A Comparative Study of Water Quality Index (WQI) and Air Quality Index (AQI) Exploring Environmental Interactions

Chimakurthy Mounika Begum
Department of Computer Science
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Bengaluru, India
bl.en.u4cse22014@bl.students.amrita.edu

Kamalapuram Vigneswara Reddy
Department of Computer Science
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Bengaluru, India
bl.en.u4cse22030@bl.students.amrita.edu

Adhi Neeraja
Department of Computer Science
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Bengaluru, India
bl.en.u4cse22004@bl.students.amrita.edu

Sayam Dheeraj
Department of Computer Science
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Bengaluru, India
bl.en.u4cse22172@bl.students.amrita.edu

Pooja Gowda
Department of Computer Science
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Bengaluru, India
g_pooja@blr.amrita.edu

***Abstract*—This research examines the relationship between air and water quality in the Mahanadi River Basin in Odisha, India. This region has experienced environmental degradation due to pollution from industrial and agricultural activities over a decade. Air Quality Index (AQI) and Water Quality Index (WQI) are traditionally analyzed independently, overlooking their interdependencies. Therefore, having this study propose a data-driven approach to filling the gap by building a predictive model that uses AQI parameters for an advance forecast of WQI, one would have the needed tool in environmental management. Using machine learning algorithms, it identifies a few key features from the AQI that significantly impact WQI. This study proposes a data science tool, that will predict the air quality with much better accuracy of 0.959 and will help in proactive decision-making by environmental managers in terms of air and water resource management for that region. The formulation of policy and control strategy will likely benefit from understanding the interrelationship between AQI and WQI, thereby upholding public health and environmental sustainability.**

***Index Terms*—Air Quality Index (AQI), Water Quality Index (WQI), Environmental Management, Mahanadi River Basin, Environmental Management**

I. Introduction

Increasing air pollution is one of the greatest challenges to environmental health, particularly in densely populated and industrial regions. The Mahanadi River in Odisha, India is home to vital water bodies that are negatively affected by degraded air quality, which can have direct consequences on water quality. Pollutants such as particulate matter (both PM2.5 and PM10), sulphur dioxide, and nitrogen dioxide can easily enter these water bodies. [11].

It is vital to understand the interrelation between the attributes of air and water. The Mahanadi River is one of the most important sources for numerous communities, so any dilution of water quality will have wider health implications for the public, as well as for the economy at large. This piece of work intends to bridge the gap regarding the impacts of air pollution on water quality, providing vital information for policymakers and environmental managers.

This work will focus on designing predictive models by integrating the existing data gathered from 55 monitoring stations along with WQI values. The study will discover interdependencies involving air quality parameters and WQI using machine learning techniques; hence, it will be generally strong for understanding their interdependencies with each other.

There is limited direct research indicating the effect of WQI on AQI. Some studies usually take these indices separately because their focus areas differ—WQI regarding water pollution and AQI regarding air pollution. Lately, it has been observed that most recent studies in industry and agricultural regions start to investigate the interdependence of pollutants in water bodies and the air quality as well [9].

A particular study has highlighted that the industrial waste in water bodies releasing Volatile organic compounds (VOCs) will evaporate and cause localized air pollution and influence AQI in the nearby regions [8], [11]. Similarly, agricultural runoff with high nitrates and phosphates can easily run into water sources that produce nutrient-rich environments causing gases like ammonia or methane through chemical processes to come into the atmosphere and eventually affect air quality [10], [16], These are mainly observed interactions in the study of industrial zones where specific parameters of air quality have been proven to have a relation with pollutants found in water bodies, like those seen along the sides of the Mahanadi River [4].

Another study on pollution in urban-industrial regions found that stagnant water bodies with a high organic matter content produce greenhouse gases, including methane, through anaerobic bacterial processes. Such gases may degrade air quality, hence associating high WQI contamination levels with elevated AQI readings in the area [1], [18]. On another hand Heavy metals and particulate matter released from polluted water sources through evaporation or aerosolization contribute to air pollution, hence affecting AQI in residential and industrial zones in the near vicinity [14]. This Work was motivated by the urgent need to solve interlinked environmental issues, especially between air and water pollution. In heavily industrialized and populated areas, the pollutants in the air can easily get into water bodies, affecting water quality directly and posing threats to public health and ecological balance. Previous studies have separately analyzed the air quality index (AQI) and water quality index (WQI) [7], ignoring their interrelationship. It would help in formulating more holistic and effective strategies in environmental management to understand the relationship between AQI and WQI. The proposed project aims to develop a predictive data science tool, which will make use of the AQI parameters for the prediction of WQI, and then derive the AQI features most relevant to water quality for the provision of actionable insights to policymakers and environmental managers. These insights can aid in the development of focused interventions to protect the air and water resources in regions like Odisha's Mahanadi Basin so that sustainable management with environmental and health-risk-minimization can be enhanced. A few authors have applied analytical and AI-based approaches to understand both the dynamics of the Air Quality Index (AQI) and Water Quality Index (WQI) independently or partly to study their interdependences [15]. Traditional approaches applied are based on Geographic Information Systems (GIS) and other approaches of multivariate statistical analyses about spatial and temporal patterns found in air and water [12]. For example, the integration of GIS with factor analysis has enabled the determination of the effects of industrial and agricultural pollutants on water quality in areas like Mahanadi Basin [8], [11]. Some work used WQI along with environmental models to show how much the impact of pollutants affects both water as well as air quality indexes. These traditional approaches do offer spatial mapping and are valuable but lack predictive power and dynamic adaptability to the real-time environmental conditions [19].

Recent developments in AI and ML brought out more sophisticated prediction tools in environmental analysis, more precisely, for AQI and WQI. Machine learning models like Decision Trees, Random Forest, and Logistic Regression have shown good performance in predicting water and air quality indicators due to their ability to handle complex environmental datasets with multiple variables [4], [10], [16]. For instance, the studies employing logistic regression were found to be significantly accurate in terms of the ability to predict the patterns related to air pollution in the Mahanadi Basin and be utilized to furnish insight in terms of AQI as well as WQI parameters. In this study, logistic regression has shown to be somewhat straightforwardly implemented and interpreted; thereby it is quite effective where the transparency of the outcome of the model is crucial.

The AI-based techniques exhibited the best accuracy with neural networks, particularly for Recurrent Neural Networks and hybrid models such as PCA-RNN and PSR-SVM-FFA, capturing nonlinear and time-dependent data characteristics. For example, by combining PCA and RNN, the dimensional space reduces but retains its most vital features, making predictions much more accurate on the complexities of the system, and water quality in the Mahanadi river basin [14], [18]. Even Hybrid models such as PSR-SVM-FFA have shown exemplary accuracy for prediction with a maximum index value up to 0.984 on Willmott's Index, hence a robust application for monthly discharge and flow and AQI, ([9]). These suffer from problems of computational cost, requirements for data, and complexity in interpretation.

Accuracy comparisons among models suggest that hybrid AI techniques generally outperform conventional statistical and simpler ML models [13]. Decision Trees (DT) have shown accuracy levels of around 96.6% in water quality prediction, making them highly efficient and faster to implement. However, in terms of handling large-scale, high-dimensional environmental datasets, neural network models, particularly when combined with feature selection techniques like PCA, tend to deliver higher precision and reliability [5], [17].

Despite these advancements, several challenges remain, especially in scaling models for broader environmental applications and adapting them to region-specific conditions like those in Odisha's industrial zones. Many studies have called for the need to include region-specific environmental variables and parameters in AI models to enhance their applicability and accuracy [6], [9]. In particular, the lack of sufficient training data for highly localized pollutants or the seasonality of specific environmental events limits the robustness of AI models, necessitating further research into tailored solutions for predictive accuracy and generalizability.

Thus, even though AI-based models have accelerated environmental monitoring significantly, it is blatant that the current gap lies in models specially calibrated to predict WQI based on AQI parameters, especially in the case of regions like Odisha. This work does not only improve the predictive accuracy but also helps in further identifying the most relevant features of AQI that impact WQI, providing more actionable insights for sustainable air and water quality management.
The findings on how air pollutants affect water quality can inform better regulation strategies and interventions through the minimization of pollution so that public health and important water resources are protected. This project is meant to use data-driven answers that predict WQI by AQI parameters, to make it possible to understand better dynamics in the environment and promote effective, responsible decision-making in ensuring that there is sustainable management of water and

air quality.

II. WorkFlow

A. FlowChart

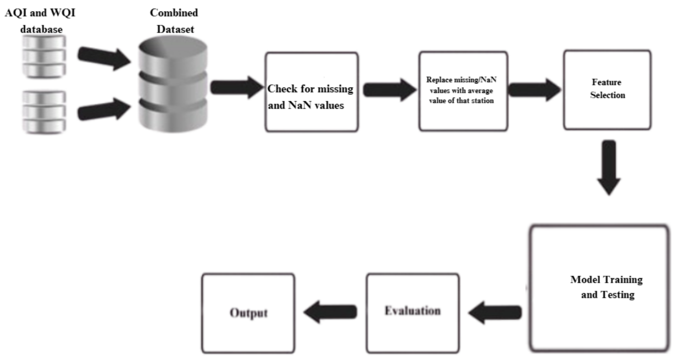

Fig. 1. Flowchart

B. Data Collection

Data collection for this study involved gathering Water Quality Index (WQI) data from the official website of the Odisha State Pollution Control Board [3]. The dataset spans eight months from January 2024 to August 2024 and includes measurements from 55 monitoring stations located along the Mahanadi River. For each of these WQI stations, we identified the nearest Air Quality Index (AQI) station and subsequently collected AQI data from the Central Pollution Control Board's official website [2]. The WQI-related features in the dataset include temperature (°C), pH, dissolved oxygen (DO, mg/L), biochemical oxygen demand (BOD, mg/L), chemical oxygen demand (COD, mg/L), total coliforms (TC, MPN/100 mL), fecal coliforms (FC, MPN/100 mL), and fecal streptococci (FS, MPN/100 mL), along with the calculated WQI. The AQI-related features encompass PM2.5, PM10, nitrogen dioxide (NO2), ammonia (NH3), sulfur dioxide (SO2), carbon monoxide (CO), ozone levels, and their corresponding AQI values. The final step involved merging these datasets based on proximity, ensuring a comprehensive analysis of the environmental conditions affecting the river ecosystem. This integrated dataset facilitates a robust examination of the interrelationship between water quality and air quality in the region, contributing valuable insights for environmental management and policy-making.

C. Data Preprocessing

We required data preprocessing in preparing gathered datasets for analysis. To begin with, we resolved the problem of missing values by the strategy of mean imputation wherein missing entries are replaced with the mean value of that respective column of columns within the Water Quality Index dataset as well as Air Quality Index dataset. We adopted this approach since maintaining integrity within the dataset along with minimization of bias may result due to excluding records based on incompleteness. We standardized the two datasets by appropriately scaling and formatting all features so that they could be used consistently in the dataset. We made sure to convert units into a standard format whenever possible, and we encoded categorical variables accurately. These preprocessing steps improved the quality of the data, providing a more reliable foundation for analyzing the relationship between water quality and air quality indicators in the study area.

III. Methodology

A. Exploratory Data Analysis (EDA)

1) AQI: Correlation heatmap Shown in Figure 2 portrays the relationship between various polluters and their impact on the Air Quality Index, AQI. The critical analysis reveals that PM10 and PM2.5 and NO are the most correlated with AQI values but with a high correlation score of 0.89 for PM10, at 0.75 and 0.70 on NO_2. This represents the fact that the amount of increase in these elements leads to a significant AQI increase, hence poor quality air. Ozone exhibits a moderate correlation strength with AQI that means 0.41 with the latter, therefore meaning it is second but well worthwhile during photochemical smog. The other two most massive air pollutants in addition SO2 and CO showed correlations are negative strength however and in lesser magnitudes and the impact NH3 manifested a statistically non-significantly weak. PM10, PM2.5; high degree inter-correlate as the both draw the almost the similar kind of source and behave same ways within the atmospheric processes. These insights are highly useful in developing data science tools for AQI forecast predictions since they focus one's attention on high impact pollutants for increasing the predictive power of models. Their knowledge also enables more precisely accurate predictive models and targeted intervention towards the management of the air quality.

2) WQI: This correlation heatmap shown in Figure 2 of Water Quality Index (WQI) parameters highlights the interrelationships between key water quality indicators and their influence on overall WQI. Strong positive correlations are observed between Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) (0.94), indicating that areas with high organic pollution tend to have elevated COD levels. Similarly, Total Coliform (TC) and Fecal Coliform (FC) are also highly correlated (0.97), reflecting their common origin from sewage contamination. BOD, COD, TC, and FC all show moderate to strong correlations with WQI (0.68 to 0.76), suggesting their significant contributions to water quality degradation. Dissolved Oxygen (DO) shows a strong negative correlation with both BOD (-0.75) and COD (-0.69), aligning with the fact that higher organic pollution reduces oxygen levels in water. In contrast, parameters such as water temperature and pH exhibit weaker correlations with WQI, indicating a lesser direct impact on water quality in this context. This analysis provides valuable insights into the main drivers of WQI, guiding targeted interventions for water quality improvement and assisting in predictive modeling

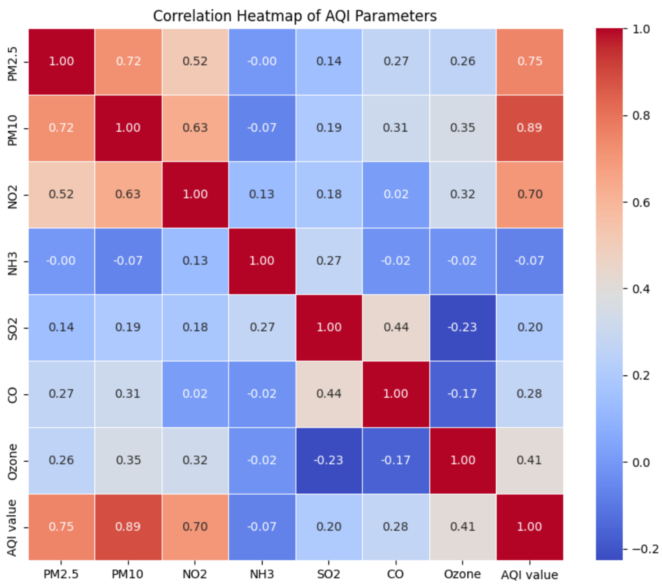

Fig. 2. AQI Correlation for Features PM2.5 , PM10 , NO2 , SO2 , CO , OZONE , AQI VALUE

efforts.

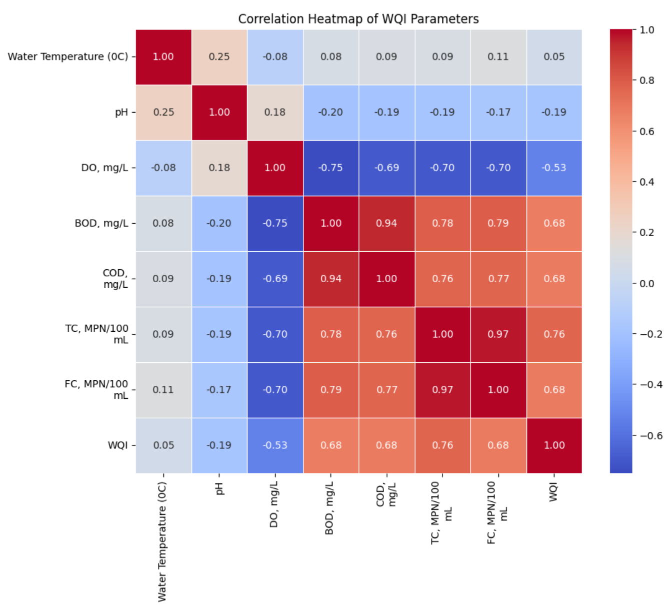

Fig. 3. WQI Corelation For Features PH,DO,BOD,COD,TC,FC,WQI,Water Temperture,WQI

3) Feature Importance: The feature importance plot shown in Figure 4 of a random forest model used with water quality data collected in Odisha show microbial contamination, particularly the total (TC, MPN/100 mL), and fecal coliform (FC, MPN/100 mL) as strongly determining factors for the whole variation of the water quality index, followed by COD in MG/L. The crucial role microbial pollution plays in defining the quality of water in Odisha and hence calls for specific intervention measures to reduce the level of coliforms and to upgrade the quality of water especially in areas with a higher level of fecal contamination.

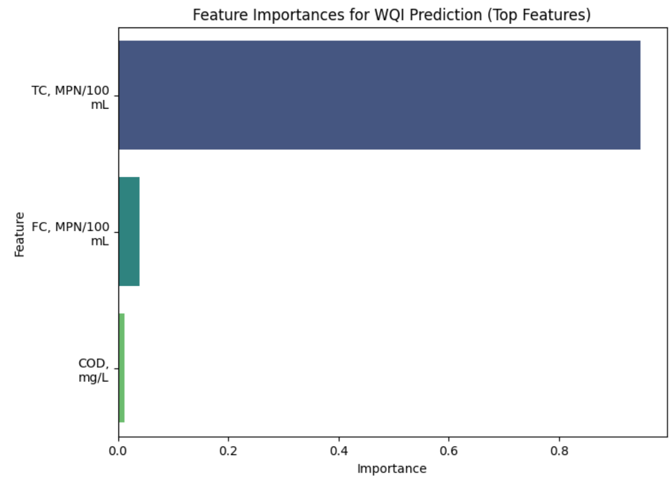

Fig. 4. Feature Selection for WQI With all WQI Features

From the plot shown in Figure 5 of feature importance, a random forest model trained on air quality data, the most significant variable for overall AQI prediction is particulate matter, followed by PM2.5 and PM10. NO2 is also a significant variable, but not as much as the former two. These findings underscore the critical role of particulate matter pollution in determining air quality and underscore the importance of targeted interventions to reduce emissions of PM10 and PM2.5, especially from vehicular and industrial sources, to improve air quality and public health.

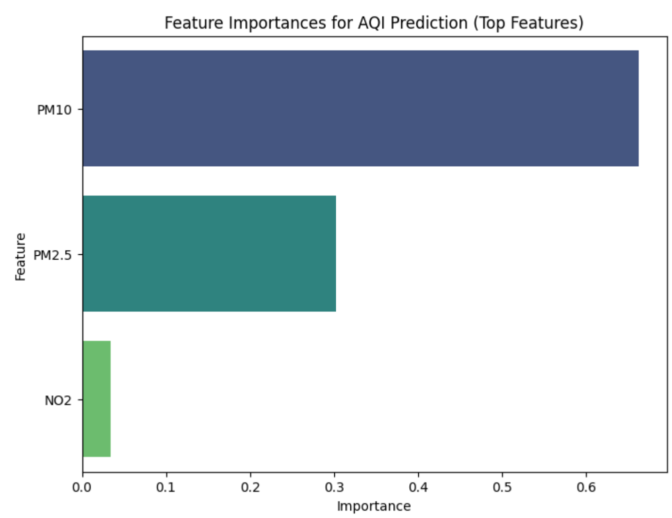

Fig. 5. Feature Selection for AQI With all AQI Features

The feature importance plot for Water Quality Index (WQI) prediction as shown in Figure 6 highlights the relative contribution of various parameters after applying feature selection

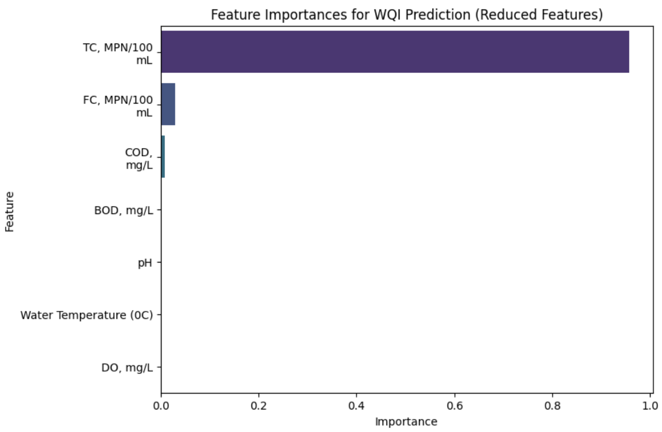

Fig. 6. Feature Importance for WQI Prediction Considering All AQI and WQI Features

on a combined Air Quality Index (AQI) and WQI dataset. The analysis reveals that Total Coliform (TC), measured in MPN/100 mL, is the most significant predictor of WQI, far surpassing other features in importance. Fecal Coliform (FC) also demonstrates a notable, albeit smaller, contribution. Minor contributions are observed for Chemical Oxygen Demand (COD), while other features such as Biochemical Oxygen Demand (BOD), pH, Dissolved Oxygen (DO), and Water Temperature show negligible importance in the prediction model.

This finding indicates that microbial contamination parameters (TC and FC) dominate the predictive capability for WQI, emphasizing their critical role in assessing water quality. These insights can guide the prioritization of monitoring efforts, particularly in regions where microbial contamination poses significant health risks. The reduction in features also enhances the model's efficiency while maintaining predictive accuracy, aligning with the principles of explainability and computational optimization in data science applications.

Figure 7 illustrates the relationship between PM10, PM2.5, NO2 concentrations, and the WQI through scatter plots. While the WQI appears to be largely invariant over a significant concentration range of the pollutants, the data points are random without a trend and indicate poor or no correlation between these pollutants and the WQI. This indicates that other factors should be more significant influencers of water quality in the region that has been studied.

The plots in Figure 8 indicate the correlation between water quality parameters such as TC, FC, COD, BOD, and DO with Air Quality Index (AQI). The plot indicates weak or nil correlation between water quality parameters and AQI. There is no clear pattern in any plot which enables one to speculate that probably other factors are dominating while establishing AQI in the studied region.

B. Model Evaluation

1) Decision Tree: It is an interpretable model, and one can predict using a tree-like structure. Every node will be a feature, like PM2.5, and the decision rules will be on branches. This project will predict WQI based on the selected AQI features, which could easily be explained as how these parameters affect water quality.

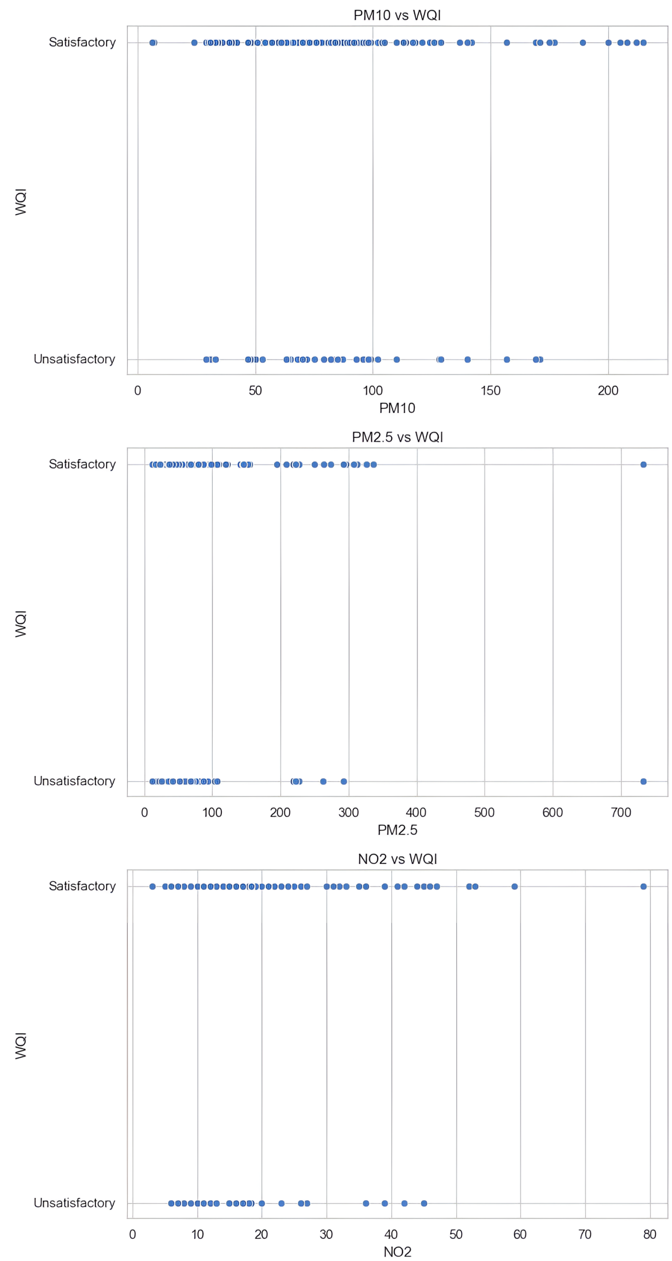

Fig. 7. Scatter Plot for Pollutant Concentration (units) Water Quality Index (WQI)

2) Random Forest: Random Forest is one such ensemble method; it aggregates multiple Decision Trees, improving accuracy. It fits several of these trees with different subsets of training data and averages their predictions to generate the outputs. This approach improves reliability and helps in capturing intricate patterns in AQI data better by implying a better predictive model for WQI.

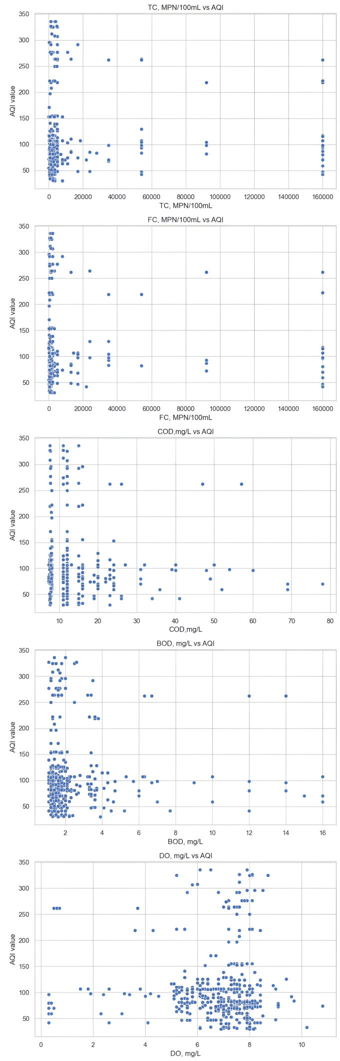

Fig. 8. Scatter Plot for Water Quality Parameter Concentration (units) Air Quality Index (AQI)

3) Training and Testing Sets:: This splitting will give the dataset: a 70% training set, and the remaining 30% as the test set. The models will be trained on the training set and validated on the test set to test how well the model performs. This splitting in this manner ensures that the evaluation is done on unseen data, thus, the models will be generalized and further provide a more accurate assessment of real-world performance.

4) Cross Validation Scores: This paper will use 5-fold cross-validation as shown in Table I to test the performance of three ensemble models: Random Forest, Gradient Boosting, and AdaBoost while predicting both Air Quality Index (AQI) and Water Quality Index (WQI) values. Features such as PM2.5, PM10, NO2, NH3, SO2, CO, and Ozone are used to predict AQI values, while features such as water temperature, pH, DO, BOD, COD, and microbial counts are used to predict WQI values. The Random Forest, Gradient Boosting, and AdaBoost models are scored with negative mean squared error (MSE) as the scoring metric. The cross-validation results indicate the average MSE for each model, which helps identify the best model for each prediction task based on their performance.

TABLE I
COMPARING CROSS-VALIDATION SCORES FOR AQI AND WQI PREDICTIONS

Model	**AQI Average Cross-Validation Score (5-fold)**	**WQI Average Cross-Validation Score (5-fold)**
Random Forest	1006.3308	0.0056
Gradient Boosting	795.4727	0.0071
AdaBoost	1064.6121	0.0068

C. Metrics

The two models, Random Forest and Decision Tree, will be evaluated using the following performance metrics:

Mean Squared Error (MSE): It is the average squared difference between the values predicted and the actual WQI values.

R^2 Coefficient of Determination: Proportion of variation in WQI to that of AQI features. This can be interpreted as the goodness of fit of the model.

IV. RESULTS AND CONCLUSION

In 2021, Mathonsi and van Zyl [3]proposed the MES-LSTM model as a hybrid approach to blend statistical and deep learning techniques for multivariate time series forecasting. With regard to the datasets on COVID-19 morbidity, MES-LSTM demonstrated higher performance than both traditional statistical and pure deep learning models in terms of forecasting accuracy and prediction interval construction. Hybrid models capture complex time dependencies and interactions among multiple variables efficiently, according to the study, promising considerable developments in predictive modeling. Chidiac et al. (2023) [2] have made a comprehensive review concerning the historical development of water quality indices, which trace the path from aggregative methods into more sophisticated models and applications, including statistical techniques and machine learning ones. They emphasized that the selection of parameters, data transformation, weight schemes, and aggregation methods are crucial components for an appropriate WQI and suggested combining sophisticated methodologies with it to increase both accuracy and flexibility in the assessment of water quality. Kumar [13]analyzed AQI models from 1960 to 2021 critically. Challenges in the development of AQI are: selection of relevant pollutants to be included, determination and aggregation of sub-index functions into a balanced index. It suggested that the application of more advanced statistical and machine learning methods would enhance the reliability and accuracy of AQIs that could be used to develop water quality indexes. In their 2021 review, Uddin et al. [19]reviewed a wide range of WQI models applied for the surface water quality assessment, reviewing the merits and demerits of different methods. According to them, most of the conventional WQIs lack flexibility and variability in handling complicated environmental conditions and interactions between pollutants. The authors suggested that machine learning be assimilated into the construction of WQIs to enhance flexibility in water quality predictions. For that purpose, according to them, more vibrant models sensitive to change must be developed for carrying out water-quality assessments.

TABLE II
MODEL COMPARISON

Model	**Mean MSE**	**Mean R^2**
Linear Regression	0.054276	0.633423
Random Forest Regressor	0.005607	0.959725
Decision Tree Regressor	0.006818	0.953338
Gradient Boosting Regressor	0.007099	0.951041
Support Vector Regressor	0.029607	0.795945
XGBoost Regressor	0.008241	0.943568

The model comparison results shown in Table II will reveal that the Random Forest Regressor outperforms other models in making predictions for both AQI and WQI. It means that the Mean Squared Error is lower and the R-squared value is higher than the previous models. In other words, Random Forest can efficiently capture complex interactions between parameters of air and water qualities, thus achieving more precise predictions. Other models, too, can be used such as Gradient Boosting Regressor, and XGBoost Regressor, which work as well, but Random Forest performs much better than others in our choices for the AQI and WQI prediction.

V. CONCLUSION

This study highlights the effectiveness of machine learning, particularly the Random Forest model, in predicting Water Quality Index (WQI) based on Air Quality Index (AQI) indicators. Random Forest consistently outperformed other models, showing high accuracy in predicting both AQI and WQI, with feature importance analysis underscoring the role of specific pollutants such as PM10, PM2.5, and microbial contaminants in determining air and water quality.

The findings suggest that machine learning models, especially when tailored to regional pollutant profiles, offer valuable insights for environmental management. The strong relationship between particulate matter, microbial pollution, and environmental quality indicators provides a foundation for targeted interventions that address specific pollution sources. Future work can build upon this study by incorporating hybrid models, which have shown promise in recent research for capturing complex interactions and improving predictive accuracy.

In conclusion, this study contributes to a deeper understanding of the relationships between air and water quality, presenting a robust, data-driven approach to environmental monitoring that can be adapted to other regions facing similar environmental challenges. The results underscore the potential of machine learning to enhance predictive accuracy, supporting more informed decision-making for sustainable environmental management.

VI. ACKNOWLEDGEMENTS

Sincere gratitude to all the agencies that contributed data to the Central Pollution Control Board (CPCB) [2] and Odisha State Pollution Control Board [3]

REFERENCES

[1] SK Bhoi, C Mallick, and CR Mohanty. Estimating the water quality class of a major irrigation canal in odisha, india: a supervised machine learning approach. *Nature Environment and Pollution Technology*, 21(2):433–446, 2022.

[2] Central Pollution Control Board. Air quality india, 2024.

[3] Odisha State Pollution Control Board. Environmental monitoring data, 2024.

[4] Abhijeet Das. Water criteria evaluation for drinking purposes in mahanadi river basin, odisha. In *International Conference on Trends and Recent Advances in Civil Engineering*, pages 237–260. Springer, 2022.

[5] Abhijeet Das. 5 identification of surface water contamination zones and its sources on mahanadi river, odisha using entropy-based wqi and mcdm techniques. *Optimization and Computational Case Studies in Civil Engineering*, page 67, 2023.

[6] Abhijeet Das. Delineation of surface water potential zones and identification of sites in mahanadi river basin, odisha, india, using gis. In *International Conference on Environmental Geotechnology, Recycled Waste Materials and Sustainable Engineering*, pages 11–38. Springer, 2023.

[7] Abhijeet Das. Identification of surface water contamination zones and its sources on mahanadi river, odisha using entropy-based wqi and mcdm techniques. *Engineering Research Transcripts*, 4:67–92, 2023.

[8] Abhijeet Das. Modelling of surface water quality and spatial mapping in mahanadi basin, odisha. *Journal of Mines, Metals & Fuels*, 71(9), 2023.

[9] Abhijeet Das. Assessment of surface water quality using entropy-wqi, fuzzy-topsis analysis, irrigation indices and spatial interpolation approaches in mahanadi river basin, odisha, india. *J. Environ. Nanotechnol*, 13(1):182–212, 2024.

[10] Abhijeet Das. A comparative assessment and its characterization of the integrated novel water pollution index and its statistical approach for the evaluation of spatial variations using factor analysis: A geospatial approach in mahanadi river, odisha. In *MATEC Web of Conferences*, volume 400, page 02007. EDP Sciences, 2024.

[11] Abhijeet Das. Investigation and comparative assessment of surface water quality for drinking purposes by using relief algorithm, gis, and machine learning: A case study of mahanadi river basin, odisha (india). In *MATEC Web of Conferences*, volume 400, page 02006. EDP Sciences, 2024.

[12] Ch. Sanjeev Kumar Dash, Ajit Kumar Behera, Sarat Chandra Nayak, Satchidananda Dehuri, and Jasaswi Prasad Mohanty. Estimation of air quality index of brajarajnagar and talcher industrial region of odisha state: A higher order neural network approach. In *2022 OITS International Conference on Information Technology (OCIT)*, pages 176–180, 2022.

[13] Thabang Mathonsi and Terence L. van Zyl. A statistics and deep learning hybrid method for multivariate time series forecasting and mortality modeling. 2022.

[14] Manas Ranjan Naik, Biswanath Mahanty, Sunil Kumar Sahoo, Viveka Nand Jha, and Naresh Kumar Sahoo. Assessment of groundwater geochemistry using multivariate water quality index and potential health risk in industrial belt of central odisha, india. *Environmental Pollution*, 303:119161, 2022.

[15] Rajashree Pany, Ashutosh Rath, and Prakash Chandra Swain. Water quality assessment for river mahanadi of odisha, india using statistical techniques and artificial neural networks. *Journal of Cleaner Production*, 417:137713, 2023.

[16] Shradhanjalee Pradhan, Janhabi Meher Sahoo, and Bibhuti Bhusan. Application of machine learning models in mahanadi river basin, odisha for prediction of flood. In *Industry 4.0 with Modern Technology*, pages 108–112. CRC Press, 2024.
[17] Sandeep Samantaray and Abinash Sahoo. Prediction of flow discharge in mahanadi river basin, india, based on novel hybrid svm approaches. *Environment, Development and Sustainability*, 26(7):18699–18723, 2024.
[18] RB Singh and KC Patra. Integrated pca–rnn approach for surface water quality assessment in the mahanadi river system. *International Journal of Environmental Science and Technology*, pages 1–16, 2024.
[19] Md Galal Uddin, Stephen Nash, and Agnieszka I Olbert. A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122:107218, 2021.