RADINKA JOURNAL OF SCIENCE
AND SYSTEMATIC LITERATURE REVIEW

ISSN 2988-5825

https://www.rjupublisher.com/ojs/index.php/RJSLR

Spatiotemporal Water Quality Modeling Using Deep Learning Architectures

Ejenarhome Otega PROSPER[a], Godfrey Perfectson OISE[b], Agwam Gladys IFEOMA[c]

[a]*Department of Computer Science, Delta State University, Delta State, Nigeria*
[b]*Department of Computing, Wellspring University, Edo State, Nigeria*
[c]*Department of Computer Science, Western Delta University, Delta State, Nigeria*

***Corresponding author:** godfrey.oise@wellspringuniversity.edu.ng

Abstract

Water quality monitoring is crucial for protecting public health and maintaining aquatic ecosystems, but traditional methods remain costly, labor-intensive, and limited in their spatial and temporal coverage. This study explores the use of machine learning and deep learning models for spatiotemporal water quality prediction using historical environmental data. The goals were to evaluate both supervised and unsupervised models for Water Quality Index (WQI) prediction, address challenges posed by data imbalance and poor precision for minority classes, and highlight the importance of interpretable approaches that aid decision-making. After conducting exploratory data analysis and feature visualization, models such as Support Vector Regression (SVR), Random Forest, XGBoost, Decision Trees, Logistic Regression, and DBSCAN were employed. SVR achieved high predictive accuracy ($R^2 = 0.9693$), while XGBoost and Decision Trees attained over 94% accuracy in classification tasks, with Decision Trees excelling at detecting minority pollution classes. DBSCAN showed moderate clustering ability, affected by noise and overlapping class boundaries. Major challenges identified include severe class imbalance in water quality data, lower precision for rare pollution events, and limited interpretability of complex models. The study suggests that combining class-balancing methods, explainable AI (XAI), and scalable architectures like edge or federated learning can greatly enhance future water-quality monitoring systems. These findings support the development of transparent, trustworthy, and SDG-aligned water management solutions.

Keywords: Water Quality Index (WQI), Spatiotemporal Modeling, Exploratory Data Analysis (EDA), Explainable AI (XAI), Deep Learning (DL), DBSCAN, XGBoost

Acknowledgments: Thank you to all parties who have helped with this research

For Citation:

Prosper, E, O., Oise, G. P, Ifeoma, A. G. (2025). Spatiotemporal Water Quality Modeling Using Deep Learning Architectures. *Radinka Journal of Science and Systematic Literature Review, 3*(3), 761-

Introduction

Water quality plays a central role in safeguarding public health, maintaining ecological balance, and supporting socio-economic development. Across many regions of the world, freshwater systems are increasingly stressed by rapid urban expansion, agricultural intensification, industrial pollution, and the impacts of climate variability (Archit Nandan, 2024). These pressures heighten the need for reliable, timely, and scalable water-quality monitoring practices capable of capturing both short-term fluctuations and long-term environmental trends (Oise & Konyeha, 2025), (LCRA, 2025). Conventional water-quality monitoring methods, though scientifically rigorous, remain constrained by high operational costs, labour-intensive procedures, and limited spatial and temporal coverage (Bhuria et al., 2024), (Malakar et al., 2021). Manual sampling and laboratory analysis often introduce delays that hinder the rapid detection of contamination events and reduce the responsiveness of water-resource management (Huang et al., 2024). As environmental conditions evolve more rapidly, these traditional approaches struggle to deliver the continuous, high-resolution insights required for modern decision-making (Peterson et al., 2020), (Shukla & Verma, 2024)

Advances in artificial intelligence, particularly machine learning (ML) and deep learning (DL), have introduced new possibilities for automated, data-driven water-quality assessment (Belyadi & Haghighat, 2021) . These models can process large and heterogeneous datasets, extract complex relationships among physical, chemical, and biological parameters, and generate predictive or classification outputs with high accuracy (Cai et al., 2020), (More & Wolkersdorfer, 2022). Despite these advantages, recent research also highlights important limitations, including the prevalence of class imbalance in environmental datasets, reduced model precision for minority pollution events (Oise, Konyeha, et al., 2025), and the lack of interpretability in deep learning models that often function as “black boxes.” Recognizing these limitations, the present study does not aim to introduce a novel algorithm or modeling framework. Instead, it focuses on systematically applying and comparing well-established ML and DL models, including Support Vector Regression, Random Forest, XGBoost, Decision Trees, Logistic Regression, and DBSCAN (Varela Pérez et al., 2022), to a publicly available historical dataset. The goal is to provide a clear assessment of their performance under realistic data conditions, examine the challenges that arise in practice, and explore how issues like data imbalance and limited minority-class precision affect the reliability of water-quality predictions (Leggesse et al., 2023).

Accordingly, this study aims to evaluate the effectiveness of supervised and unsupervised learning models for Water Quality Index (WQI) prediction and water-quality classification; investigate how data imbalance influences model behavior, particularly for rare pollution events; assess the interpretability of different modeling approaches in the context of water-resource management; and discuss practical recommendations for improving the transparency, scalability, and real-world applicability of AI-based water-quality monitoring systems.

Methodology

The methodology of this study was designed to systematically evaluate the performance of established machine learning (ML) models for predicting and classifying water quality using historical environmental data. The workflow consisted of data acquisition, cleaning, preprocessing, exploratory analysis, supervised and unsupervised modeling, hyperparameter tuning, and model evaluation. This structured approach ensures clarity, reproducibility, and rigorous comparison across models.

Data Collection and Preparation

The dataset was obtained from a government repository on Kaggle and includes physicochemical parameters commonly used in water-quality assessment: temperature, pH, dissolved oxygen (DO), conductivity, and the Water Quality Index (WQI) (Archit Nandan, 2024). . Extensive data cleaning was performed before modeling. Missing values in continuous variables were handled using mean imputation, while categorical values (if present) were used with mode imputation. Erroneous values were identified using physical plausibility thresholds, for example, pH outside 0–14, DO exceeding 20 mg/L, and conductivity above 30,000 µS/cm. Values beyond these ranges were removed to preserve dataset integrity.

Exploratory Data Analysis (EDA)

An exploratory analysis was conducted to understand feature distributions and correlations. Visual tools such as histograms, boxplots, Kernel Density Estimates (KDE), and correlation heatmaps were used to examine patterns, detect outliers, and guide feature treatment. Dimensionality reduction techniques, including Principal Component Analysis (PCA) and t-SNE, were applied to visualize variance structure and potential groupings among water-quality samples.

Preprocessing and Feature Scaling

Since many ML algorithms are sensitive to feature magnitude, all continuous variables were standardized using Z-score normalization (StandardScaler). This step was essential for models such as Support Vector Regression (SVR), Logistic Regression, and DBSCAN, which rely on distance-based computations. For DBSCAN clustering specifically, DO and pH were independently standardized to ensure balanced contribution to distance metrics.

Supervised Learning Models

Both regression and classification tasks were performed.
Regression: SVR was used to predict continuous WQI values.
Classification: Logistic Regression, Random Forest, XGBoost, and Decision Tree classifiers were trained to categorize water quality into discrete classes.
Since WQI is originally continuous, it was categorized using recognized environmental thresholds:
0–50: Good
51–100: Moderate
101–200:Poor
These thresholds produced three classes, though the distribution was imbalanced, with fewer samples in the Poor class.

Hyperparameter Tuning and Model Validation

To ensure robust evaluation, all supervised models were trained using 5-Fold Cross-Validation. Model optimization was performed through Grid Search, with key hyperparameters tuned as follows:
Random Forest: number of trees (100–500), max depth (5–20)
XGBoost: learning rate (0.01–0.3), estimators (100–300), max depth (3–10)
SVR: C (1–100), kernel type (RBF/linear), gamma (scale/auto)
Logistic Regression: regularization (C = 0.1–10), L1/L2 penalties
Regression models were evaluated using R^2 and MSE, while classification models used accuracy, precision, recall, macro-F1, and weighted-F1 scores to capture both majority and minority class performance.

Unsupervised Learning Using DBSCAN

DBSCAN was employed to identify natural clusters within the dataset. Parameter sensitivity was addressed by examining k-distance plots and iteratively tuning values. The final parameters selected were:
ε (epsilon) = 0.5
MinPts = 5

These parameters balanced cluster density and noise detection. Clustering quality was measured using the Silhouette Score, calculated after excluding noise points.

	DO	PH	Conductivity	WQI	WQI clf
0	6.7	7.5	203.0	46.730530	3
1	5.7	7.2	189.0	46.330319	3
2	6.3	6.9	179.0	38.233597	3
3	5.8	6.9	64.0	40.613732	3
4	5.8	7.3	83.0	47.561912	3
5	5.5	7.4	81.0	50.740733	0
6	6.1	6.7	308.0	35.752151	3
7	6.4	6.7	414.0	34.332210	3
8	6.4	7.6	305.0	49.932230	3
9	6.3	7.6	77.0	50.363157	0

Figure 1. First 10 items in the dataset

Results and Discussion

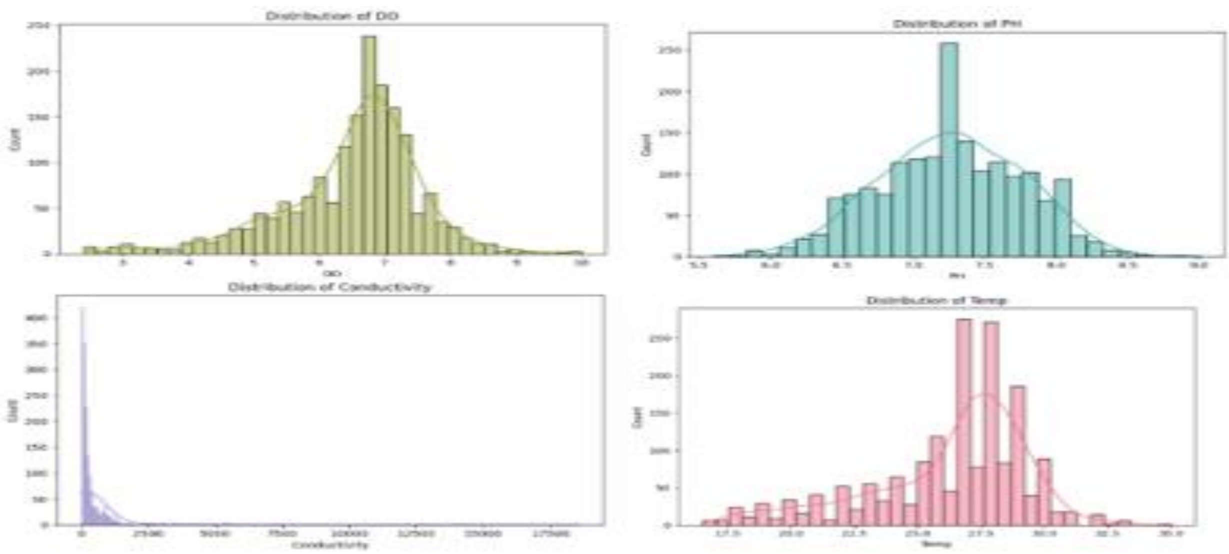

Figure 2. Water Quality Parameter Distributions

Figure 2 shows the temperature, which has a mildly right-skewed distribution with most readings around 27–28°C. Dissolved oxygen is roughly normally distributed, peaking at 7–8 mg/L.

pH values form a nearly symmetrical, neutral-centered distribution (7.2–7.4). Conductivity is strongly right-skewed, with most values near zero but occasional very high readings indicating sporadic spikes in ionic concentration.

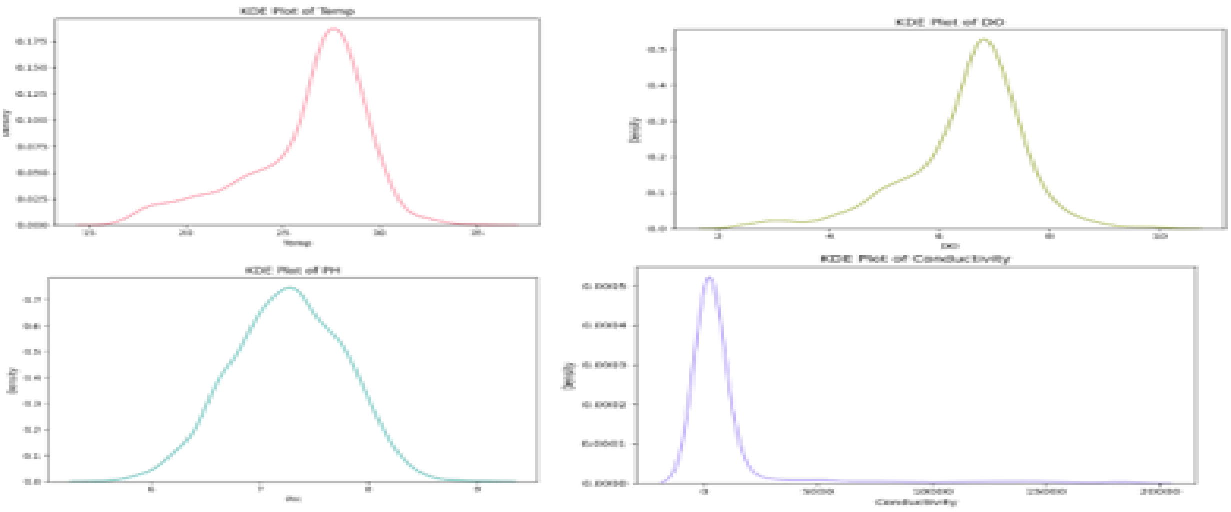

Figure 3. Kernel Density Estimates of Water Quality Parameters

Figure 2 displays the KDE plot for temperature, showing two main clusters: a primary peak near 28°C and a secondary peak around 23°C, within a 15–35°C range. Dissolved oxygen has a dominant peak at 7 mg/L, with a smaller shoulder at 4–5 mg/L, indicating occasional lower DO levels. pH shows a symmetrical, neutral-centered curve around 7.3, with decreasing likelihood toward acidic or alkaline extremes (5.5–9.0). Conductivity peaks sharply near zero, with rare but notable high-value spikes above 20,000 µS/cm, suggesting isolated pollution or mineral influx events.

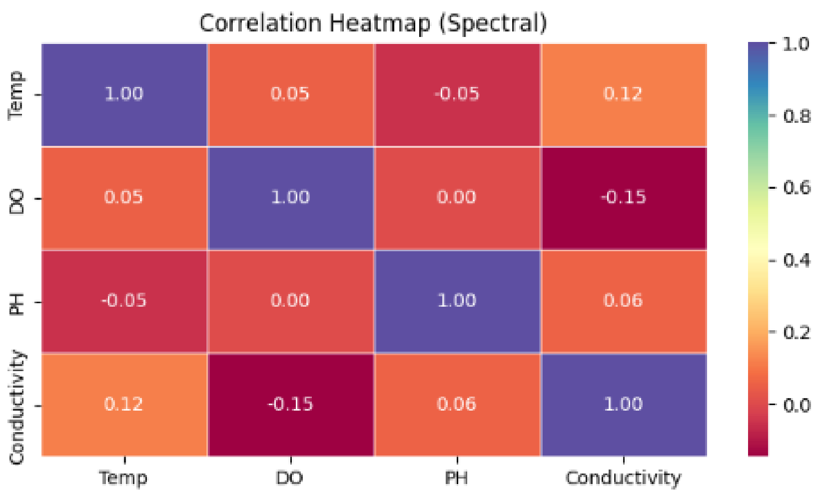

Figure 4. Correlation Heatmap of Temperature, Dissolved Oxygen (DO), pH, and Conductivity.

The correlation heatmap reveals generally weak linear relationships among the key water quality parameters: Temperature, Dissolved Oxygen (DO), pH, and Conductivity. Temperature has a very weak positive correlation with DO and Conductivity, and a very weak negative correlation with pH. DO shows minimal linear association with both Temperature and pH, but has a weak negative correlation with Conductivity. Similarly, pH appears to have almost no linear correlation with any of the other variables, including DO, Temperature, and Conductivity. Conductivity itself demonstrates a weak positive correlation with Temperature and a weak negative correlation with DO, but lacks significant correlation with pH

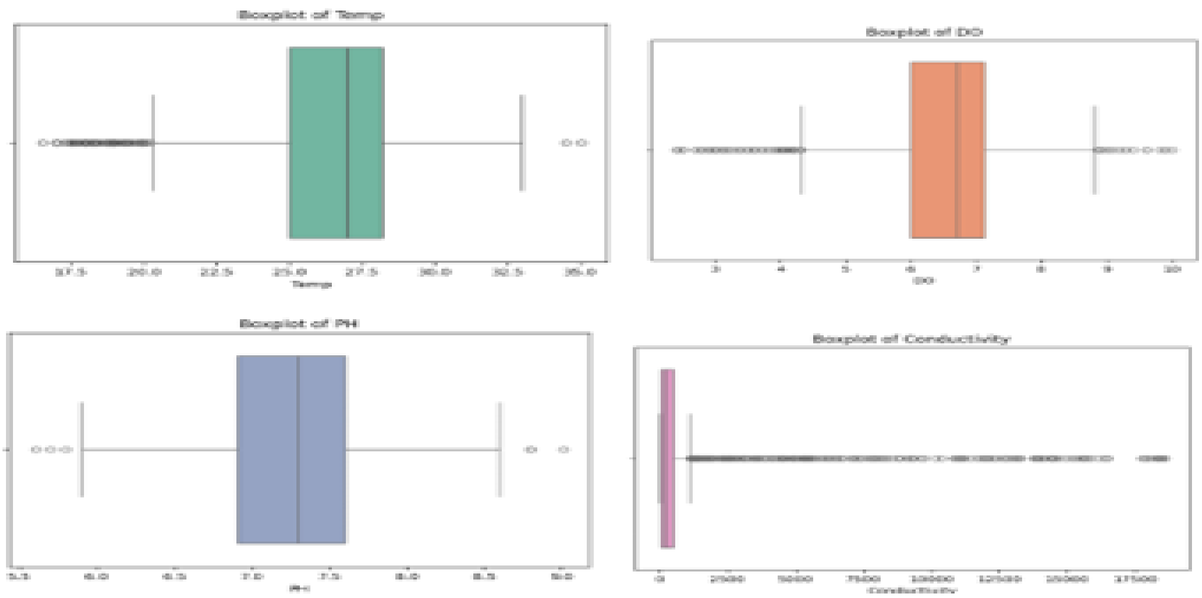

Figure 5. Water Quality Parameter Boxplots

Figure 5 depicts box plots of water quality parameters. Temperature (median ≈26.5°C) and DO (median ≈6.7 mg/L) show moderate spread with some outliers. pH (median 7.3) is tightly clustered near neutrality. Conductivity is right-skewed with several high outliers, indicating occasional spikes in dissolved ions.

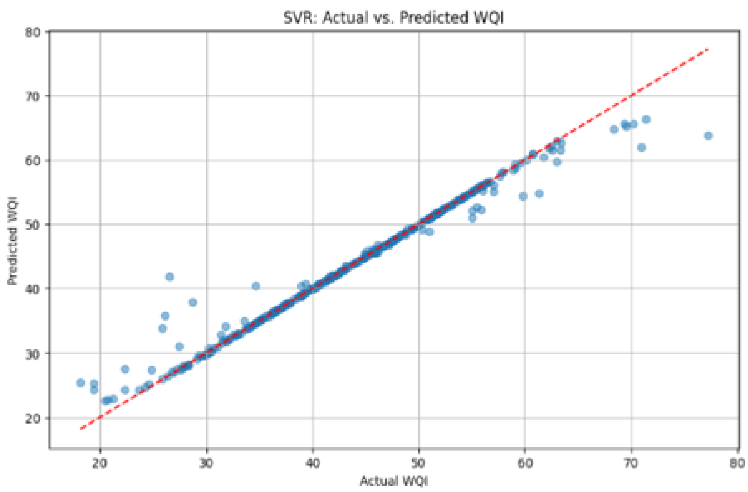

Figure 6. SVR: Actual vs. Predicted WQI.

This scatter plot evaluates the performance of a Support Vector Regressor (SVR) model by comparing its predicted Water Quality Index (WQI) values against the actual WQI values. The close clustering of the blue data points around the red y=x line indicates a strong agreement between the predictions and the actual values. This is further supported by the high R-squared value of 0.9693, signifying that the model explains a large proportion of the variance in the actual

WQI, and the relatively low Mean Squared Error of 3.1749, suggesting small average prediction errors. Overall, the plot visually confirms the SVR model's accurate prediction of the Water Quality Index.

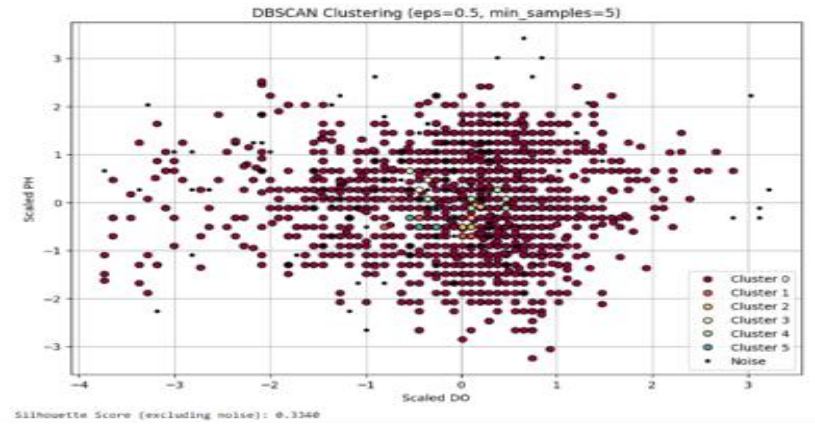

Figure 7. DBSCAN Clustering of Scaled Dissolved Oxygen and pH Data

This scatter plot visualizes the results of DBSCAN clustering on scaled Dissolved Oxygen and pH data, using parameters epsilon=0.5 and minimum samples=5. The algorithm identified six distinct clusters, each represented by a different color, and classified 106 data points as noise (shown in black). The plot reveals the spatial distribution of these clusters in the two-dimensional space of scaled DO and pH. However, the silhouette score (excluding noise) of -0.1568 suggests that the resulting clusters might not be well-separated or internally cohesive, indicating potential suboptimal parameter selection or an inherent challenge in clustering this specific dataset with the current DBSCAN configuration.

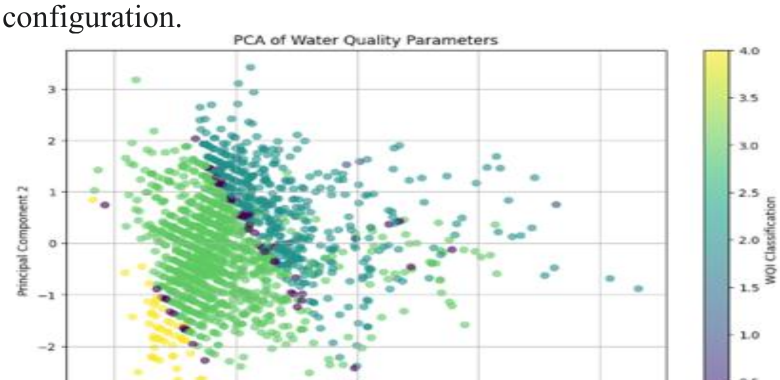

Figure 8. PCA of Water Quality Parameters

This scatter plot visualizes the results of a Principal Component Analysis (PCA) applied to water quality parameters, reducing the data to two principal components (PC1 and PC2), which together explain approximately 72% of the original variance. Each point represents a water sample, colored according to its Water Quality Index (WQI) classification. The plot reveals some clustering of samples based on WQI, particularly in the bottom-left, suggesting that the principal components capture some of the variance related to water quality categories. However, there is also overlap between different WQI classifications, indicating that while PC1 and PC2 capture a significant portion of the data's variability, they don't perfectly separate the defined water quality categories.

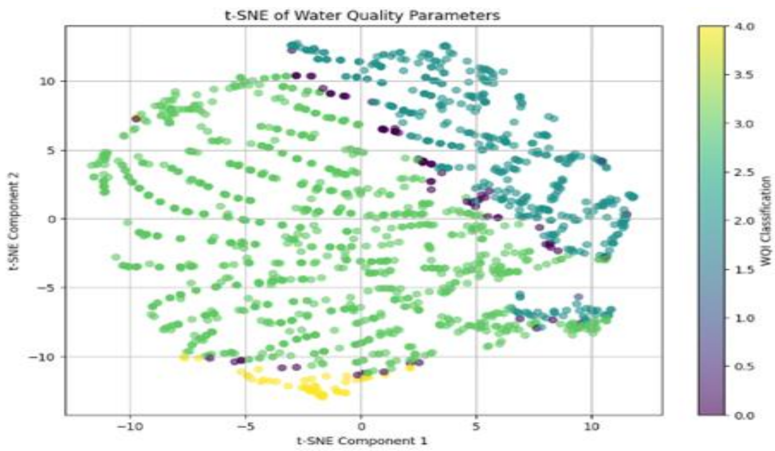

Figure 9. t-Distributed Stochastic Neighbor Embedding (t-SNE)

This scatter plot displays the results of t-Distributed Stochastic Neighbour Embedding (t-SNE) applied to water quality parameters, reducing the data to two dimensions to visualize potential underlying structure, with each point coloured by its Water Quality Index (WQI) classification. The t-SNE technique aims to preserve the local relationships in the high-dimensional data, and the resulting plot shows some degree of clustering of points with similar WQI classifications, such as a distinct cluster of yellow points at the bottom. However, there is also considerable overlap between different colour groups, indicating that while t-SNE reveals some separation based on WQI, the water quality parameters do not perfectly distinguish these classifications in the reduced two-dimensional space.

```
--- Logistic Regression ---
Accuracy:  0.952112676056338
Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        12
           1       0.00      0.00      0.00         1
           2       0.96      1.00      0.98        92
           3       0.95      1.00      0.97       238
           4       1.00      0.67      0.80        12

    accuracy                           0.95       355
   macro avg       0.58      0.53      0.55       355
weighted avg       0.92      0.95      0.93       355
```

```
--- Random Forest Classifier ---
Accuracy: 0.9492957746478873
Classification Report:
              precision    recall  f1-score   support

           0       0.64      0.58      0.61        12
           1       0.00      0.00      0.00         1
           2       0.95      0.95      0.95        92
           3       0.96      0.99      0.98       238
           4       1.00      0.58      0.74        12

    accuracy                           0.95       355
   macro avg       0.71      0.62      0.65       355
weighted avg       0.95      0.95      0.95       355
```

```
--- XGBoost Classifier ---
Accuracy: 0.9549295774647887
Classification Report:
              precision    recall  f1-score   support

           0       0.64      0.58      0.61        12
           1       0.00      0.00      0.00         1
           2       0.96      0.93      0.95        92
           3       0.97      0.99      0.98       238
           4       0.91      0.83      0.87        12

    accuracy                           0.95       355
   macro avg       0.69      0.67      0.68       355
weighted avg       0.95      0.95      0.95       355
```

```
--- Decision Tree Classifier ---
Accuracy: 0.9436619718309859
Classification Report:
              precision    recall  f1-score   support

           0       0.41      0.58      0.48        12
           1       1.00      1.00      1.00         1
           2       0.98      0.95      0.96        92
           3       0.97      0.98      0.97       238
           4       0.88      0.58      0.70        12

    accuracy                           0.94       355
   macro avg       0.85      0.82      0.82       355
weighted avg       0.95      0.94      0.95       355
```

Figure 10. Classification Report of the models

The classification reports highlight high overall accuracy for Logistic Regression (0.9521), Random Forest Classifier (0.9493), XGBoost Classifier (0.9549), and Decision Tree Classifier (0.9437). However, performance on individual classes varies, with low precision and recall observed for classes 0 and 1 across most models. XGBoost generally shows a strong balance with a macro F1-score of 0.68, while the Decision Tree Classifier achieves the highest macro precision at 0.85. Logistic Regression has lower macro precision (0.58) and recall (0.53). The weighted F1-scores are consistently high (around 0.93-0.95) for all models, reflecting good performance on the majority of classes. In essence, while all models achieve similar high accuracy, their ability to correctly classify minority classes and their balance between precision and recall differ. The exploratory analysis provided important context for interpreting the model performances. Temperature and conductivity displayed right-skewed distributions, indicating episodic spikes likely attributable to environmental or pollution events (Yin et al., 2021). Dissolved oxygen and pH exhibited more symmetric distributions, reflecting natural variability around typical environmental ranges (He et al., 2024).

The weak correlations observed among all parameters explain why linear models struggled to classify minority categories reliably (Chauhan et al., 2024). PCA accounted for approximately 72% of the variance in its first two components, while t-SNE revealed partial but incomplete separation among water quality categories. These findings confirm the multivariate and nonlinear nature of the dataset, thereby justifying the superior performance of nonlinear models such as XGBoost and Decision Trees. (Oise, Nwabuokei, et al., 2025). Taken together, the results demonstrate that machine learning methods are highly effective for both predictive and classification-based assessment of water quality (Pereira et al., 2024). SVR proved to be a reliable model for continuous WQI prediction, while XGBoost and Decision Tree models provided strong classification performance, particularly in handling class imbalance. The performance limitations of DBSCAN and linear classifiers further highlight the importance of selecting model architectures aligned with the intrinsic characteristics of environmental datasets (Getachew & Manjunatha, 2022). The insights from EDA reinforce the necessity of nonlinear, ensemble-based, and density-aware techniques for robust water quality modeling. These findings collectively position machine learning as a powerful tool for modernizing water quality monitoring, contributing to the broader objectives of sustainable water resource management.

Conclusion

This study highlights the potential of machine learning (ML) and deep learning (DL) for improving spatiotemporal water quality monitoring. Using parameters such as temperature, pH, dissolved oxygen, conductivity, and Water Quality Index (WQI), the study evaluated multiple supervised and unsupervised models. The Support Vector Regressor (SVR) demonstrated excellent WQI prediction ($R^2 = 0.9693$), while XGBoost achieved the highest classification accuracy (95.5%) with balanced performance. Decision Trees were effective for identifying minority classes, and DBSCAN provided moderate clustering, though limited by noise and overlapping clusters. Exploratory analysis tools like KDE, PCA, and t-SNE proved useful in revealing data patterns and guiding model selection. The findings indicate that AI-driven approaches can enhance water quality prediction, classification, and monitoring, supporting informed decision-making. Limitations include class imbalance, model interpretability, and adaptability to dynamic environmental conditions, which may affect practical deployment. Future research should focus on integrating explainable AI (XAI), edge computing, federated learning, and hybrid models to improve transparency, scalability, and performance, particularly in resource-limited settings. Advancing clustering methods and adaptive models will further strengthen real-time monitoring capabilities. Overall, this study provides a data-driven framework for intelligent water quality assessment, contributing to sustainable water management and the objectives of Sustainable Development Goal 6.

Conflicts of Interest

There were no Conflicts of Interest

Author Contributions

Ejenarhome Otega Prosper: conceptualization, methodology, validation; Agwam Gladys Ifeoma: formal analysis, investigation; Godfrey Perfectson Oise: resources, data curation; Godfrey Perfectson Oise and Agwam Gladys Ifeoma: original draft writing; Ejenarhome Otega Prosper, Godfrey Perfectson Oise, and Agwam Gladys Ifeoma: review and editing, visualization. All authors have read and approved the final manuscript.

Funding

The research and writing of this article were funded by personal funds.

References

Archit Nandan. (2024). *Water Quality Prediction Data Analytics Project* [Dataset]. Kaggle online data repository. https://www.kaggle.com/code/architnandan/water-quality-prediction-data-analytics-project/input

Belyadi, H., & Haghighat, A. (2021). Supervised learning. In *Machine Learning Guide for Oil and Gas Using Python* (pp. 169–295). Elsevier. https://doi.org/10.1016/B978-0-12-821929-4.00004-4

Bhuria, R., Gill, K. S., Upadhyay, D., & Devliyal, S. (2024). Predicting Water Purity by Riding the Ensemble Waves with Gradient Boosting Classification Technique. *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 1365–1368. https://doi.org/10.1109/ICSCSS60660.2024.10624936

Cai, J., Xu, K., Zhu, Y., Hu, F., & Li, L. (2020). Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. *Applied Energy*, *262*, 114566. https://doi.org/10.1016/j.apenergy.2020.114566

Chauhan, R., Shukla, S., Pokhriyal, S., Devliyal, S., & Kumar, R. R. (2024). Enhancement of Usability of AI Chatbots using Machine Learning. *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*, 1–6. https://doi.org/10.1109/APCIT62007.2024.10673475

Getachew, B., & Manjunatha, B. R. (2022). Impacts of Land-Use Change on the Hydrology of Lake Tana Basin, Upper Blue Nile River Basin, Ethiopia. *Global Challenges*, *6*(8), 2200041. https://doi.org/10.1002/gch2.202200041

He, M., Qian, Q., Liu, X., Zhang, J., & Curry, J. (2024). Recent Progress on Surface Water Quality Models Utilizing Machine Learning Techniques. *Water*, *16*(24), 3616. https://doi.org/10.3390/w16243616

Huang, R., Yu, C., Wang, H., Zhang, S., Wang, L., Li, H., Zhang, Z., & Zhou, Z. (2024). Spatial and Temporal Modeling on Energy Consumption of Wastewater Treatment Based on Machine Learning Algorithms. *ACS ES&T Water*, *4*(3), 1119–1130. https://doi.org/10.1021/acsestwater.3c00430

LCRA. (2025). *Water Quality Indicators*. Lower Colorado River Authority. https://www.lcra.org/water/quality/colorado-river-watch-network/water-quality-indicators/

Leggesse, E. S., Zimale, F. A., Sultan, D., Enku, T., Srinivasan, R., & Tilahun, S. A. (2023). Predicting Optical Water Quality Indicators from Remote Sensing Using Machine Learning Algorithms in Tropical Highlands of Ethiopia. *Hydrology*, *10*(5), 110. https://doi.org/10.3390/hydrology10050110

Malakar, P., Sarkar, S., Mukherjee, A., Bhanja, S., & Sun, A. Y. (2021). Use of machine learning and deep learning methods in groundwater. In *Global Groundwater* (pp. 545–557). Elsevier. https://doi.org/10.1016/B978-0-12-818172-0.00040-2

More, K. S., & Wolkersdorfer, C. (2022). Predicting and Forecasting Mine Water Parameters Using a Hybrid Intelligent System. *Water Resources Management*, *36*(8), 2813–2826. https://doi.org/10.1007/s11269-022-03177-2

Oise, G. P., Konyeha, C. C., Onwuzo, C. J., Otega, E. P., Akilo, B. E., Comfort, O. T., Odimayomi, J. A., & Belinda, U. N. (2025). Interpretable air quality classification for public health using machine learning. *Advances in Computing and Engineering*, *5*(2), 75. https://doi.org/10.21622/ACE.2025.05.2.1428

Oise, G. P., & Konyeha, S. (2025). Optimization of E-Waste Sorting Process Using Deep Learning. *Radinka Journal Of Science And Systematic Literature Review*, *3*(2), 612–622. https://doi.org/10.56778/rjslr.v3i2.503

Oise, G. P., Nwabuokei, O. C., Akpowehbve, O. J., Eyitemi, B. A., & Unuigbokhai, N. B. (2025). Towards Smarter Cyber Defense: Leveraging Deep Learning For Threat Identification And Prevention. *Fudma Journal Of Sciences*, *9*(3), 122–128. https://doi.org/10.33003/fjs-2025-0903-3264

Pereira, E., Santos, J., & Barboza, E. (2024). Comparing TinyML Models and Libraries for On-Device Water Potability Classification. *2024 XIV Brazilian Symposium on Computing Systems Engineering (SBESC)*, 1–6. https://doi.org/10.1109/SBESC65055.2024.10771818

Peterson, K. T., Sagan, V., & Sloan, J. J. (2020). Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. *GIScience & Remote Sensing*, *57*(4), 510–525. https://doi.org/10.1080/15481603.2020.1738061

Shukla, Ms. G., & Verma, S. B. (2024). An In-Depth Analysis of Current Development in Energy Efficient Method for MAC Protocol to Extend Life Span of WSN. *2024 International Conference on Cybernation and Computation (CYBERCOM)*, 20–24. https://doi.org/10.1109/CYBERCOM63683.2024.10803259

Varela Pérez, P., Greiner, B. E., & Von Cossel, M. (2022). Socio-Economic and Environmental Implications of Bioenergy Crop Cultivation on Marginal African Drylands and Key Principles for a Sustainable Development. *Earth*, *3*(2), 652–682. https://doi.org/10.3390/earth3020038

Yin, J., Medellín-Azuara, J., Escriva-Bou, A., & Liu, Z. (2021). Bayesian machine learning ensemble approach to quantify model uncertainty in predicting groundwater storage change. *Science of The Total Environment*, *769*, 144715. https://doi.org/10.1016/j.scitotenv.2020.144715