Water Pollution Prediction Using Machine Learning : Water is Potable or Not

[1]Dr. Santosh Kumar Singh, [2]Mr.Amit Pandey, [3]Deepa Jaiswal
[1]HOD (Information Technology), [2]Guide, [3]PG Student
[1,2,3,4]Department of Information Technology
Thakur College of Science and Commerce, Mumbai:400101, Mumbai, India

[1]sksingh14@gmail.com, [2]amitpandey8089@gmail.com,[3]jaiswaldeepa735@gmail.com

***Abstract*—Clean and safe drinking water is a fundamental human need, but water pollution is a severe environmental and public health problem. Laboratory chemical and biological analysis is the traditional method for estimating water quality, which is expensive, time-consuming, and not available at remote locations away from the analytical laboratory. Applying quantifiable physicochemical properties such as pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and turbidity, this research advocates for a machine learning method reliant on artificial intelligence to predict water potability. Supervised machine learning algorithms like Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Classifier, Logistic Regression, and Decision Tree were evaluated for prediction performance. Models were highly accurate, and Support Vector Classifier contains a maximum of 68.85%. The article assumes the potential of AI-based models being efficient, scalable, and cost-effective solutions to water quality estimation with immense possibility for real-time measurement and sustainable management of water resources.**

***Index Terms*—Water Quality Prediction, Machine Learning, Artificial Intelligence, Water Potability, Support Vector Classifier, Random Forest Classifier.**

I. INTRODUCTION

The most important substance for human survival is water, yet pure and safe drinking water is hard to find in most parts of the globe. Fecal-contaminated water may cause serious illnesses like gastrointestinal infections, neurologic diseases, and even life-threatening diseases such as typhoid and cholera. It is estimated by the World Health Organization (WHO) that approximately 2 billion individuals drink feces-polluted water, resulting in disease transmission and mortality. Determining whether water is safe to drink or not is thus an environment and public health imperative.

Conventional methods for determining the quality of water include the use of chemical and biological tests, which, if accurate, are normally expensive, time-consuming, and professional in nature. In addition, for distant and developing regions, infrastructure does not allow for regular testing of water. Machine learning can usher in the potential to make cost-effective, automated, and effective approaches to estimating potability in water with quantifiable physicochemical properties.

Water pollution is defined by analyzing the most important physicochemical parameters, which determine whether water is safe or not for drinking. Some of the most important parameters are **PH, Hardness, Solids, Chloramines, Sulfate,Conductivity, Trihalomethanes, Turbidity.**

Several machine learning techniques are applied to the water potability study such as the **support vector machines, random forests, naive Bayes, k-nearest neighbors, logistic regression and decision tree** that are commonly applied for the classification. **SVM classifies** water potability on the basis of the selection of the best hyperplane to separate potable and non-potable samples on the basis of physicochemical properties. **SVM** is able to handle complicated, non-linear relationships using kernel functions.**Random Forest** is different in that it creates numerous decision trees based on water parameters and averages them out to provide correct classification. It reduces overfitting and increases model stability.**Naïve Bayes** predicts water potability based on probability estimation of water being potable with independent attribute assumptions. It works effectively if data is normally distributed.**KNN** predicts water potability based on most similar water samples in data and classifying the most common class. It uses distance measures like Euclidean distance for classification. **Logistic Regression** predicts drinkability of water as a probability from input attributes with the help of a sigmoid function. **Logistic Regression** is used in binary classification problems like potable or not potable water. **Decision Tree** categorizes potability of water by inducing rule-based divisions of physicochemical parameters to categorize examples as potable or not potable. It displays decision rules graphically as trees.

II. Problem Statement

It cannot predict the pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Trihalomethanes, and Turbidity of water since it does not analyze the most important contaminants such as heavy metals (arsenic, mercury, lead), microbial contaminants (coliform bacteria, E. coli), and organic indicators (chemical oxygen demand (COD) and dissolved oxygen (DO)). These traits are necessary to include in order to mark the safety of the water, and without them, the model may not be able to identify dangerous contamination.

Furthermore, overcorrelation between features will add redundancy and decrease predictive ability. A few parameters by themselves will not be significant but will affect potability when interacted. Potability is dependent on several interacting variables, and thresholding or linear relationships may not always result in the optimum classification. Water of a proper pH may prove to be poisonous to some harmful agents beyond or below the range of data. Higher feature selection and model complexity will increase predictability.

III. Literature Review

Mustafa Yurtsever and Murat Emeç in his research work titled as "Potable Water Quality Prediction Using Artificial Intelligence and Machine Learning Algorithms for Better Sustainability" had utilized machine learning (ML) models for predicting drinking water quality. Conventional methods are costly and time-consuming for laboratory testing whereas AI models are timely and efficient. For water quality classification Neural networks, boosting models like XGBoost and AdaBoost, and decision trees have been applied. Support Vector Machines (SVM), LSTM, and Random Forest have been utilized extensively as well with accuracy of 70% to 97%. A hybrid SVR + XGBoost model with 99.54% accuracy has been suggested in the study that beats all the other models.[1]

Yingyi Chen, Lihua Song, Yeqi Liu, Ling Yang and Daoliang Li in his research work titled as "A Review of the Artificial Neural Network Models for Water Quality Prediction" The focus is on the use of Fake Neural Systems (ANNs) in the context of crave for water quality. Different Straight Backslide (MLR) and ARIMA are unable to explain nonlinear water quality plans. It focuses on the emerging usage of feedforward, flat, and half-breed ANN models for lack of water quality measures of fragmented oxygen (DO), biochemical oxygen request (BOD), pH, and turbidity. It abstracts 151 articles scattered between 2008 and 2019 and states that ANNs are more accurate and universal compared to regular models for water quality expectation.[2]

Muhammad Sani Gaya, Sani Isah Abba, Aliyu Muhammad Abdu, Abubakar Ibrahim Tukur, Mubarak Auwal Saleh, Parvaneh Esmaili, Norhaliza Abdul Wahab in his research work titled as "Estimation of Water Quality Index Using Artificial Intelligence Approaches and Multi-Linear Regression" applied Multi-Linear Relapse (MLR) to estimate Water Quality Record (WQI) of India's Yamuna Waterway's Palla station. Traditional models lack the capability to make precise estimations and thus the need for the use of Counterfeit Neural Systems (ANN) and Flexible Neuro-Fuzzy Induction Framework (ANFIS) has been necessary. AI models outperform MLR in consensus to the consider as ANN and ANFIS increase precision up to 10% during the testing handle. AI-based methods are efficient for water pollution level prediction and for effective water quality administration. [3]

Mohammed Falah Allawi, Sinan Q. Salih, Murizah Kassim, Majeed Mattar Ramal, Abdulrahman S. Mohammed, Zaher Mundher Yaseen in research work titled as "Application of Computational Model Based Probabilistic Neural Network for Surface Water Quality Prediction" The study examines the applicability of utilizing Artificial Intelligence (AI) models for estimation of surface water quality parameters such as Dissolved Oxygen (DO) and Biochemical Oxygen Demand (BOD5). Probabilistic Neural Network (PNN) and Multi-Layer Perceptron Neural Network (MLPNN) were utilized as AI models for water quality estimation using various physical and chemical parameters such as turbidity, pH, temperature, total dissolved solids, and sulfate content. It is established by the results that PNN outperforms MLPNN with higher accuracy ($R^2 = 0.94$ for DO and $R^2 = 0.93$ for BOD5) and can be viewed as a reliable model for water quality estimation.[4]

Nur Najwa Mohd Rizal, Gasim Hayder, Mohammed Mnzool, Bushra M. E. Elnaim, Adil Omer Yousif Mohammed, Manal M. Khayyat in his research work titled as "Comparison between Regression Models, Support Vector Machine (SVM), and Artificial Neural Network (ANN) in River Water Quality Prediction" The study compares machine learning (ML) models of water quality prediction for Malaysia's Langat River. It compares six regression models with two Support Vector Machines (SVM) and one Artificial Neural Network (ANN) for Total Suspended Solids (TSS) prediction, Total Solids (TS) prediction, and Dissolved Solids (DS) prediction. The models are compared based on Root Mean Square Error (RMSE), Mean Square Error (MSE), and Coefficient of Determination (R^2). Results show that ANN outperformed all the models with SVM and GPR models being overfitting. The research indicates that the most suitable model for water quality prediction is ANN.[5]

Nallakaruppan, E. Gangadevi, M. Lawanya Shri, Balamurugan Balusamy, Sweta Bhattacharya, Shitharth Selvarajan in his research work titled as "Reliable Water Quality Prediction and Parametric Analysis Using Explainable AI Models" The effort here is focused on Explainable Artificial Intelligence (XAI) for the prediction of water quality to offset the drawback of conventional black-box AI models. Different machine learning models like Logistic Regression, Support Vector Machine (SVM), Gaussian Naïve Bayes, Decision Tree (DT), and Random Forest (RF)

are used for drinking water classification. It is focused on the SHAP and LIME explainers' role for feature importance and interpretability analysis. RF accomplished with highest accuracy of 99.99% and therefore is the best option in classification. Total Dissolved Solids (TDS) was found to be the most impactful factor for water quality identification.[6]

Mosleh Hmoud Al-Adhaileh, Fawaz Waselallah Alsaade in his research work titled as "Modelling and Prediction of Water Quality by Using Artificial Intelligence" The study presents some of the artificial intelligence (AI) methods applied to model and predict water quality. It presents AI models such as Artificial Neural Networks (ANNs), Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Machines (SVMs), and K-Nearest Neighbors (KNN). It also refers to the improvement in accuracy with deep learning techniques such as CNN-LSTM. Some of the past studies established the feasibility of machine learning models in predicting water quality index (WQI) and determining the classification of the water quality condition. The literature also addresses the coupling of AI methodologies for real-time monitoring and water management sustainability.[7]

Afaq Juna, Muhammad Umer, Saima Sadiq, Hanen Karamti, Ala' Abdulmajid Eshmawi, Abdullah Mohamed, Imran Ashraf in his research work titled as "Water Quality Prediction Using KNN Imputer and Multilayer Perceptron" The paper contrasted different approaches to predicting water quality based on machine learning and deep learning techniques. The standard water quality models of classification will not work due to the occurrence of missing data. Artificial Intelligence models like ANN, SVM, and deep networks have been demonstrated by research studies to be well-suited to predict water quality. It is responsible for XGBoost and LSTM model applications in time-series forecasting. It recommends the utilization of multilayer perceptron (MLP) for more precision and KNN imputer for filling missing values, which is better than traditional models. [8]

Ozgur Kisi,Kulwinder Singh Parmar, Amin Mahdavi-Meymand, Rana Muhammad Adnan, Shamsuddin Shahid, Mohammad Zounemat-Kermani in his research work titled as "Water Quality Prediction of the Yamuna River in India Using Hybrid Neuro-Fuzzy Models" The study is on various models of water quality prediction with special focus on artificial intelligence (AI) techniques such as Adaptive Neuro-Fuzzy Inference System (ANFIS) and Least Square Support Vector Machine (LSSVM). It contrasted the performance of meta-heuristic optimization techniques such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Harmony Search (HS), and Teaching–Learning-Based Optimization (TLBO) for improvement in prediction accuracy. Traditional statistical models are not able to handle nonlinearities and therefore are less accurate in comparison to AI-based models. It was already established by research that AI-based hybrid models are more accurate in prediction due to parameter tuning optimization. The paper elucidates why AI-based neuro-fuzzy models are superior for predicting water quality over conventional prediction techniques.[9]

Nur Afyfah Suwadi, Morched Derbali, Nor Samsiah Sani, Meng Chun Lam, Haslina Arshad, Imran Khan, Ki-Il Kim in his research work titled as " An Optimized Approach for Predicting Water Quality Features Based on Machine Learning" The paper discusses the shortcomings of conventional water quality analysis and increasing dependence on machine learning (ML) techniques for predictive modeling. The paper covers Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), and Naïve Bayes (NB) for the classification of water quality. The paper also covers the use of feature selection techniques such as information gain, correlation, and symmetrical uncertainty in the direction of achieving maximum model accuracy. It has already been established through past research that hybrid ML models improve the predictability, particularly in the treatment of complicated data. It targets the function of WQI parameters such as dissolved oxygen (DO) and biochemical oxygen demand (BOD) towards markedly enhancing ML-based water quality forecast.[10]

IV. METHODOLOGY

- **Dataset and Preprocessing** The Water Potability Dataset consists of 3,276 water samples and nine numeric attributes and one target attribute (Potability), indicating whether the water is potable (1) or not (0). The dataset contains a number of physicochemical parameters such as pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, and Turbidity. These parameters influence water quality in that they determine its acidity, mineral content, organic matter, and levels of disinfectant byproducts.The data set is of the type where it can be employed for a binary classification problem, and the goal is to classify if the sample water is potable (1) or not (0)..

➢ **Preprocessing Steps include :**

- **Handling missing values :** Missing values for quantitative attributes such as pH, sulfate, and trihalomethanes were handled via mean or median imputation depending on whether they have a normal or skewed distribution.

- **Normalization :** All numerical attributes were normalized using Min-Max scaling to scale the values in the range of 0 to 1 with an attempt to improve model performance.

- **Feature Selection:** Correlation analysis was employed to remove features that had high correlation. Overfitting can be minimized and efficiency enhanced by choosing fundamental features like pH, conductivity, and turbidity.
- **Data Splitting :** The data was divided into the train (80%) and test (20%) sets in a bid to accurately measure model performance.

V. Machine Learning Models

➢ **Logistic Regression**

It forecasts water potability with 61% accuracy for different chemical and physical parameters using logistic regression which is not unusual given the dataset's nature. It includes 3,276 examples with 9 input variables such as pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and turbidity. There are, however, quite substantial missing values—necessarily in the pH, sulfate, and trihalomethanes columns. Logistic regression, as a linear classifier, is highly sensitive to missing values, feature scale changes, and class imbalance, all of which have adverse impacts on prediction performance.

Additionally, feature sizes of data range from tens of thousands for solids but others like turbidity are single-digit features. Such type of disparity without scaling would over-weight some features in the model and reduce accuracy. Also, in case of target classes (non-potable or potable water) being class-imbalanced, logistic regression would be even more out of place because logistic regression doesn't natively handle class imbalance.

Linearity of logistic regression also means that it is possibly unable to detect high-order interactions or nonlinear relationships among features. This may be the explanation for the satisfactory performance at 61% accuracy. For the model to be improved, imputation to deal with missing values, scaling of feature values, and usage of stronger classification models like Random Forest or Gradient Boosting are recommended. These models are also more capable of handling missing data and non-linear relationships, with the possibilities of even greater accuracy for the water potability prediction task.

➢ **K - Nearest Neighbors**

Following the application of the K-Nearest Neighbors (KNN) algorithm to the potability of water dataset, a 63.49% accuracy is achieved. KNN is a neighbor algorithm that is data point similarity-dependent and very sensitive to missing data as well as feature scaling. In this data set, certain of the features like pH, sulfate, and trihalomethanes have missing values, which, if not properly treated, would adversely affect the model to calculate good distances between points. The range for differences in features is also highly disparate in the data set—e.g., solids values are in tens of thousands whereas turbidity is in single numbers. Without standardization or normalization, high-value features swamp the distance calculation and get poor performance.

The relatively modest gain in accuracy over logistic regression suggests that KNN does get some advantage out of its capability to learn non-linear relationships in the data. But the accuracy is still bounded, maybe by the data set's complexity, unbalanced classes, and potential noise in the data. Apart from adjusting the performance of KNN, missing values must be treated properly, features are scaled using tools such as Standard Scaler or Min - Max Scaler, and several values of K are tried in an attempt to reach the best number of neighbors. Even predictive precision can be enhanced by employing distance-weighted KNN in such a manner that the choice of classification would become increasingly influenced by more proximate neighbors.

➢ **Support Vector Classifier**

Using the water potability data, the Support Vector Machine (SVM) algorithm was applied, it yielded a better accuracy of 68.85%, which is a significant difference compared to K-Nearest Neighbors (KNN) and logistic regression models. SVM is especially beneficial in classification problems with high and non-linear relationships because it works based on finding the optimal hyperplane that best separates classes based on the largest margin. In this case, the model would have been improved by SVM's ability to process high-dimensional data and separate classes more effectively even when the data set has overlapping features.

That the accuracy was improved is an indication that SVM was able to identify the finer patterns that are associated with the chemical and physical characteristics of water, i.e., pH, hardness, sulfate concentration, and concentration of trihalomethanes. In addition, SVM is also quite insensitive to the feature scaling issues after the proper normalization has been applied, and this enables more generalization to new data. Proper normalization of missing values would have further enhanced the performance of the SVM model. 68.85% accuracy achieved by this test is a representation of the moderate predictability, and it is comforting that SVM is capable of performing better on this particular issue of classification than simple models. With further tuning, such as experimenting with various kinds of kernels and other

regularization parameters, SVM model can be taken to a still higher level of accuracy in the prediction of water potability.

- **Naive Bayes**

While being used on the water potability dataset, Naive Bayes algorithm has achieved a classification rate of 63.68%. Naive Bayes is a probabilistic classifier based on Bayesian statistics and implements Bayes' theorem with the assumption that none of the features are correlated with each other. Although this never occurs using actual datasets—especially environmental datasets like water quality parameters—the model does very well due to its being low cost and easy in dealing with classification problems.

Correct performance is proof of the model's ability to learn from data for missing-value and intercorrelation situations for attributes. Notwithstanding this, very poor performance for Naive Bayes here is because of the intercorrelated features of pH, sulfate, hardness, and solids that break the independence assumption of the model. Furthermore, the fact that most of the features are continuous cannot be well explained by model assumptions, especially where data are not normally distributed.

Overall, despite the fact that Naive Bayes is a fast and simple baseline model, the 63.68% accuracy is telling us maybe it's not the ideal algorithm to utilize in modeling these intricate relationships in this data. That being said, it is insightful and a valuable benchmark in evaluating the performance of more sophisticated models such as Support Vector Machines or Random Forests.

- **Decision Tree**

58.96% accuracy is reached when Decision Tree algorithm has been applied with the data of water potability, which is considerably low in relation to algorithms such as Support Vector Machine (SVM) or Naive Bayes. Decision Trees divide the data set into subsets based upon feature value that best reduces impurity and thereby are interpretable and intuitive. Decision Trees are computationally expensive and significantly overfit, particularly on noisy data, missing data, or high feature dependencies.

Lower accuracy here is that Decision Tree was unable to generalize across the test set and that may be due to overfitting the training set or due to the introduction of biases due to missing data and scale variation across attributes like solids, sulfate, and pH. If water quality variables having non-linear relationships are considered, a basic Decision Tree may not be able to make decisions on such complexities and hence result in misclassifications. In addition, Decision Trees also suffer from minor data fluctuations inducing tree structure instability and volatile model behavior.In summary, although Decision Trees are useful for revealing feature importance and illustrating decision paths, the accuracy of 58.96% suggests that more sophisticated models or ensemble methods like Random Forest or Gradient Boosting should be employed as a final effort to predict water drinkability in this data set.

- **Random Forest**

68.21% accuracy was achieved with the use of the Random Forest algorithm when applied to data on water potability, way above that from the basic decision trees, Naive Bayes, and the logistic regression algorithms. Random Forest is an ensemble learning technique by which predictions are made using several decision trees and aggregating them in a try to prevent overfitting as well as derive the best possible generalization. This method enables the model to learn the detail and non-linear pattern of the data better than the application of one decision tree.

Greater accuracy demonstrates the power of Random Forest to handle datasets with features of varying scales and missing values because it will still be able to function perfectly even in the presence of missing data. The model is also not as reliant on a single feature because it randomly samples features at training time, hence stabilizing its predictions. In water potability prediction, Random Forest may have learned the subtle interaction between chemical properties such as pH, hardness, sulfate, and trihalomethanes concentration, hence providing improved classification performance.

In conclusion, the accuracy of 68.21% indicates that Random Forest is an appropriate model for this type of data and a good balance between accuracy and interpretability. With some additional tuning of tree depth or trees number, the performance of the model will be slightly further improved.

VI.Model Results

Model	Accuracy
Logistic Regression	**62.85%**
K- Nearest Neighbors	**63.49%**
Support Vector Classifier	**68.85%**
Naive Bayes	**63.68%**
Decision Tree	**58.96%**
Random Forest	**68.21%**

VII. Conclusion

Comparison of some different machine learning models in the prediction of water potability based on chemical parameters has given insightful results. Out of six models compared, Support Vector Classifier (SVC) performed best with a 68.85% accuracy followed by the Random Forest model with a 68.21% accuracy. The outcome is that models specifically designed to handle the kind of complex, non-linear relationships between data in this dataset perform best here. Nonetheless, the overall accuracy levels conform to the problem of classifying water potability being a moderately difficult one based on data complexity, missing data, and feature distributions overlapping. Easier models such as Logistic Regression, K-Nearest Neighbors, Naive Bayes, and Decision Tree yielded lower accuracies, which once again confirms the requirement to be more complex algorithms or richer datasets for improved prediction.

VII. Future Work

Comparison of some different machine learning models in the prediction of water potability based on chemical parameters has given insightful results. Out of six models compared, Support Vector Classifier (SVC) performed best with a 68.85% accuracy followed by the Random Forest model with a 68.21% accuracy. The outcome is that models specifically designed to handle the kind of complex, non-linear relationships between data in this dataset perform best here. Nonetheless, the overall accuracy levels conform to the problem of classifying water potability being a moderately difficult one based on data complexity, missing data, and feature distributions overlapping. Easier models such as Logistic Regression, K-Nearest Neighbors, Naive Bayes, and Decision Tree yielded lower accuracies, which once again confirms the requirement to be more complex algorithms or richer datasets for improved prediction.

References

1. Yurtsever, M., & Emeç, M. (2023). Potable Water Quality Prediction Using Artificial Intelligence and Machine Learning Algorithms for Better Sustainability.
DOI: 10.21121/eab.1252167.

2. Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). **A Review of the Artificial Neural Network Models for Water Quality Prediction**. *Applied Sciences*, **10**(17), 5776.
DOI: 10.3390/app10175776.

3. Gaya, M. S., Abba, S. I., Abdu, A. M., Tukur, A. I., Saleh, M. A., Esmaili, P., & Wahab, N. A. (2020). Estimation of Water Quality Index Using Artificial Intelligence Approaches and Multi-Linear Regression. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 9(1), 126-134.
DOI: 10.11591/ijai.v9.i1.pp126-134.

4. Allawi, M.F., Salih, S.Q., Kassim, M., Ramal, M.M., Mohammed, A.S., & Yaseen, Z.M. (2022). **Application of Computational Model Based Probabilistic Neural Network for Surface Water Quality Prediction**. *Mathematics*, **10**(3960).
DOI: 10.3390/math10213960.

5. Rizal, N. N. M., Hayder, G., Mnzool, M., Elnaim, B. M. E., Mohammed, A. O. Y., & Khayyat, M. M. (2022). **Comparison between Regression Models, Support Vector Machine (SVM), and Artificial Neural Network (ANN) in River Water Quality Prediction**. *Processes*, **10**(1652).
DOI: 10.3390/pr10081652.

6. Nallakaruppan, M. K., Gangadevi, E., Shri, M. L., Balusamy, B., Bhattacharya, S., & Selvarajan, S. (2024). **Reliable Water Quality Prediction and Parametric Analysis Using Explainable AI Models**. *Scientific Reports, 14*(7520).
DOI: 10.1038/s41598-024-56775-y.

7. Al-Adhaileh, M. H., & Alsaade, F. W. (2021). *Modelling and Prediction of Water Quality by Using Artificial Intelligence*. Sustainability, **13**(8), 4259

8. Juna, A., Umer, M., Sadiq, S., Karamti, H., Eshmawi, A. A., Mohamed, A., & Ashraf, I. (2022). *Water Quality Prediction Using KNN Imputer and Multilayer Perceptron*. Water, **14**(17), 2592.

9. Kisi, O., Parmar, K. S., Mahdavi-Meymand, A., Adnan, R. M., Shahid, S., & Zounemat-Kermani, M. (2023). *Water Quality Prediction of the Yamuna River in India Using Hybrid Neuro-Fuzzy Models*. Water, **15**(6), 1095.

10. Suwadi, N. A., Derbali, M., Sani, N. S., Lam, M. C., Arshad, H., Khan, I., & Kim, K.-I. (2022). *An Optimized Approach for Predicting Water Quality Features Based on Machine Learning*. Wireless Communications and Mobile Computing, **2022**, Article ID 3397972.