Results in Engineering 32 (2026) 112205

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Research paper

Explainable soft-sensing of dissolved oxygen in a multi-station aquatic monitoring network using ensemble machine learning and neural networks

Md. Abdullah Al Mamun Hridoy [a], Matteo Bodini [b], Chiara Bordin [c,*], Khan Md Arafat Adnan [d], Mohd. Saifur Rahman [e], Abdullah Ibna Shawkat [f], Pakorn Ditthakit [g], Petra Schneider [h], Andleeb Masood [i], Md. Abdullah Al Mamun [j,k], Paolo Pastorino [l], Leonardo Goliatt [m]

[a] *Faculty of Fisheries, Sylhet Agricultural University, Sylhet 3100, Bangladesh*
[b] *Department of Pathophysiology and Transplantation, University of Milan, Via Francesco Sforza 35, Milan, 20122, Italy*
[c] *Department of Computer Science, UiT The Arctic University of Norway, 6050, Langnes, 9037, Tromsø, Norway*
[d] *Department of Chemical Engineering, Bangladesh University of Engineering & Technology, Dhaka 1200, Bangladesh*
[e] *Department of Agricultural Construction and Environmental Engineering, Faculty of Agricultural Engineering and Technology, Sylhet Agricultural University, Sylhet 3100, Bangladesh*
[f] *Faculty of Agricultural Engineering and Technology, Sylhet Agricultural University, Sylhet 3100, Bangladesh*
[g] *Center of Excellence in Sustainable Disaster Management, School of Engineering and Technology, Walailak University, Nakhon Si Thammarat, 80161, Thailand*
[h] *Department Water, Environment, Civil Engineering and Safety, Magdeburg-Stendal, University of Applied Sciences, Breitscheidstr. 2, D-39114 Magdeburg, Germany*
[i] *School of Physical Science and Technology, Southwest Jiaotong University, Chengdu, 611756, China*
[j] *Department of Fish Health Management, Sylhet Agricultural University, Sylhet 3100, Bangladesh*
[k] *Faculty of Veterinary Medicine, Universitas Brawijaya, Malang, East Java, 65151, Indonesia*
[l] *Experimental Zooprophylactic Institute of Piedmont, Liguria and Aosta Valley Via Bologna, 148, 10154 Turin, Italy*
[m] *Department of Computational and Applied Mechanics, Federal University of Juiz de Fora, Juiz de Fora, 36036-900, Brazil*

ARTICLE INFO

Keywords:
Dissolved oxygen
Soft-sensing
Ensemble machine learning
ExtraTrees
Multi-station monitoring
SHAP
Water quality

ABSTRACT

Dissolved oxygen (DO) is a critical water-quality indicator because it reflects aquatic system condition and directly informs monitoring and management decisions. This study develops and compares interpretable machine-learning and deep-learning models for multi-station DO soft-sensing from co-measured hydro-physical and water-quality variables. Model development was evaluated under a primary random split, a strict chronological train-past → test-future split, and leave-one-station-out (LOSO) validation to assess both temporal and spatial generalization. Baselines based on linear regression and seasonal mean prediction were also included. Among the evaluated models, ExtraTrees provided the strongest performance under the primary split ($R^2 = 0.999706$, RMSE = 0.020752 mg L^{-1}, MAE = 0.008211 mg L^{-1}), while chronological validation produced substantially lower but more realistic performance ($R^2 = 0.539698$, RMSE = 0.363559 mg L^{-1}, MAE = 0.268316 mg L^{-1}). Under LOSO validation, ExtraTrees retained comparatively strong station-transfer performance, with mean ± standard deviation values of $R^2 = 0.9295 \pm 0.1270$, RMSE = 0.2432 ± 0.1436 mg L^{-1}, and MAE = 0.1591 ± 0.0950 mg L^{-1}. Ablation analysis showed that oxygen saturation was highly informative, but strong predictive performance remained possible even after its exclusion. SHAP interpretation consistently identified temperature as the dominant negative predictor of DO, with seasonal and salinity/conductivity effects providing secondary explanatory value. The present work contributes an interpretable and deployment-oriented framework for multi-station DO soft-sensing, in which validation design, target-linked predictor assessment, and feature explanation are integrated to support more defensible model selection and environmental monitoring applications

* Corresponding author.
E-mail addresses: aamhridoy.fisheries@student.sau.ac.bd (Md.A.A.M. Hridoy), chiara.bordin@uit.no (C. Bordin), dpakorn@mail.wu.ac.th (P. Ditthakit), paolo.pastorino@izsplv.it (P. Pastorino).

https://doi.org/10.1016/j.rineng.2026.112205
Received 12 April 2026; Received in revised form 23 June 2026; Accepted 25 July 2026
Available online 26 July 2026

1. Introduction

Water is considered one of the most crucial natural resources for household needs, agricultural irrigation, and a wide range of industrial activities, and its quality status is measured based on three different fundamental properties, such as physical, chemical, and biological [1, 2]. Variations in these properties directly influence aquatic ecosystem health, public safety, and the efficiency of water-dependent economic sectors. Different types of water quality parameters are measured for each of the mentioned property to evaluate water quality. DO is among the most important parameters in measuring water quality status [3]. DO is the most critical variable in river systems, indicating key biogeochemical processes and directly influence the survival, distribution, and metabolic activity of aquatic organisms [4]. Effective freshwater ecosystem protection and water quality management are not feasible without an accurate prediction of DO concentrations [5,6]. DO is also an essential parameter for the health of aquatic ecosystems, and it is critical for monitoring water quality [7].

Recent DO studies have increasingly adopted machine-learning and deep-learning approaches for aquatic prediction tasks, particularly in structured water-quality datasets where nonlinear relationships among temperature, salinity, conductivity, turbidity, and seasonal forcing are expected. However, many of these studies emphasize predictive accuracy under a single validation design, making it difficult to distinguish between distribution-matched performance and true temporal or spatial generalization [8,9]. In addition, variables such as oxygen saturation may be strongly linked to DO and can potentially inflate predictive performance if not evaluated carefully through feature-ablation analysis. These issues indicate that strong accuracy alone is insufficient to establish methodological robustness in multi-station DO modeling.

Traditional DO prediction approaches primarily depend on in-situ measurements and laboratory analyses, including empirical temperature -DO relationships and biochemical oxygen demand-based models [10]. These models are constrained by limitations including sampling frequency, laboratory dependent conditions, and high environmental variability to provide accurate real-time and continuous DO predictions. Moreover, the relationship between DO concentration and its driving factors such as water temperature, pH, and turbidity is much complex and nonlinear, which makes it difficult for traditional statistical methods to evaluate these interactions accurately [10–12]. Accurate DO soft-sensing not only essential to the protection of water resources but also enables the early detection of potential environmental risks [13].

The contribution of the present study therefore lies not in proposing a fundamentally new algorithm but in providing a transparent and deployment-oriented evaluation of dissolved oxygen soft-sensing in a multi-station aquatic monitoring network. Specifically, the study compares ensemble machine-learning and deep-learning models under a common framework and evaluates their behavior using a primary random split, a strict chronological split, leave-one-station-out station transfer, oxygen-saturation ablation, baseline comparison, and SHAP-based interpretation. This combination of multi-regime validation and interpretability provides a more realistic basis for model comparison [14,15], feature assessment, and practical deployment than a single high-accuracy result alone.

DO dynamics are non-linear, reflect strong inter-variable relations, and significant temporal variability, which makes simple modeling approaches inadequate [16]. In recent years, ML techniques have been increasingly used to model complex processes in water resource systems [17]. Thus ML techniques have been proposed as an appropriate technique to capture the nonlinearity in any complex system [18]. These approaches including algorithms such as Random Forest, KNN, and Extra Trees have been increasingly adopted for the continuous prediction of water quality indicators, providing more accurate and dynamic environmental monitoring [13,19].

In recent years, DL has emerged as a widely adopted and very effective methodology across diverse scientific and engineering disciplines [20]. For instance, a hybrid model integrating gradient boosting decision trees (GBDT) with long short-term memory (LSTM) networks has been developed for DO prediction in pond systems, demonstrating high predictive accuracy and strong generalization capability [21]. Recent studies further highlight the increasing adoption of ML techniques in water quality prediction and assessment, reflecting a broader shift toward data-driven approaches in water resource management. These studies also emphasize the integration of Water Quality Index (WQI) frameworks with ML models for real-time evaluation of aquatic ecosystems, which is able to enhance predictive accuracy through advanced ML methodologies [19]. Another effective approach is station embedding, where each monitoring station is encoded as a learned vector that captures site-specific characteristics, also include sensor calibration biases and local climatic conditions [22]. Batch Normalization also takes a step towards reducing internal covariate shift, and in doing so surprisingly accelerates the training of deep neural nets. Thus, Batch Normalization makes it possible to use saturating non-linearities by preventing the network from getting stuck in the saturated modes [23].

The proposed model enhances predictive accuracy by integrating an ensemble of regression learners and incorporates SHapley Additive exPlanations (SHAP) to provide transparent and interpretable insights into the factors driving DO dynamics [24]. Station-specific effects can be effectively modeled through station embedding layers, which learn distinct vector representations for each monitoring station, enabling the model to account for heterogeneous site characteristics and systematic biases across different locations [22]. SHAP has been widely applied in water quality studies to identify key drivers such as temperature, pH, and conductivity, and to quantify their interactions within complex models, thereby enhancing interpretability and yielding actionable ecological insights [21,25].

Accordingly, the objectives of this study are fourfold: (i) to compare machine-learning and deep-learning models for dissolved oxygen estimation in a multi-station monitoring dataset; (ii) to assess model performance under both temporal and station-based generalization settings; (iii) to evaluate the predictive role of oxygen saturation through feature-ablation analysis and baseline benchmarking; and (iv) to interpret the dominant drivers of dissolved oxygen variability using SHAP-based explainable artificial intelligence. Through this design, the study aims to contribute a more rigorous and practically relevant perspective to recent DO soft-sensing research.

2. Materials and methods

2.1. Study context, monitoring network, and variables

This study used a multi-station aquatic monitoring dataset to estimate DO, mg L^{-1} from co-measured environmental variables. The final dataset contained 10,079 matched observations collected at multiple monitoring stations from 2023-07-17 to 2024-06-17. Because the source data represent a monitoring network rather than a controlled aquaculture pond experiment, the study is framed as a multi-station aquatic monitoring analysis with practical implications for aquaculture-related early warning rather than as a direct farm-scale intervention study. Candidate predictors included sampling depth (m), water temperature (°C), salinity, specific conductivity (mS cm^{-1}), oxygen saturation (%), turbidity (FNU), chlorophyll-a (μg L^{-1}), and station identity. Timestamp-derived seasonal variables, including month and day of year, were also generated for models that explicitly included seasonality. Station-level metadata are presented in Table 1, while variable-level summary statistics are provided in Supplementary Table S1.

2.2. Data preprocessing and feature preparation

All records were screened for consistency, duplicate entries, and missing values before model development. Rows with missing target

Table 1
Station metadata and observation coverage across the monitoring network.

n_observations	Start Date	End Date	Mean DO (mg L^{-1})	Mean Temperature (°C)	Mean Depth (m)	Mean Salinity
318	2023-08-18	2024-06-17	10.465	15.95	1.03	1.936
388	2023-07-17	2024-06-17	10.484	16.52	1.12	2.024
539	2023-07-17	2024-06-17	10.503	16.01	1.38	2.035
388	2023-07-17	2024-06-17	10.105	16.49	1.08	2.077
497	2023-07-17	2024-06-17	10.179	16.58	1.44	2.066
380	2023-07-17	2024-06-17	10.351	15.65	1.02	2.066
604	2023-07-17	2024-06-17	10.179	16.59	1.51	2.124
615	2023-07-17	2024-06-17	10.387	16.20	1.67	2.132
373	2023-07-17	2024-06-17	10.364	15.77	1.18	2.114
620	2023-07-17	2024-06-17	10.317	16.45	1.57	2.124
709	2023-07-17	2024-06-17	10.373	16.75	1.63	2.159
649	2023-07-17	2024-06-17	10.373	16.31	1.65	2.127
412	2023-07-17	2024-06-17	10.461	16.43	1.27	2.085
574	2023-07-17	2024-06-17	10.431	16.24	1.71	2.151
392	2023-07-17	2024-06-17	10.315	17.17	1.14	2.126
582	2023-07-17	2024-06-17	10.350	16.47	1.70	2.152
387	2023-07-17	2024-06-17	10.412	17.44	1.28	2.147
587	2023-07-17	2024-06-17	10.278	16.55	1.81	2.182
457	2023-07-17	2024-06-17	10.453	16.48	1.34	2.150
608	2023-07-17	2024-06-17	10.361	16.64	1.92	2.153

values were excluded from the analysis, and predictor completeness was checked before model fitting. Exploratory analysis was conducted using correlation matrices, boxplots, frequency distributions, cumulative distribution plots, and temporal trend visualizations in order to characterize the variable structure and identify potential skewness, outliers, and inter-variable dependencies. For scale-sensitive models, including KNN and neural networks, numerical predictors were standardized using statistics derived only from the training data, and the same scaling parameters were subsequently applied to validation and test subsets. Tree-based models were fitted on the original variable scale. Station identity was incorporated either as a categorical input or, in the station-aware neural network, as a learned embedding.

Because oxygen saturation may be physically and, in some monitoring systems, operationally linked to DO, its role was treated with particular caution. In addition to the full predictor set, dedicated ablation experiments were designed to evaluate model performance after excluding oxygen saturation and when using oxygen saturation as the sole predictor.

2.3. Training, validation, and testing strategy

To ensure transparent evaluation, model development was conducted under explicitly defined training, validation, and testing procedures. The principal analysis used an 80:20 random split, where 80% of the observations were used for model development and 20% were reserved for external testing. Within the training partition, hyperparameter tuning for machine-learning models was conducted using 5-fold cross-validation, whereas deep-learning models used a validation holdout derived from the training portion. No information from the final test set was used during preprocessing, scaling, or model selection. The complete evaluation design is summarized in Table 2.

To assess more realistic generalization, two additional validation regimes were used. First, a chronological split was implemented in a train-past → test-future manner. In this setting, 8,063 observations collected from 2023-07-17 to 2024-04-29 were used for training, while 2,016 later observations from 2024-04-29 to 2024-06-17 were reserved for testing. Second, a leave-one-station-out (LOSO) design was applied to assess spatial transferability, where each station was withheld in turn as an external test set and all remaining stations were used for model training. This procedure was repeated across 20 unique stations. These complementary validation settings were introduced to evaluate both temporal and spatial generalization more rigorously than a single random split alone.

2.4. Baseline models and feature-ablation experiments

To contextualize model performance, simple benchmark models were included alongside the advanced machine-learning and deep-learning approaches. Two baseline models were considered: a multiple linear regression model and a seasonal mean predictor based on monthly average DO values. These baselines were selected to determine whether the more flexible nonlinear models provided meaningful gains beyond

Table 2
Evaluation strategy used for model development and testing.

Evaluation strategy	Train description	Validation description	Test description	Condition/parameters
Primary random split	80% random sample	5-fold cross-validation for machine-learning hyperparameter tuning; validation holdout from the training partition for deep-learning models	20% random sample	Random shuffle applied to the full dataset
Chronological split	First 8,063 observations ordered by date (2023-07-17 to 2024-04-29)	None	Last 2,016 observations ordered by date (2024-04-29 to 2024-06-17)	Train-past → test-future evaluation for temporal generalization
Leave-one-station-out (LOSO)	All stations except one withheld station	None	Single left-out station	Repeated iteratively across 20 unique stations to assess spatial transferability
Ablation: full features	80% random sample	None	20% random sample	All predictors included, including oxygen saturation (%)
Ablation: without oxygen saturation	80% random sample	None	20% random sample	Oxygen saturation (%) excluded from the predictor set
Ablation: oxygen saturation only	80% random sample	None	20% random sample	Oxygen saturation (%) is used as the sole predictor

transparent parametric and climatological approaches.

To evaluate the predictive role of oxygen saturation, three feature configurations were considered. The first configuration used the full feature set, including oxygen saturation. The second excluded oxygen saturation entirely. The third used oxygen saturation as the sole predictor. These ablation experiments were designed to test whether the apparent predictive skill depended primarily on oxygen saturation or whether useful DO estimation could still be achieved from the remaining hydro-physical and seasonal variables. Final hyperparameter settings and training configurations of the evaluated machine-learning and deep-learning models are provided in Supplementary Table S3.

2.5. ML models

2.5.1. Random forest

Random Forest is an ensemble regression method that constructs multiple decision trees from bootstrap samples of the training data and aggregates their predictions by averaging. At each split, only a random subset of predictors is considered, which reduces correlation among trees and improves generalization for nonlinear tabular datasets. In the present study, Random Forest was used as a benchmark ensemble learner because of its widespread use in environmental prediction problems and its ability to represent nonlinear predictor–response relationships without extensive feature engineering. Final hyperparameter settings and tuning details should be reported in Table S3. We denote the model as RF_k, with K = d as the default and K = * indicating the best value selected from K = 1,…,n [26].

2.5.2. Extra-Trees regressor

The Extra-Trees regressor is an ensemble method that builds many highly randomized regression trees, each trained on different random samples and random splits. For a new input x, every tree $t$in the ensemble produces its own prediction $\hat{y}^{(t)}(x)$. The final output of the model is obtained by averaging these individual tree predictions, which helps to reduce variance and improve generalization compared with a single decision tree. For a single-target regression problem, the ensemble prediction is given by

$$\hat{y}(x) = \frac{1}{T}\sum_{t=1}^{T} \hat{y}^{(t)}(x),$$

where $T$is the total number of trees in the forest [27].

2.5.3. KNN regression

KNN regression predicts the target value of a new sample from the average of its nearest neighbors in feature space. Because KNN is sensitive to variable scale, all numerical predictors used in this model were standardized before fitting. KNN was included as a distance-based nonlinear benchmark to compare with the ensemble tree models. Final hyperparameter settings and tuning details should be reported in Table S3.

2.6. DL models

2.6.1. Baseline MLP

The baseline DL model is a standard MLP trained with backpropagation. In this architecture, several fully connected layers transform the input features through nonlinear activation functions, and the network weights are iteratively updated to minimize a loss function using gradient descent. This follows the classic backpropagation framework, where hidden layers learn internal representations that map inputs to outputs effectively [28].

2.6.2. Deep MLP + BatchNorm

The second model extends the baseline MLP by increasing depth and inserting Batch Normalization layers between linear transformations and nonlinear activations. Batch Normalization stabilizes the distribution of activations within each mini-batch, which reduces internal covariate shift and allows for faster training with higher learning rates and improved convergence in deep networks. This design follows the Batch Normalization method proposed by Ioffe and Szegedy for accelerating deep neural network training [23].

2.6.3. Station embedding MLP

The third model incorporates a learnable embedding vector for each measurement station, which is concatenated with the numerical input features and passed through an MLP. These station embeddings allow the network to capture station-specific characteristics in a low-dimensional latent space while still sharing information across all stations. This idea is inspired by recent work on station-embedding-based neural networks for station-level demand soft-sensing in carsharing systems, where station embeddings significantly improved predictive accuracy [29].

2.7. Model evaluation and interpretability analysis

Model performance was evaluated using the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), and, where relevant, mean squared error (MSE). For the LOSO analysis, mean ± standard deviation across withheld stations was reported to summarize spatial transferability. Performance under the primary random split, chronological split, baseline comparisons, and ablation settings was reported separately in order to avoid conflating in-sample and out-of-distribution validation regimes.

To improve interpretability, SHAP were used to characterize the direction and magnitude of predictor contributions for the best-performing model. Because oxygen saturation exhibited a strong predictive influence in the ablation analysis, SHAP results for the full-feature model were interpreted alongside the reduced-feature model results to avoid over-attributing target-linked information. The main SHAP interpretation in the present study focused on the ExtraTrees model, which provided the strongest overall performance.

3. Results and Interpretations

3.1. Descriptive statistics and station metadata

The monitoring dataset comprised 10,079 matched observations collected from 2023-07-17 to 2024-06-17 across 20 stations. Summary statistics for the measured variables are presented in The monitoring dataset comprised 10,079 matched observations collected from 2023-07-17 to 2024-06-17 across 20 stations. Across stations, mean DO values were centered around approximately 10.1–10.5 mg L^{-1}, while mean temperatures ranged from approximately 15.6 to 17.4°C, indicating a relatively coherent but spatially heterogeneous monitoring network. Variable-level summary statistics of the analysis dataset are provided in Supplementary Table S1, while Table 1 summarizes station-level metadata and observation coverage across the monitoring network.

3.2. Exploratory analysis

The exploratory assessment of the dataset provided an overview of the sampling intensity and the distributional behavior of key water-quality parameters across all monitoring stations.Salinity and specific conductivity showed an almost perfect positive correlation (r 1.0), while turbidity and chlorophyll-a were also strongly positively correlated (r 0.9). Temperature was positively correlated with oxygen saturation (r 0.8) but strongly negatively correlated with DO (r −0.8) (Fig. 1). Fig. 2 presents the station-wise frequency of observations for depth, temperature, salinity, specific conductivity, oxygen saturation, DO, turbidity, and chlorophyll-a. The frequencies show that all parameters were

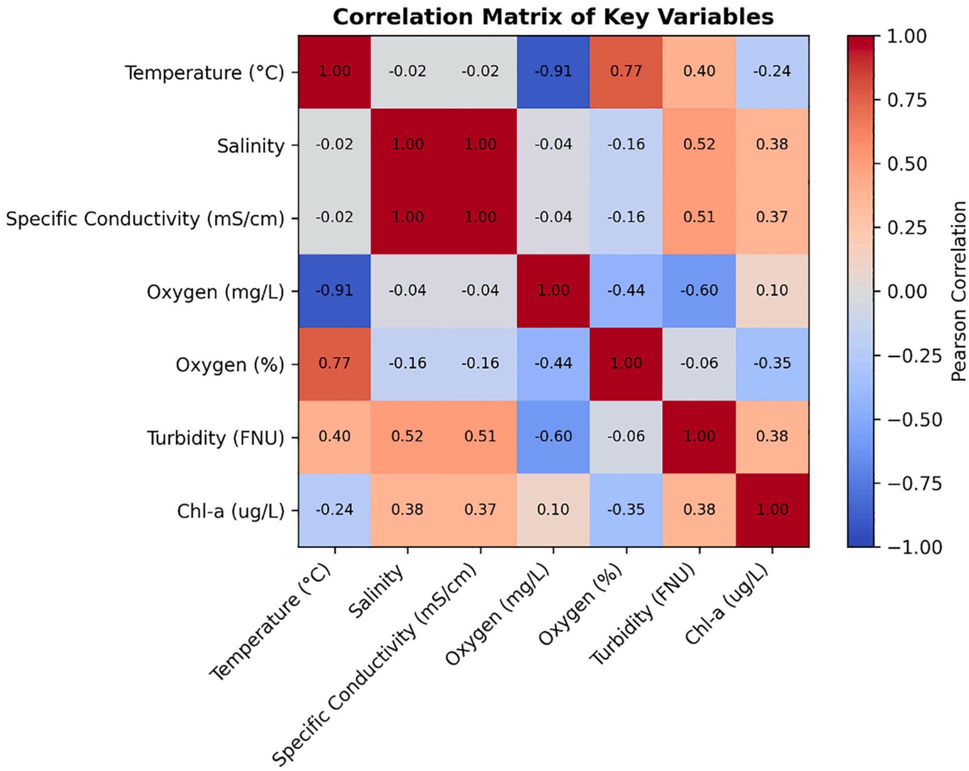

Fig. 1. Correlation matrix of key parameters.

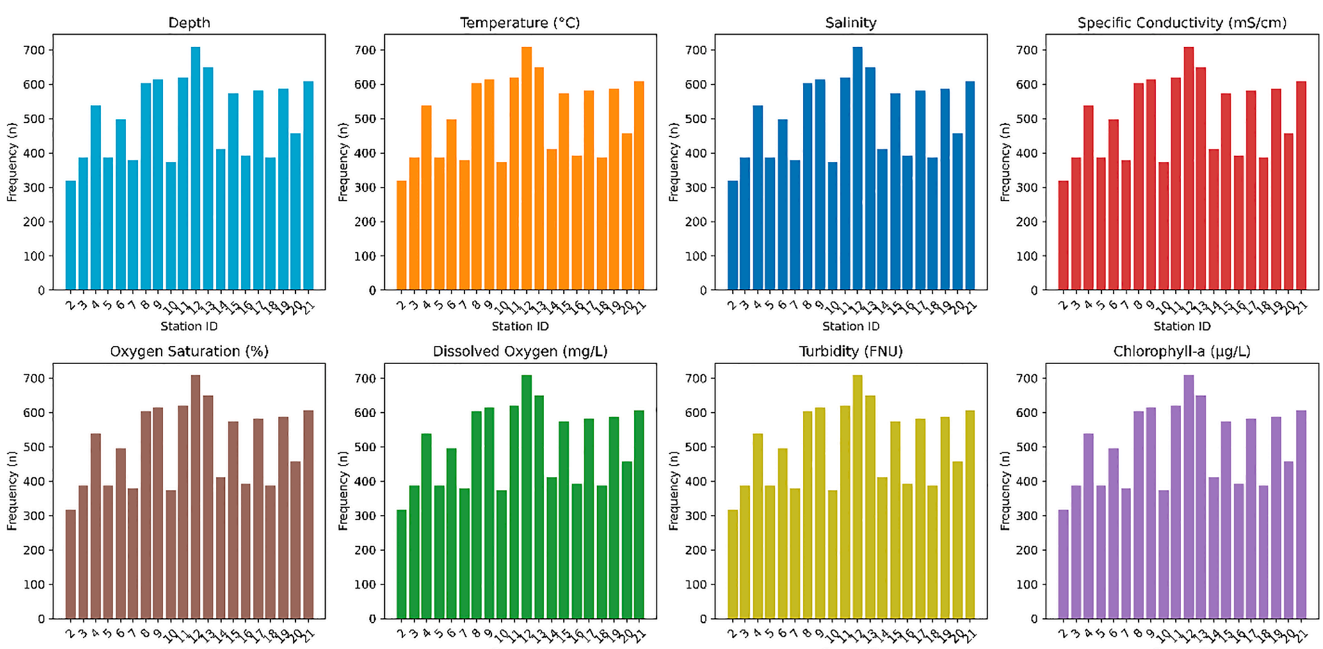

Fig. 2. Station-wise frequency of observations for all measured water-quality parameters (depth, temperature, salinity, specific conductivity, oxygen saturation, DO, turbidity, and chlorophyll-a).

measured consistently across stations, with only minor differences in the number of observations. This even distribution indicates balanced spatial coverage and supports robust comparisons among stations without concerns of sampling bias.

To further characterize the central tendencies and variability of the measured variables, Fig. 3 summarizes their distributions using boxplots. Depth exhibited a wide range, reflecting variation in water column structure and station-specific sampling depths. Temperature showed moderate variability, with most values clustered around the central range and fewer extreme values. Salinity and specific conductivity displayed comparatively broader spreads, indicating spatial heterogeneity in freshwater–marine mixing conditions. Oxygen saturation showed a wider distribution with several upper outliers, whereas DO (mg L^{-1}) was more tightly distributed around its median, suggesting stable oxygen

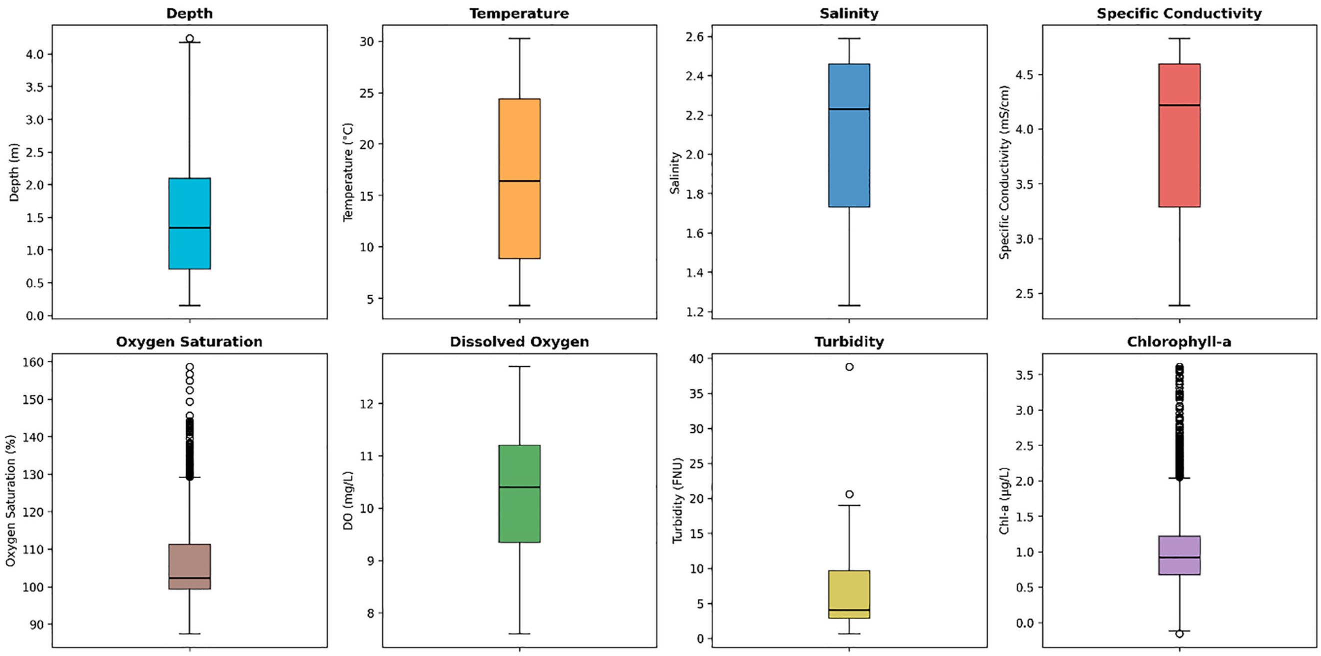

Fig. 3. Boxplots summarizing the distribution of depth and key water-quality parameters (temperature, salinity, specific conductivity, oxygen saturation, DO, turbidity and chlorophyll-a).

concentrations relative to saturation dynamics.

Turbidity and chlorophyll-a demonstrated moderate variability with occasional high outliers, indicating episodic increases in suspended particulate matter and phytoplankton biomass. These elevated values likely reflect localized biological or physical processes such as phytoplankton growth events or sediment resuspension.

The frequency distributions of the measured variables revealed clear patterns in the physical and chemical structure of the study area. Depth values were strongly skewed toward shallow waters, with most observations concentrated between 0.5 and 2.5 m, indicating that the monitoring stations were situated within a predominantly shallow estuarine system. Temperature exhibited a wide and multimodal distribution ranging from approximately 8 to 30 °C, reflecting sampling across multiple seasons and the strong seasonal thermal variability typical of coastal environments (Fig. 4). Salinity values varied between 1.25 and 2.5 PSU and displayed several distinct modes, suggesting fluctuating freshwater–marine mixing. Specific conductivity showed a similar multimodal pattern, consistent with the salinity trend, indicating that ionic composition was primarily driven by variations in salinity. Oxygen saturation was largely concentrated around 90–120%, and DO ranged from 8 to 12 mg L^{-1}, illustrating generally well-oxygenated conditions throughout the sampling period. Turbidity was mostly low (<10 FNU) with occasional peaks at higher values, suggesting episodic resuspension or sediment-rich inflows. Chlorophyll-a concentrations showed a

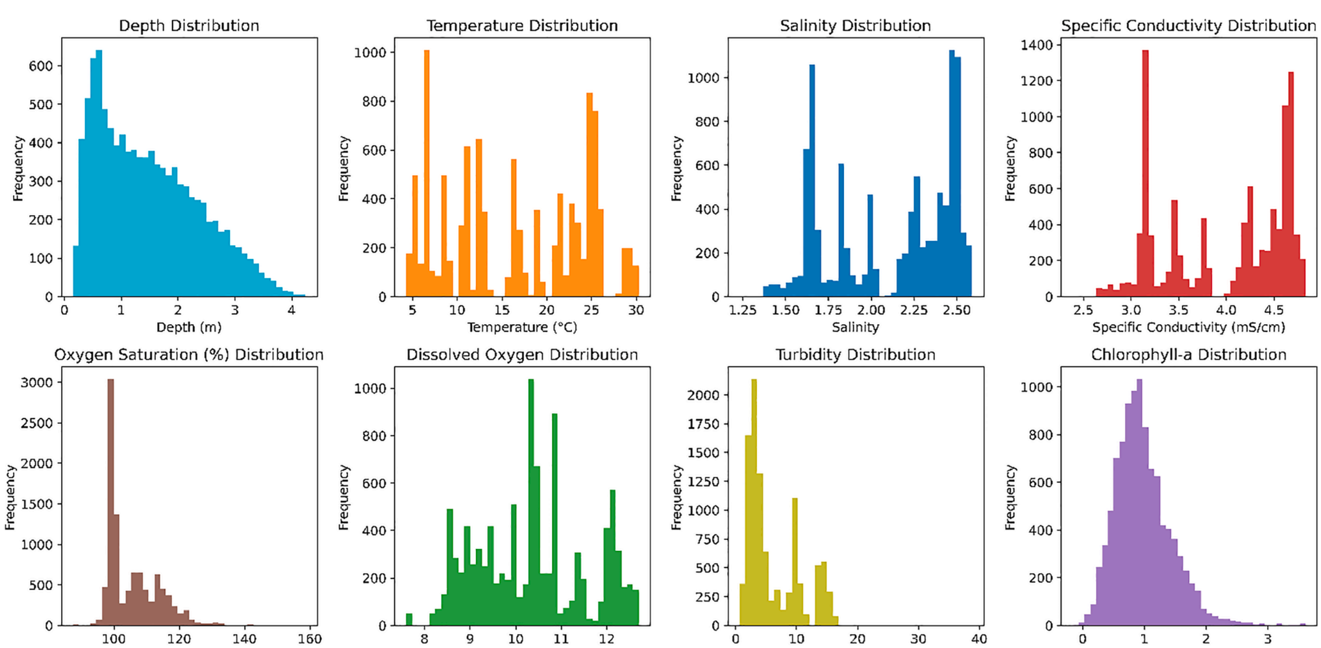

Fig. 4. Frequency distributions of depth and physico-chemical parameters.

narrow, unimodal distribution centered around 0.5–1.5 μg L^{-1}, indicating relatively low phytoplankton biomass typical of light-limited or turbid estuarine systems.

Depth fluctuated moderately through time, varying between 1.2 and 1.7 m, which likely reflected tidal influence and seasonal changes in water availability. Temperature decreased steadily from around 30 °C in July 2023 to roughly 12 °C in January 2024 before rising again, mirroring the regional transition from summer to winter and the subsequent pre-monsoon warming. Salinity exhibited a pronounced decline from approximately 2.4 PSU in July 2023 to around 1.5 PSU in early 2024, indicating progressive freshening of the system during the post-monsoon and dry seasons. Specific conductivity showed an identical decreasing trend, reaffirming its direct relationship with salinity. Oxygen saturation fluctuated considerably, with values exceeding 120%, likely due to elevated photosynthetic activity or enhanced solubility during cooler periods, while lower values around 80–90% in intervening months suggested reduced biological oxygen production or increased respiration. DO followed a seasonal pattern opposite to temperature, increasing to nearly 12 mg L^{-1} during cooler months and declining slightly thereafter, consistent with the temperature dependence of oxygen solubility. Turbidity varied irregularly through time, with the highest values recorded during mid-2023, likely reflecting monsoon-related sediment inputs, and reduced values in subsequent months (Fig. 5). Chlorophyll-a showed a distinct post-monsoon peak in October–November 2023 before declining toward early 2024, indicating seasonal enhancement of phytoplankton productivity associated with improved nutrient availability and moderate mixing conditions.

The cumulative distribution plot of DO shows a gradual increase in cumulative probability from approximately 8 to 12.5 mg L^{-1}, indicating that DO concentrations within this range account for the full spectrum of observed values. The lower tail of the curve remains nearly flat until around 8 mg L^{-1}, signifying that only a small fraction of measurements fall near the lower end of the distribution. The curve then rises steadily between 9 and 11 mg L^{-1}, reflecting that most observations cluster within this interval, which represents the dominant DO regime of the system (Fig. 6). A steeper incline near 10.5–11.5 mg L^{-1} suggests a dense concentration of measurements in this range, consistent with the well-oxygenated conditions shown in the frequency distribution. The curve approaches a cumulative probability of 1 near 12.5 mg L^{-1}, indicating that the highest DO values rarely exceed this level.

The 3D scatter plot illustrates the combined variation of chlorophyll-a, DO, and sampling stations, with depth represented through the colour gradient. A broad spread of points is evident across all stations, indicating considerable spatial heterogeneity in both chlorophyll-a, and DO. Higher chlorophyll-a values tend to cluster with moderately elevated DO concentrations, suggesting that phytoplankton biomass contributes to oxygen production through photosynthesis (Fig. 7). This pattern is reflected in the slightly upward trend in DO with increasing chlorophyll-a across stations. The colour coding by depth reveals that shallower samples, represented by lighter colours, generally correspond to higher chlorophyll-a values and, in many cases, higher DO. This is consistent with the enhanced light availability in shallow waters, which favours photosynthetic activity and subsequent oxygen generation. In contrast, deeper samples, indicated by darker colours, are more frequently associated with lower chlorophyll-a, and moderately lower DO, suggesting reduced primary production at depth. Spatial variability across stations is also apparent, with some stations showing a wider range of chlorophyll-a, and DO values, indicating localized differences in environmental conditions such as nutrient availability, hydrodynamics, or freshwater influence.

A clear seasonal signal is visible, with lower DO values (typically 8–10 mg L^{-1}) dominating the early sampling months from July to September 2023, particularly across the upper and mid-estuary stations (Fig. 8). From October onward, DO progressively increases, with the highest concentrations (12–13 mg L^{-1}) observed during December 2023 to January 2024 across nearly all stations.

Bivariate relationships indicate that DO exhibits a clear negative association with temperature, with the densest clusters showing lower DO at higher temperatures and higher DO at cooler conditions, consistent with temperature-dependent oxygen solubility. DO also shows a weak to moderate negative trend with salinity and specific conductivity, with lower DO values more frequent at the high end of both gradients, suggesting freshening generally coincides with enhanced oxygenation. In contrast, depth displays no consistent monotonic relationship with DO across the sampled range; DO remains comparatively uniform from very shallow to deeper casts, implying limited stratification within the observed depths and/or effective vertical mixing. As expected, oxygen saturation and DO are strongly and positively correlated, reflecting their

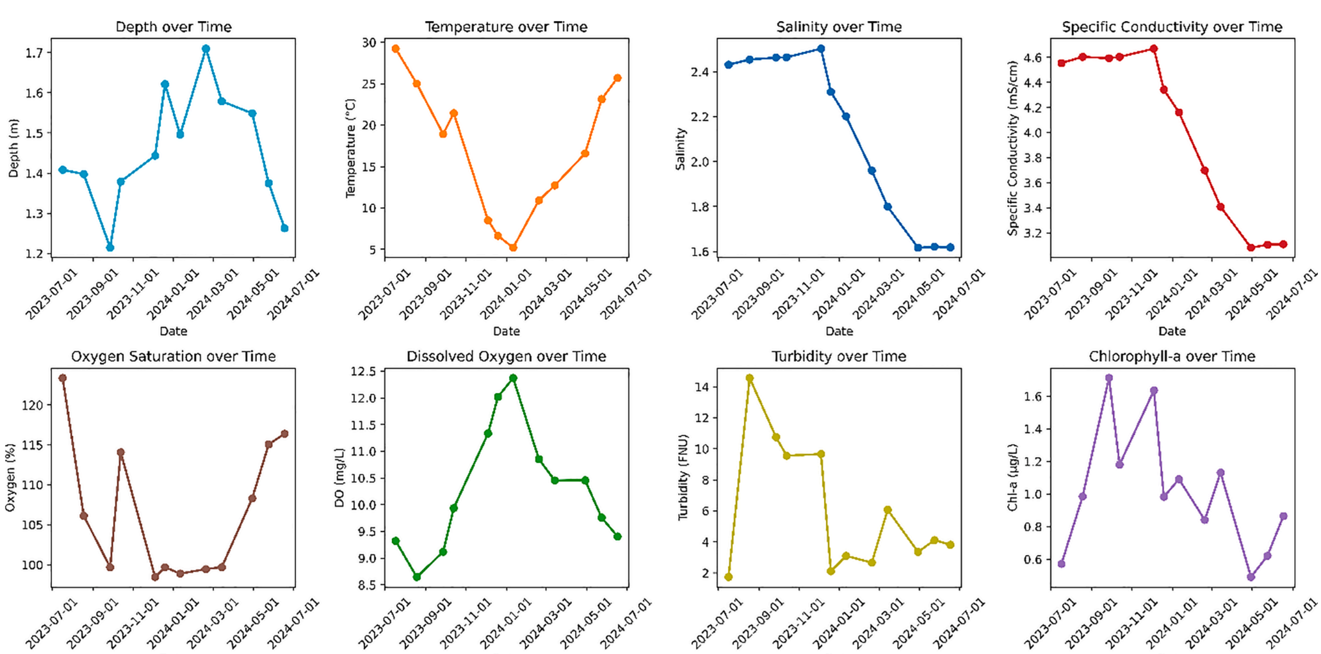

Fig. 5. Temporal variation of depth and key water-quality parameters, averaged across all stations.

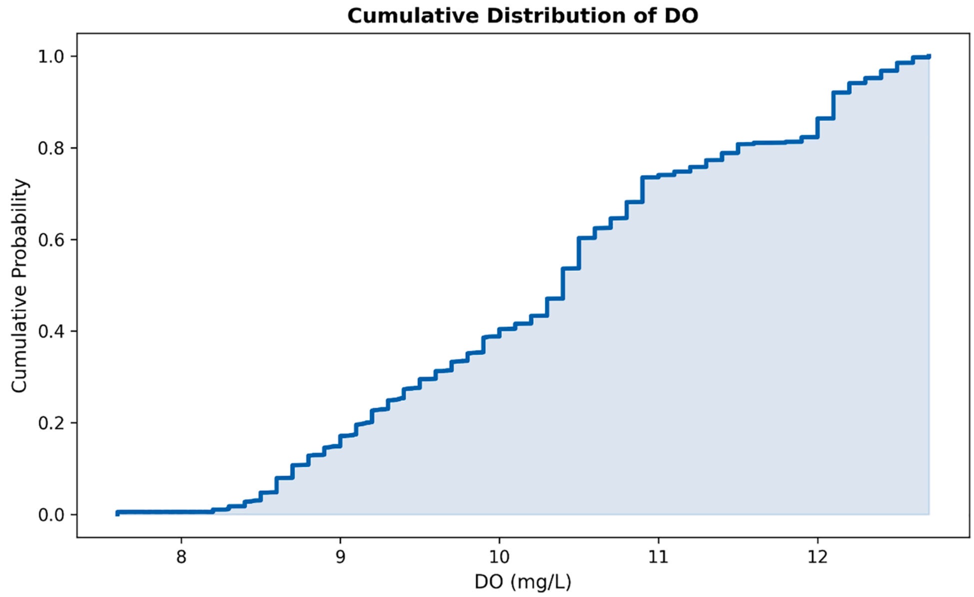

Fig. 6. Cumulative distribution of DO.

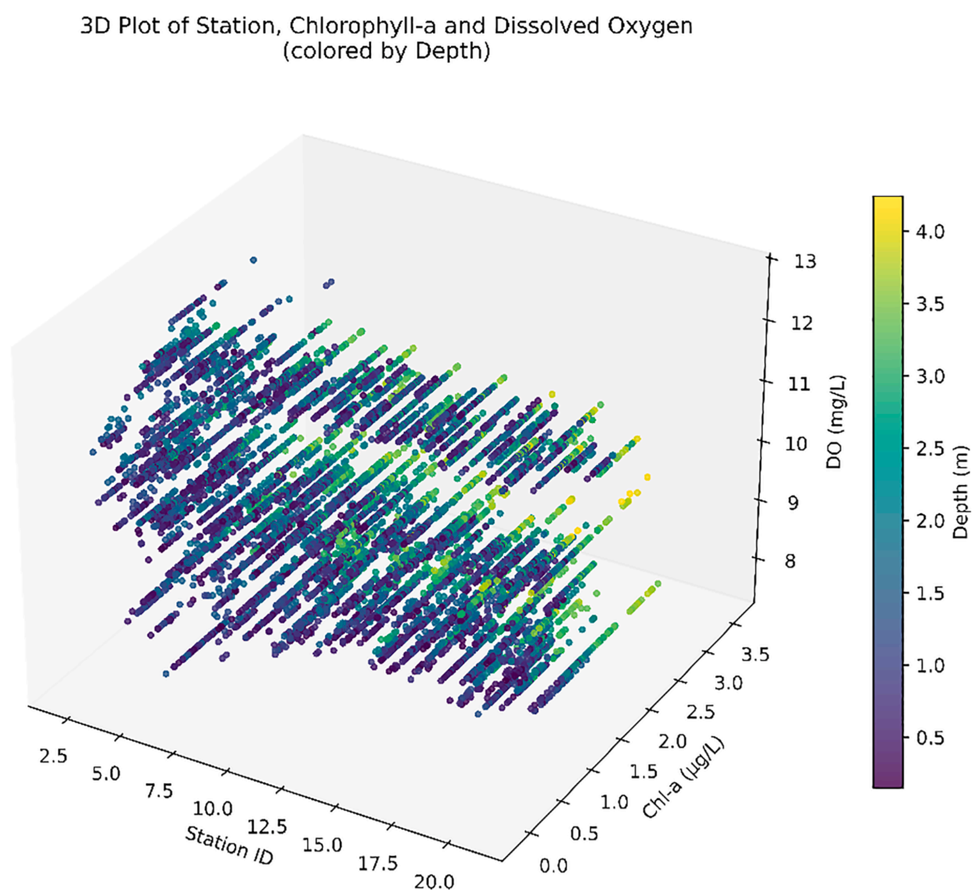

Fig. 7. 3D scatter plot of Station ID, chlorophyll-a and DO, coloured by sampling depth.

direct physical linkage. The turbidity–DO panel presents a broad scatter with substantial overlap at low turbidity and a few higher-turbidity observations tending toward moderate DO, indicating at most a weak, non-systematic relationship that likely reflects episodic resuspension rather than sustained oxygen demand (Fig. 9).

On the other hand, salinity and specific conductivity covary tightly, forming nearly parallel high-to-high and low-to-low trajectories, and observations with higher values on these axes tend to cross the oxygen axis at comparatively lower normalized levels, consistent with the DO depression under more saline, high-conductivity conditions (Fig. 10). Turbidity exhibits episodic peaks that intersect a wide range of oxygen values and often coincide with lower chlorophyll-a, indicating that

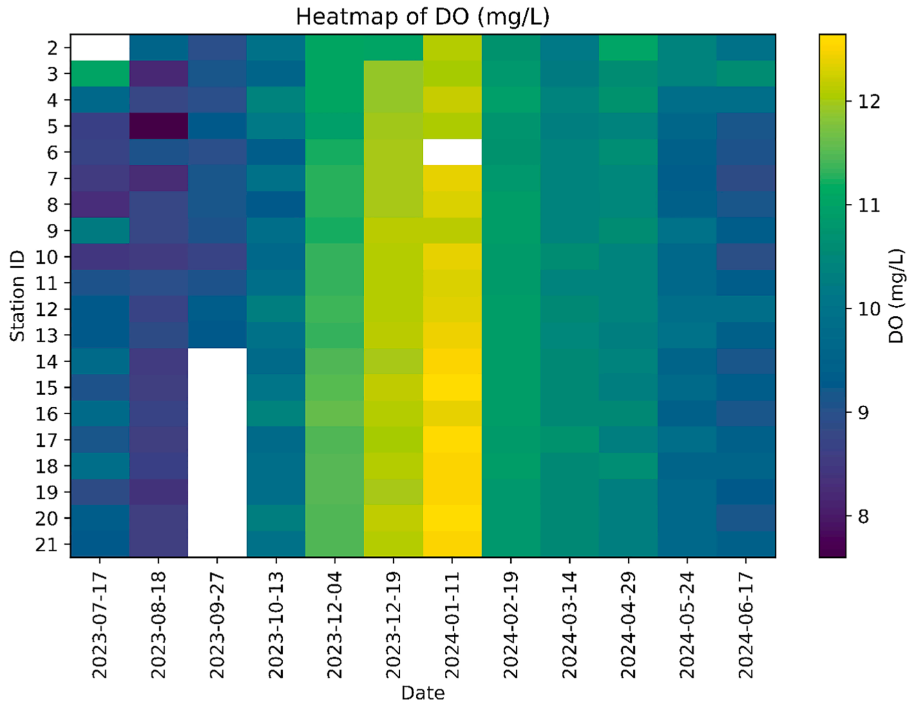

Fig. 8. Heatmap of DO.

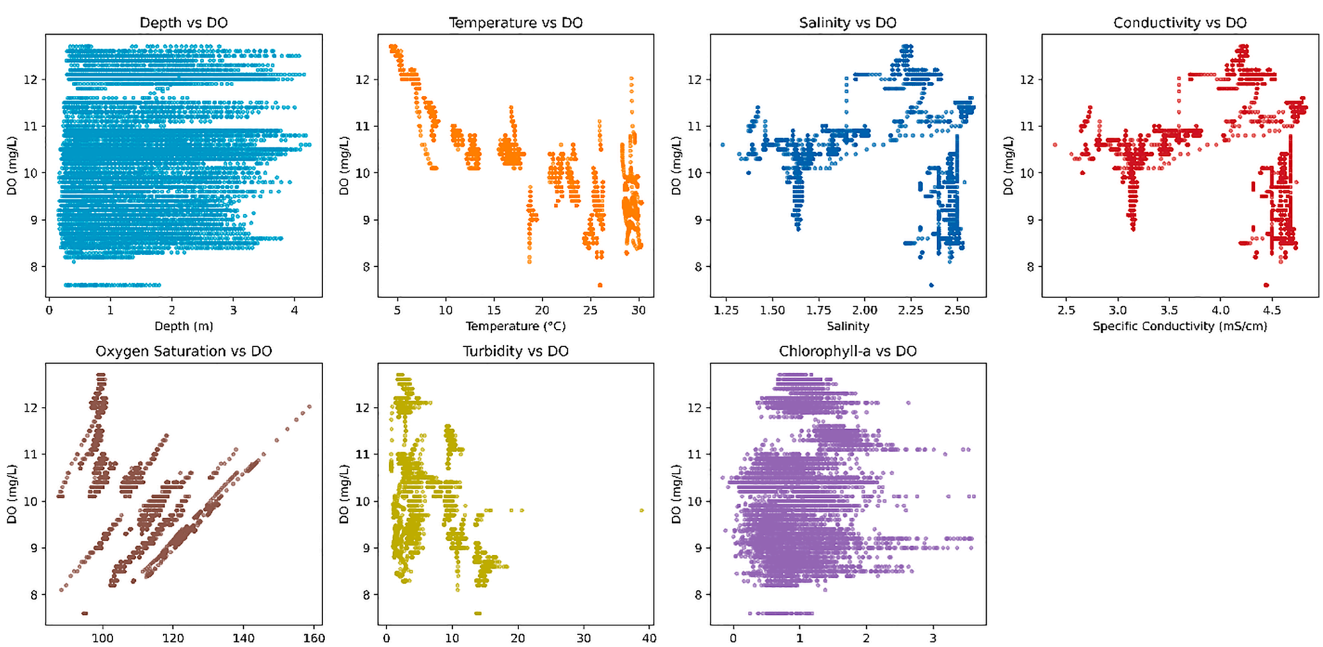

Fig. 9. Relationships between DO and key hydro-physical parameters (depth, temperature, salinity, specific conductivity, oxygen saturation, turbidity, and chlorophyll-a) across all stations and sampling dates.

events of elevated suspended matter are not consistently associated with enhanced oxygenation or phytoplankton biomass. Chlorophyll-a values cluster toward the lower half of the normalized range for most observations, and crossings show diverse oxygen outcomes, underscoring that chlorophyll-a alone does not strongly predict DO when temperature and salinity vary concurrently.

3.3. Model performance under the primary split

Under the primary random split, the ensemble machine-learning models clearly outperformed the deep-learning architectures. Extra-Trees provided the strongest performance, achieving an R^2 of 0.999706, RMSE of 0.020752 mg L^{-1}, MAE of 0.008211 mg L^{-1}, and MSE of 0.000431. Random Forest and KNN also performed strongly, with R^2 values of 0.999310 and 0.999004, respectively. Among the neural-network models, the best-performing architecture was the Deep MLP with Batch Normalization, which achieved an R^2 of approximately 0.975 with a substantially higher error profile than the leading tree-based models. These primary random-split results indicated that the tabular structure of the dataset was highly favorable for ensemble tree models.

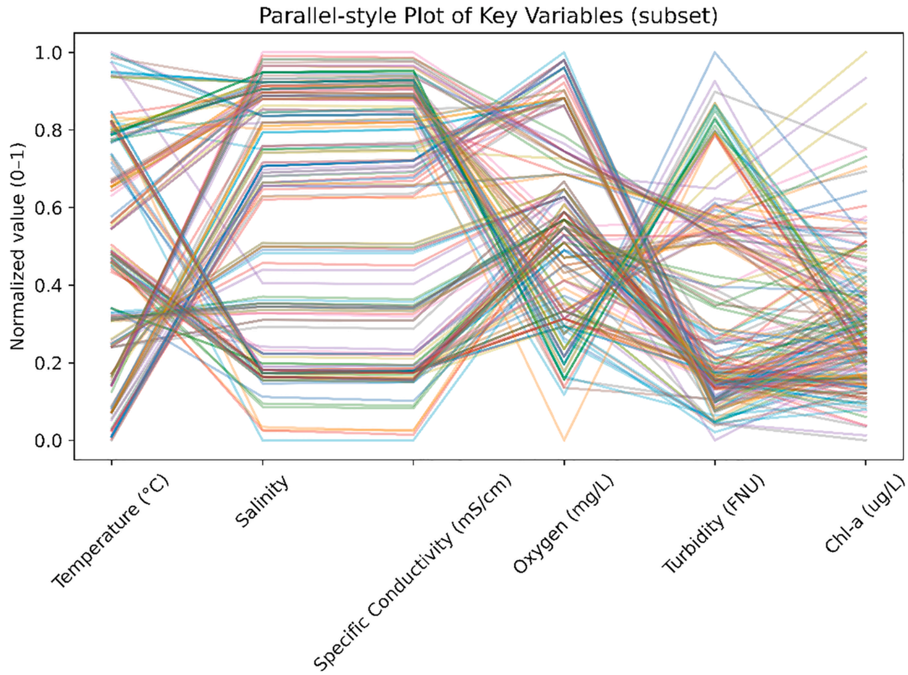

Fig. 10. Parallel-coordinates plot of normalized hydro-physical variables, for a subset of observations, showing joint variation among parameters.

The corresponding comparative metrics are summarized in Table 3.

3.4. Chronological split results

To evaluate temporal generalization under a stricter and more deployment-relevant setting, the models were re-evaluated using a chronological train-past → test-future design. Under this evaluation, predictive performance decreased substantially relative to the random-split results, indicating that the original high performance under the primary split scores were optimistic for future-time generalization. Among the tested machine-learning models, ExtraTrees remained the best-performing approach, achieving an R^2 of 0.5397, an RMSE of 0.3636 mg L^{-1}, and an MAE of 0.2683 mg L^{-1}. In contrast, Random Forest and KNN performed poorly under the same temporal split, with negative R^2 values of −0.8531 and −0.8113, respectively, indicating weak extrapolative performance under future observations (Fig. 11). These findings highlight the importance of reporting strict chronological validation in DO modeling and show that random mixed-data partitions can substantially overestimate predictive skill in temporally structured monitoring datasets.

Table 3
Main performance results under the primary, chronological, and LOSO evaluation settings.

Evaluation split	Model	R^2	RMSE (mg L^{-1})	MAE (mg L^{-1})
Primary random split	ExtraTrees	0.999706	0.020752	0.008211
	RandomForest	0.999310	0.031778	0.012637
	KNN	0.999004	0.038192	0.012148
	Deep MLP + BatchNorm	0.9749	0.1916	0.1385
Chronological split	ExtraTrees	0.539698	0.363559	0.268316
	RandomForest	-0.853120	0.729467	0.608426
	KNN	-0.811259	0.721180	0.578049
LOSO (mean ± std)	ExtraTrees	0.9295 ± 0.1270	0.2432 ± 0.1436	0.1591 ± 0.0950

Note: The primary random split produced substantially higher performance estimates than the stricter chronological and leave-one-station-out evaluations, indicating that validation design strongly affected apparent model skill.

3.5. LOSO validation

To assess spatial transferability, a LOSO validation scheme was applied in which each station was withheld in turn as an external test set while all remaining stations were used for model training. Under this setting, the ExtraTrees model showed generally strong but heterogeneous station-wise performance. Across the 20 withheld stations, LOSO performance ranged from moderate to very high, with station-level R^2 values spanning 0.4425 to 0.9948. The overall LOSO mean ± standard deviation for ExtraTrees was $R^2 = 0.9295 \pm 0.1270$, RMSE $= 0.2432 \pm 0.1436$ mg L^{-1}, and MAE $= 0.1591 \pm 0.0950$ mg L^{-1}. The strongest generalization was observed at Station 21 ($R^2 = 0.9948$), whereas the weakest was observed at Station 2 ($R^2 = 0.4425$), indicating that spatial transferability was not uniform across the monitoring network. Nevertheless, the overall LOSO results remained substantially stronger than the chronological split results, suggesting that the model generalized more reliably across stations than across future time periods. This pattern indicates that temporal extrapolation represented a greater challenge than spatial transfer within the present dataset. Detailed station-wise LOSO performance metrics for the ExtraTrees model are provided in Supplementary Table S2.

Fig. 12 illustrates the variation in LOSO performance across stations and highlights that, although generalization was strong overall, transferability was not uniform across the monitoring network.

3.6. Baseline comparison

To contextualize the predictive skill of the proposed models, baseline comparisons were performed using linear regression and a seasonal mean predictor. In the Table 4 baseline comparison setting, linear regression achieved an R^2 of 0.9308, an RMSE of 0.3182 mg L^{-1}, and an MAE of 0.2406 mg L^{-1}, while the seasonal mean model produced a similar performance with an R^2 of 0.9289, an RMSE of 0.3226 mg L^{-1}, and an MAE of 0.2305 mg L^{-1}. By comparison, the best-performing machine-learning model, ExtraTrees, achieved markedly lower errors and substantially higher explanatory power, with an R^2 of 0.9996, an RMSE of 0.0240 mg L^{-1}, and an MAE of 0.0109 mg L^{-1}. The best deep-

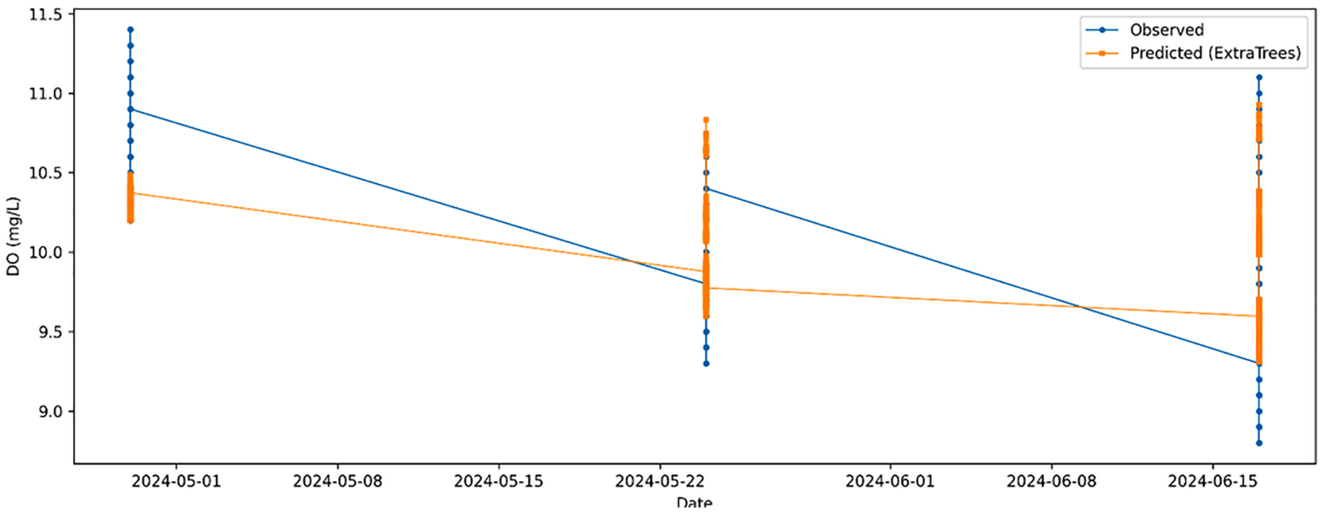

Fig. 11. Observed and predicted DO during the chronological test period using the ExtraTrees model.

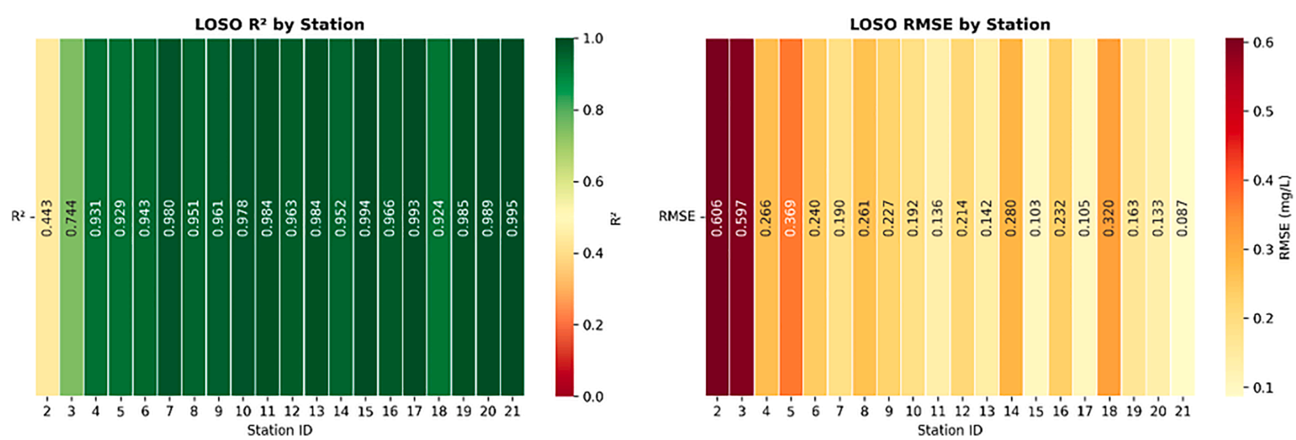

Fig. 12. Station-wise performance of the ExtraTrees model under LOSO validation.

Table 4
Baseline comparison against linear, seasonal, and advanced predictive models.

Model	R^2	RMSE	MAE
Linear Regression	0.930831	0.318242	0.240641
Seasonal Mean	0.928930	0.322585	0.230468
ExtraTrees	0.999606	0.024020	0.010949
Deep MLP + BatchNorm	0.9749	0.1916	0.1385

learning model, the Deep MLP with Batch Normalization, also outperformed the simple baselines, with an R^2 of 0.9749, RMSE of 0.1916 mg L^{-1}, and MAE of 0.1385 mg L^{-1}, although it remained less accurate than the ensemble tree approach (Fig. 13). These comparisons confirm that the strong performance of the leading models was not merely a reflection of seasonal structure alone and that advanced models, particularly ExtraTrees, provided meaningful gains over simpler benchmark approaches.

3.7. Ablation analysis

Because oxygen saturation (%) may be physically and operationally linked to DO, an ablation analysis was conducted to evaluate its influence on model performance. Using the ExtraTrees model, the full-feature configuration achieved an R^2 of 0.9998, RMSE of 0.0170 mg L^{-1}, and MAE of 0.0066 mg L^{-1}. When oxygen saturation was excluded from the predictor set, performance remained very strong, with $R^2 = 0.9996$, RMSE = 0.0240 mg L^{-1}, and MAE = 0.0109 mg L^{-1}, indicating that the model retained substantial predictive capacity from the remaining hydro-physical and seasonal variables (Table 5). However, the oxygen-saturation-only model also performed extremely strongly, with $R^2 = 0.9994$, RMSE = 0.0289 mg L^{-1}, and MAE = 0.0106 mg L^{-1}. These results show that oxygen saturation contributed considerable predictive information and therefore required careful interpretation. At the same time, the small decrease in performance after removing oxygen saturation indicates that the model's predictive ability was not solely dependent on that variable. The ablation analysis therefore supports a more balanced interpretation: oxygen saturation is highly informative, but strong DO estimation remains possible even when this potentially target-linked predictor is excluded.

Given the strong predictive influence of oxygen saturation observed in Table 5, the interpretability analysis presented below was considered together with the ablation results to avoid over-interpreting potentially target-linked information.

3.8. Feature importance and SHAP interpretation

The feature-importance results show which factors are most important for predicting DO. In all the ML models, temperature is clearly the most important factor. When temperature values are changed, the model's accuracy drops the most, which means DO is strongly controlled by water temperature. Turbidity is the second most important factor in RandomForest and KNN, but in the ExtraTrees model, almost all

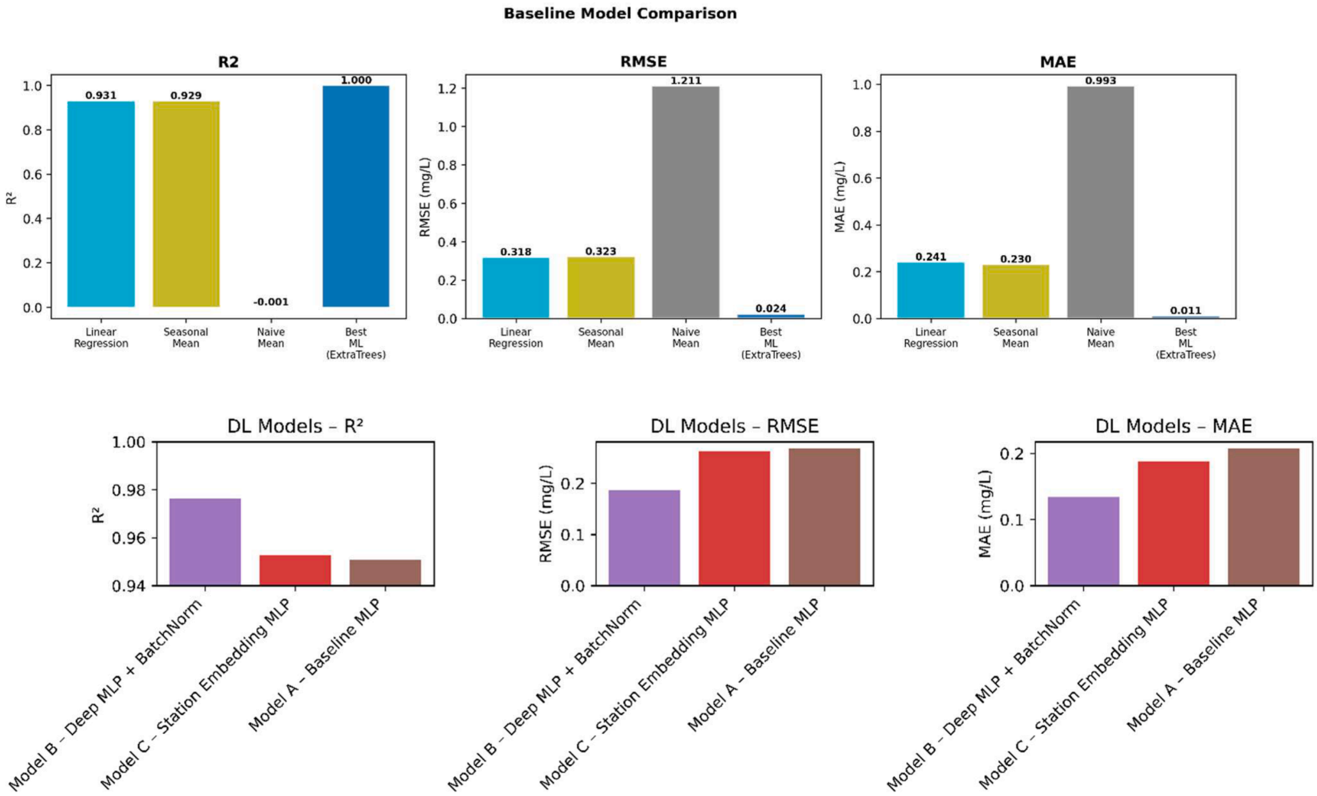

Fig. 13. Comparative performance of baseline and advanced models for DO estimation.

Table 5
Ablation results for the ExtraTrees model under different oxygen saturation feature configurations.

Configuration	R^2	RMSE (mg L^{-1})	MAE (mg L^{-1})
Full features (with oxygen saturation)	0.999803	0.016975	0.006553
Without oxygen saturation	0.999606	0.024020	0.010949
Oxygen saturation only	0.999429	0.028918	0.010589

importance is concentrated in temperature, and other features matter only a little. Features such as station ID, month, specific conductivity, day-of-year, salinity, chlorophyll-a, and depth have very small importance scores, meaning they only slightly change the model's prediction when they are shuffled (Fig. 14).

In the DL models, temperature is also the most important feature, but other variables play a bigger role compared with the ML models. For example, in Model B (Deep MLP with BatchNorm), temperature, turbidity, salinity, and day-of-year all contribute to the prediction. The Baseline MLP (Model A) relies more on month and day-of-year, meaning it uses seasonal patterns to make predictions. The Station-Embedding model (Model C) spreads its attention across temperature, conductivity, turbidity, and seasonal variables. In all DL models, chlorophyll-a and depth remain among the least important features.

The SHAP analysis of the ExtraTrees model showed that temperature was the dominant predictor of DO and that higher temperature values consistently exerted negative contributions to predicted DO. Seasonal variables provided a second tier of explanatory value, while salinity and specific conductivity contributed smaller but still measurable effects. Depth and chlorophyll a remained comparatively weak contributors in the present dataset. Because oxygen saturation also showed strong predictive influence in the ablation analysis, SHAP interpretation of the full-feature model was considered together with the ablation results to avoid over-interpreting a potentially target-linked variable (Fig. 15). Overall, the interpretability analysis remained consistent with the broader ecological and physical understanding of DO variability and supported the use of ExtraTrees as an interpretable soft-sensing model for the monitoring network.

4. Discussion

4.1. Relative performance of ML and DL models

Tree-based ensemble methods outperformed the evaluated deep networks for multi-station DO prediction, with ExtraTrees achieving the highest accuracy (R^2 0.9997; RMSE 0.021 mg L^{-1}; MAE 0.008 mg L^{-1}), followed closely by Random Forest and then KNN, whereas the best DL model (deep MLP with BatchNorm) reached a lower R^2 (0.976) with markedly higher errors. These findings are consistent with the well-recognized strength of tree ensembles on structured, tabular hydro-physicochemical data where nonlinear interactions are present, but sample sizes (10,000 observations) may be modest for training deeper neural networks without architectural inductive biases. Furthermore, the close alignment between observed and predicted seasonal trajectories across stations (Fig. 11) indicates that the best ML models not only achieve pointwise accuracy but also preserve phenological patterns relevant to management decision-making.

A likely reason for the DL underperformance is that generic feed-forward MLPs, even with Batch Normalization or station embeddings, may not fully exploit temporal ordering, seasonality, or site-specific structure unless explicitly encoded (e.g., sequence models or attention with spatio-temporal priors), whereas ensembles learn strong nonlinear tabular mappings with minimal feature engineering [30]. In addition, the relatively limited feature set and the dominance of a few highly predictive variables (notably temperature) can favor ensembles whose greedy partitioning isolates key thresholds efficiently, while MLPs may require more data and regularization to approximate sharp,

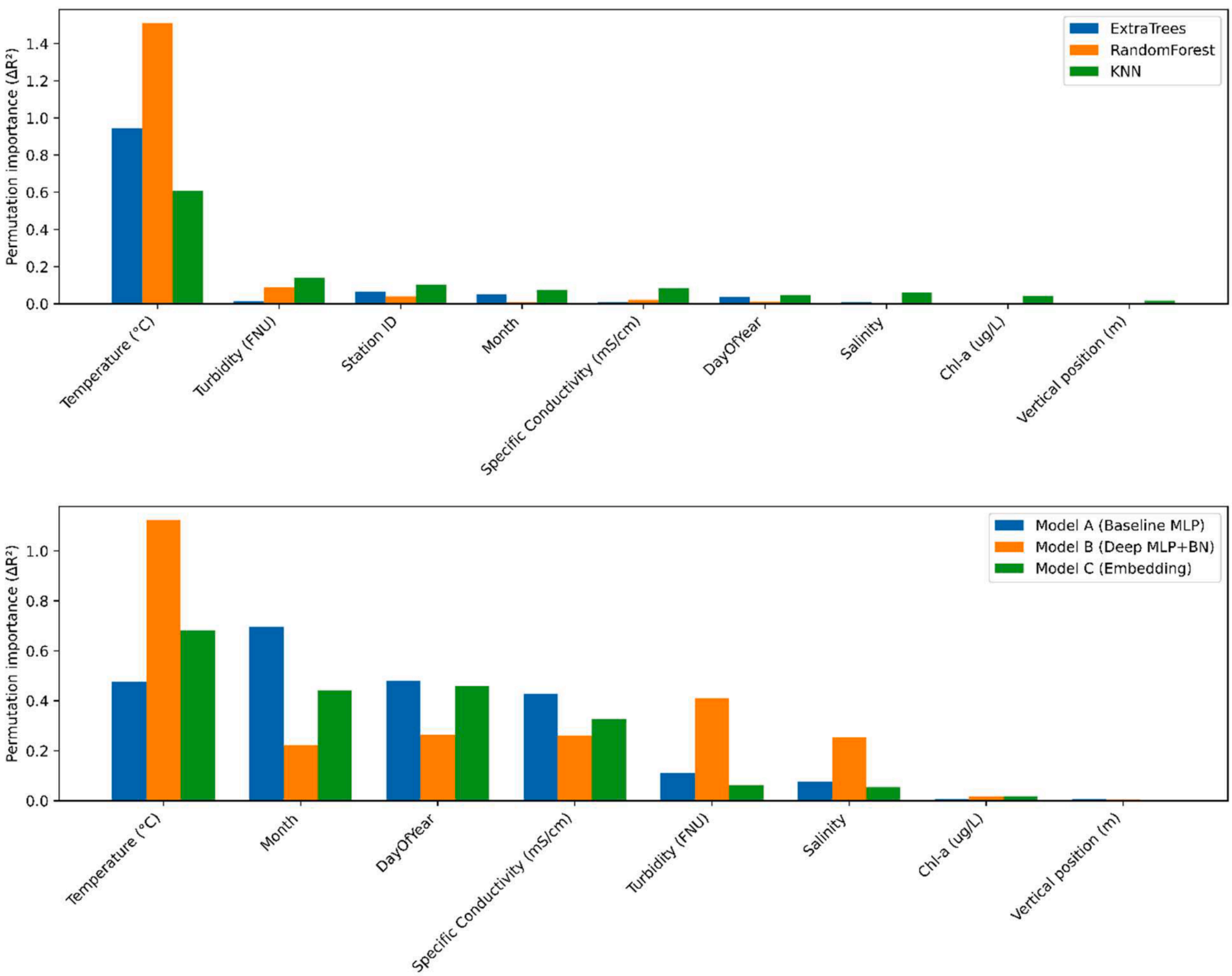

Fig. 14. Feature Importance – ML and DL models.

low-dimensional decision boundaries [9,31].

The contribution of the present study does not lie in the introduction of a fundamentally new predictive architecture, but in the transparent and deployment-oriented comparison of machine-learning and deep-learning models within a common multi-station DO framework. In contrast to many recent DO studies that report high predictive skill under a single validation design, the revised manuscript evaluates model behavior under a primary random split, a strict chronological split, a leave-one-station-out transfer, an oxygen-saturation ablation, and SHAP-based interpretation. This broader evaluation design provides a more realistic basis for model selection and interpretation than a single high R^2 value alone. It therefore constitutes an important part of the study's novelty and practical relevance.

4.2. Temporal and spatial generalization

A central finding of the revised analysis is that model performance depended strongly on the validation regime. Under the primary random split, the machine-learning models achieved high performance under the primary split accuracy, with ExtraTrees reaching $R^2 = 0.999706$. However, when the same model class was evaluated under a strict chronological train-past → test-future split, predictive performance decreased substantially, with ExtraTrees falling to $R^2 = 0.5397$ and both Random Forest and KNN yielding negative R^2 values. This contrast indicates that the original random-split results were optimistic for future-time generalization and confirms that random mixed-data partitions can substantially overestimate predictive skill in temporally structured monitoring data [32–34].

The station-based analysis provided a complementary perspective on robustness. Under leave-one-station-out validation, ExtraTrees retained a generally strong overall performance, with a mean R^2 of 0.9295 ± 0.1270 across withheld stations, although clear variability remained among individual locations. This indicates that spatial transferability across stations was comparatively more stable than temporal extrapolation across future periods [35]. In practical terms, the monitoring network appeared to share sufficient common DO structure for station-to-station transfer to remain feasible in most cases, whereas future-time prediction was more sensitive to changing seasonal and environmental conditions. This distinction is important because it shows that temporal and spatial generalization should be evaluated separately in multi-station water-quality studies rather than inferred from a single pooled performance estimate.

4.3. Oxygen saturation and interpretability

The ablation analysis directly addressed one of the most important methodological concerns raised during review: the possibility that oxygen saturation could inflate DO prediction performance. The revised results confirmed that oxygen saturation was highly informative, as evidenced by the very strong performance of the oxygen-saturation-only

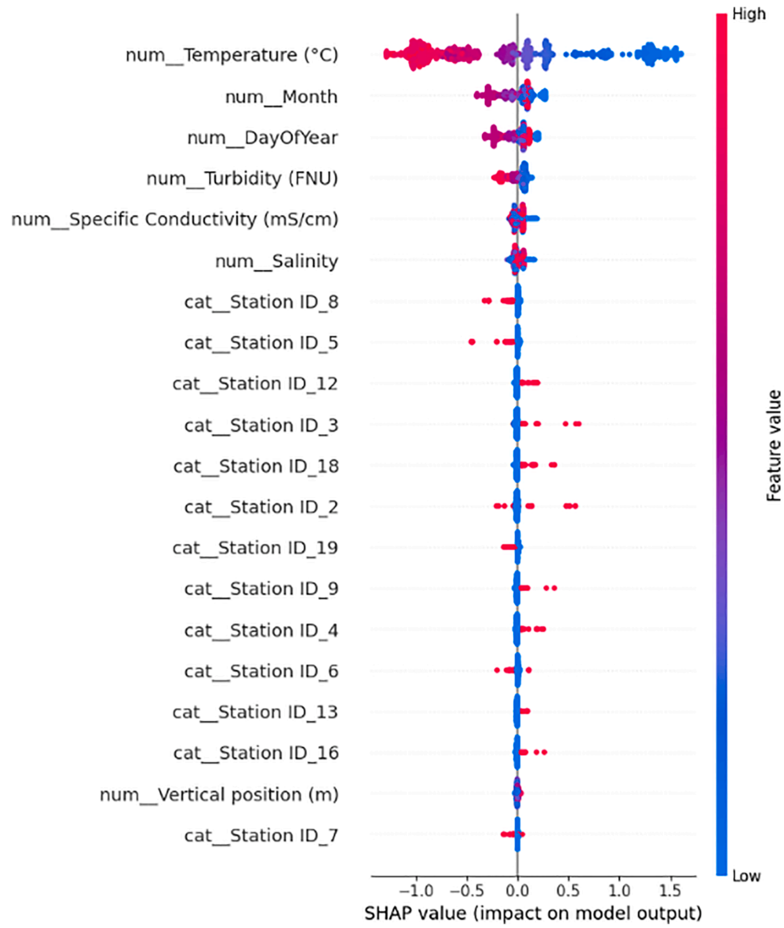

Fig. 15. SHAP beeswarm summary for the ExtraTrees model used in DO soft sensing.

model ($R^2 = 0.999429$). At the same time, the model remained highly accurate after excluding oxygen saturation, with only a modest reduction in performance ($R^2 = 0.999606$). These findings show that oxygen saturation contributed substantial predictive information, but that the remaining hydro-physical and seasonal variables also contained sufficient signal to support strong DO estimation.

The interpretability analysis remained broadly consistent with known environmental controls of DO [36]. SHAP results for the ExtraTrees model identified temperature as the dominant negative predictor, while seasonal variables formed a second tier of explanatory value and salinity/conductivity contributed smaller but still measurable effects. Depth and chlorophyll-a had comparatively lower influence in the present dataset. This agreement between the model explanation and known physical relationships supports the ecological plausibility of the model behavior. Importantly, the interpretability contribution of the present study lies not only in applying SHAP, but in combining SHAP with feature-ablation testing to distinguish physically meaningful drivers from variables that may carry strongly target-linked information [37]. In that sense, the study contributes an interpretability framework that is more robust than feature-importance.

The relatively strong station-to-station transferability and the limited added value of station-aware architectures suggest that shared physicochemical drivers dominated over fixed spatial effects across this monitoring network, reducing the relative contribution of station-specific signatures to the predictive task.

4.4. Practical implications for monitoring and soft-sensing

From an applied standpoint, the revised results support the use of interpretable ensemble models as practical tools for DO soft-sensing in aquatic monitoring networks [38]. Even after more stringent validation was introduced, ExtraTrees remained the strongest-performing model overall and showed comparatively stable station-to-station transferability under leave-one-station-out evaluation. These characteristics make ensemble trees attractive for operational monitoring systems where multiple stations must be evaluated efficiently, where computational simplicity is valuable, and where transparent model behavior is preferred over less interpretable alternatives. The study, therefore, has a clear engineering contribution in the context of environmental sensing and monitoring system design [38,39].

The revised framing also clarifies the application domain of the work. Although the original manuscript emphasized aquaculture-related early warning, the dataset and station descriptions are more accurately represented as a multi-station aquatic monitoring network. Accordingly, the practical relevance to aquaculture is best interpreted as indirect support for environmental surveillance, early warning, and condition awareness rather than as a directly validated pond-scale intervention framework. This change improves the conceptual consistency of the manuscript and better aligns the data source, modeling task, and application claims. At the same time, the ablation results suggest that strong DO soft-sensing may still be achievable using reduced predictor sets, which is relevant for monitoring systems where sensor availability, maintenance cost, and variable independence are practical deployment constraints.

4.5. Limitations

Several limitations should be acknowledged. First, although the revised manuscript now includes stricter chronological and station-based validation, not all models were benchmarked equally comprehensively under every evaluation regime [40], and the deep-learning architectures were not assessed under the full set of strict validation conditions to the same extent as ExtraTrees. Second, the task studied here is more accurately described as DO soft-sensing or contemporaneous estimation than as true lead-time soft-sensing, because explicit lag-based soft-sensing horizons were not modeled [41]. Third, the very high performance of the oxygen-saturation-only model indicates that some predictors may carry strongly target-linked information, which reinforces the importance of careful feature documentation and cautious interpretation in future applications.

Methodologically, DL architectures tailored to spatio-temporal learning (e.g., temporal convolution, gated recurrent units, or attention with station embeddings) [22] could be benchmarked against ensembles using identical splits, with careful hyperparameter optimization and data augmentation to probe longer lead times and seasonal regime shifts. Moreover, robust experiments, such as controlled noise injection, missing-data stress tests, and rebalancing of uneven station histories, would clarify failure modes and inform deployment under imperfect field conditions.

At the same time, these limitations help define the real contribution of the study. The present manuscript does not claim universal near-perfect DO prediction, nor does it propose a fundamentally new algorithmic architecture. Rather, its contribution lies in showing that when a multi-station DO problem is evaluated through multiple validation regimes, baseline benchmarking, oxygen-saturation ablation, and SHAP-based interpretation, the resulting evidence becomes more transparent, more scientifically defensible, and more informative for deployment than a single high-accuracy result alone. Future work should extend this framework to true sequence-aware soft-sensing models, more extensive cross-station benchmarking of deep architectures, repeated temporal resampling, and uncertainty-aware evaluation. By identifying both the capabilities and limitations of current model performance, the revised manuscript offers a more realistic and methodologically rigorous perspective within recent DO machine-learning research.

5. Conclusion

This study evaluated ensemble machine-learning and deep-learning models for dissolved oxygen estimation in a multi-station aquatic monitoring network and showed that ExtraTrees was the most accurate model under the primary evaluation setting. Under the primary random split, ExtraTrees achieved near-perfect performance, but the revised analyses demonstrated that these estimates were overly optimistic for future-time prediction. When evaluated under a strict chronological split, model skill decreased substantially, with ExtraTrees remaining the best-performing method but at a more realistic level of temporal generalization. In contrast, leave-one-station-out validation showed generally strong spatial transferability, indicating that the model generalized more reliably across stations than across future time periods.

The ablation analysis showed that oxygen saturation contributed substantial predictive information, as demonstrated by the strong performance of the oxygen-saturation-only model. However, the model also maintained very high performance after oxygen saturation was removed, indicating that useful dissolved oxygen estimation remained possible from the remaining hydro-physical and seasonal variables. SHAP interpretation consistently identified temperature as the dominant negative driver of dissolved oxygen, with seasonal variables and salinity/conductivity providing secondary explanatory value. Future work should extend this framework to true lead-time forecasting, additional uncertainty-aware evaluation, and broader cross-station benchmarking.

Overall, the revised manuscript demonstrates that transparent validation, feature ablation, and explainable modeling are essential for credible dissolved oxygen prediction studies. The principal contribution of the work lies in providing a multi-station, interpretable dissolved oxygen soft-sensing framework that is better aligned with realistic monitoring deployment and that remains relevant to aquaculture-related early warning as an indirect application domain. Future work should extend this framework to true lead-time forecasting, additional uncertainty-aware evaluation, and broader cross-station benchmarking of neural-network architectures.

Data and code availability for reproducibility

https://github.com/MHridoy-123/Multi-station-DO-prediction-modelling-Aquaculture-early-warning-system_8032.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Software usage

Artificial intelligence (AI) software, including machine learning libraries such as TensorFlow, was utilized for data analysis and modelling in this study. While no AI tools were used to write or prepare the manuscript text, AI-assisted tools were employed solely to support ideation and conceptual thinking during the research process.

CRediT authorship contribution statement

Md. Abdullah Al Mamun Hridoy: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Matteo Bodini:** Writing – review & editing, Validation, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Chiara Bordin:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Khan Md Arafat Adnan:** Writing – review & editing, Validation, Formal analysis. **Mohd. Saifur Rahman:** Writing – review & editing, Validation, Investigation, Formal analysis. **Abdullah Ibna Shawkat:** Writing – review & editing, Validation, Investigation, Formal analysis. **Pakorn Ditthakit:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Conceptualization. **Petra Schneider:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Andleeb Masood:** Writing – review & editing, Validation, Investigation, Formal analysis. **Md. Abdullah Al Mamun:** Writing – review & editing, Validation, Investigation, Formal analysis, Conceptualization. **Paolo Pastorino:** Writing – review & editing, Validation, Investigation, Formal analysis. **Leonardo Goliatt:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

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

Acknowledgments

The author expresses sincere gratitude to the Mendeley Data repository for providing a platform to share and access the dataset, which greatly supported the development of this research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rineng.2026.112205.

References

[1] S.D. Latif, M.S.B.N. Azmi, A.N. Ahmed, C.M. Fai, A. El-Shafie, Application of artificial neural network for forecasting nitrate concentration as a water quality parameter: a case study of feitsui reservoir, Taiwan, Int. J. Des. Nat. Ecodynamics 15 (5) (2020) 647–652, https://doi.org/10.18280/ijdne.150505.

[2] A. Parsaie, A.H. Haghiabi, Numerical routing of tracer concentrations in rivers with stagnant zones, Water Sci. Technol.: Water Supply 17 (3) (2017) 825–834, https://doi.org/10.2166/ws.2016.175.

[3] M. Ay, Ö. Kişi, Estimation of DO by using neural networks and neuro fuzzy computing techniques, KSCE J. Civ. Eng. 21 (5) (2017) 1631–1639, https://doi.org/10.1007/s12205-016-0728-6.

[4] F. Granata, S. Zhu, F. Di Nunno, DO forecasting in the mississippi river: advanced ensemble MLmodels, Environ. Sci.: Adv. 3 (11) (2024) 1537–1551, https://doi.org/10.1039/d4va00119b.

[5] Z.C. Oleiwi, K.K. Obayes, N.K. Hadi, R.Q. Yaseen, A.J Sahib, Forecasting DO in lakes using different AI models, J. Al-Qadisiyah Comput. Sci. Math. 17 (1) (2025), https://doi.org/10.29304/jqcsm.2025.17.11959.

[6] S. Pinthong, N. Salaeh, Q.B. Pham, M.A.A.M. Hridoy, C. Bordin, P. Ditthakit, Coupled spiking neural network–extra trees framework for daily streamflow imputation with ERA5-land integration, Eng. Sci. 41 (2026) 2204, https://doi.org/10.30919/es2204.

[7] R. Bhateria, D. Jain, Water quality assessment of lake water: a review, Sustain. Water Resour. Manag. 2 (2) (2016) 161–173, https://doi.org/10.1007/s40899-015-0014-7.

[8] M.N. Assaf, Q. Abdelal, N.M. Hussein, G. Halaweh, A.J. Alzubaidi, Water quality monitoring and management: integration of machine learning algorithms and Sentinel-2 images for the estimation of Chlorophyll-a, Model. Earth Syst. Environ. 11 (5) (2025) 348, https://doi.org/10.1007/s40808-025-02543-4.

[9] M.A.A.M. Hridoy, C. Bordin, P. Pastorino, K.N.A. Maulud, M.M. Pathan, A.O. Baki, Neural network-based prediction of water contamination dynamics in a rapidly developing tropical region, Case Stud. Chem. Environ. Eng. 13 (2026) 101342, https://doi.org/10.1016/j.cscee.2026.101342.

[10] Othman, Z., & Wu, Y. (2025). DO Prediction in Rivers Based on RF-LSTMModel: A Case Study of Dujiangyan. doi:10.20944/preprints202504.2594.v1.

[11] Tiyasha, T.M. Tung, Z.M Yaseen, A survey on river water quality modelling using artificial intelligence models: 2000–2020, J. Hydrol. 585 (2020) 124670, https://doi.org/10.1016/J.JHYDROL.2020.124670.

[12] M.A.A.M. Hridoy, P. Pastorino, C. Bordin, M. Bodini, N. Dhar, P. Schneider, K.N. A. Maulud, Operational agricultural water management in river-based aquaculture: A machine-learning approach to predict dissolved oxygen in the Halda River, Bangladesh, Aquac. Eng. 113 (2026) 102693, https://doi.org/10.1016/j.aquaeng.2026.102693.

[13] Ç. Arisoy, E. Süzgen, G. Yildiz, Y.B. Üniversitesi, M.-M. Fakültesi, M. Bölümü, Bozok journal of engineering and architecture predicting DO levels in aquatic ecosystems using MLmodels, Bozok J. Eng. Arch. 4 (1) (2025) 56–65, https://doi.org/10.70700/bjea.1590677.

[14] M.A.A.M. Hridoy, A.I. Shawkat, C. Bordin, M.R. Acharjee, A. Masood, A.O. Baki, M. A. Al Mamun, Advanced machine learning models for accurate water quality classification and WQI prediction: implications for aquatic disease risk management, Sci. Total Environ. 1008 (2025) 180965, https://doi.org/10.1016/j.scitotenv.2025.180965.

[15] M.A.A.M. Hridoy, M. Bodini, M.R. Mahin, P. Schneider, P. Pastorino, C. Bordin, L. Goliatt, Operational early-warning forecasts of aquaculture water quality: interpretable ML for TDS under walk-forward validation, Ecol. Inform. 96 (2026) 103804, https://doi.org/10.1016/j.ecoinf.2026.103804.

[16] L. Yin, L. Fu, H. Wu, Q. Xia, Y. Jiang, J. Tan, Y. Guo, Modeling DO in a crab pond, Ecol. Model. 440 (2021) 109385, https://doi.org/10.1016/J.ECOLMODEL.2020.109385.

[17] B.F. Ziyad Sami, S.D. Latif, A.N. Ahmed, M.F. Chow, M.A. Murti, A. Suhendi, B. H. Ziyad Sami, J.K. Wong, A.H. Birima, A. El-Shafie, MLalgorithm as a sustainable tool for DO prediction: a case study of Feitsui Reservoir, Taiwan. Sci. Rep. 12 (1) (2022), https://doi.org/10.1038/s41598-022-06969-z.

[18] S. Niloy, M.S. Rahman, S. Izlal, F.M. Ruhad, A. Chanda, A.K. Saha, M.A. Hossain, Spatial mapping of cropland soil load bearing capacity in northeastern bangladesh: a multi-feature-based prediction using machine learning-remote sensing fusion, Earth Syst. Environ. (2026) 1–25, https://doi.org/10.1007/s41748-026-01118-2.

[19] R. Choudhary, A. Kumar, P. C, M.M. Naik, M. Choudhury, N.A Khan, Predicting water quality index using stacked ensemble regression and SHAP based explainable artificial intelligence, Sci. Rep. 15 (1) (2025), https://doi.org/10.1038/s41598-025-09463-4.

[20] Y. Ma, Q. Fang, S. Xia, Y. Zhou, Prediction of the DO content in aquaculture based on the CNN-GRU hybrid neural network, Water (Switz.) 16 (24) (2024) 3547, https://doi.org/10.3390/w16243547.

[21] M.A.A.M. Hridoy, C. Bordin, A. Masood, K. Masood, Predictive modelling of aquaculture water quality using IoT and advanced machine learning algorithms, Results Chem. 16 (2025) 102456, https://doi.org/10.1016/j.rechem.2025.102456.

[22] K. Abdalgader, M. Mbarak, M. Alam, Deep neural network with attention and station embeddings for robust spatio-temporal multisensor temperature forecasting, Agri Eng. 7 (12) (2025) 399, https://doi.org/10.3390/agriengineering7120399.

[23] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International conference on ML, 2015, pp. 448–456, https://doi.org/10.48550/arXiv.1502.03167, pmlr.

[24] K. Merabet, F. Di Nunno, F. Granata, S. Kim, R.M. Adnan, S. Heddam, M. Zounemat-Kermani, Predicting water quality variables using gradient boosting machine: global versus local explainability using SHapley Additive Explanations (SHAP), Earth Sci. Inform. 18 (3) (2025) 298, https://doi.org/10.1007/s12145-025-01796-y.

[25] H.A. Madni, M. Umer, A. Ishaq, N. Abuzinadah, O. Saidani, S. Alsubai, M. Hamdi, I. Ashraf, Water-quality prediction based on H2O AutoML and explainable AI techniques, Water (Switz.) 15 (3) (2023) 475, https://doi.org/10.3390/w15030475.

[26] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42, https://doi.org/10.1007/s10994-006-6226-1.

[27] D. Kocev, M. Ceci, T. Stepišnik, Ensembles of extremely randomized predictive clustering trees for predicting structured outputs, Mach. Learn. 109 (11) (2020) 2213–2241, https://doi.org/10.1007/s10994-020-05894-4.

[28] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536, https://doi.org/10.1038/323533a0.

[29] F. Zhao, W. Wang, H. Sun, H. Yang, J. Wu, Station-level short-term demand forecast of carsharing system via station-embedding-based hybrid neural network, Transp. B: Transp. Dyn. 10 (1) (2022) 1–19, https://doi.org/10.1080/21680566.2021.1951885.

[30] X. Liang, Z. Hu, X. You, ENSONet: a data-driven ENSO forecasting model with concise spatial location learning parameters and temporal embedding, Clim. Dyn. 62 (5) (2024) 4081–4098, https://doi.org/10.1007/s00382-024-07119-z.

[31] K.M. Jablonka, D. Ongari, S.M. Moosavi, B. Smit, Big-data science in porous materials: materials genomics and machine learning, Chem. Rev. 120 (16) (2020) 8066–8129, https://doi.org/10.1021/acs.chemrev.0c00004.

[32] T.M. Detmer, J.J. Parkos III, D.H Wahl, Long-term data show effects of atmospheric temperature anomaly and reservoir size on water temperature, thermal structure, and DO, Aquat. Sci. 84 (1) (2022) 3, https://doi.org/10.1007/s00027-021-00835-2.

[33] B. Goh, Y. Wang, P. Nelaturu, H. Zhang, M. Moorehead, T. Duong, A. Couet, Nobility vs. mobility: insights into molten salt corrosion mechanisms of high-entropy alloys via high-throughput experiments and machine learning, Matter 7 (6) (2024) 2313–2335, https://doi.org/10.1016/j.matt.2024.05.004.

[34] X. Chen, C. Zhao, J. Chen, H. Jiang, D. Li, J. Zhang, C. Wang, Water quality parameters-based prediction of DO in estuaries using advanced explainable ensemble machine learning, J. Environ. Manag. 380 (2025) 125146, https://doi.org/10.1016/j.jenvman.2025.125146.

[35] A.A.M. Hridoy, S. Neogi, R. Ujjaman, M. Hasan, Water quality interactions and their synergistic effects on aquaculture performance in Bangladesh: a critical review, Results Chem. 16 (2025) 102306, https://doi.org/10.1016/j.rechem.2025.102306.

[36] M.A.A.M. Hridoy, P. Ditthakit, A. Hosen, Machine learning–driven prediction of heavy metal pollution for aquatic health surveillance and disease control, Mach. Learn.–Driven Predict. Heavy Met. Pollut. Aquat. Health Surveill. Dis. Control (2025), https://doi.org/10.2139/ssrn.5930557. ISCIENCE-D-25-19889.

[37] Hridoy, M. A. A. M., Pastorino, P., Bordin, C., Goliatt, L., Schneider, P., Shawkat, A. I., & Macêdo, B. S. (2026d). Machine Learning Enhanced Prediction of TDS for Strengthening Aquatic Disease Early Warning Systems. doi:10.21203/rs.3.rs-8199017/v1.

[38] E. Eslami, A.K. Salman, Y. Choi, A. Sayeed, Y. Lops, A data ensemble approach for real-time air quality Forecasting using extremely randomized trees and deep neural networks, Neural Comput. Appl. 32 (11) (2020) 7563–7579, https://doi.org/10.1007/s00521-019-04287-6.

[39] Hridoy, M. A. A. M., & Paul, P. B. (2024). Assessing Total Dissolved Oxygen and Electrical Conductivity Using Sentinel-2 Remote Sensing: A Multivariable Approach to Flood Detection in Flood-Prone Urban Area Sylhet. doi:10.21203/rs.3.rs-5048730/v1.

[40] F.M. Megahed, Y.J. Chen, L.A. Jones-Farmer, S.E. Rigdon, M. Krzywinski, N. Altman, Comparing classifier performance with baselines, Nat. Methods 21 (4) (2024) 546–548, https://doi.org/10.1038/s41592-024-02234-5.

[41] M. Kruk, Prediction of environmental factors responsible for chlorophyll a-induced hypereutrophy using explainable machine learning, Ecol. Inform. 75 (2023) 102005, https://doi.org/10.1016/j.ecoinf.2023.102005.