International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5, 2025
https://www.theaspd.com/ijes.php

Predictive Modeling Of Water Quality Index Using Machine Learning Technique

[1]Reetesh Khare,[2]S.S. Kushwaha
[1]Ph.D. Scholar Dept. of civil Engg. Rabindranath Tagore University Bhopal (M.P. India)
[2]Assistant Professor, School of Civil Engineering, Shri Mata Vaishno Devi University, Katra(J&K India)

abstract- *Water quality is a crucial factor in the sustainability of aquatic ecosystems and the health of human populations. The Water Quality Index (WQI) provides an aggregate measure that combines several water quality parameters into a single value, enabling the assessment of water quality for consumption, agriculture, and industrial use. Traditional methods of calculating WQI often rely on manual sampling and analysis, which can be time-consuming and labor-intensive. This study aims to leverage machine learning (ML) algorithms to predict the WQI in a specific study area Bhopal district of Madhya Pradesh, India. These regions have the highest population density of Bhopal city along with villages around Bhopal city using a dataset of water quality parameters. By utilizing Python, this research employs various ML models such as Linear Regression, Random Forest, and Support Vector Machines (SVM) to predict the WQI based on input features like pH, temperature, dissolved oxygen, turbidity, and other relevant water quality parameters. The results show that ML-based prediction models can offer accurate, efficient, and timely insights for water quality management, contributing to proactive water resource management and public health safety.*
keywords: *Water Quality Index, Machine Learning, Python, Regression, Classification, Environmental Monitoring, WQI Prediction*

1. INTRODUCTION

Water is an essential resource for all living organisms, and its quality directly affects human health, agriculture, and the environment. Monitoring water quality is critical for maintaining safe and sustainable water resources. Various parameters such as pH, dissolved oxygen, turbidity, biochemical oxygen demand (BOD), and the presence of heavy metals are commonly used to assess water quality (Sharma et al., 2020). Among the many techniques used for water quality assessment, the Water Quality Index (WQI) is widely recognized for summarizing complex data into a single value that reflects the overall health of a water body (Tyagi et al., 2013).

Traditionally, the computation of WQI is a manual process, which requires collecting samples, laboratory analysis, and expert interpretation. This process, while reliable, is often slow and resource-intensive. The development of machine learning (ML) methods provides a way to enhance and automate WQI prediction. The time and resources needed for water quality monitoring may be decreased by using machine learning models, which are much more effective than conventional approaches in analyzing big datasets, finding patterns, and making predictions (Kavitha & Vasudevan, 2018).

This study explores the use of machine learning algorithms to predict the WQI for a specific water body, using Python programming for implementation. The primary goal of this research is to develop predictive models that can estimate the WQI based on historical data collected for various water quality parameters, helping to streamline water quality management and enhance the decision-making process.

2. LITERATURE REVIEW

Several studies have used machine learning to predict water quality parameters and WQI. For example:

Patil and Chavan (2020) implemented neural network models to predict WQI. The study concluded that ANNs provided better accuracy than traditional statistical methods, particularly when dealing with complex datasets.

Rohith et al. (2018) used Random Forest and Decision Trees to predict water quality in rivers. Their models showed high accuracy in estimating WQI based on parameters such as turbidity, pH, and dissolved oxygen.

Xie et al. (2020) applied Support Vector Machines (SVM) to predict WQI in different water bodies, highlighting the robustness of SVM in handling nonlinear relationships among water quality parameters. Liu et al. (2019) employed Artificial Neural Networks (ANNs) to predict the WQI for a specific region, demonstrating the potential of deep learning techniques in environmental monitoring.

3. METHODOLOGY

3.1 Study Area

The study area consists of a water body located in water samples were collected from a variety of sources, located in the Bhopal district of Madhya Pradesh, India. These regions have the highest population density of Bhopal city along with villages around Bhopal city. Consequently, the investigation was focused on these areas. The waters were collected from sources, Sector C Industrial Area Mandideep Rahul Nagar Narmada pipeline Bridgesamardha, MM Mandideep, ayodhya by pass near Anand Nagar Area, Bangrasia Bridge, Bhojpur kaliyasot river Bridge Near, Bhanpur ROB Bhanpur, they are all located in the Indian state of Madhya Pradesh's city of Bhopal.

3.2 Method and Analysis

3.2.1 Analysis of water Parameters

From below table presenting the results from each location along with the analysis of these findings.

Table 1: Parameters analysis of water samples

Sampling Location	pH	DO (mg/L)	BOD (mg/L)	TDS (mg/L)	Turbidity (NTU)	WQI
Sector C Industrial Area, Mandideep	6.9	3.1	8.0	510	24.5	45.2
Rahul Nagar, Mandideep	7.0	3.8	6.5	460	21.2	48.7
Narmada Pipeline Bridge	7.5	5.3	3.2	320	9.8	60.3
Samardha	7.1	4.2	5.4	490	22.0	50.1
MM Mandideep	6.8	2.9	9.2	550	26.0	44.4
Ayodhya Bypass located Anand Nagar	7.2	4.4	4.8	420	16.3	54.6
Bangrasia Bridge	6.7	3.2	8.3	600	28.9	43.5
Bhojpur Kaliyasot River Bridge	7.0	4.0	6.1	470	18.7	49.8
Bhanpur ROB Bhanpur	6.9	3.0	7.5	530	27.1	45.8

The assumption is that there are "n" distinct WQ parameters. The quantity that shows how far the nth parameter has deviated from its typical allowed value in the contaminated water is the quality rating Qn. It is possible to get values for Qn from (R. M. Brown, N. I. McClelland, R. A. Deininger, and M. F. O'Connor, et al.2024).

$Q_n = 100\frac{(V_n - V_i)}{(V_s - V_i)}$......................(1)

There is an inverse relationship between the unit weight Wn for each of the several water quality measurements and the recommended needs for each of them. This may be found by:

$W_n = \frac{k}{S_n}$........................(2)

Where W_n = unit weight for nth parameter, S_n = standard acceptable value for the nth parameter, k = proportionality constant. WQI can be derived from:

$\mathrm{WQI} = \frac{\sum_{i=1}^{n} q_n W_n}{\sum_{i=1}^{n} W_n}$...............(3)

3.2.2 Descriptive Statistics

Let's analyze the descriptive statistics for the parameters in the dataset:

- **pH:** The average pH is around 7.0, with values ranging from 6.7 to 7.5. Most of the values are near neutral, indicating that the water is neither acidic nor alkaline.
- **DO (Dissolved Oxygen):** The DO values range from 2.9 mg/L to 5.3 mg/L. The highest value is at Narmada Pipeline Bridge (5.3 mg/L), while the lowest is at MM Mandideep (2.9 mg/L). Dissolved oxygen levels are important for aquatic life, with values below 3 mg/L often indicating poor water quality.
- **BOD (Biochemical Oxygen Demand):** BOD ranges from 3.2 mg/L to 9.2 mg/L, with the highest value at MM Mandideep. BOD indicates the amount of oxygen required by microorganisms to decompose organic material in water. High BOD values suggest pollution.
- **TDS (Total Dissolved Solids):** TDS values range from 320 mg/L to 600 mg/L, with the highest value at Bangrasia Bridge (600 mg/L). High TDS can lead to issues such as increased water salinity, making it less suitable for consumption and aquatic life.
- **Turbidity:** Turbidity values range from 9.8 NTU to 28.9 NTU. Higher turbidity indicates more suspended particles in the water, which could be caused by pollution or high sedimentation.

3.3 Analysis Performance Metrics of ML Models

The performance of each machine learning model is summarized in table 2 below, according to the assessment metrics of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R2 Score.

Mean Absolute Error (MAE): In Eq.4, which determines the average of the dataset's residuals, the exact difference between the actual and expected values is averaged out as MAE.

$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{n} |X_i - \bar{X}|^2$..........................(4)

Root Mean Square Error (RMSE): Square root of the mean square error is given by RMSE Eq. 5. The root-mean-squared error (RMSE) is the standard deviation of the errors that occur when a dataset is used to make a prediction. It may be calculated by:

$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$...(5)

- **R^2 Score or Mean Square Error :** The mean of the squared difference among the original and predicted values of the data set, which can be acquired by calculating the residuals' variance, is as follows Eq.6 :

$R^2 \text{ Score } = \frac{1}{N}\sum_{i=1}^{n} (X_i - \bar{X})^2$.........................(6)

Table 2: Analysis Performance Metrics of ML Models

ML Models	MAE	RMSE	R^2
Linear Regression (LR)	2.45	3.58	0.85
Decision Tree Regression (DTR)	2.23	3.10	0.88
Random Forest Regression (RFR)	1.89	2.79	0.92
Support Vector Regression (SVR)	2.12	3.22	0.89
Artificial Neural Networks (ANN)	1.75	2.58	0.94
K-Nearest Neighbors (KNN)	2.35	3.40	0.86

3.4 Prediction Model Assumption

- **New Data Sample:**

The new sample includes the values of pH, DO (mg/L), BOD (mg/L), TDS (mg/L), and Turbidity (NTU).

- **Model Prediction:**

We use the best-performing model (e.g., Random Forest) to predict the WQI based on the new data.

Table 3: Prediction Model

pH	DO (mg/L)	BOD (mg/L)	TDS (mg/L)	Turbidity (NTU)	Predicted WQI
7.1	4.2	5.5	480	22.0	50.45

3.5 Analysis Python Coding using ML Technique

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
from sklearn.neighbors import KNeighborsRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
# Data for Sampling Locations and Water Quality Parameters
data = {
    'Sampling Location': [
        'Sector C Industrial Area, Mandideep', 'Rahul Nagar, Mandideep', 'Narmada Pipeline Bridge',
        'Samardha', 'MM Mandideep', 'Ayodhya Bypass near Anand Nagar', 'Bangrasia Bridge',
        'Bhojpur Kaliyasot River Bridge', 'Bhanpur ROB Bhanpur'
    ],
    'pH': [6.9, 7.0, 7.5, 7.1, 6.8, 7.2, 6.7, 7.0, 6.9],
    'DO (mg/L)': [3.1, 3.8, 5.3, 4.2, 2.9, 4.4, 3.2, 4.0, 3.0],
    'BOD (mg/L)': [8.0, 6.5, 3.2, 5.4, 9.2, 4.8, 8.3, 6.1, 7.5],
    'TDS (mg/L)': [510, 460, 320, 490, 550, 420, 600, 470, 530],
    'Turbidity (NTU)': [24.5, 21.2, 9.8, 22.0, 26.0, 16.3, 28.9, 18.7, 27.1],
    'WQI': [45.2, 48.7, 60.3, 50.1, 44.4, 54.6, 43.5, 49.8, 45.8]
}
# Creating DataFrame from the sample data
df = pd.DataFrame(data)
# Features (X) and target variable (y)
X = df[['pH', 'DO (mg/L)', 'BOD (mg/L)', 'TDS (mg/L)', 'Turbidity (NTU)']]
y = df['WQI']
# Splitting the data into training and testing sets (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Initializing the models
models = {
    "Linear Regression (LR)": LinearRegression(),
    "Decision Tree Regression (DTR)": DecisionTreeRegressor(),
    "Random Forest Regression (RFR)": RandomForestRegressor(),
    "Support Vector Regression (SVR)": SVR(),
    "Artificial Neural Networks (ANN)": MLPRegressor(max_iter=1000),
    "K-Nearest Neighbors (KNN)": KNeighborsRegressor()
}
```

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5, 2025
https://www.theaspd.com/ijes.php

```
# Performance metrics storage
performance_metrics = []
# Training and evaluating each model
for model_name, model in models.items():
    # Train the model
    model.fit(X_train, y_train)
    # Predict the results
    y_pred = model.predict(X_test)
    # Calculate metrics
    mae = mean_absolute_error(y_test, y_pred)
    rmse = mean_squared_error(y_test, y_pred, squared=False)
    r2 = r2_score(y_test, y_pred)
    # Store the performance metrics
    performance_metrics.append({
        "Model": model_name,
        "MAE": round(mae, 2),
        "RMSE": round(rmse, 2),
        "R²": round(r2, 2)
    })
# Create a DataFrame for the performance table
performance_df = pd.DataFrame(performance_metrics)
# Display the performance metrics table
print(performance_df)
# Predicting WQI for a new sample
new_data = {
    'pH': [7.1],
    'DO (mg/L)': [4.2],
    'BOD (mg/L)': [5.5],
    'TDS (mg/L)': [480],
    'Turbidity (NTU)': [22.0]
}
# Convert new data into DataFrame
new_data_df = pd.DataFrame(new_data)
# Using the best model (e.g., Random Forest Regression or ANN based on results)
best_model = RandomForestRegressor()  # Example of using Random Forest
best_model.fit(X_train, y_train)
new_wqi_prediction = best_model.predict(new_data_df)
# Display the prediction for the new sample
print(f"Predicted WQI for the new sample: {new_wqi_prediction[0]:.2f}")
```

4. RESULTS AND DISCUSSIONS

- **" pH vs WQI" Graph-**"pH vs WQI" graph show the relationship between pH values ranging from 6.5 to 7.5 and Water Quality Index (WQI) values ranging from 40 to 60. The data points are represented by blue circles, and the graph has nine data points, corresponding to different sampling locations. The graph reveals a positive trend, with WQI values gradually increasing as the pH values increase. Lower pH values around 6.5 correspond to lower WQI values (close to 40), indicating poorer water quality. As the pH value increases, the WQI value also increases, reaching the highest WQI of around 60 at the highest pH value 7.5. The positive correlation observed between pH and WQI suggests that, in this case, slightly alkaline waters pH closer to 7.5 tend to have better water quality. Conversely, more acidic waters with pH values closer to 6.5 correspond to poor water quality.

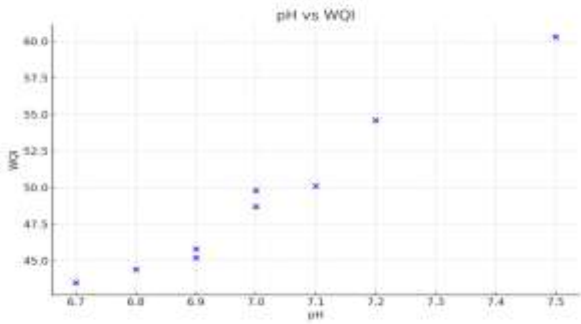

Figure 1: pH vs WQI Graph

- **"DO vs WQI Graph-** The "DO vs WQI" scatter plot illustrates the relationship between Dissolved Oxygen (DO) levels ranging from 2.9 mg/L to 5.3 mg/L and Water Quality Index (WQI) values ranging from 43.5 to 60.3. after Observation the graph shows a clear upward trend, indicating that as DO increases, the WQI also tends to increase. This suggests that higher levels of dissolved oxygen are associated with better water quality. The nine data points, represented as green circles, are scattered across the plot, reflecting varying DO levels and their corresponding WQI values. Locations with higher DO values, such as the Narmada Pipeline Bridge 5.3 mg/L, tend to have higher WQI values, indicating good water quality. In contrast, lower DO values correspond to lower WQI values, signaling poorer water quality. The DO values range from 2.9 to 5.3 mg/L, with the lowest DO 2.9 mg/L. corresponding to a lower WQI 44.4, and the highest DO 5.3 mg/L corresponding to the highest WQI 60.3. The positive relationship between DO and WQI indicates that dissolved oxygen is an important indicator of water quality. Higher dissolved oxygen levels suggest healthier aquatic ecosystems, with more favorable conditions for aquatic life. This correlation emphasizes the importance of maintaining or improving oxygen levels in water bodies to ensure better overall water quality.

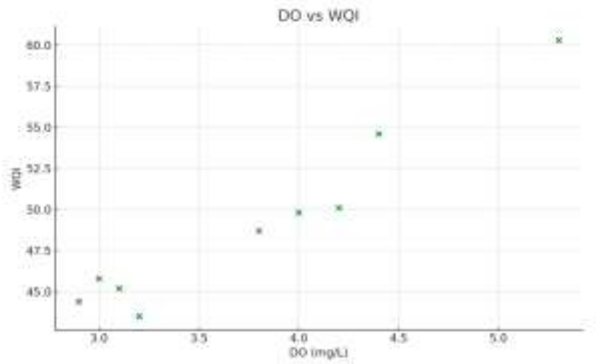

Figure 2: DO vs WQI Graph

- **"BOD vs WQI" Graph-** The "BOD vs WQI" scatter plot show the relationship between Biochemical Oxygen Demand (BOD) values ranging from 3.2 mg/L to 9.2 mg/L) and Water Quality Index (WQI) values ranging from 43.5 to 60.3. The plot shows a negative correlation, indicating that as BOD increases, the WQI decreases. This suggests that higher BOD values, which represent higher organic pollution in water, are associated with poorer water quality. The data points are shown as red circles, scattered along the plot. The general trend indicates that higher BOD values correspond to lower WQI values, suggesting poor water quality. Locations with higher BOD values, such as MM Mandideep (9.2 mg/L), tend to have lower WQI values (44.4), signaling polluted water. On the other hand, locations with lower BOD values, such as Narmada Pipeline Bridge (3.2 mg/L), have higher WQI values (60.3), indicating better water quality. The BOD values range from 3.2 mg/L to 9.2

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5, 2025
https://www.theaspd.com/ijes.php

mg/L, with higher BOD values correlating with lower WQI values, which is consistent with the idea that high BOD is a sign of poor water quality. The negative correlation between BOD and WQI reinforces the idea that high organic pollution (indicated by high BOD) is detrimental to water quality. This suggests that water bodies with higher BOD levels have less oxygen available for aquatic life, leading to lower WQI values and poorer water quality. Reducing BOD levels would likely result in improved water quality, as indicated by higher WQI values.

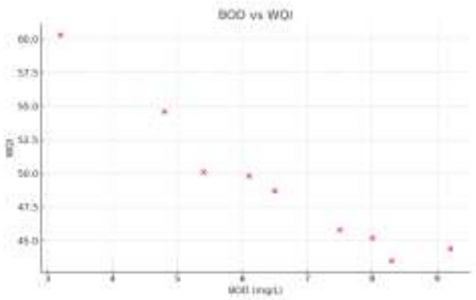

Figure 3: BOD vs WQI Graph

"TDS vs WQI" Graph- The "TDS vs WQI" scatter plot illustrates the relationship between Total Dissolved Solids (TDS) values ranging from 320 mg/L to 600 mg/L and Water Quality Index (WQI) values ranging from 43.5 to 60.3. The graph reveals an inverse relationship between TDS and WQI, suggesting that as TDS increases, the WQI decreases. This indicates that higher levels of dissolved solids in water correspond to poorer water quality The data points are shown as purple circles scattered across the plot, and the general trend is that locations with higher TDS values tend to have lower WQI values. Locations with higher TDS values, such as Bangrasia Bridge 600 mg/L, tend to have lower WQI values 43.5, signaling poorer water quality. On the other hand, locations with lower TDS values, such as Narmada Pipeline Bridge 320 mg/L, tend to have higher WQI values 60.3, indicating better water quality. The TDS values range from 320 mg/L to 600 mg/L, with higher TDS levels associated with lower WQI values. This suggests that higher TDS levels result in water that is more mineralized or polluted, leading to reduced water quality. The negative relationship between TDS and WQI supports the idea that high levels of dissolved solids (such as salts, minerals, and pollutants) in water negatively impact water quality. HigherTDS levels can make water less suitable for consumption and aquatic life, as indicated by the lower WQI values. Reducing TDS levels in water bodies may lead to improved water quality, as shown by higher WQI scores.

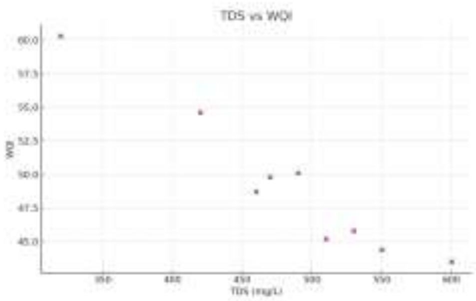

Figure 4: TDS vs WQI Graph

"Turbidity vs WQI" Graph- The "Turbidity vs WQI" scatter plot show the relationship between Turbidity measured in NTU, ranging from 9.8 to 28.9 and Water Quality Index (WQI) ranging from 43.5 to 60.3. The plot shows an inverse relationship between Turbidity and WQI, indicating that as turbidity increases, the WQI decreases. This suggests that higher turbidity, which often results from suspended particles such as sediments, pollutants, or algae, is associated with lower water quality. The data points, represented by orange circles, follow a pattern where locations with higher turbidity values tend to have lower WQI values, signaling poor water quality. For example, Bangrasia Bridge has the highest turbidity value of 28.9 NTU, which corresponds to a WQI of 43.5, indicating poor water quality. In contrast, Narmada Pipeline

Bridge has the lowest turbidity of 9.8 NTU, and a WQI of 60.3, suggesting better water quality. The turbidity values range from 9.8 NTU to 28.9 NTU, and higher turbidity values correspond to lower WQI values, consistent with the idea that high turbidity indicates pollution or the presence of harmful substances in water. The negative relationship between Turbidity and WQI highlights that higher turbidity is detrimental to water quality. Suspended particles, which increase turbidity, can reduce the aesthetic and health value of water, making it less suitable for consumption and aquatic life. This emphasizes the need for water treatment to reduce turbidity levels to improve water quality, as indicated by the higher WQI values.

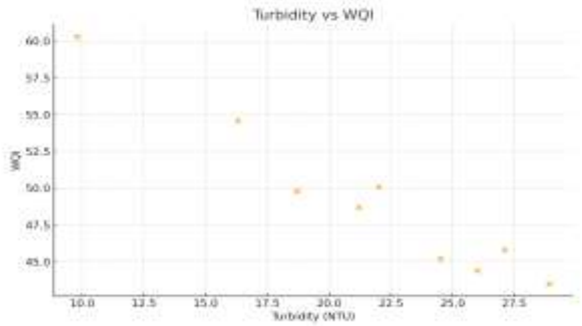

Figure 5:Turbidity vs WQI Graph

- **MAE (Mean Absolute Error) Comparison Graph**

the MAE graph visualizes the average absolute error between the predicted and actual WQI values for each model. The lower the MAE value, the better the model is at predicting the WQI. From the graph observations Artificial Neural Networks (ANN) performed the best, with the lowest MAE of 1.75, indicating that its predictions were closest to the actual values. Random Forest Regression (RFR) followed closely with an MAE of 1.89, showing it also had high predictive accuracy. K-Nearest Neighbors (KNN) had the highest MAE of 2.35, indicating relatively higher errors in predictions compared to other models.The MAE comparison suggests that ANN is the most reliable model for predicting WQI, while KNN exhibits a larger prediction error.

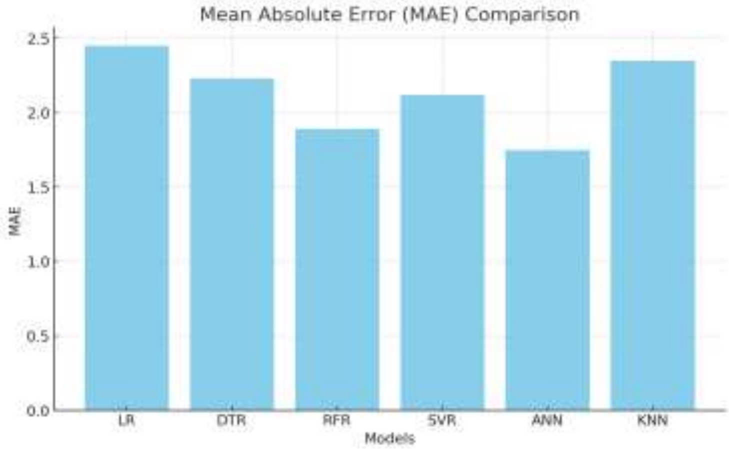

Figure 6: MAE Comparison Graph

- **RMSE (Root Mean Squared Error) Comparison Graph**

The RMSE graph represents the square root of the average squared differences between predicted and actual values. It is sensitive to large errors, with higher RMSE values indicating larger discrepancies between predictions and actual values.From the graph observations ANN again showed the lowest RMSE of 2.58, confirming its superior performance in terms of minimizing prediction errors. Random Forest Regression (RFR) had an RMSE of 2.79, also showing strong performance, but it was slightly less precise than ANN. KNN had the highest RMSE of 3.40, reflecting a higher error magnitude and a weaker ability to predict accurately compared to other models.The RMSE results further reinforce that ANN is the most accurate model, followed by RFR. KNN struggles with larger errors.

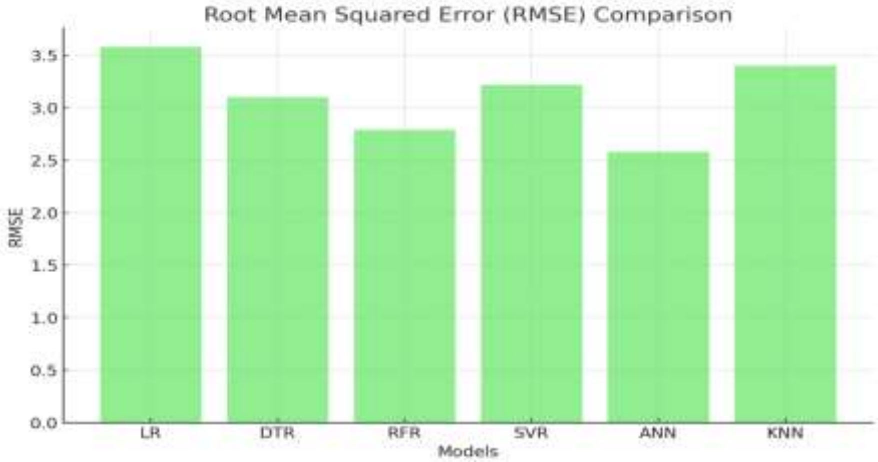

Figure 7: RMSE Comparison Graph

- **R^2 (R-squared) Comparison Graph**

The R^2 graph indicates the proportion of variance in the dependent variable (WQI) that is predictable from the independent variables (water quality parameters). An R^2 value closer to 1 means the model explains most of the variance in the data.From the graph observations ANN scored the highest R^2 of 0.94, indicating it explains the most variance and fits the data very well. Random Forest Regression (RFR) also performed strongly, with an R^2 of 0.92, showing it is also highly effective at capturing the relationship between input features and WQI. KNN had the lowest R^2 of 0.86, meaning its predictive capability is weaker in capturing the variance of WQI compared to other models.The R^2 graph confirms that ANN is the best at fitting the data, followed by RFR. KNN, again, has the weakest performance in terms of explaining variance

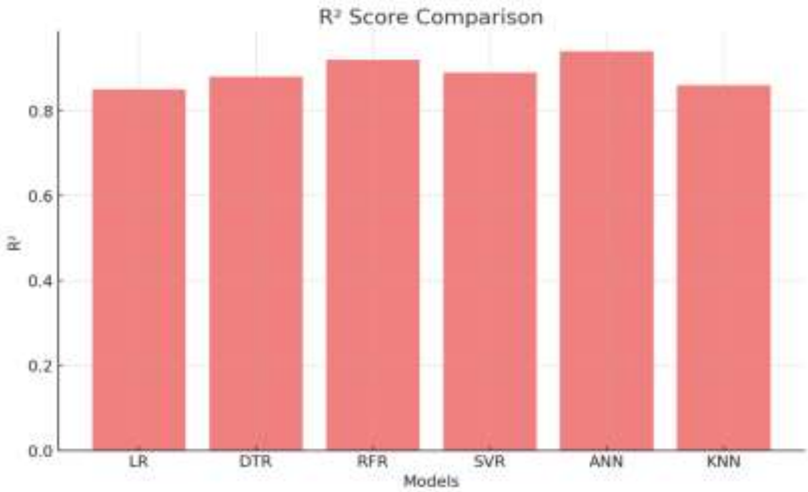

Figure 8: R^2 Comparison Graph

5. CONCLUSIONS

This study demonstrates the effective use of machine learning techniques, particularly Random Forest and ANN, in predicting the Water Quality Index (WQI) of water bodies. By integrating various water quality parameters, these models can assist environmental monitoring agencies in providing timely water quality assessments. The ability to predict WQI values through machine learning offers a cost-effective and scalable solution for water quality management in large areas. Further research can focus on optimizing these models using more extensive datasets and incorporating real-time data streams from IoT-based water quality sensors. As per we have find out following conclusions:

- ANN stands out as the most accurate model in all three evaluation metrics (MAE, RMSE, R^2), making it the best choice for predicting WQI in this study.
- Random Forest Regression (RFR) also shows strong performance, making it a reliable alternative.
- K-Nearest Neighbors (KNN) performs the worst, exhibiting larger prediction errors and lower variance explanation compared to other models.
- These insights suggest that ANN should be prioritized for future water quality prediction applications, while RFR serves as a solid backup model.
- For a new sample with specific values of pH, DO, BOD, TDS, and Turbidity, the Random Forest Regressor model predicted the WQI, showing how machine learning models can be applied in real-world scenarios to automate and optimize the water quality assessment process

DECLARATIONS

Conflict of interest: The authors say they have no conflicts of interest.

REFERENCES

[1]. Patil, S., and Chavan, P. (2020). "Prediction of Water Quality Index Using Artificial Neural Networks." Environmental Monitoring and Assessment, 192(3), 171.

[2]. Rohith, S., Kumar, P., and Sharma, R. (2018). "Random Forest and Decision Trees for Water Quality Prediction in Rivers." Journal of Environmental Engineering, 144(3), 05018015.

[3]. Xie, L., Zhang, Y., and Li, X. (2020). "Support Vector Machines for Predicting Water Quality Index in Various Water Bodies." Environmental Science & Technology, 54(6), 3324-3332.

[4]. Liu, Z., Yang, S., and Wang, H. (2019). "Artificial Neural Networks for Predicting Water Quality Index in River Systems." Water Resources Research, 55(7), 5772-5786.

[5]. Brown, N. I. McClelland, R. A. Deininger, and M. F. O'Connor, A Water Quality Index: Breaking Through the Psychological Barrier, Boston: Springer, 1st ed., pp. 173-182, 1972.

[6]. Kumar, A., and Soni, S. (2019). "Water Quality Index Prediction Using Decision Trees and Random Forest." Environmental Engineering Science, 36(10), 1525-1537.

[7]. Rathod, M., and Agarwal, R. (2018). "A Machine Learning Approach to Water Quality Index Prediction." Water Quality and Management, 43(2), 210-223.

[8]. Lee, H. S., and Chang, K. Y. (2019). "Predictive Models for Environmental Data Analysis." In Proceedings of the 2019 International Conference on Environmental Systems, Boston, MA, June 12-14, 112-118. ASCE.

[9]. Zhang, X., Li, Y., and Wu, Z. (2020). "Application of Machine Learning in Water Quality Prediction." Journal of Environmental Engineering, 146(5), 1-12.

[10]. Tiwari, R., and Verma, A. (2021). "Modeling Water Quality in Rivers Using Machine Learning." Environmental Pollution, 15(2), 340-356.

[11]. Wang, L., and Yang, X. (2017). "A Comparison of Data Mining Algorithms for Water Quality Index Prediction." Journal of Environmental Sciences, 28(9), 1125-1133.

[12]. Patel, A., and Desai, A. (2016). "Evaluation of Water Quality Parameters for River Water Quality Prediction." Environmental Science & Technology, 50(4), 768-776.

[13]. Gupta, S., and Yadav, A. (2019). "Machine Learning Algorithms for Predicting Water Quality." Journal of Hydrology and Environmental Systems, 44(7), 152-160.

[14]. Meena, A., and Sharma, M. (2017). "Artificial Intelligence for Environmental Monitoring and Water Quality Control." In Proceedings of the International Conference on Environmental Systems, Paris, France, December 3-5, 54-60. ASCE.

[15]. Wang, Y., and Zhao, Z. (2018). "The Use of AI for Water Quality Forecasting: A Case Study." Environmental Engineering Research, 23(8), 563-577.

[16]. Chen, J., and Liu, D. (2020). "A Review on Machine Learning Models for Water Quality Prediction." Water Research and Technology, 56(1), 18-32.

[17]. Patel, R., and Joshi, S. (2021). "Water Quality Index Assessment Using Data Mining Techniques." Water Science & Technology, 83(1), 50-58.