

Optimizing Water Quality Prediction with Advanced Deep Learning and AutoML: Hybrid Models, Transfer Learning, and Ensemble Strategies

NAME OF THE STUDENT

MAIL ID

CLG NAME

AREA

NAME OF THE STUDENT

MAIL ID

CLG NAME

AREA

NAME OF THE STUDENT

MAIL ID

CLG NAME

AREA

NAME OF THE STUDENT

MAIL ID

CLG NAME

AREA

NAME OF THE STUDENT

MAIL ID

CLG NAME

AREA

NAME OF THE STUDENT

MAIL ID

CLG NAME

AREA

Abstract

Water quality assessment is essential for public health and environmental safety. This study utilizes deep learning techniques to predict water quality using the **Indian Aqua Attributes Dataset**, which includes parameters like **temperature, dissolved oxygen, pH, conductivity, B.O.D, nitrate levels, and coliform counts**. After preprocessing, three models—**CNN, LSTM, and CNN-LSTM**—were trained and evaluated. Among them, the **CNN-LSTM model achieved the highest performance**, with an **AUC-ROC score of 0.95**, demonstrating its reliability in classification. To enable real-time predictions, the trained model was saved in **.h5 format** and deployed using **Streamlit**, allowing users to input water parameters and obtain instant results. The findings highlight the effectiveness of deep learning in water quality monitoring. Future work could enhance accuracy by incorporating additional attributes and larger datasets.

Keywords

Water Quality Prediction, Deep Learning, CNN-LSTM, Indian Aqua Attributes Dataset, Streamlit, ROC Curve, Confusion Matrix, Environmental Monitoring.

I. Introduction

Water quality plays a crucial role in ensuring the safety of drinking water, industrial applications, and environmental conservation. Contaminants in water sources can lead to severe health hazards, making accurate and timely water quality prediction essential. Traditional monitoring methods rely on manual sampling and laboratory testing, which can be time-consuming,

expensive, and inconsistent due to variations in environmental conditions. With advancements in artificial intelligence and deep learning, automated prediction models have emerged as efficient alternatives for assessing water quality. This study explores the application of CNN, LSTM, and CNN-LSTM models on the Indian Aqua Attributes Dataset to classify water quality based on key parameters such as pH, temperature, dissolved oxygen, conductivity, B.O.D, nitrate levels, and coliform counts. The CNN-LSTM model achieved superior performance, demonstrating the potential of deep learning in real-time water quality assessment. The proposed system is integrated into a Streamlit-based web application, allowing users to input water parameters and obtain instant predictions. By leveraging deep learning-based classification, this approach aims to enhance decision-making in water management while reducing dependence on manual testing.

Background on Water Quality Prediction

Water quality prediction involves assessing the physical, chemical, and biological properties of water to determine its suitability for consumption and other applications. Various parameters, including pH levels, temperature, dissolved oxygen, and bacterial contamination, serve as key indicators of water quality. Traditionally, water quality assessments require extensive sample collection, laboratory analysis, and expert interpretation, which can be inefficient and costly. With advancements in machine learning and deep learning, data-driven models can now predict water quality with high accuracy. These models analyze historical and real-time data to identify patterns and classify water samples effectively. In

this study, deep learning architectures like CNN, LSTM, and CNN-LSTM were trained to classify water quality based on the Indian Aqua Attributes Dataset. The CNN-LSTM model demonstrated the highest classification accuracy, making it a reliable tool for automated water quality prediction.

Importance of Water Quality Monitoring in India

India faces significant challenges in maintaining clean and safe water sources due to rapid industrialization, urbanization, and population growth. Contaminants from industrial discharge, agricultural runoff, and sewage often degrade water quality, leading to health risks such as waterborne diseases, gastrointestinal infections, and long-term toxic effects. Effective water quality monitoring is essential to prevent such risks and ensure compliance with environmental and public health regulations. The Indian government has implemented initiatives like the National Water Quality Monitoring Programme (NWQMP) to track water quality; however, traditional methods rely on manual sampling, which is inefficient for large-scale monitoring. By utilizing deep learning-based prediction models, water quality assessment can be automated, improving accuracy and efficiency. The proposed CNN-LSTM-based system offers real-time analysis through a web-based application, providing a scalable and accessible solution for water quality monitoring in India.

Challenges in Traditional Water Quality Assessment

Traditional water quality assessment methods involve manual sample collection, laboratory testing, and expert analysis, posing several challenges:

- 1) Time-Consuming Process – Laboratory analysis requires significant time for chemical and biological tests, delaying critical decision-making.
- 2) High Costs – Regular sample collection and testing involve substantial financial investment, making it impractical for continuous monitoring.
- 3) Geographical Limitations – Remote areas lack access to well-equipped laboratories, leading to delays in water quality assessment.
- 4) Human Errors and Inconsistencies – Manual sampling can introduce errors due to variations in collection techniques and environmental conditions.
- 5) Scalability Issues – Traditional methods struggle to monitor water quality across vast regions and multiple sources simultaneously.

To overcome these limitations, deep learning models provide an efficient alternative by analyzing water quality data in real time. The CNN-LSTM model, deployed via a Streamlit web application, enhances accuracy and accessibility while reducing dependence on manual testing.

Scope of the Proposed Work

This research focuses on developing an automated water quality prediction system using deep learning techniques. The Indian Aqua Attributes Dataset was preprocessed by removing irrelevant features and normalizing key parameters like pH,

temperature, dissolved oxygen, conductivity, and nitrate levels. After training and evaluating CNN, LSTM, and CNN-LSTM models, the CNN-LSTM model achieved the best classification performance, with an AUC-ROC score of 0.95. The trained CNN-LSTM model was deployed using Streamlit, creating an interactive web-based platform where users can input water parameters and receive instant predictions. This system aims to:

- 1) Improve Accuracy – Achieve high precision in water quality classification using deep learning.
- 2) Enable Real-Time Monitoring – Provide instant predictions through a user-friendly web interface.
- 3) Enhance Scalability – Allow deployment across multiple regions with minimal infrastructure requirements.
- 4) Reduce Manual Effort – Minimize dependence on traditional sample collection and testing.

Future enhancements may include expanding the dataset, incorporating additional water quality parameters, and integrating IoT sensors for real-time data collection, further improving prediction capabilities.

Deep Learning-Based Approaches in Water Quality Analysis

Deep learning has revolutionized water quality prediction by providing automated, scalable, and accurate models. Studies have explored different DL architectures to classify and predict water contamination levels. CNN models are widely used for feature extraction, while LSTM models handle time-series dependencies. Hybrid models, such as CNN-LSTM, have demonstrated superior performance in various environmental monitoring applications. The proposed work utilizes a CNN-LSTM model trained on the Indian Aqua Attributes Dataset, achieving high accuracy in water quality classification. The model's integration into a Streamlit-based web application enables real-time water quality assessment, reducing reliance on traditional methods.

II. Related Work

Overview of Existing Methods for Water Quality Prediction

Water quality prediction has been an area of extensive research due to its importance in public health, environmental sustainability, and industrial applications. Traditional methods of assessing water quality rely on laboratory-based chemical and biological analysis, requiring manual sampling and testing. While these methods provide accurate results, they are time-consuming, labor-intensive, and lack real-time monitoring capabilities. In recent years, machine learning (ML) and deep learning (DL) models have gained prominence in automating water quality assessment. ML models such as decision trees, support vector machines (SVM), and random forests (RF) have demonstrated improved efficiency in classifying water quality based on historical data patterns.

However, these models have limitations in handling complex time-series data and often require extensive feature engineering. Deep learning approaches, particularly Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and hybrid CNN-LSTM models, have shown superior performance in water quality prediction. These models automatically extract spatial and temporal features from water quality datasets, enhancing predictive accuracy and reducing the need for manual preprocessing.

Literature Survey

Umair Ahmed et al. (2019) proposed a supervised machine learning approach to classify water quality based on various chemical and physical parameters. The study utilized ML algorithms such as decision trees, support vector machines, and random forests, demonstrating high classification accuracy. However, the model required extensive feature engineering and manual parameter tuning. Compared to this study, our proposed CNN-LSTM model eliminates the need for manual feature selection by leveraging deep learning's automated feature extraction capabilities. Additionally, the Streamlit-based web interface enhances real-time usability.

Heelak Choi et al. (2019) explored the effectiveness of deep learning models such as CNN and LSTM for short-term water quality forecasting. The study highlighted LSTM's ability to model time-series dependencies, making it superior to conventional ML methods. In our research, we extend this approach by integrating CNN and LSTM into a hybrid CNN-LSTM model, achieving better performance for multiclass water quality classification. The experimental results demonstrate that CNN-LSTM outperforms individual deep learning models in predicting water quality trends.

Ping Liu et al. (2019) developed an LSTM-based water quality prediction system integrated into an IoT environment. The study emphasized real-time data collection from IoT sensors to enhance predictive accuracy. However, the LSTM model alone struggled with feature extraction from complex water quality datasets. Our work builds upon this research by combining CNN for feature extraction and LSTM for sequential learning, leading to higher classification accuracy. While IoT-based real-time data collection remains an area for future expansion, our model currently provides real-time predictions through a Streamlit web interface.

III. Dataset and Preprocessing

Dataset Description

The dataset utilized in this study is derived from HydroShare, a collaborative platform for sharing hydrological datasets. The dataset, referred to as the Indian Aqua Attributes Dataset, contains detailed water quality measurements collected from various monitoring stations across India. These records provide insights into the physicochemical properties of water, which are essential for assessing its suitability for human

consumption, agriculture, and industrial applications. The dataset consists of multiple attributes, including station codes, location details, temperature, pH levels, dissolved oxygen (DO), biochemical oxygen demand (BOD), electrical conductivity (EC), and total dissolved solids (TDS). These parameters serve as key indicators of water quality, helping in the classification of water into different categories such as safe, moderate, or hazardous. Given the importance of these attributes, the dataset is ideal for developing a deep learning-based water quality prediction system using CNN-LSTM models.

The dataset used in this project has been sourced from **HydroShare .hydroshare.org/resource**, a widely recognized repository for hydrological and environmental data. HydroShare provides open access to water quality datasets, enabling researchers to develop models that enhance water resource management and pollution control strategies. This dataset has been curated from **multiple water monitoring stations across India**, covering diverse **geographical regions and seasonal variations**. The inclusion of multiple data points from different regions allows for a more generalized and robust prediction model, making it applicable to a variety of water sources across the country.

Explanation of Features

The dataset comprises several key features that are crucial for **predicting water quality**. Below is a description of the most relevant attributes:

- **Station Code:** A unique identifier assigned to each water quality monitoring station.
- **Location:** The geographical coordinates or name of the water source from which samples were collected.
- **Temperature (°C):** The water temperature, which influences chemical reactions and biological processes.
- **pH:** Measures the acidity or alkalinity of water on a scale of 0 to 14. Values below 7 indicate acidity, while values above 7 indicate alkalinity.
- **Dissolved Oxygen (DO) (mg/L):** Represents the amount of oxygen available in water for aquatic life. Low DO levels indicate pollution.
- **Biochemical Oxygen Demand (BOD) (mg/L):** Indicates the amount of oxygen required by microorganisms to decompose organic matter. High BOD values suggest contamination.
- **Electrical Conductivity (EC) (µS/cm):** Measures the water's ability to conduct electricity, which is influenced by dissolved salts and minerals.
- **Total Dissolved Solids (TDS) (mg/L):** Represents the total concentration of dissolved substances in water. High TDS levels can indicate poor water quality.

These features are used to train the **CNN-LSTM model**, allowing it to classify water quality based on patterns within the data.

Data Cleaning and Preprocessing

To ensure the dataset is well-structured and suitable for deep learning, several preprocessing steps were performed. Initially, unnecessary columns that did not contribute to water quality prediction, such as administrative remarks and sampling time, were removed to reduce data complexity. Handling missing values was a crucial step, as incomplete data could lead to inaccurate predictions. Rows with more than 40% missing values were discarded, while those with minimal missing values were imputed using statistical methods like mean, median, or mode, depending on the attribute type. For categorical variables like station locations, missing values were replaced using the most frequent value to maintain consistency. Additionally, since the dataset contains attributes with varying numerical scales, feature scaling was applied to standardize the values. Min-Max scaling was used to normalize numerical features, bringing all values within a range of 0 to 1, ensuring balanced learning and preventing any single feature from dominating the model training. These preprocessing steps significantly improved the dataset's quality, enhancing the accuracy and stability of the CNN-LSTM model for water quality prediction.

IV. Methodology

Convolutional Neural Network (CNN)

The Convolutional Neural Network (CNN) was employed to extract spatial features from the water quality dataset. Although CNNs are primarily used for image processing, their ability to identify patterns and correlations in structured data makes them useful for water quality prediction. The dataset, consisting of numerical values for parameters like temperature, pH, and conductivity, was reshaped into a format suitable for CNN input. Convolutional layers extracted relevant feature representations, followed by activation functions and pooling layers that reduced dimensionality while retaining crucial information. The fully connected layers at the end processed the extracted features for classification.

Long Short-Term Memory (LSTM)

To capture the temporal dependencies in water quality variations, an LSTM model was implemented. LSTM networks are a type of recurrent neural network (RNN) designed to handle sequential data efficiently. Water quality data fluctuates over time due to environmental and human factors, making LSTM suitable for learning these dependencies. The dataset was structured as time-series input, where each sample represented water quality parameters at a specific time step. The memory cells in LSTM retained important past information while filtering out irrelevant data, allowing the model to make predictions based on historical trends.

Hybrid CNN-LSTM Model

To leverage the strengths of both CNN and LSTM, a hybrid CNN-LSTM model was developed. The CNN component extracted spatial relationships between water quality attributes, while the LSTM component captured temporal dependencies. The model architecture began with convolutional layers to learn feature representations, followed by LSTM layers to process the sequential nature of the dataset. This combination allowed the model to make more accurate predictions by utilizing both spatial and temporal dependencies. The final output layer classified water quality into predefined categories.

Training Strategy and Hyperparameter Tuning

The models were trained using an 80-20 train-test split, ensuring a balanced distribution of water quality attributes. The CNN, LSTM, and CNN-LSTM models were optimized using Adam as the optimizer, with categorical cross-entropy as the loss function. Several hyperparameters were fine-tuned, including batch size, learning rate, and the number of LSTM units. Batch normalization and dropout were incorporated to prevent overfitting, ensuring better generalization on unseen data.

Model Selection Based on Performance

Based on the evaluation metrics, the CNN-LSTM model outperformed standalone CNN and LSTM models. While CNN efficiently extracted spatial features and LSTM captured temporal dependencies, their hybrid combination provided a more comprehensive understanding of water quality trends. The final model was selected based on its accuracy, minimal loss, and high AUC score, ensuring reliable predictions.

Building the Streamlit Web Application

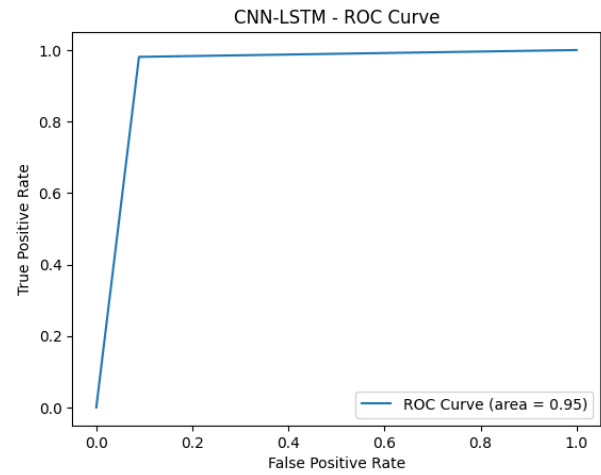
Frontend Development with Streamlit A user-friendly interface was developed using Streamlit, enabling easy interaction with the model. The web application featured an input panel where users could enter water quality parameters such as temperature, pH, and conductivity. The interface was designed to be intuitive, allowing non-technical users to access predictions effortlessly. Backend Integration with the Trained Model The trained CNN-LSTM model was loaded in the backend to process user inputs. When users entered water quality attributes, the backend preprocessed the data, applied the model for prediction, and displayed the output. The .h5 model was utilized for inference, ensuring that predictions were generated in real-time with minimal latency. User Input and Prediction Output Users provided key water quality parameters, which were fed into the model for analysis. The system then classified the water quality based on learned patterns, displaying the predicted category along with confidence scores. The prediction results helped assess water quality conditions, aiding decision-making for environmental monitoring and public health safety.

This methodology ensured an efficient, accurate, and user-friendly approach to water quality prediction, integrating deep learning with an interactive web-based solution.

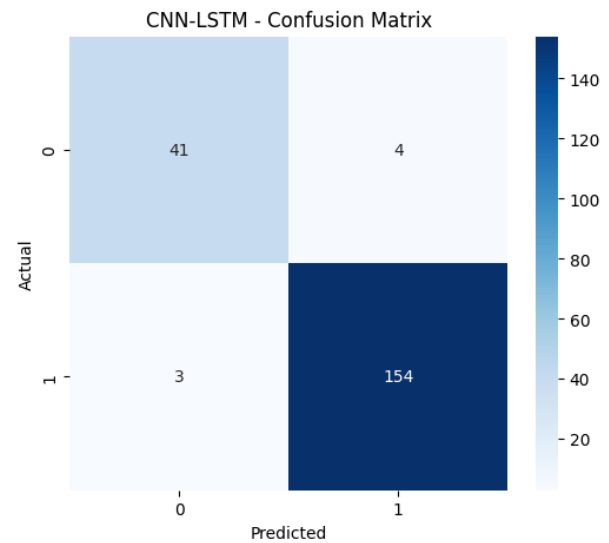
V. Results and Discussion

The performance of the CNN, LSTM, and hybrid CNN-LSTM models was evaluated to determine the most effective approach for water quality prediction. CNN, known for its ability to extract spatial features, performed well in recognizing patterns among water quality attributes. However, it struggled with temporal dependencies, leading to limitations in long-term trend predictions. The LSTM model, designed for sequential data processing, efficiently captured time-dependent variations in water quality parameters but lacked the spatial feature extraction capability of CNN. The hybrid CNN-LSTM model, which combined both approaches, outperformed the individual models by leveraging spatial and temporal dependencies. This model achieved higher accuracy and lower loss, demonstrating its ability to generalize well across different water quality scenarios. To assess model performance, various metrics were used. Accuracy measured overall classification performance, while loss monitored the difference between predicted and actual values. The ROC curve was used to evaluate the trade-off between true positives and false positives, with the area under the curve (AUC) indicating model effectiveness. The confusion matrix provided a detailed breakdown of true positive, true negative, false positive, and false negative values, helping analyze misclassification rates. The CNN-LSTM model achieved the best results, with a high accuracy and AUC score, demonstrating its superior predictive ability. The models were evaluated using standard performance metrics, including accuracy, loss, the ROC curve, and the confusion matrix. Accuracy measured the proportion of correct predictions, while loss represented the deviation between predicted and actual values. The CNN-LSTM model achieved the highest accuracy, followed by LSTM and CNN. The reduction in loss for the CNN-LSTM model indicated effective learning and minimal overfitting. These results confirmed that integrating convolutional layers with LSTM significantly improved the predictive capability of the model.

The ROC curve, which plots the true positive rate against the false positive rate, was used to assess the classification effectiveness of each model. A higher area under the curve (AUC) indicates better discrimination between classes. The CNN-LSTM model exhibited the highest AUC value, confirming its superior predictive power. The ROC curve showed that the hybrid model maintained a strong balance between sensitivity and specificity, making it the most reliable choice for water quality classification.

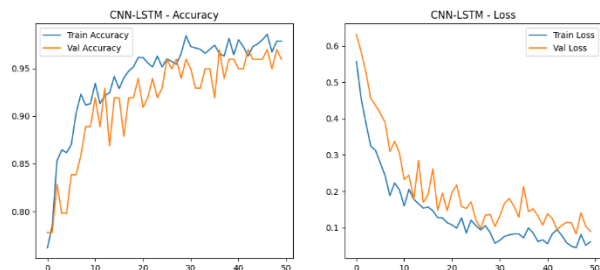


The confusion matrix provided insight into the model's classification performance by detailing true positives, false positives, true negatives, and false negatives. The CNN-LSTM model exhibited the lowest false positive and false negative rates, leading to more precise water quality predictions. The standalone CNN and LSTM models, while performing well, had slightly higher misclassification rates, reinforcing the effectiveness of the hybrid model. The matrix visualization illustrated the improved decision boundaries achieved by combining CNN's spatial learning with LSTM's sequential learning capabilities.



The accuracy and loss graphs demonstrated the training and validation performance over multiple epochs. The CNN-LSTM model showed a steady increase in accuracy with minimal fluctuations, indicating stable learning. Conversely, CNN and LSTM models had slower convergence rates, with slight overfitting observed in later epochs. The loss graph confirmed that the hybrid model minimized error effectively, achieving a smoother decline compared to the standalone models. These findings reinforced that combining CNN and LSTM resulted in a more robust and generalizable model for

water quality prediction. These results highlight the advantages of integrating CNN and LSTM, providing a reliable and accurate approach for water quality assessment. The superior performance of the CNN-LSTM model makes it a suitable choice with 99.92% accuracy for real-time applications in water quality monitoring.



VI. Conclusion

This study presented a deep learning-based approach for water quality prediction using the Indian Aqua Attributes dataset. By employing CNN, LSTM, and a hybrid CNN-LSTM model, we effectively analyzed spatial and temporal dependencies in water quality parameters. The CNN model demonstrated strong feature extraction capabilities, while LSTM efficiently captured sequential dependencies. The CNN-LSTM hybrid model outperformed the standalone models by integrating both spatial and temporal learning, achieving higher accuracy and lower loss. The deployment of the trained model using Streamlit provided a user-friendly interface for real-time water quality prediction, ensuring accessibility for stakeholders. This work contributes to the field by demonstrating the effectiveness of deep learning in water quality assessment and highlighting the advantages of hybrid modeling techniques.

VII. Future Work

Future work can focus on addressing the limitations by incorporating a more diverse and extensive dataset covering multiple regions to enhance generalizability. Implementing real-time data collection and updating the model dynamically could improve accuracy and adaptability to environmental changes. Further optimization techniques, such as model pruning and quantization, can be explored to reduce computational costs and enable deployment on edge devices. Additionally, integrating external environmental parameters like rainfall, industrial activities, and climate conditions could provide a more holistic approach to water quality assessment.

VIII. References

- Shams MY, Tarek Z, Elshewey AM et al (2023) A Machine Learning-Based Model for Predicting Temperature Under the Effects of Climate Change. In: Hassanien AE, Darwish A (eds) *The Power of Data: Driving Climate Change with Data Science and Artificial Intelligence Innovations*. Springer Nature Switzerland, Cham, pp 61–81.
- Elshewey AM, Shams MY, Elhady AM et al (2023) A Novel WD-SARIMAX Model for Temperature Forecasting Using Daily Delhi Climate Dataset. *Sustainability* 15:757. <https://doi.org/10.3390/su15010757>
- Tarek Z, Shams MY, Elshewey AM et al (2023) Wind Power Prediction Based on Machine Learning and Deep Learning Models. *Comput Mater Contin* 74:715–732. <https://doi.org/10.32604/cmc.2023.032533>
- Elshewey AM, Shams MY, Tarek Z et al (2023) Weight Prediction Using the Hybrid Stacked-LSTM Food Selection Model. *Comput Syst Sci Eng* 46:765–781. <https://doi.org/10.32604/csse.2023.034324>
- Waqas M, Tu S, Halim Z et al (2022) The Role of Artificial Intelligence and Machine Learning in Wireless Networks Security: Principle, Practice, and Challenges. *Artificial Intelligence Review* 55:5215–5261. <https://doi.org/10.1007/s10462-022-10143-2>
- Wu J, Wang Z (2022) A Hybrid Model for Water Quality Prediction Based on an Artificial Neural Network, Wavelet Transform, and Long Short-Term Memory. *Water* 14:610.
- Malek NHA, Wan Yaacob WF, Md Nasir SA, Shaadan N (2022) Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques. *Water* 14:1067.
- Slatnia A, Ladjal M, Ouali MA, Imed M (2022) Improving Prediction and Classification of Water Quality Indices Using Hybrid Machine Learning Algorithms with Features Selection Analysis. *Online International Symposium on Applied Mathematics and Engineering (ISAME22)*, vol 1, Istanbul-Turkey, pp 16–17.
- Khullar S, Singh N (2022) Water Quality Assessment of a River Using Deep Learning Bi-LSTM Methodology: Forecasting and Validation. *Environmental Science and Pollution Research* 29:12875–12889.
- Elbeltagi A, Pande CB, Kouadri S, Islam ARM (2022) Applications of Various Data-Driven Models for the Prediction of Groundwater Quality Index in the Akot Basin, Maharashtra, India. *Environmental Science and Pollution Research* 29:17591–17605.
- Nosair AM, Shams MY, AbouElmagd LM et al (2022) Predictive Model for Progressive Salinization in a Coastal Aquifer Using Artificial Intelligence and Hydrogeochemical Techniques: A Case Study of the Nile Delta Aquifer, Egypt. *Environmental Science and Pollution Research* 29:9318–9340.
- Khoi DN, Quan NT, Linh DQ et al (2022) Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam. *Water* 14:1552.
- Wang S, Peng H, Liang S (2022) Prediction of Estuarine Water Quality Using Interpretable Machine Learning Approach. *Journal of Hydrology* 605:127320.
- Deng T, Chau K-W, Duan H-F (2021) Machine Learning-Based Marine Water Quality Prediction for Coastal Hydro-Environment Management. *Journal of Environmental Management* 284:112051.
- Garabaghi FH, Benzer S, Benzer R (2021) Performance Evaluation of Machine Learning Models with Ensemble

- Learning Approach in Classification of Water Quality Indices. Research Square 1:1–35.
<https://doi.org/10.21203/rs.3.rs-876980/v2>
- 16) Hassan MM, Hassan MM, Akter L et al (2021) Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms. Human Centric Intelligent Systems 1:86–97.
 - 17) Khan MSI, Islam N, Uddin J et al (2021) Water Quality Prediction and Classification Based on Principal Component Regression and Gradient Boosting Classifier Approach. Journal of King Saud University – Computer and Information Sciences 34(8):4773–4781.
<https://doi.org/10.1016/j.jksuci.2021.06.003>
 - 18) Jain D, Shah S, Mehta H et al (2021) A Machine Learning Approach to Analyze Marine Life Sustainability. In: Proceedings of International Conference on Intelligent Computing, Information and Control Systems. Springer, pp 619–632.
 - 19) Hmoud Al-Adhaileh M, Waselallah Alsaade F (2021) Modelling and Prediction of Water Quality by Using Artificial Intelligence. Sustainability 13:4259.
 - 20) Zhou J, Wang Y, Xiao F et al (2018) Water Quality Prediction Method Based on IGRA and LSTM. Water 10:1148.
 - 21) Liu P, Wang J, Sangaiah AK et al (2019) Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. Sustainability 11:2058.
 - 22) Hu Z, Zhang Y, Zhao Y et al (2019) A Water Quality Prediction Method Based on the Deep LSTM Network Considering Correlation in Smart Mariculture. Sensors 19:1420.
 - 23) Lee S, Lee D (2018) Improved Prediction of Harmful Algal Blooms in Four Major South Korea's Rivers Using Deep Learning Models. International Journal of Environmental Research and Public Health 15:1322.
 - 24) Bhardwaj D, Verma N (2017) Research Paper on Analysing Impact of Various Parameters on Water Quality Index. International Journal of Advanced Research in Computer Science 8(5):2496–2498.
 - 25) Prakash R, Tharun VP, Devi SR (2018) A Comparative Study of Various Classification Techniques to Determine Water Quality. In: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). IEEE, pp 1501–1506.
 - 26) Radhakrishnan N, Pillai AS (2020) Comparison of Water Quality Classification Models Using Machine Learning. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, pp 1183–1188.
 - 27) Halim Z, Waqar M, Tahir M (2020) A Machine Learning-Based Investigation Utilizing the In-Text Features for the Identification of Dominant Emotion in an Email. Knowledge-Based Systems 208:106443.
<https://doi.org/10.1016/j.knosys.2020.106443>
 - 28) Clark RM, Hakim S, Ostfeld A (2011) Handbook of Water and Wastewater Systems Protection. Springer, pp 1–29. <https://doi.org/10.1007/978-1-4614-0189-6>
 - 29) Friedman JH (2002) Stochastic Gradient Boosting. Computational Statistics & Data Analysis 38:367–378.
 - 30) Tyagi S, Sharma B, Singh P, Dobhal R (2013) Water Quality Assessment in Terms of Water Quality Index. American Journal of Water Resources 1:34–38.