

GSJ: Volume 13, Issue 10, October 2025, Online: ISSN 2320-9186

www.globalscientificjournal.com

A Comprehensive Comparison Between Traditional and Deep Learning Models in Predicting Dissolved Oxygen

Tanvi Lodhavia

October 16, 2025

Abstract

Water is one of the most important resources for all life on Earth. However, with industrialization, urbanization, and climate change, water bodies around the world are in danger. An important component of water is the amount of oxygen present in it (also known as Dissolved Oxygen). Predicting DO levels can be difficult, especially due to the non-linear nature of its factors, ranging from photosynthesis and respiration to temperature-dependent solubility. Previous models span mechanistic DO-BOD formulations and data-driven methods (SVR, tree ensembles, LSTM), yet inconsistent datasets, non-temporal splits, and missing persistence baselines hinder fair comparison. Thus, this paper explores how various Traditional and Machine Learning Models (Ridge Regression, Random Forest, HistGradientBoosting Regressor, LSTM, GRU and TCN) compare in predicting Dissolved Oxygen.

Keywords: Dissolved Oxygen, Machine Learning, Deep Learning, Water Quality Prediction

1. Introduction

Water is essential to all life on Earth. However, water quality is complex, impacted by pollution, weather tides, human activities, and many other factors [1]. Several threats to the state of water exist, including, pollution, over-extraction, climate change, and habitat destruction. These threats can lead to the deterioration of water quality and quantity, negatively impacting human health, wildlife and the environment [2]. Due to rapid urbanization, industrialization and inefficient agricultural practices, water sources are being contaminated, rendering them unsafe for consumption. The World Health Organization reports that approximately 2.2 Billion People (22% of the global population) lack access to safely managed drinking water services, underscoring the gravity of the global water crisis. [3][4]

1.1. Role of Dissolved Oxygen in Water Bodies

An important component of water quality is Dissolved Oxygen. Dissolved Oxygen (DO) is the amount of oxygen that is present in water. Sources of DO include atmospheric oxygen, aquatic plants, and running water. DO is considered an important measure of water quality as it is a direct indicator of an aquatic resource's ability to support aquatic

life. This is because all animals need Oxygen to breathe, which they get from the DO present in water. If DO levels drop, some sensitive animals may move away, decline in health or even die. [5]

1.2. Dissolved Oxygen Dynamics and Prediction Challenges

Dissolved Oxygen levels are influenced by various physical, biological and chemical processes, including photosynthesis, respiration, atmospheric re-aeration, temperature-dependent solubility, etc. Thus, accurately predicting DO taking into account all the various factors that affects it can be difficult. [6]

1.3. Modeling Paradigms

There are two main types of dissolved oxygen concentration prediction models: 1. Traditional models based on physics, and 2. Data-drive models based on artificial intelligence. Traditional prediction models simulate the physical mechanisms in rivers and describe the transfer process of pollutants in water. In recent years, with the development of machine learning algorithms, the application of AI-based data-driven models in dissolved oxygen prediction has gradually attracted attention. Compared with models based on physical mechanisms, data-driven models have a simple modeling process and do not require complex mathematical formulas and parameter settings. [7]

1.4. Traditional Machine Learning vs Deep Learning Models

When considering machine learning models, a key distinction lies between traditional and deep learning approaches. Traditional machine learning, such as regression models, and decision trees, excels with smaller datasets, and rely on manually engineered features, while deep learning methods use multi-layered neural networks to automatically extract patterns directly from the data. [8] The difference is especially significant in predicting DO in water quality, which involves complex, non-linear, and time-dependent interactions.

This paper compares 6 different machine learning models: Ridge Regression, Random Forest, HistGradient-Boosting Regressor, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Temporal Convolutional Network (TCN)

1.5. Research Topic

This paper will delve into **A comprehensive comparison between traditional ML and Deep Learning Models on predictive Dissolved Oxygen in Water Quality.**

2. Related Work

Recent works show that classical, tabular ML, especially tree ensembles and margin methods remain a very competitive option for predicting Dissolved Oxygen. This is also strengthened by the research carried out in this paper. A Mississippi River study that Bayesian-optimized SVR, a regression tree, and a boosting ensemble reported BO-SVR as best, and used SHAP to show temperatures, discharge, and gage height as the dominant drivers. Although, performance did degrade with longer horizons, it remained acceptable out to around 30 days. This paper sets a strong benchmark because it couples accuracy with interpretability and principled tuning.

At the same time, deep learning models are also being increasingly used for DO. LSTM usecases range from continental-scale applications to graphs and transformer architectures. These models are especially useful when high-frequency, multi-source data is available and when the goal is multi-step forecasting that exploits temporal dependencies. But the evidence is mixed across sites: several studies still report Traditional Models to be better than the Deep Learning Models.

Evidence from smaller systems points the same way. For an urban lake, a simple ANN outperformed multilinear regression, but this was on a narrow predictor set (temp, pH, conductivity and ORP), reinforcing that even shallow nets can help once nonlinearity enters but also that results don't automatically transfer to riverine settings or longer horizons.

Across the papers reviewed, the recurring limitation is comparability and generalization: many papers are single-sit, optimize different metrics/horizons, and rarely report skill versus naive persistence. The Mississippi study explicitly warns that cross-paper comparisons are brittle because of dataset differences, which could potentially also affect the comparison carried out in this research paper.

3. Theoretical Framework

Machine Learning is a subfield of Artificial Intelligence (AI), which is broadly defined as the capability of a machine to imitate intelligent human behaviour. Artificial Intelligence systems are used to perform complex tasks in a way that is similar to how humans solve problems. [11]

There are three subcategories of Machine Learning, each with its own use case: Supervised Learning, Unsupervised Learning, and Reinforcement Learning.

The prediction of dissolved oxygen falls into the category of **Supervised Machine Learning**. This is because for DO, the dataset consists of historical, time-stamped examples where both the inputs and the desired outputs (DO) are known. Each training case consists of an input window (Recent DO, temperature, depth, salinity, etc., plus seasonal indicators) and a label: the measured DO H weeks later (y_{t+H} in mg/L). The goal is to minimize the error prediction in the hand-out data.

Furthermore, this setup supports clear objective functions (e.g., MSE), straightforward evaluation (RMSE, MAE, NSE), and fair baselines (persistence). Additionally, alternatives do not fit as well: **unsupervised** methods do not have target labels and they answer different questions (eg: regime discovery, or anomaly detection), while **reinforcement learning** is appropriate only when choosing control actions (e.g., aeration) that influence future DO. Given labeled weekly records and an operational need for numeric forecasts, supervised time-series regression is the appropriate formulation

3.1. Ridge Regression Model

A ridge regression, also known as L2 regularization, is one of the several types of regularization for linear regression models. Regularization is a statistical method to reduce errors caused by overfitting on training data. Ridge regression specifically corrects for multicollinearity in regression analysis. [10]

The problem of *multicollinearity* arises when in a regression model, two or more independent variables are highly correlated to each other. This is an issue because linear

regression models rely on the assumption that independent variables are truly independent. If they are not, it becomes difficult for the model to accurately estimate the individual effect of each variable on the dependent variable. [12] For instance, in the case of predicting DO, several key independent variables like water temperature, salinity and conductivity, are naturally and highly correlated. Hence, Linear Regression could potentially cause high levels of inaccuracies.

The Ridge Regression Model adds a penalty term to the model's cost function. This penalty, which is proportional to the square of the magnitude of the coefficients, shrinks the coefficients towards zero. [13] Hence, the use of Ridge Regression Models is especially useful when dealing with large amounts of highly related data such as Water Quality and Dissolved Oxygen.

3.2. Random Forest

Random Forest is a commonly-used machine learning algorithm, trademarked by Leo Breiman and Adele Cutler, that combines the output of multiple *decision trees* to reach a single result. It's easy to use, and flexible, which is why it handles both classification and regression problems.

Decision Trees help mapping out different choices and their possible outcomes. It's used in machine learning for tasks like classification, and prediction. The Tree consists of one question. Each question helps an individual to arrive at a final decision, which would be denoted by the leaf node. Through this, decision trees seek to find the best split to subset the data, and they are typically trained through the Classification and Regression Tree (CART) algorithm. [14]

This paper includes *Random Forest* because it offers a strong, sample-efficient baseline for weekly dissolved oxygen regression. By averaging many CART trees built on bootstrap samples and random feature subsets, it greatly reduces variance and overfitting while naturally capturing nonlinear relationships and interactions among drivers (e.g. temperature, salinity, depth, seasonality). These properties make Random Forest a practical and credible comparator alongside Ridge and the deep learning models in the study.

3.3. HistGradient-Boosting Regressor

HGBR is a gradient-boosted decision tree model that builds an additive ensemble of shallow CART trees, each new tree correcting the residual errors of the current ensemble. The scikit-learn implementation uses histogram binning to discretize continuous features into a small number of bins, which makes training fast and memory-efficient while retaining accuracy. Regularization is controlled through the learning rate, number/depth of trees, and minimum leaf size, with optional early stopping on a validation set; it also handles missing values (NaNs) natively.

We include HGBR because it is a strong tabular baseline for nonlinear regression with limited, noisy, weekly data. Compared with a single tree, boosting reduces bias; compared with Random Forest, boosting can achieve higher accuracy by iteratively fitting residual structure rather than averaging independent trees. It captures thresholds and interactions among drivers (e.g., temperature $\times$ salinity $\times$ seasonality) without heavy feature engineering, requires minimal preprocessing, and remains sample-efficient relative to deep sequence models. Finally, it provides straightforward feature importance diagnostics, giving interpretable insight into which lags/covariates matter across horizons.

These properties make HGBR a natural complement to Ridge and Random Forest in our traditional-ML baseline set. [15][16]

3.4. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are an extension of RNNs mainly introduced to handle situations where RNNs fail. It was designed to model chronological sequences and their long range dependencies more precisely than conventional RNNs. The basic difference between the architectures of RNNs and LSTMs is that the hidden layer of LSTM is a gated unit or gated cell. [17] At the core of an LSTM, there are three gates: Forget Gate, Input Gate and Output Gate. This gating mechanism allows LSTMs to retain important patterns over long horizons while ignoring noise, making them especially effective for modeling nonlinear, time-dependent interactions in environmental data. [18]

For this study, LSTM is chosen as the deep learning model because of its:

1. Sequence learning ability - it can use raw sequences of water quality variables without requiring manual lag feature engineering
2. Nonlinear modeling power - it can capture complex relationships between environmental variables that linear models like Ridge cannot
3. Proven effectiveness - LSTMs are widely used in forecasting problems ranging from weather and climate to air pollution, making them a natural choice for water quality prediction.

3.5. Gated Recurrent Unit (GRU)

Gated Recurrent Units (GRUs) are a type of RNN introduced by Cho et al. in 2014. The core idea behind GRUs is to use gating mechanisms to selectively update the hidden state at each time step allowing them to remember important information while discarding irrelevant details. GRUs aim to simplify the LSTM architecture by merging some of its components and focusing on just two main gates: the update gate and the reset gate. [19]

Compared to LSTM, GRU has fewer parameters (because there is no separate cell state), which makes it easier to train and less data hungry. At the same time, it still mitigates the vanishing gradient problem. The GRU Model is integrated to test whether a leaner recurrent memory can capture DO dynamics better than LSTM under a modest weekly dataset. Practically, GRU tends to converge faster, is stable with early stopping and dropout, and offers a complementary recurrent baseline alongside LSTM in the deep-learning set.

3.6. Temporal Convolutional Network (TCN)

Temporal Convolutional Networks (TCNs) are deep neural network architectures that are used in trajectory prediction tasks. They are trained on historical trajectory data and are capable of predicting the future trajectory of a vehicle or pedestrian. TCNs have been shown to outperform recurrent networks in tasks such as handwritten recognition, audio synthesis, and time-series data. One advantage of TCNs is their ability to handle variable-length sequences without information leakage. They offer a powerful approach for modeling temporal dependencies in trajectory prediction tasks, capturing short-term

and long-term dynamics efficiently and with interpretable receptive fields. However, additional considerations may be required for spatial relationships and long-term memory. [20]

TCN was included in this paper to provide a different inductive bias from RNNs. It is particularly good at learning local temporal motifs and seasonal patterns from fixed windows without explicit feature engineering. With relatively few parameters, TCN offers a parameter-efficient deep model that can fairly test whether convolutional temporal patterns improve DO forecasts over recurrent and traditional ML approaches.

3.7. Comparison between the 6 models

Model	Inductive Bias	Nonlinearity & Interactions	Needs Lag/Seasonality Features	Works Well with Small Weekly Data
Ridge	Additive, smooth	X (linear)	High (lags + sin/cos)	✓
Random Forest	Piecewise rules	✓	Medium	✓
HistGradient	Additive trees	✓✓	Medium	✓
LSTM	State memory	✓✓	Low (uses raw windows)	△ (needs more data)
GRU	State memory (lean)	✓✓	Low	△
TCN	Local temporal patterns	✓✓	Low	△ – ✓ (parameter-efficient)

Table 1: Comparison of model properties for time series forecasting. Checkmarks (✓) indicate strong suitability, triangles (△) indicate partial suitability, and X denotes limitations.

4. Protocol Design & Research Method

4.1. Data

The dataset used, *by Supriyo Ain*, comprises detailed records of water quality parameters collected once every two weeks. These parameters include turbidity, pH, dissolved oxygen (DO), salinity, and temperature. According to the data source, the sampling was conducted at designated locations within various water bodies such as Bay, D-Pool (fishing pond), C-Pool, and A-Pool. [9] Here's the link to the dataset: <https://www.kaggle.com/datasets/supriyoain/water-quality-data>.

4.2. Models Considered

This paper compares 6 different models: 3 Traditional Models and 3 Deep Learning Models. The traditional models include Ridge Regression, Random Forest, and HistGradient-Boosting Regressor. Furthermore, the Deep Learning Models involve Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Network (TCN). The theory and reason behind choosing each of the models is explained in the Theoretical Framework Section.

4.3. Evaluation Metrics

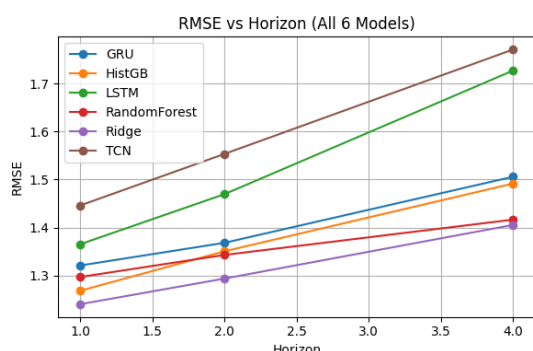
The models were evaluated using the RMSE and MAE (mg/L) for accuracy, and R^2 and NSE for goodness-of-fit. To ground results operationally, we also report Skill vs Persistence (improvement over a naïve **next=last** baselines; values 0 indicate the model beats persistence).

- **RMSE:** Root Mean Square Error is the measure of how well a regression line fits the data points. RMSE can also be construed as Standard Deviation in the residuals. [21]
- **MAE:** Mean Absolute Error (MAE) is a simple yet powerful metric used to evaluate the accuracy of regression models. It measures the average absolute difference between the predicted values and the actual target values. [23]
- **NSE:** The NSE (Nash-Sutcliffe efficiency) is a widely used metric in hydrological modeling to assess the performance of a model, it is a normalized statistic that compares the residual variance of the model to the variance of the measured data. [22]

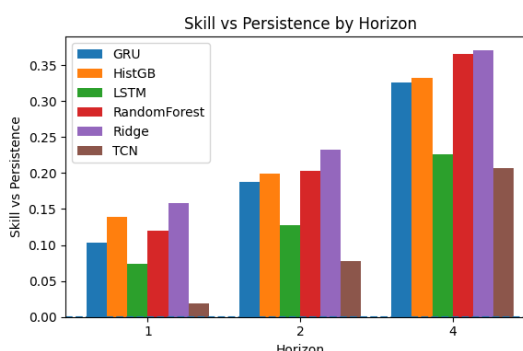
Here's the link to the code that was used to train all the models:

<https://colab.research.google.com/drive/1KTQZQshghCJJiGo1pI5Hn7x0kR1r6Rc2?usp=sharing>

5. Findings and Conclusion



(a) RMSE vs Horizon



(b) Skill vs Persistence

According to the RMSE Evaluation Metrics, the higher the RMSE value, the worse the model is. From Graph (a), generated by Python itself, it is visible that Ridge Regression has the least RMSE value, and TCN has the highest. Furthermore, the Skill vs Persistence

supports the theory, displaying the Ridge Regression models displaying the highest Skill vs Persistence value. At the same time, the lowest values throughout are displayed by the TCN Model. This corroborates the initial works studied further giving support to the idea that classical, tabular ML models are great at predicting Dissolved Oxygen and Water Quality.

However, a limitation to this study could be the limited dataset: Deep learning models typically require large samples, and richer, higher-frequency covariates to learn long-range dependencies and nonlinear interactions. With weekly sampling, we had to restrict capacity which likely led to underfitting and shrinkage toward the mean. Thus showing that Traditional Models Deep Learning Models. This could also be a potential for future research, where heavier datasets with larger samples could be considered for model comparisons.

© GSJ

Bibliography

- [1] E. Hogue, “The Importance of Water Quality,” *sanctuaries.noaa.gov*, Aug. 2021.
<https://sanctuaries.noaa.gov/news/aug21/water-quality-month.html>
- [2] S. W. M. Team, “What are the main threats to good water status?,” *Smart Water Magazine*, Apr. 18, 2023. <https://smartwatermagazine.com/q-a/what-are-main-threats-good-water-status>
- [3] M. Sankaran, “Why water quality is a global concern - KETOS,” *KETOS*, Oct. 06, 2022.
<https://ketos.co/why-water-quality-is-a-global-concern>
- [4] WHO, “Water supply, sanitation and hygiene monitoring,” *www.who.int*, 2023.
<https://www.who.int/teams/environment-climate-change-and-health/water-sanitation-and-health/monitoring-and-evidence/wash-monitoring>
- [5] Environmental Protection Agency, “Indicators: Dissolved Oxygen,” *United States Environmental Protection Agency*, Jan. 10, 2025.
<https://www.epa.gov/national-aquatic-resource-surveys/indicators-dissolved-oxygen>
- [6] Z. Hao, “A dissolved oxygen prediction model based on GRU–N-Beats,” *Frontiers in marine science*, vol. 11, May 2024, doi: <https://doi.org/10.3389/fmars.2024.1365047>.
- [7] Y. Zhao and M. Chen, “Prediction of river dissolved oxygen (DO) based on multi-source data and various machine learning coupling models,” *PLoS ONE*, vol. 20, no. 3, pp. e0319256–e0319256, Mar. 2025, doi: <https://doi.org/10.1371/journal.pone.0319256>.
- [8] G. Boesch, “What’s the difference between Machine Learning and Deep Learning?,” *viso.ai*, Jul. 05, 2022. <https://viso.ai/deep-learning/deep-learning-vs-machine-learning/>
- [9] “Water Quality Data ,” *www.kaggle.com*.
<https://www.kaggle.com/datasets/supriyoain/water-quality-data>
- [10] IBM, “Ridge Regression,” *Ibm.com*, Sep. 21, 2023.
<https://www.ibm.com/think/topics/ridge-regression>
- [11] S. Brown, “Machine Learning, Explained,” *MIT Sloan*, Apr. 21, 2021.
<https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>
- [12] M. P. Allen, “The problem of multicollinearity,” *Understanding Regression Analysis*, pp. 176–180, 1997, doi: https://doi.org/10.1007/978-0-585-25657-3_37.
- [13] M. El-Dereny and N.I. Rashwan, “Solving multicollinearity problem using ridge regression models,” vol. 6, no. 9, Jan. 2011, Available:

https://www.researchgate.net/publication/267660465_Solving_multicollinearity_problem_using_ridge_regression_models

[14] IBM, “Random Forest,” *Ibm.com*, Oct. 20, 2021. <http://ibm.com/think/topics/random-forest>

[15] “sklearn.ensemble.HistGradientBoostingRegressor,” *scikit-learn*.

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.HistGradientBoostingRegressor.html>

[16] GeeksforGeeks, “HistGradientBoostingClassifier in Sklearn,” *GeeksforGeeks*, Jun. 06, 2024. <https://www.geeksforgeeks.org/machine-learning/histgradientboostingclassifier-in-sklearn/>

[17] GeeksforGeeks, “Understanding of LSTM Networks,” *GeeksforGeeks*, May 10, 2020. <https://www.geeksforgeeks.org/machine-learning/understanding-of-lstm-networks/>

[18] GeeksforGeeks, “What is LSTM Long Short Term Memory?,” *GeeksforGeeks*, Jan. 16, 2019. <https://www.geeksforgeeks.org/deep-learning/deep-learning-introduction-to-long-short-term-memory/>

[19] GeeksforGeeks, “Gated Recurrent Unit Networks,” *GeeksforGeeks*, Jul. 09, 2019. <https://www.geeksforgeeks.org/machine-learning/gated-recurrent-unit-networks/>

[20] Vibha Bharilya and N. Kumar, “Machine learning for autonomous vehicle’s trajectory prediction: A comprehensive survey, challenges, and future research directions,” *Vehicular Communications*, vol. 46, pp. 100733–100733, Jan. 2024, doi: <https://doi.org/10.1016/j.vehcom.2024.100733>.

[21] GeeksforGeeks, “ML | Mathematical explanation of RMSE and Rsquared error,” *GeeksforGeeks*, Aug. 2019. <https://www.geeksforgeeks.org/machine-learning/ml-mathematical-explanation-of-rmse-and-r-squared-error/> (accessed Sep. 22, 2025).

[22] Sohrab Kolsoumi, “Nash Sutcliffe Model Efficiency Coefficient Calculator,” *Agrimetsoft.com*, 2023. https://agrimetsoft.com/calculators/Nash%20Sutcliffe%20model%20Efficiency%20coefficient?srsltid=AfmBOoqlcHdVisnt92xn7ezwDpzHUCqdHsBznn-Dt7ADKSCceXMO_NwQ#google_vignette (accessed Sep. 22, 2025).

[23] M. W. Ahmed, “Understanding Mean Absolute Error (MAE) in Regression: A Practical Guide,” *Medium*, Aug. 24, 2023. <https://medium.com/@m.waqar.ahmed/understanding-mean-absolute-error-mae-in-regression-a-practical-guide-26e80ebb97df>