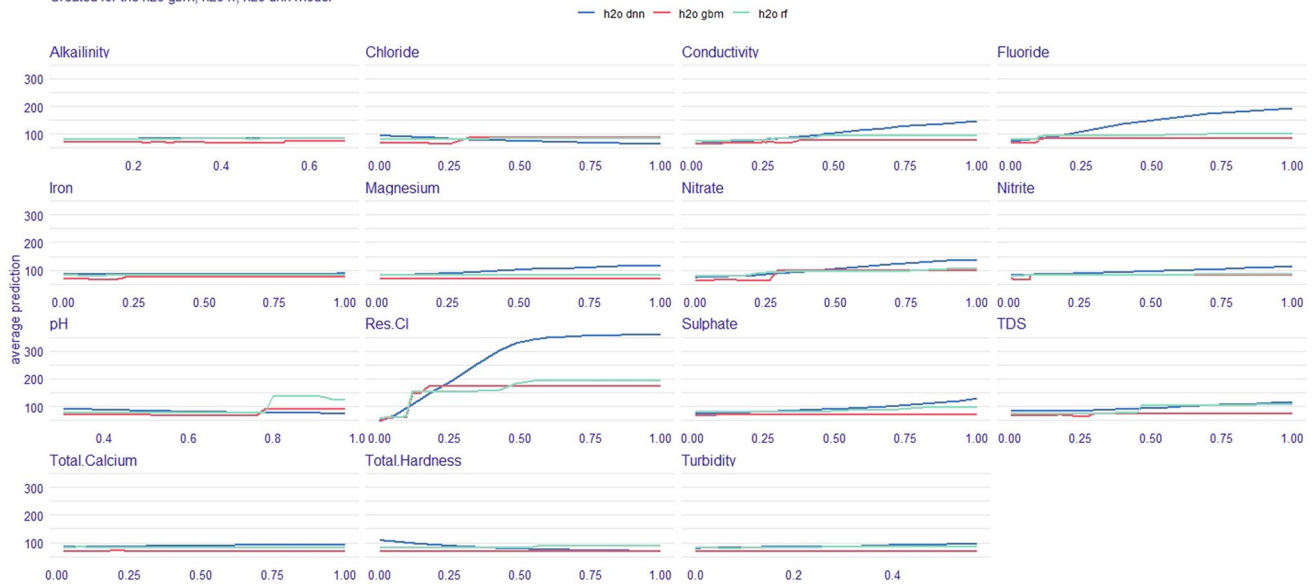


Fig. 9 Partial dependence plots for predictive features in water quality index modelling: The multi-line graphs represent the influence of individual water quality parameters on the average predictions of h2o dnn, h2o gbm and h2o rf models, with distinct trends

from the graphs. For example, the feature ‘Res.Cl’ (residual chlorine) in the h2o gbm model shows a sharp increase in the average predicted value as the feature value increases, indicating a strong dependence on this feature for predicting the WQI. Similarly, ‘sulphate’ shows a remarkable increase in predicted WQI with increasing feature value for the h2o dnn model. On the other hand, conductivity shows a relatively flat line for all three models, suggesting that changes in conductivity have a less pronounced effect on the model’s WQI prediction. This may indicate that the range of conductivity values in the dataset does not vary significantly or that the model does not consider it a strong predictor in conjunction with other characteristics. The PDPs can be particularly useful for managing water pollution by indicating which characteristics should be prioritised for monitoring and control. For example, if ‘Res.Cl’ and ‘sulphate’ are found to have a significant impact on WQI, as suggested by their steep PDP slopes for certain models, then measures to control these parameters in water bodies could be important to maintain good water quality. In contrast, features with shallow PDPs may be of lower priority in terms of immediate impacts on water quality, but could be important in a cumulative or contextual sense. These findings can inform water quality management strategies and enable targeted interventions that can be more cost-effective and focussed on the most influential water quality parameters. This targeted approach can help to mitigate the effects of water pollution more efficiently and effectively.

ALE plots are used to show how features influence the prediction of a model on average. They are an alternative to PDPs, which can treat correlated features more accurately

by considering the local effects of the features (Fig. 10). Using the ALE plots, we can observe the average change in WQI prediction as a function of the different feature values. For example, the feature ‘Res.Cl’ (residual chlorine) shows a clear positive slope for the h2o gbm and h2o dnn models, indicating that higher values of residual chlorine have an increasingly positive effect on the predicted WQI. Conversely, features such as ‘iron’, ‘magnesium’ and ‘turbidity’ show almost flat lines for all models, suggesting that these features have little to no effect on the average predicted WQI when other factors are taken into account. This lack of impact could be due to the fact that these features do not vary greatly within the dataset or their effects are masked by correlations with other features. These ALE plots can significantly improve the management of water pollution by identifying the most influential factors affecting the water quality predicted by the models. For example, if you know that residual chlorine has a significant positive impact on WQI predictions, you can adjust water treatment practises to ensure that chlorine concentrations are maintained at optimal levels. Similarly, understanding that iron and magnesium have minimal average impact on WQI predictions can shift attention from these parameters to more influential parameters, optimising resource allocation and intervention strategies. This targeted approach, based on robust machine learning analyses, can lead to more effective water management practises that result in better water quality and improved environmental and public health outcomes.

ICE charts are a refinement of PDPs and provide a more detailed view by plotting the predicted outcome against a feature for individual instances, thus taking into account

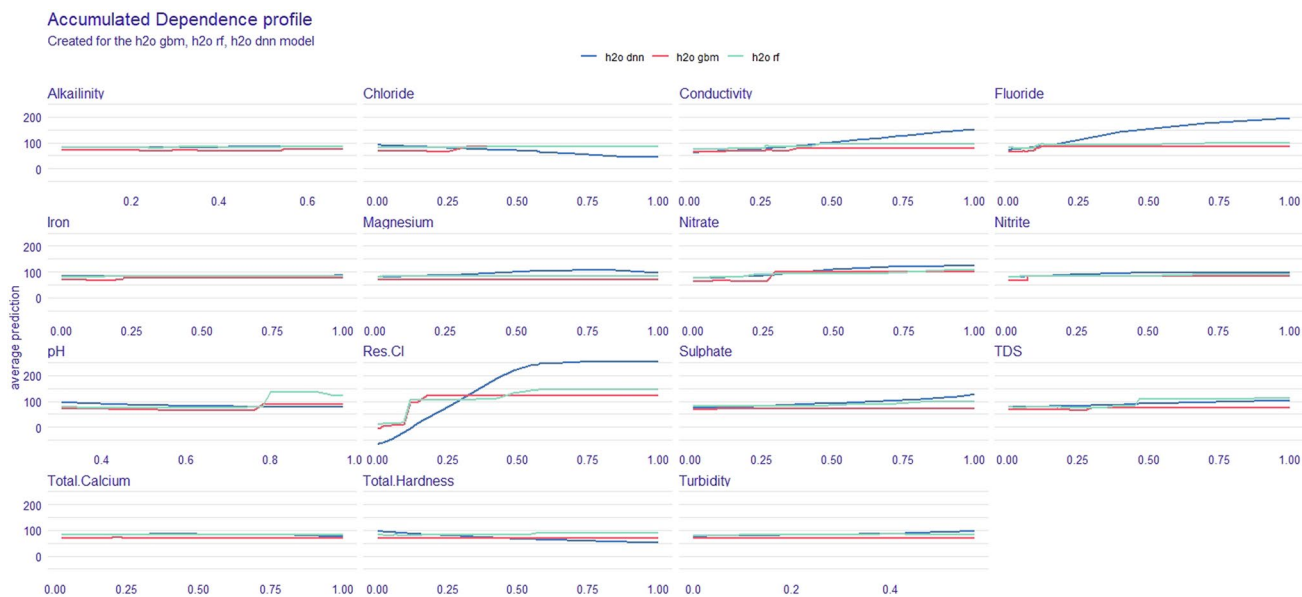


Fig. 10 ALE plots for key water quality parameters in h2o dnn, h2o gbm and h2o rf models

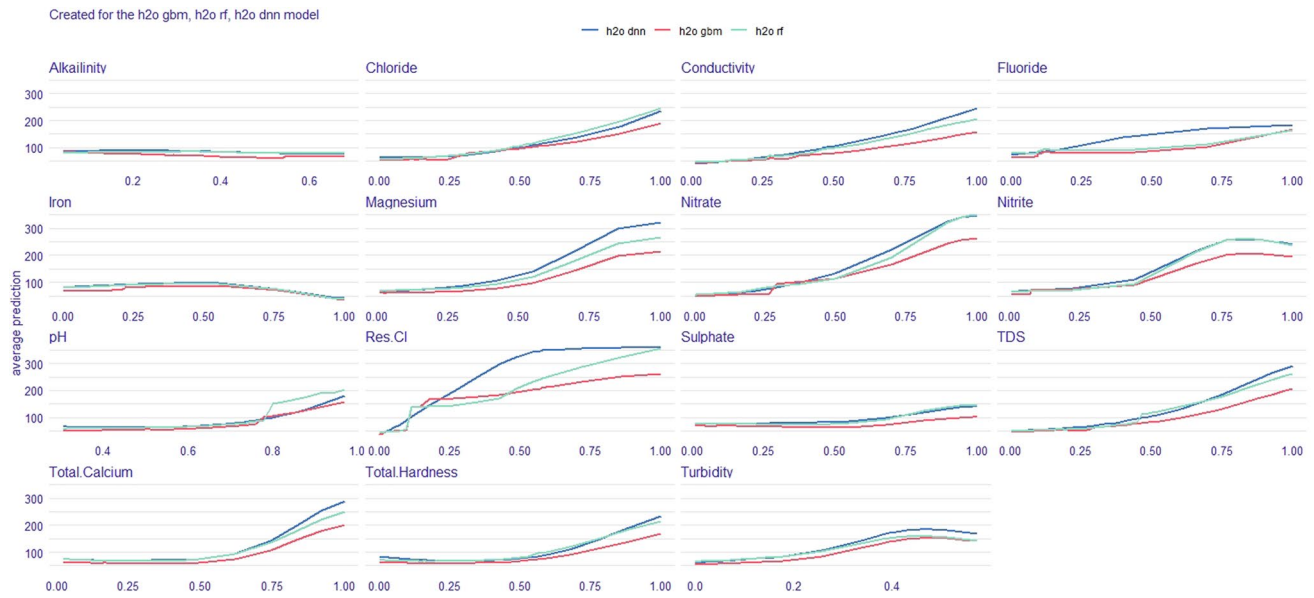


Fig. 11 ICE plots for key water quality parameters in h2o dnn, h2o gbm and h2o rf models

the heterogeneity of the data set (Fig. 11). Each line in an ICE diagram represents an instance from the data set and shows how the prediction changes with different values of the feature. Using the ICE plots, we can see that certain features such as ‘Res.Cl’ (residual chlorine), ‘conductivity’, ‘nitrate’ and ‘total hardness’ have a significant positive relationship with WQI prediction in all three models. This indicates that as the values for these characteristics increase, the predicted WQI generally increases, meaning that these characteristics are positively correlated with water quality. In particular, ‘Res.Cl’ and ‘conductivity’ appear to have a particularly strong and consistent influence on the model predictions, as shown by the steep slope of their ICE lines. On the other hand, features such as ‘alkalinity’, ‘chloride’ and ‘iron’ show relatively flat ICE lines, suggesting that changes in these feature values do not noticeably alter the predicted WQI, at least not within the observed range of the dataset. The insights gained from the ICE plots can be invaluable for decision making in water pollution management. For example, the strong positive correlation of ‘Res.Cl’ and ‘conductivity’ with high WQI predictions emphasises the importance of these parameters in water quality assessment. Water treatment plants and pollution control agencies may prioritise the monitoring and regulation of these parameters to ensure water safety and compliance with quality standards. In addition, the relatively flat ICE lines for ‘alkalinity’ and ‘iron’ suggest that they are less critical as control points for improving water quality within the monitored areas, allowing for more targeted and resource-efficient management strategies. Overall, ICE plots can help identify the

features that need to be closely monitored and actively managed to maintain or improve water quality.

The PDPs, ALE plots and ICE plots collectively provide a comprehensive overview of how different features influence WQI predictions for the machine learning models of h2o dnn, h2o gbm and h2o rf. These visualisations show that residual chlorine, conductivity, nitrate, total hardness and pH are the five most influential parameters that consistently affect WQI predictions. Residual chlorine and conductivity stand out in all models, indicating their strong predictive relationship with water quality, which could result in targeted management actions such as precise chlorination practices and monitoring of ion concentrations. The influence of nitrate points to the need to control agricultural runoff and industrial waste, while total hardness and pH changes may be crucial indicators of mineral balance and acid–base equilibrium in water bodies. Using these insights from ML-based WQI modelling enables data-driven decision making in water pollution management by focusing remediation efforts on the factors that have the greatest impact on water quality. This optimises resource allocation, improves the effectiveness of measures and ultimately ensures the safety and cleanliness of water resources.

Discussion

In our study, we conducted a comprehensive assessment of water quality in Saudi Arabia using a range of advanced ML models with XAI techniques. Our main focus was on developing a WQI with an entropy-weighted approach that

is aligned with WHO standards. We carefully analysed the performance of various ML models, including GBM, DNN and RF, to accurately predict the WQI. An important aspect of our study was the application of XAI, which allowed us to interpret the complex decision-making processes of these ML models. This approach provided deep insights into the most influential water quality parameters, enabling targeted and effective strategies for water pollution management. Our research not only provides a new perspective for assessing water quality in a region struggling with environmental challenges but also sets a precedent for the application of advanced technologies in environmental management.

The approach of our study to assess water quality in Saudi Arabia involved a comprehensive analysis using an entropy-weighted WQI in conjunction with World Health Organization standards. This methodology allowed us to capture the overall state of water quality in different locations with a single value. Our results showed a mean WQI of 188.14 and a high standard deviation of 794.92, indicating considerable variability in water quality across the sampled locations. Compared to previous studies in Saudi Arabia, our results show a more differentiated picture of water quality. For example, Alsubih et al. (2022) reported that the water from the dams was generally suitable for irrigation, except for problems with sodium content and adsorption ratio. This is in contrast to our results, where a significant proportion of water samples (over 35%) fell into the ‘unsuitable’ category for drinking water consumption, emphasising the severity of water quality problems for drinking water use. Masoud et al. (2022) found a similar variation in Drinking Water Quality Index (DWQI) scores, with a significant number of samples requiring treatment before consumption. Our study confirms this, with almost 40% of samples falling into the ‘excellent’ and ‘good’ categories, while a significant proportion still required treatment. The large differences in water quality can be attributed to several factors. Natural causes such as the presence of minerals and radionuclides, as highlighted by Haider et al. (2017), and anthropogenic activities such as the discharge of industrial and domestic wastewater and agricultural runoff have a significant impact on water quality. Alharbi et al. (2021) reported elevated concentrations of ions such as TDS, Ca^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} and F^- in central Saudi Arabia exceeding WHO drinking water standards. This indicates a combined influence of natural mineral dissolution and anthropogenic activities. Alfaleh et al. (2023) used an entropy-weighted WQI similar to our approach and found that wastewater discharge was a critical factor in reducing water quality in Ha'il, Saudi Arabia. This is consistent with our results, where the high variability in water quality indicates the influence of both natural and anthropogenic factors. Possible reasons for the deteriorating water quality in Saudi Arabia include over abstraction of groundwater leading to increased concentration of

pollutants, discharge of industrial and municipal wastewater, inadequate wastewater treatment plants and agricultural runoff leading to nutrient pollution. These factors, combined with the arid climate and limited renewable water resources, pose a major challenge to the preservation of water quality. Therefore, our study, in conjunction with previous research, emphasises the complex and multi-faceted nature of water quality problems in Saudi Arabia. The results emphasise the need for comprehensive water management strategies that take into account both natural and anthropogenic factors affecting water quality. These include improving wastewater treatment infrastructure, regulating industrial discharges, sustainable agricultural practises and careful groundwater management to ensure the long-term availability and quality of water resources in the region.

In our research, we have thoroughly investigated the performance of various machine learning models in predicting WQI in Saudi Arabia. The main motivation for developing a robust WQI model was to achieve high similarity with the WQI based on laboratory analytical parameters. A high-precision, data-driven model offers significant advantages: First, it requires only input data to automatically calculate the overall WQI, eliminating manual weighting calculations, integration of WHO standards, generation of standard index values, data aggregation and final WQI calculation. This process ensures both rapidity and precision in water quality assessment. Secondly, the development of robust data-driven models optimised by the interaction of various parameters facilitates a comprehensive understanding of these interactions at different sites. This aspect is usually difficult to recognise in the standard WQI calculation and the XAI is not generally applicable. However, in our optimised data-driven WQI prediction system, XAI can be used effectively to interpret these interactions and provide valuable insights into the behaviour of parameters contributing to WQI at all sites. This capability significantly improves decision-making processes and enables targeted pollution reduction strategies. In our study, the RF model (RF_grid1_model_79) showed the highest training R^2 of 0.99, indicating near-perfect prediction accuracy, which decreased slightly to 0.96 during testing. Although this decrease indicates a robust model, it is also an indication of possible problems with overfitting. The DNN model (DNNs_model_81) showed R^2 values of 0.98 during training and 0.97 during testing, indicating exceptional performance and generalisation ability. The GBM model (gbm_grid1_model_38) was robust but had a higher probability of overfitting with R^2 values of 0.92 for training and 0.90 for testing. Our results are consistent with the findings of Uddin et al. (2023) and Raheja et al. (2022), where DNN models performed better in predicting water quality. The lower error values and better accuracy of DNN models for WQI in these studies confirm our findings on the effectiveness of DNN models. In contrast, the studies by Lee

et al. (2022) and Sheikh Khozani et al. (2022) reported that LSTM models showed excellent performance in predicting water quality, which is consistent with our observations and emphasises the potential of advanced neural network models in this area. Regarding RF models, the results of our study are consistent with the research of Devi (2019), Hassan et al. (2021) and Mosavi et al. (2021). These studies reported high accuracy and robustness of RF models in predicting water quality, which is consistent with our observations. The differences in model performance can be attributed to several factors. First, the inherent characteristics of each model, such as the ability of RF to handle non-linear relationships and the ability of DNN to capture complex patterns, play a crucial role. Second, the type and quality of the data set, including the size, diversity and distribution of water quality parameters, significantly affect the performance of the model. Third, models such as GBM may overfit due to their high complexity and sensitivity to the training dataset. Therefore, our research in conjunction with previous studies emphasises the effectiveness of advanced ML models such as RF and DNN in accurately predicting WQI. The findings from these studies are crucial for the development of efficient water quality monitoring systems and enable policy makers and environmental managers to make informed decisions based on reliable predictions. The consistent performance of these models in various studies emphasises their potential in addressing complex environmental problems such as water quality assessment.

In our study, the application of XAI facilitated the interpretation of optimised ML models for WQI and improved decision making in water pollution management. In particular, the use of XAI enabled a clear understanding of how these models make their predictions, which increased confidence in their results. Our diagnostic analysis has shown that the RF model is particularly robust, showing the most accurate and consistent WQI predictions with the lowest average prediction errors. This consistency suggests that the RF model should play a central role in monitoring as it provides reliable assessments to guide environmental policy. The identification of conductivity and nitrate as the most influential parameters in determining water quality is particularly significant. Based on these results, policy makers should prioritise the regulation and continuous monitoring of nitrate levels and conductivity in water bodies. Introducing stricter controls on agricultural runoff and industrial discharges, which are the main sources of nitrates and salts that affect conductivity, could be an effective strategy. In addition, this prioritisation helps to optimise resource allocation and ensures that monitoring and mitigation measures are targeted where they are most needed. In addition, our model profiling using techniques such as PDP, ALE and ICE plots provided detailed insights into the impact of various parameters on the WQI. These

insights should help in the development of targeted pollution reduction strategies, such as tailored treatments for specific pollutants identified as critical at different locations. The practical application of XAI in our study is in line with recent research findings such as those of Park et al. (2022b), Alshehri and Rahman (2023) and Mia et al. (2023), which collectively emphasise the utility of XAI in improving the interpretability of ML models for more informed water management decisions. Building on our findings, it is recommended that similar XAI applications be integrated into national water quality monitoring systems to provide a framework for continuous improvement of environmental management practices. These systems should not only focus on routine assessments, but also enable adaptive management strategies that can respond to real-time data to optimise interventions to effectively improve water quality.

The novelty of our study lies in the comprehensive application of XAI to interpret ML models specifically optimised for WQI prediction in Saudi Arabia. Our approach not only identified the most influential water quality parameters, but also enabled a deeper understanding of their interactions and impacts on water quality at various locations. This level of detailed analysis is particularly novel given the unique environmental challenges in Saudi Arabia. Furthermore, our study advances the field by demonstrating the practical application of XAI in environmental management and setting a precedent for future research and application in similar contexts. Therefore, the application of XAI in the interpretation of WQI-ML models in our study represents a significant advance for the management of water pollution. By providing a clear understanding of how different parameters influence WQI, we have paved the way for more targeted and effective strategies to improve water quality, ultimately contributing to sustainable water resources management. Given the significant differences in water quality and the identification of key parameters such as conductivity and nitrate as critical factors, policy makers can focus on specific areas that require immediate attention. Implementing strict regulations on industrial discharges and agricultural runoff, which are the primary contributors to the elevated levels of these parameters, is critical. Furthermore, our study argues in favour of integrating advanced machine learning models into national water quality monitoring systems. This would enable real-time, data-driven decision making to respond quickly to scenarios of deteriorating water quality. Investing in the infrastructure to support these technological implementations, as well as educating and engaging the public in efforts to protect water and prevent pollution, could significantly improve the effectiveness of these measures. Ultimately, these strategies are in line with Saudi Arabia's Vision 2030 for environmental sustainability and ensure the protection of water resources for future generations.

Conclusion

This study represents a crucial step in improving water quality assessment methods in Saudi Arabia, an area severely affected by water pollution and scarcity. Our approach, which uses advanced data-driven modelling including machine learning and XAI, offers a paradigm shift in the interpretation and use of water quality data for sustainable management. The use of XAI in conjunction with advanced machine learning models such as random forest and deep neural networks introduces a new level of transparency and reliability to environmental science. This methodological innovation is fundamental as it enables a deeper understanding of prediction mechanisms and increases stakeholder confidence by making model decisions clear and understandable. Furthermore, the integration of these advanced technologies supports the development of robust, evidence-based strategies for water management. It enables policy makers and environmental managers to make more informed decisions tailored to address both the immediate and long-term challenges of water sustainability in arid regions. Our research methodology and its application provide valuable insights that could be adapted and replicated in other regions around the world facing similar environmental challenges.

While this study provides valuable insights into water quality management, it has limitations due to the small sample size and the influence of seasonal and geographical variability on water quality. Future research should aim to collect more diverse data across different seasons and regions to improve the model's robustness and predictive accuracy. Incorporating real-time data together with satellite imagery and IoT-based sensors could further refine our understanding of water quality dynamics. The potential of these approaches in Saudi Arabia, with its arid climate and scarce water resources, is considerable. Our findings provide an important framework for policy makers and environmental managers to make informed decisions for sustainable water management. This research paves the way for more sophisticated, data-driven strategies to address water scarcity and pollution, which are critical for environmental sustainability and public health in the region.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11356-024-33921-7>.

Acknowledgements The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Research Group under grant number RGP2/442/44. The authors are also thankful to the undergraduate students of college of engineering, King Khalid University to assist in data collections.

Author contribution Conceptualization: Saeed Alqadhi, Javed Mallick, Hoang Thi Hang, Majed Alsubih; data curation, formal analysis: Javed Mallick, Saeed Alqadhi, Hoang Thi Hang; funding acquisition: Saeed Alqadhi; methodology: Javed Mallick, Hoang Thi Hang; project

administration: Saeed Alqadhi, Javed Mallick; resources: Hoang Thi Hang, Majed Alsubih; software: Javed Mallick; supervision: Javed Mallick; validation: Saeed Alqadhi and Javed Mallick; writing—original draft: Saeed Alqadhi, Javed Mallick; writing—review and editing: Hoang Thi Hang, Majed Alsubih.

Funding Funding for this research was given under award numbers RGP2/442/44 by the Deanship of Scientific Research; King Khalid University, Ministry of Education, Kingdom of Saudi Arabia.

Availability of data and materials The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Ahmed IA, Talukdar S, Shahfahad, Parvez A, Rihan M, Baig MRI, Rahman A (2022) Flood susceptibility modeling in the urban watershed of Guwahati using improved metaheuristic-based ensemble machine learning algorithms. *Geocarto Int* 37(26):12238–12266
- Ahmed IA, Talukdar S, Baig MRI, Ramana GV, Rahman A (2024) Quantifying soil erosion and influential factors in Guwahati's urban watershed using statistical analysis, machine and deep learning. *Remote Sens Appl: Soc Environ* 33:101088
- Ahmed IA, Talukdar S, Islam ARMT, Rihan M, Malafaia G, Bera S, Ramana, GV, Rahman A (2023) Contribution and behavioral assessment of physical and anthropogenic factors for soil erosion using integrated deep learning and game theory. *J Clean Prod* 416:137689
- Alam R, Ahmed Z, Seefat SM, Nahin KTK (2021) Assessment of surface water quality around a landfill using multivariate statistical method, Sylhet, Bangladesh. *Environmental Nanotechnology, Monitoring & Management* 15:100422
- Aldhyani TH, Al-Yaari M, Alkahtani H, Maashi M (2020) Water quality prediction using artificial intelligence algorithms. *Appl Bionics Biomech* 2020
- Aldrees A, Awan HH, Javed MF, Mohamed AM (2022) Prediction of water quality indexes with ensemble learners: bagging and boosting. *Process Saf Environ Prot* 168:344–361
- Alfaleh A, Ben Khedher N, Alviz-Meza A (2023) Is the entropy-weighted water quality index a suitable index for evaluating the groundwater quality in Ha'il, Saudi Arabia? *Water Sci Technol* 88(3):778–797
- Alghamdi AG, Aly AA, Aldhumri SA, Al-Barakaha FN (2020) Hydrochemical and quality assessment of groundwater resources in Al-Madinah City, Western Saudi Arabia. *Sustainability* 12(8):3106
- Al-Hammad BA, Abd El-Salam MM (2016) Evaluation of heavy metal pollution in water wells and soil using common leafy green plant indicators in the Al-Kharj region, Saudi Arabia. *Environ Monit Assess* 188:1–10
- Alharbi TG (2018) Identification of hydrogeochemical processes and their influence on groundwater quality for drinking and agricultural usage in Wadi Nisah, Central Saudi Arabia. *Arab J Geosci* 11:1–12

- Alharbi TG, Zaidi FK (2018) Hydrochemical classification and multivariate statistical analysis of groundwater from Wadi Sahba Area in Central Saudi Arabia. *Arab J Geosci* 11:1–10
- Alharbi T, El-Sorogy AS, Qaysi S, Alshehri F (2021) Evaluation of groundwater quality in central Saudi Arabia using hydrogeochemical characteristics and pollution indices. *Environ Sci Pollut Res* 28(38):53819–53832
- Al-Omran AM, Aly AA, Al-Wabel MI, Sallam AS, Al-Shayaa MS (2016) Hydrochemical characterization of groundwater under agricultural land in arid environment: a case study of Al-Kharj, Saudi Arabia. *Arab J Geosci* 9:1–17
- Alshehri F, Rahman A (2023) Coupling machine and deep learning with explainable artificial intelligence for improving prediction of groundwater quality and decision-making in Arid Region, Saudi Arabia. *Water* 15(12):2298
- Alsubih M, Mallick J, Islam ARMT, Almesfer MK, Kahla NB, Talukdar S, Ahmed M (2022) Assessing surface water quality for irrigation purposes in some dams of Asir Region, Saudi Arabia using multi-statistical modeling approaches. *Water* 14(9):1439
- Al-Sulttani AO, Al-Mukhtar M, Roomi AB, Farooque AA, Khedher KM, Yaseen ZM (2021) Proposition of new ensemble data-intelligence models for surface water quality prediction. *IEEE Access* 9:108527–108541
- APHA (American Public Health Association) (1995) Standard methods for the examination of water and wastewater, 19th ed. American Public Health Association, Washington, DC
- Asadollah SBHS, Sharafati A, Motta D, Yaseen ZM (2021) River water quality index prediction and uncertainty analysis: a comparative study of machine learning models. *J Environ Chem Eng* 9(1):104599
- Bhagat SK, Tiyasha T, Tung TM, Mostafa RR, Yaseen ZM (2020) Manganese (Mn) removal prediction using extreme gradient model. *Ecotoxicol Environ Saf* 204:111059
- Dabbagh AE, Abderrahman WA (1997) Management of groundwater resources under various irrigation water use scenarios in Saudi Arabia. *Arab J Sci Eng* 22:247–264
- Devi SVSG (2019) Random forest advice for water quality prediction in the regions of Kadapa district. *Int J Innov Technol Explor Eng* 8:1–3
- Dodangeh E, Panahi M, Rezaie F, Lee S, Bui DT, Lee CW, Pradhan B (2020) Novel hybrid intelligence models for flood-susceptibility prediction: meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search. *J Hydrol* 590:125423
- Elbeltagi A, Pande CB, Kouadri S, Islam ARMT (2022) Applications of various data-driven models for the prediction of groundwater quality index in the Akot basin, Maharashtra, India. *Environ Sci Pollut Res* 29:17591–17605
- El-Shebli M, Sharrab Y, Al-Fraihat D (2023) Prediction and modeling of water quality using deep neural networks. *Environ Dev Sustain* 26:11397–11430
- Erickson ML, Elliott SM, Brown CJ, Stackelberg PE, Ransom KM, Reddy JE, Cravotta CA III (2021) Machine-learning predictions of high arsenic and high manganese at drinking water depths of the glacial aquifer system, northern continental United States. *Environ Sci Technol* 55(9):5791–5805
- Fallatah OA (2020) Groundwater quality patterns and spatiotemporal change in depletion in the regions of the Arabian Shield and Arabian Shelf. *Arab J Sci Eng* 45(1):341–350
- Fang Z, Wang Y, Peng L, Hong H (2021) Predicting flood susceptibility using LSTM neural networks. *J Hydrol* 594:125734
- Haider H, Al-Salamah IS, Ghumman AR (2017) Development of groundwater quality index using fuzzy-based multicriteria analysis for Buraydah, Qassim, Saudi Arabia. *Arab J Sci Eng* 42:4033–4051
- Hameed M, Sharqi SS, Yaseen ZM, Afan HA, Hussain A, Elshafie A (2017) Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Comput Appl* 28:893–905
- Hassan MM, Hassan MM, Akter L, Rahman MM, Zaman S, Hasib KM, Jahan N, Smrity RN, Farhana J, Raihan M, Mollick S (2021) Efficient prediction of water quality index (WQI) using machine learning algorithms. *Human-Centric Intell Syst* 1(3–4):86–97
- Hmoud Al-Adhaileh M, Wasellallah Alsaade F (2021) Modelling and prediction of water quality by using artificial intelligence. *Sustainability* 13(8):4259
- Ho JY, Afan HA, El-Shafie AH, Koting SB, Mohd NS, Jaafar WZB, Sai HL, Malek MA, Ahmed AN, Mohtar WHMW, El-Shafie A (2019) Towards a time and cost effective approach to water quality index class prediction. *J Hydrol* 575:148–165
- Irwan D, Ali M, Ahmed AN, Jacky G, Nurhakim A, Ping Han MC, AlDahoul N, El-Shafie A (2023) Predicting water quality with artificial intelligence: a review of methods and applications. *Arch Comput Methods Eng* 30(8):4633–4652
- Islam ARMT, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, Kuriqi A, Linh NTT (2021) Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci Front* 12(3):101075
- Khanfar AR (2008) Groundwater investigation in bel-ahmar, Assir, Kingdom of Saudi Arabia. *Saudi J Biol Sci* 15:289–296
- Khoi DN, Quan NT, Linh DQ, Nhi PTT, Thuy NTD (2022) Using machine learning models for predicting the water quality index in the La Buong River, Vietnam. *Water* 14(10):1552
- Khuan LY, Hamzah N, Jailani R (2002) Prediction of water quality index (WQI) based on artificial neural network (ANN). In: Student Conference on Research and Development. IEEE, Shah Alam, Malaysia, p 157–161. <https://doi.org/10.1109/SCORED.2002.1033081>
- Kim J, Seo D (2024) Three-dimensional augmentation for hyperspectral image data of water quality: an integrated approach using machine learning and numerical models. *Water Res* 251:121125
- Kim SB, Lee JS, Kim KT (2022) WQI class prediction of Sihwa Lake using machine learning-based models. *Sea J Korean Soc Oceanogr* 27(2):71–86
- Kogekar AP, Nayak R, Pati UC (2021) A CNN-BiLSTM-SVR based deep hybrid model for water quality forecasting of the river ganga. In: 2021 IEEE 18th India Council International Conference (INDICON). Guwahati, India, p 1–6. <https://doi.org/10.1109/INDICON52576.2021.9691532>
- Kouadri S, Kateb S, Zegait R (2021) Spatial and temporal model for WQI prediction based on back-propagation neural network, application on EL MERK region (Algerian southeast). *J Saudi Soc Agric Sci* 20(5):324–336
- Lap BQ, Du Nguyen H, Hang PT, Phi NQ, Hoang VT, Linh PG, Hang BTT (2023) Predicting water quality index (WQI) by feature selection and machine learning: a case study of An Kim Hai irrigation system. *Eco Inform* 74:101991
- Lee HW, Kim M, Son HW, Min B, Choi JH (2022) Machine-learning-based water quality management of river with serial impoundments in the Republic of Korea. *J Hydrol: Reg Stud* 41:101069
- Leong WC, Bahadori A, Zhang J, Ahmad Z (2021) Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). *Int J River Basin Manag* 19(2):149–156
- Mahato S, Pal S, Talukdar S, Saha TK, Mandal P (2021) Field based index of flood vulnerability (IFV): a new validation technique for flood susceptible models. *Geosci Front* 12(5):101175
- Mallick J, Singh CK, AlMesfer MK, Kumar A, Khan RA, Islam, Rahman A (2018) Hydro-geochemical assessment of groundwater quality in Aseer Region, Saudi Arabia. *Water* 10(12):1847
- Masoud MH, Rajmohan N, Basahi JM, Niyazi BA (2022) Application of water quality indices and health risk models in the arid

- coastal aquifer, Southern Saudi Arabia. *Environ Sci Pollut Res* 29(46):70493–70507
- Mei P, Li M, Zhang Q, Li G (2022) Prediction model of drinking water source quality with potential industrial-agricultural pollution based on CNN-GRU-Attention. *J Hydrol* 610:127934
- Mia MY, Haque ME, Islam ARMT, Jannat JN, Jion MMMF, Islam MS, Siddique MAB, Idris AM, Senapathi V, Talukdar S, Rahman A (2023) Analysis of self-organizing maps and explainable artificial intelligence to identify hydrochemical factors that drive drinking water quality in Haor region. *Sci Total Environ* 904:166927
- Mosavi A, Sajedi Hosseini F, Choubin B, Goodarzi M, Dineva AA, Rafiei Sardooi E (2021) Ensemble boosting and bagging based machine learning models for groundwater potential prediction. *Water Resour Manag* 35:23–37
- Nayan AA, Kibria MG, Rahman MO, Saha J (2020) River water quality analysis and prediction using GBM. In: 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT). IEEE, Dhaka, Bangladesh, p 219–224. <https://doi.org/10.1109/ICAICT51780.2020.9333492>
- Palkar S, Usgaonkar S, Ansari S (2022) Wq-Net: a deep neural network model for water quality prediction. In: OCEANS 2022-Chennai. IEEE, Chennai, India, p 1–6. <https://doi.org/10.1109/OCEANSChennai45887.2022.9775235>
- Park J, Ahn J, Kim J, Yoon Y, Park J (2022a) Prediction and interpretation of water quality recovery after a disturbance in a water treatment system using artificial intelligence. *Water* 14(15):2423
- Park J, Lee WH, Kim KT, Park CY, Lee S, Heo TY (2022b) Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. *Sci Total Environ* 832:155070
- Raheja H, Goel A, Pal M (2022) Prediction of groundwater quality indices using machine learning algorithms. *Water Pract Technol* 17(1):336–351
- Raheja H, Goel A, Pal M (2024) A novel approach for prediction of groundwater quality using gradient boosting-based algorithms. *ISH J Hydraul Eng* 1–12
- Saha TK, Pal S, Talukdar S, Debanshi S, Khatun R, Singha P, Mandal I (2021) How far spatial resolution affects the ensemble machine learning based flood susceptibility prediction in data sparse region. *J Environ Manag* 297:113344
- Sajib AM, Diganta MTM, Moniruzzaman M, Rahman A, Dabrowski T, Uddin MG, Olbert AI (2024) Assessing water quality of an ecologically critical urban canal incorporating machine learning approaches. *Ecol Inf* 80:102514
- Sakaa B, Elbeltagi A, Boudibi S, Chaffai H, Islam ARMT, Kulimushi LC, Choudhari P, Hani A, Brouziyne Y, Wong YJ (2022) Water quality index modeling using random forest and improved SMO algorithm for support vector machine in Saf-Saf river basin. *Environ Sci Pollut Res* 29(32):48491–48508
- Šandera J, Štych P (2024) Mapping changes of grassland to arable land using automatic machine learning of stacked ensembles and H2O library. *Eur J Remote Sens* 57(1):2294127
- Sarafaraz J, Kaleybar FA, Karamjavan JM, Habibzadeh N (2024) Predicting river water quality: an imposing engagement between machine learning and the QUAL2Kw models (case study: Aji-Chai, river, Iran). *Results Eng* 21:101921
- Saud AG, Abdullah SA (2009) Water resources and reuse in Al-Madinah. In: Proceedings of the International Conference on Water Conservation in Arid Regions (ICWCAR'09), Water Research Center-King Abdulaziz University, Jeddah, Saudi Arabia, pp 12–14
- Sheikh Khozani Z, Iranmehr M, Wan Mohtar WHM (2022) Improving Water Quality Index prediction for water resources management plans in Malaysia: application of machine learning techniques. *Geocarto Int* 37(25):10058–10075
- Sidek LM, Mohiyaden HA, Marufuzzaman M, Noh NSM, Heddam S, Ehteram M, Kisi O, Sammen SS (2024) Developing an ensemble machine learning model for predicting water quality index in Johor River Basin. *Environ Sci Eur* 36(1):67
- Singh KR, Dutta R, Kalamdhad AS, Kumar B (2019) Information entropy as a tool in surface water quality assessment. *Environ Earth Sci* 78:1–12
- Sinha S (2023) Water quality forecasting using a Novel Hybrid DNN-MBGD optimization and WAWQI technique for assessment of surface water quality index in 10 districts of Uttar Pradesh. *J Earth Syst Sci* 132(3):117
- Talukdar S, Ghose B, Shahfahad, Salam R, Mahato S, Pham QB, Linh NT, Costache R, Avand M (2020a) Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stoch Environ Res Risk Assess* 34:2277–2300
- Talukdar S, Singha P, Mahato S, Pal S, Liou YA, Rahman A (2020b) Land-use land-cover classification by machine learning classifiers for satellite observations—a review. *Remote Sens* 12(7):1135
- Talukdar S, Ahmed S, Naikoo MW, Rahman A, Mallik S, Ningthoujam S, Bera S, Ramana GV (2023a) Predicting lake water quality index with sensitivity-uncertainty analysis using deep learning algorithms. *J Clean Prod* 406:136885
- Talukdar S, Pal S, Naikoo MW, Rahman A (2023b) Exploring the flooding under damming condition in Punarbhaba River of India and Bangladesh. Floods in the Ganga–Brahmaputra–Meghna Delta. Springer International Publishing, Cham, pp 161–182
- Talukdar S, Bera S, Naikoo MW, Ramana GV, Mallik S, Kumar PA, Rahman A (2024) Optimisation and interpretation of machine and deep learning models for improved water quality management in Lake Loktak. *J Environ Manag* 351:119866
- Tiyasha, Tung TM, Yaseen ZM (2021) Deep learning for prediction of water quality index classification: tropical catchment environmental assessment. *Nat Resour Res* 30(6):4235–4254
- Uddin MG, Nash S, Rahman A, Olbert AI (2023) Assessing optimization techniques for improving water quality model. *J Clean Prod* 385:135671
- Verma M, Loganathan VA, Bhatt VK (2022) Development of entropy and deviation-based water quality index: case of River Ganga, India. *Ecol Ind* 143:109319
- Wu J, Liu H, Wei G, Song T, Zhang C, Zhou H (2019) Flash flood forecasting using support vector regression model in a small mountainous catchment. *Water* 11(7):1327
- Yusri HHH, Hassan SLM, Halim ISA, Abdullah NE (2022) Water quality classification using SVM And XGBoost method. In: 2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC). IEEE, Shah Alam, Malaysia, p 231–236. <https://doi.org/10.1109/ICSGRC55096.2022.9845143>
- Zali MA, Retnam A, Juahir H, Zain SM, Kasim MF, Abdullah B, Saadudin SB (2011) Sensitivity analysis for water quality index (WQI) prediction for Kinta River, Malaysia. *World Appl Sci J* 14:60–65

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.