AI-POWERED DISSOLVED OXYGEN MONITORING AND PREDICTION

K. Shekhar[1*], Ravula Sai Sreeja[2], Murahari Jaya Krishna[2], Guvva Nikesh Goud[2], Guduri Teja Sri[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering

Vaagdevi College of Engineering(UGC - Autonomus), Bollikunta, Warangal, Telangana.

*Corresponding author: K. Shekhar(shekhar_k@vaagdevi.edu.in)

ABSTRACT

Dissolved oxygen (DO) is a key indicator of river ecosystem health, essential for the survival and respiration of aquatic organisms. Traditional DO monitoring involves collecting water samples and analyzing them in laboratories, which is time-consuming, expensive, and limited in both spatial and temporal coverage. These limitations hinder real-time monitoring and accurate prediction, making it challenging to maintain optimal DO levels. In response to these challenges, this research proposes a machine learning (ML)-based approach to predict DO levels in river water. ML models can process large and diverse datasets, identifying complex patterns and correlations that traditional methods may overlook. By leveraging historical and real-time environmental data, the proposed system enables continuous, automated, and accurate prediction of DO levels. This enhances water quality monitoring, promotes effective environmental conservation, and supports sustainable water resource management. The adoption of ML not only reduces the dependency on manual sampling but also provides a scalable solution for widespread implementation. Ultimately, this research supports proactive ecosystem protection and informed decision-making in water management.

Keywords: Dissolved oxygen, river health, machine learning, water quality, prediction model, environmental monitoring.

1. INTRODUCTION

Predicting dissolved oxygen levels in river water using machine learning is a multifaceted and essential area of research with significant environmental implications. Dissolved oxygen is a critical parameter in aquatic ecosystems, as it directly influences the survival and health of aquatic organisms. When dissolved oxygen levels drop below certain thresholds, it can lead to fish kills, reduced biodiversity, and other ecological imbalances. Machine learning offers a powerful approach to tackle this challenge by leveraging historical and real-time data [1]. To achieve accurate predictions, machine learning models are trained on a diverse dataset that includes variables such as temperature, pH, turbidity, nutrient concentrations, and even weather conditions. These models use statistical algorithms and computational techniques to learn the complex relationships between these environmental factors and dissolved oxygen levels. By analyzing patterns and correlations within the data, machine learning models can generate predictions of future dissolved oxygen levels in river water [2].

The benefits of predicting dissolved oxygen levels are multifaceted. Firstly, it allows for early detection of potential water quality issues. This early warning system can be instrumental in preventing or mitigating adverse effects on aquatic life, as it provides authorities with the information needed to take timely corrective measures [3]. Furthermore, the ability to forecast dissolved oxygen levels aids in resource management and pollution control. Regulatory agencies and environmental organizations can use these predictions to make informed decisions about water allocation, pollutant discharge limits, and other conservation efforts. In addition to its practical applications, predicting dissolved oxygen levels using machine learning contributes to a deeper understanding of the complex dynamics of river ecosystems. It enables researchers to explore intricate cause-and-effect relationships among environmental variables and dissolved oxygen levels, shedding light on the underlying ecological processes. This knowledge can inform the development of more effective environmental policies and sustainable river management practices [4]. In conclusion, the utilization of machine learning for predicting dissolved oxygen levels in river water represents a crucial intersection of environmental science and technology. It empowers us to proactively address water quality challenges, safeguard aquatic ecosystems, and enhance our comprehension of the intricate interactions within riverine environments. As the field of machine learning continues to advance, so too does our capacity to protect and preserve the health of our waterways [5].

1.2 Motivation

The motivation for predicting dissolved oxygen levels in river water using machine learning stems from a pressing need to address critical environmental and ecological concerns. Firstly, freshwater ecosystems, including rivers and streams, play a vital role in supporting biodiversity and sustaining aquatic life. However, these ecosystems are increasingly under threat due to pollution, habitat destruction, and the impacts of climate change. Dissolved oxygen serves as a crucial indicator of water quality in these environments. When oxygen levels drop below certain thresholds, it can lead to oxygen-deprived "dead zones," harming fish populations and disrupting the entire food web [6]. Therefore, accurate predictions of dissolved oxygen levels are essential to proactively detect and mitigate potential crises, preventing ecological imbalances and the loss of valuable species. Furthermore, the ever-growing human population and industrial activities contribute to mounting pressures on freshwater resources. Effective water resource management is paramount for ensuring a sustainable and reliable supply of clean water for drinking, agriculture, and industry. Predicting dissolved oxygen levels allows water authorities and policymakers to make informed decisions about water allocation, pollution control measures, and environmental regulations [7]. By optimizing these processes, it becomes possible to strike a balance between human needs and environmental preservation. Machine learning's motivation in this context arises from its ability to handle complex and large-scale datasets, extract intricate patterns, and provide accurate predictions. Traditional monitoring methods often involve sporadic sampling, which may not capture rapid changes in water quality [8]. Machine learning models, on the other hand, can continuously analyze real-time data from various sensors and sources, offering a more comprehensive and proactive approach to managing water resources and safeguarding aquatic ecosystems. In summary, the motivation for predicting dissolved oxygen levels in river water using machine learning is driven by the urgent need to protect freshwater ecosystems, ensure water quality for human consumption, and enhance our understanding of the intricate dynamics within these environments [9]. It represents a fusion of environmental conservation, technological innovation, and sustainable resource management, with the overarching goal of preserving our planet's vital water resources for future generations [10].

2. LITERATURE SURVEY

Ziyad Sami, et al. [11] developed a reliable prediction model to predict D.O. in the Fei-Tsui reservoir for better water quality monitoring. The proposed model is an artificial neural network (ANN) with one hidden layer. Twenty-nine years of water quality data have been used to validate the accuracy of the proposed model. A different number of neurons have been investigated to optimize the model's accuracy. Statistical indices have been used to examine the reliability of the model. Bolick, et al. [12] proposed a Comparison of machine learning algorithms to predict dissolved oxygen in an urban stream. A multiple linear regression model was compared to machine learning algorithms k-nearest neighbor, decision tree, random forest, and gradient boosting. These algorithms were evaluated to understand which best predicted dissolved oxygen (DO) from water temperature, conductivity, turbidity, and water level change at four locations along the urban stream.

Moon, et al. [13] proposed an urban river dissolved oxygen prediction model using machine learning. To predict the optimized WQ, we selected pH, SS, water temperature, total nitrogen (TN), dissolved total phosphorus (DTP), NH3-N, chemical oxygen demand (COD), dissolved total nitrogen (DTN), and NO3-N as the input variables of the AdaBoost model. Dissolved oxygen (DO) was used as the target variable. Nair, et al. [14] proposed Analysing and Modelling Dissolved Oxygen Concentration Using Deep Learning Architectures. This work employs deep learning algorithms like Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) effective prediction models for forecasting DO levels in river water. The models are developed and validated using the river water quality data collected from eleven sampling stations during the year 2016 to 2020. Zhou, et al. [15] proposed an interpretable and explainable model that integrates the shapley additive explanations (SHAP) algorithm with the long short-term memory network model (LSTM) is to evaluate the contributions of karst spring discharge, precipitation, water temperature, and specific conductance to DO concentrations in karst spring flow. The hybrid model can predict the temporal fluctuations of DO levels and provide a robust characterization of DO behaviours.

Heddam, et al. [16] proposed the application of long short-term memory (LSTM) deep learning for dissolved oxygen (DO) prediction in rivers. The model was trained and calibrated using three predictors: (i) river water temperature (Tw), (ii) air temperature, and (iii) river discharge (Q). The variables were measured on an hourly time scale and collected from two USGS stations. The LSTM model was compared against genetic programming (GP), the group method of data handling neural network (GMDH), support vector regression (SVR), and Gaussian process regression (GPR) models. Salas, et al. [17] proposed the Potential of mapping dissolved oxygen in the Little Miami River using Sentinel-2 images and machine learning algorithms. The authors mapped the spatiotemporal changes of DO in the Little Miami River (LMR) using 10-m Sentinel-2 images. We trained two machine learning algorithms – Random Forest (RF) and Support Vector Machine (SVM) – to predict DO concentrations using spectral predictors derived from the satellite images. Moreover, we calculated several metrics, which include Root Mean Squared Error (RMSE), Amount of Variance Explained (AVE), Coefficient of Efficiency (COE), and Normalized Mean Bias (NMB) to assess the performance of the models and accuracy of the DO maps.

Garabaghi, et al. [18] proposed Modeling dissolved oxygen concentration using machine learning techniques with a dimensionality reduction approach. The authors propose an accurate prediction model for DO concentrations. The performance of the Random Forest (RF) and multilayer perceptron (MLP) algorithms was evaluated in generating the regression models. Moreover, the effect of dimensionality reduction of the data by the wrapper feature Selection method on the performance of the models was evaluated. Ahmed, et al. [19] proposed the development of a dissolved oxygen forecast model using a hybrid machine-learning algorithm with hydro-meteorological variables. This work aims to forecast dissolved oxygen (DO) concentration using a multivariate adaptive regression spline (MARS) hybrid model coupled with maximum overlap discrete wavelet transformation (MODWT) as a feature decomposition approach for Surma River water using a set of water quality hydro-meteorological variables. Heddam, et al. [20] proposed a novel hybrid model based on signal processing decomposition, extreme learning machine and parallel chaos search for forecasting DO several days in advance. The correlation between DO data at several times lags were calculated using the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The DO concentration time series were decomposed using the empirical wavelet transform technique (EWT), and the multiresolution analysis (MRA) components were then provided.

Khabusi, et al. [21] proposed A Deep Learning Approach to Predict Dissolved Oxygen in Aquaculture. This study aimed at designing a prediction model for DO in aquatic environments. To achieve the objective, time series data consisting of 70374 records and 15 attributes from Mumford Cove in Connecticut, USA collected for over 5 years was preprocessed and used to train long-short term memory (LSTM) recurrent neural network (RNN) for DO prediction. Siddik, et al. [22] proposed the Application of machine learning approaches in predicting estuarine dissolved oxygen (DO) under a limited data environment. The application of machine learning (ML) approaches to predict estuarine dissolved oxygen (DO) from a set of environmental covariates including nutrients remains unexplored due to nutrient data unavailability. Adedeji, et al. [23] proposed Predicting in-stream water quality constituents at the watershed scale using machine learning. The authors implemented five ML algorithms—Support Vector Machines, Random Forest (RF), eXtreme Gradient Boost (XGB), ensemble RF-XGB, and Artificial Neural Network (ANN) —and demonstrated our modeling framework in an inland stream—Bullfrog Creek, located near Tampa, Florida.

Kim, et al. [24] proposed Machine learning predictions of chlorophyll-a in the Han river basin, Korea. This work developed a model to predict concentrations of chlorophyll-a ([Chl-a]) as a proxy for algal population with data from multiple monitoring stations in the Han river basin, by using machine-learning predictive models, then analyzed the relationship between [Chl-a] and the input variables of the optimized model. Yan, et al. [25] proposed a new framework to predict long-term water quality by using Bayesian-optimised machine learning methods and key pollution indicators collected from monitoring stations in the Pearl River Estuary, Guangdong, China. The optimised stacked generalisation (SG-op) model achieved the best performance with the highest accuracy (0.992) and Kappa coefficient (0.987).

3. PROPOSED METHODOLOGY

Figure 1 showcases a comprehensive data analysis and machine learning workflow for water quality testing data. It encompasses data loading, cleaning, visualization, correlation analysis, linear regression, random forest regression, and evaluation. The goal is to gain insights into the dataset and develop predictive models for dissolved oxygen levels, which can be valuable for water quality assessment and environmental monitoring. The detailed operation illustrated as follows:

Importing Libraries: The work begins by importing essential Python libraries for data analysis and machine learning. These libraries include pandas for data manipulation, seaborn and matplotlib for data visualization, and modules from scikit-learn, a popular machine learning library, for regression analysis.

User Registration and Authentication System: A registration and login interface is developed using Tkinter. User data including username, password, and role (Admin/User) is stored in an SQLite database. The authentication system checks credentials and grants access based on the assigned role. This step ensures secure and controlled access to the system features.

Loading Data: The dataset is loaded from a CSV file named "Water Quality Testing.csv" into a pandas DataFrame called **df**. This step allows for easy data manipulation and analysis in subsequent steps.

Data Cleaning: Data cleaning is a critical step to ensure the quality and integrity of the dataset. The work performs several data cleaning operations:

- It checks for missing values and duplicates in the dataset, which is essential to address data quality issues.
- A summary of the dataset is provided using **df.describe()**, which provides statistics like mean, standard deviation, minimum, and maximum values for numeric columns.

- The "Sample ID" column is dropped, indicating that it may not be relevant for the analysis.

Basic Visualization: This section aims to gain insights into the data through basic visualizations:

- A count plot is created for "Conductivity (μS/cm)" to visualize the distribution of conductivity values.
- A distribution plot is generated for "pH" to understand its distribution.
- A pair plot is constructed for the entire dataset, which allows for visual exploration of relationships between variables.

Correlation Analysis: To identify relationships between variables, a heatmap is created to visualize the correlation matrix of the dataset. The heatmap provides a color-coded representation of the correlation coefficients, helping to identify which variables are positively or negatively correlated.

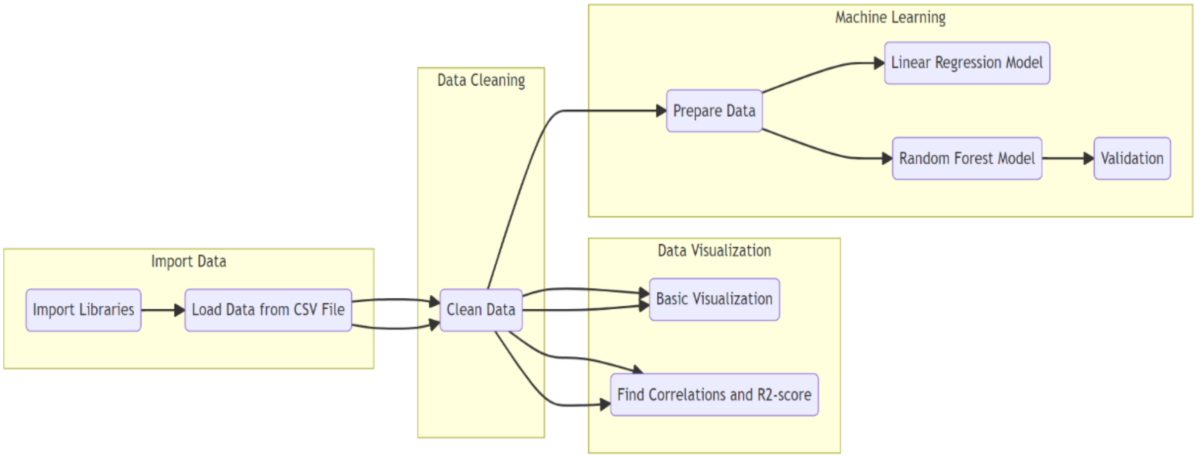

Figure 1. Proposed System Model.

Linear Regression Analysis: This section involves linear regression analysis between the target variable "Dissolved Oxygen (mg/L)" and other variables in the dataset. Linear regression is a statistical method used to model the relationship between a dependent variable (in this case, dissolved oxygen) and one or more independent variables.

- Parameters like slope, intercept, R-value (correlation coefficient), p-value, and R-squared value are calculated. These statistics help assess the strength and significance of the linear relationship.
- The work likely provides insights into whether and how other variables affect dissolved oxygen levels.

Machine Learning Part: This section involves building and evaluating machine learning models for predicting dissolved oxygen levels:

- The data is split into features (X) and the target variable (y), with "Dissolved Oxygen (mg/L)" being the target variable.
- Two regression models are trained and evaluated: linear regression and random forest regression.
- For each model, a scatter plot of predicted vs. actual values is generated, allowing for visual assessment of the model's performance.

Creating Predictions DataFrame: A DataFrame named predictions_df is created to store actual and predicted values. This DataFrame likely allows for further analysis and comparison of model predictions.

3.1 Linear regression

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc. Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable. The linear regression model provides a sloped straight line representing the relationship between the variables.

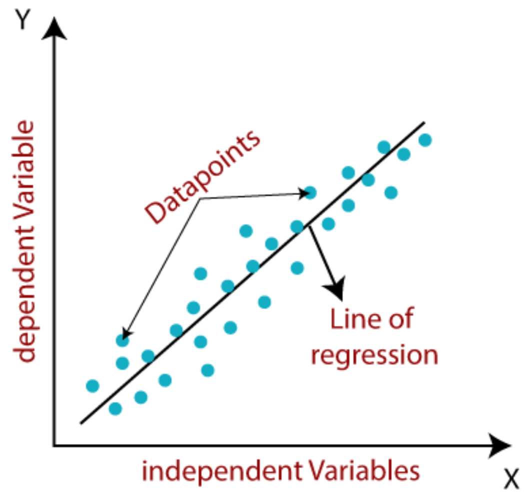

Fig 2: Linear regression model.

3.2 Random Forest Regression

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

3.2.1 Random Forest algorithm

Step 1: In Random Forest n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

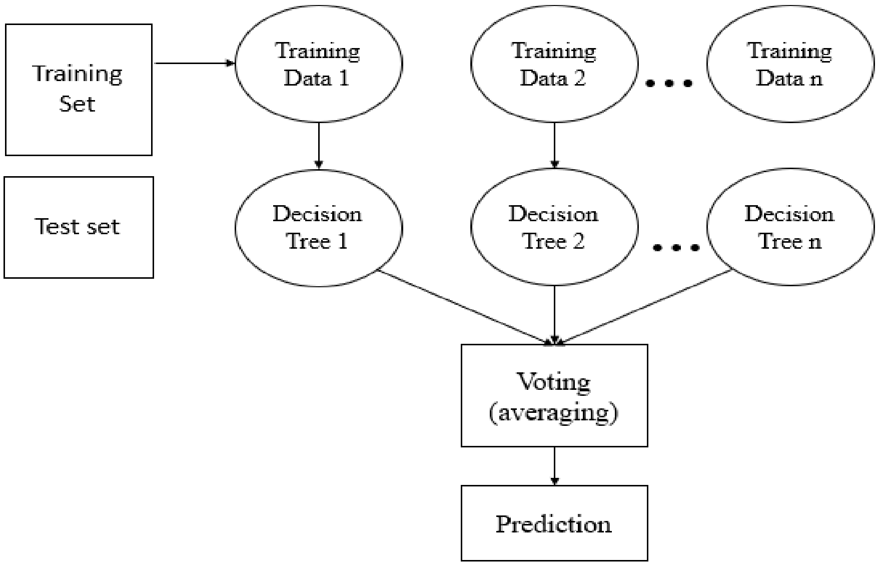

Fig. 3: Random Forest algorithm.

3.2.2 Important Features of Random Forest

- **Diversity**- Not all attributes/variables/features are considered while making an individual tree, each tree is different.
- **Immune to the curse of dimensionality**- Since each tree does not consider all the features, the feature space is reduced.
- **Parallelization**-Each tree is created independently out of different data and attributes. This means that we can make full use of the CPU to build random forests.
- **Train-Test split**- In a random forest we don't have to segregate the data for train and test as there will always be 30% of the data which is not seen by the decision tree.
- **Stability**- Stability arises because the result is based on majority voting/ averaging.

3.3.3 Assumptions for Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random Forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

Below are some points that explain why we should use the Random Forest algorithm

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

4. RESULTS AND DISCUSSION

4.1 Dataset

Water quality is a crucial aspect of environmental management, and it is essential to measure various physical, chemical, and biological parameters to monitor it effectively. This dataset of 200 rows contains measurements of six critical water quality parameters widely used in water quality monitoring and analysis. The dataset provides a representative snapshot of water quality and can be used for various research, education, and decision-making purposes.

- pH: pH measures the acidity or basicity of a liquid on a scale from 0 to 14. Values less than 7 indicate acidic conditions, values greater than 7 suggest primary needs and a value of 7 indicates a neutral state.
- Dissolved Oxygen (DO): DO measures the amount of oxygen dissolved in water and is essential for aquatic life. High DO levels are crucial for the survival of fish and other marine organisms.
- Temperature: Temperature affects various physical, chemical, and biological processes in water bodies. It is an essential factor that influences the rate of many aquatic processes.
- Biochemical Oxygen Demand (BOD): BOD measures the amount of oxygen microorganisms require to decompose organic matter in water. High BOD levels can indicate that the water is polluted with organic matter and may not be suitable for consumption or recreation.
- Total Suspended Solids (TSS): TSS measures the amount of solids suspended in water, including organic matter, sediment, and other pollutants. High TSS levels can indicate poor water quality and affect aquatic life and other uses of water.
- Nitrate-Nitrogen (NO3-N): NO3-N measures the amount of nitrate in water. Nitrate is an essential nutrient for plant growth, but high nitrate levels in drinking water can harm human health.

4.2 Simulation Analysis

	Sample ID	pH	Temperature (°C)	Turbidity (NTU)	Dissolved Oxygen (mg/L)	Conductivity (µS/cm)
0	1	7.25	23.1	4.5	7.8	342
1	2	7.11	22.3	5.1	6.2	335
2	3	7.03	21.5	3.9	8.3	356
3	4	7.38	22.9	3.2	9.5	327
4	5	7.45	20.7	3.8	8.1	352
...	...	...	...	...	...	...
495	496	7.01	20.8	4.6	7.1	327
496	497	7.31	22.5	3.8	9.4	361
497	498	7.02	21.2	4.7	7.5	334
498	499	7.25	23.0	3.9	8.7	359
499	500	7.12	20.9	4.4	8.2	339

500 rows × 6 columns

Figure 4: Sample dataset used for Predicting Dissolved Oxygen Levels in River Water using Machine Learning.

	Sample ID	pH	Temperature (°C)	Turbidity (NTU)	Dissolved Oxygen (mg/L)	Conductivity (µS/cm)
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	250.500000	7.161140	22.054400	4.169400	8.382200	344.362000
std	144.481833	0.107531	0.903123	0.397492	0.822396	13.038672
min	1.000000	6.830000	20.300000	3.100000	6.000000	316.000000
25%	125.750000	7.080000	21.200000	3.800000	7.800000	333.000000
50%	250.500000	7.160000	22.200000	4.200000	8.400000	344.000000
75%	375.250000	7.250000	22.900000	4.500000	9.100000	355.000000
max	500.000000	7.480000	23.600000	5.100000	9.900000	370.000000

Figure5: Summary statistics of a dataset's numerical columns.

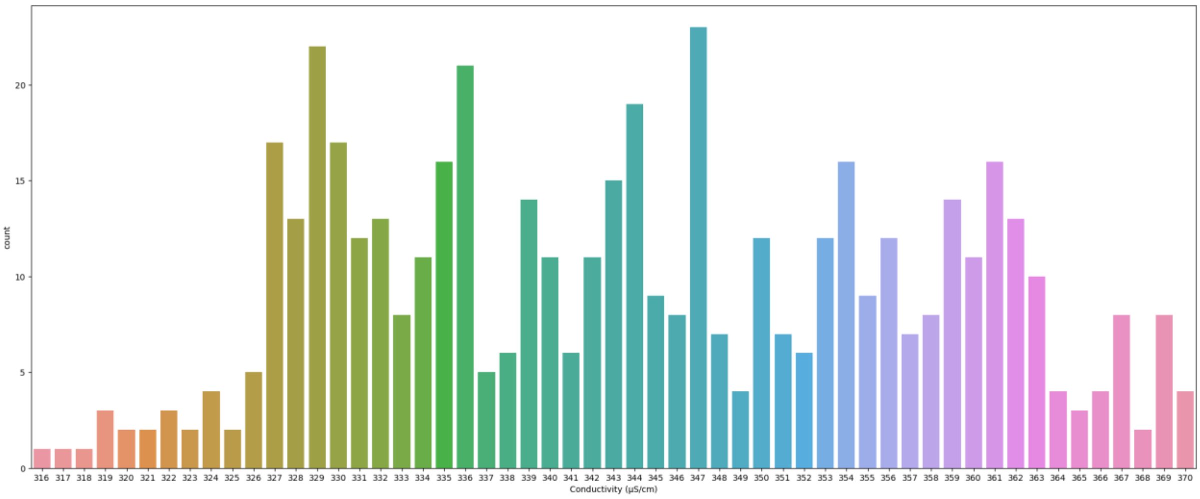

Figure 6: Count plot to visualize the distribution of values in the "Conductivity (µS/cm)"

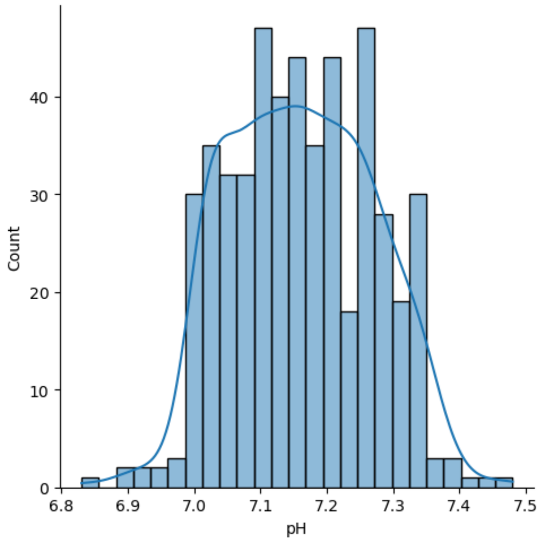

Figure 7: Distribution plot for distribution of values in the "pH" column

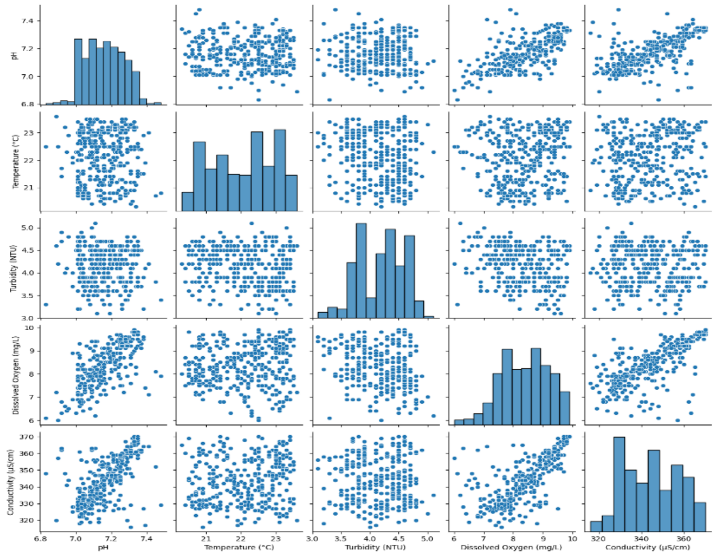

Figure 8: illustrating pairwise relationships and distributions of numerical variables in the Dataset

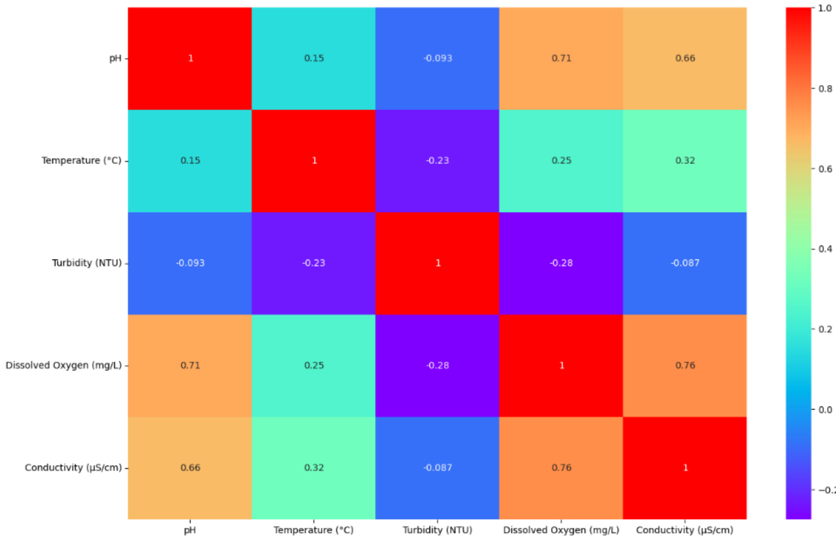

Figure 9: Visualizing the correlation matrix of numerical columns in the Data Frame.

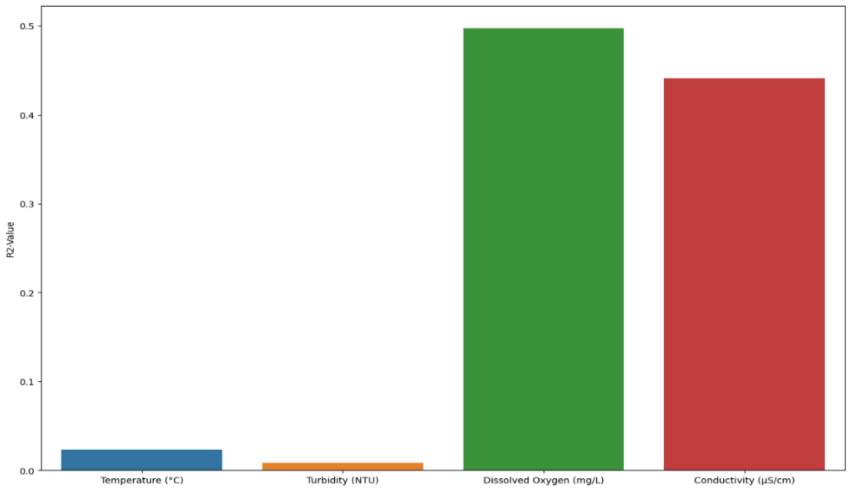

Figure 10: showcase the R2-Values ("R-squared values") from the Data Frame .

	pH	Temperature (°C)	Turbidity (NTU)	Conductivity (µS/cm)
0	7.25	23.1	4.5	342
1	7.11	22.3	5.1	335
2	7.03	21.5	3.9	356
3	7.38	22.9	3.2	327
4	7.45	20.7	3.8	352
5	6.89	23.6	4.6	320
6	7.19	21.2	4.2	350
7	6.98	22.1	3.7	325

Figure 8: features from the Data Frame

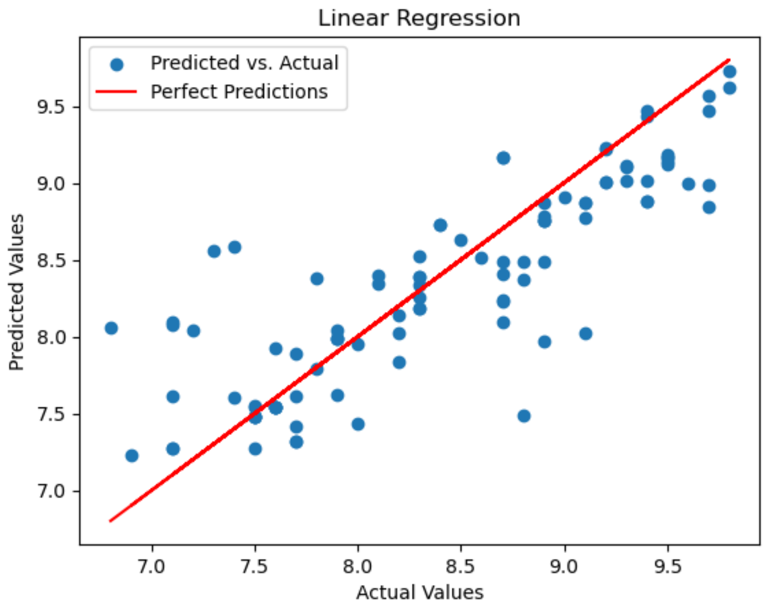

Figure 11: compare predicted values with actual values for Existing Linear Regression

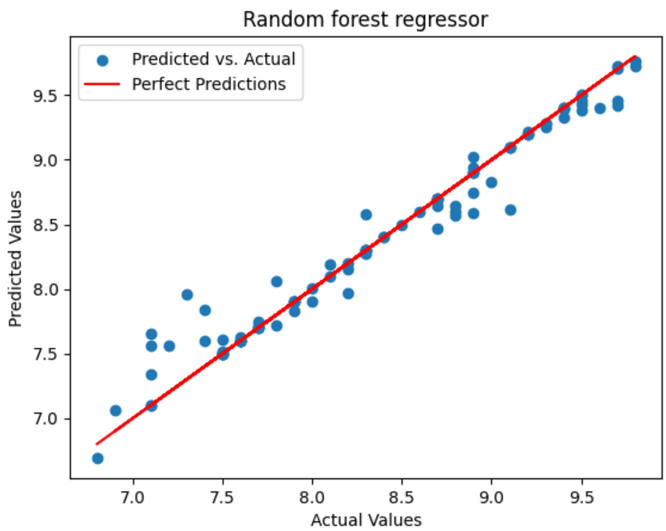

Figure 12: compare predicted values with actual values for Proposed random forest Regression

```
     Actual  Predicted
361     8.9      8.900
73      7.9      7.910
374     8.3      8.300
155     6.9      7.063
104     7.9      7.835
..      ...        ...
347     7.5      7.500
86      9.0      8.821
75      8.0      8.006
438     8.7      8.700
15      6.8      6.693

[100 rows x 2 columns]
```

Figure 13: Data Frame showing the actual and predicted values side by side.

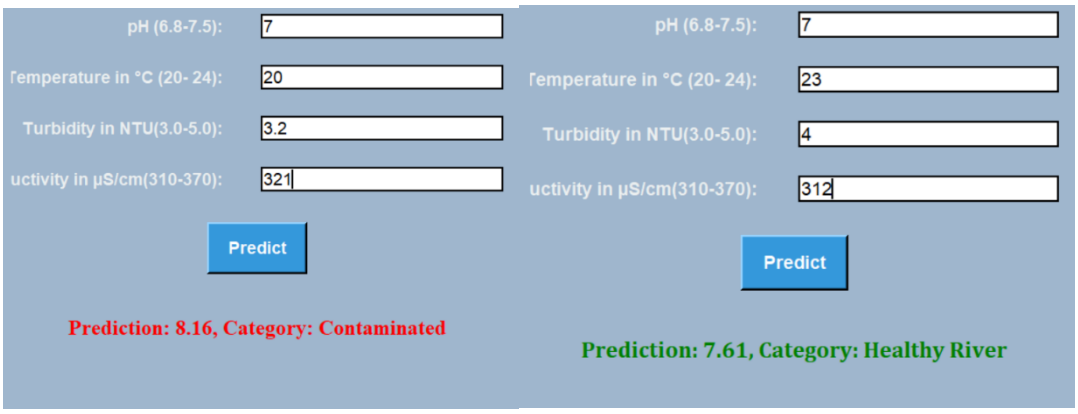

Figure 14: Model Prediction on Test Cases.

Table 1: Performance comparison of quality metrics obtained using linear regressor (LR) model and random forest regressor (RFR) model.

Model	MAE	MSE	RMSE	R2SCORE
LR model	0.308565	0.187902	0.433477	0.714892
RFR model	0.076527	0.020768	0.144113	0.968488

5. CONCLUSION

In conclusion, the comprehensive data analysis and machine learning work performed on the water quality testing dataset represent a crucial endeavour for understanding and predicting critical aspects of water quality, specifically the levels of dissolved oxygen. The work begins with data preparation, encompassing data loading and cleaning, where the identification and rectification of missing values and duplicates are essential for data integrity. Subsequently, basic visualizations provide an initial exploration of the data's distribution and interrelationships among variables. A notable highlight is the correlation analysis, visualized through a heatmap, enabling the identification of significant associations between various water quality parameters. The subsequent linear regression analysis, focusing on the relationship between pH and other variables, offers valuable insights into the dataset. In the machine learning phase, the work splits the data into features and the target variable, followed by the training and evaluation of two regression models such as linear regression and random forest regression. The scatter plots of predicted versus actual values serve as a visual gauge of model performance. Furthermore, the creation of the predictions Data frame facilitates in-depth analysis and comparison of model outcomes. Altogether, this work serves as a foundational step in leveraging data-driven insights to monitor and manage water quality effectively, which is vital for environmental preservation and ensuring the availability of clean and safe water resources. The future scope for water quality analysis and prediction using data-driven approaches is expansive and holds the potential for significant advancements in environmental monitoring, water resource management, and sustainability.

REFERENCES

[1] Rizal, Nur Najwa Mohd, Gasim Hayder, and Salman Yussof. "River Water Quality Prediction and Analysis–Deep Learning Predictive Models Approach." Sustainability Challenges and Delivering Practical Engineering Solutions: Resources, Materials, Energy, and Buildings. Cham: Springer International Publishing, 2023. 25-29.

[2] Zheng, Hang, et al. "Large-scale prediction of stream water quality using an interpretable deep learning approach." Journal of environmental management 331 (2023): 117309.

[3] Khan, Prince Waqas, and Yung Cheol Byun. "Optimized Dissolved Oxygen Prediction Using Genetic algorithm and Bagging Ensemble Learning for Smart Fish Farm." IEEE Sensors Journal (2023).

[4] Banerjee, Kakoli, et al. "A machine-learning approach for prediction of water contamination using latitude, longitude, and elevation." Water 14.5 (2022): 728.

[5] Nong, Xizhi, et al. "Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using support vector regression coupled with multiple feature engineering and optimization methods: A case study in China." Ecological Indicators 146 (2023): 109845.

[6] Zhi, Wei, et al. "Temperature outweighs light and flow as the predominant driver of dissolved oxygen in US rivers." Nature Water 1.3 (2023): 249-260.

[7] Park, Seongsik, et al. "An Antifouling Redox Sensor with a Flexible Carbon Fiber Electrode for Machine Learning-Based Dissolved Oxygen Prediction in Severely Eutrophic Waters." Water 15.13 (2023): 2467.

[8] Nair, Jitha P., and M. S. Vijaya. "River water quality prediction and index classification using machine learning." Journal of Physics: Conference Series. Vol. 2325. No. 1. IOP Publishing, 2022.

[9] Regier, Peter J., et al. "Seasonal drivers of dissolved oxygen across a tidal creek–marsh interface revealed by machine learning." Limnology and Oceanography (2023).

[10] Liang, Xu, et al. "Deep Learning Prediction of Dissolved Oxygen in a Typical Tidal Reach of Pearl River Basin for Identification of Governing Factors." Available at SSRN 4475079.

[11] Ziyad Sami, Balahaha Fadi, et al. "Machine learning algorithm as a sustainable tool for dissolved oxygen prediction: a case study of Feitsui Reservoir, Taiwan." Scientific Reports 12.1 (2022): 3649.

[12] Bolick, Madeleine M., et al. "Comparison of machine learning algorithms to predict dissolved oxygen in an urban stream." Environmental Science and Pollution Research (2023): 1-22.

[13] Moon, J., Lee, J., Lee, S., & Yun, H. (2022). Urban river dissolved oxygen prediction model using machine learning. Water, 14(12), 1899.

[14] Nair, Jitha P., and M. S. Vijaya. "Analysing And Modelling Dissolved Oxygen Concentration Using Deep Learning Architectures." International Journal of Mechanical Engineering 7 (2022): 12-22.

[15] Zhou, Renjie, and Yanyan Zhang. "Predicting and explaining karst spring dissolved oxygen using interpretable deep learning approach." Hydrological Processes 37.8 (2023): e14948.

[16] Heddam, Salim, et al. "Predicting dissolved oxygen concentration in river using new advanced machines learning: Long-short term memory (lstm) deep learning." Computers in Earth and Environmental Sciences. Elsevier, 2022. 1-20.

[17] Salas, Eric Ariel L., et al. "Potential of mapping dissolved oxygen in the Little Miami River using Sentinel-2 images and machine learning algorithms." Remote Sensing Applications: Society and Environment 26 (2022): 100759.

[18] Garabaghi, Farid Hassanbaki, Semra Benzer, and Recep Benzer. "Modeling dissolved oxygen concentration using machine learning techniques with dimensionality reduction approach." Environmental Monitoring and Assessment 195.7 (2023): 879.

[19] Ahmed, Abul Abrar Masrur, et al. "The development of dissolved oxygen forecast model using hybrid machine learning algorithm with hydro-meteorological variables." Environmental Science and Pollution Research 30.3 (2023): 7851-7873.

[20] Heddam, Salim. "Parallel Chaos Search Based Incremental Extreme Learning Machine Based Empirical Wavelet Transform: A New Hybrid Machine Learning Model for River Dissolved Oxygen Forecasting." Computational Intelligence for Water and Environmental Sciences. Singapore: Springer Nature Singapore, 2022. 355-376.

[21] Khabusi, Simon Peter, and Yo-Ping Huang. "A Deep Learning Approach to Predict Dissolved Oxygen in Aquaculture." 2022 International Conference on Advanced Robotics and Intelligent Systems (ARIS). IEEE, 2022.

[22] Siddik, Mohammad Abu Zafer. "Application of machine learning approaches in predicting estuarine dissolved oxygen (DO) under a limited data environment." Water Quality Research Journal 57.3 (2022): 140-151.

[23] Adedeji, Itunu C., Ebrahim Ahmadisharaf, and Yanshuo Sun. "Predicting in-stream water quality constituents at the watershed scale using machine learning." Journal of Contaminant Hydrology 251 (2022): 104078.

[24] Kim, Kyung-Min, and Johng-Hwa Ahn. "Machine learning predictions of chlorophyll-a in the Han river basin, Korea." Journal of Environmental Management 318 (2022): 115636.

[25] Yan, Tao, Annan Zhou, and Shui-Long Shen. "Prediction of long-term water quality using machine learning enhanced by Bayesian optimisation." Environmental Pollution 318 (2023): 120870.