IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SMART INDUSTRIAL REAL-TIME WATER QUALITY MONITORING AND PREDICTION USING MACHINE LEARNING

[1]M. Padma Sree, [2]G. Srinivasa Rao, [3]E. Lakshmi Prasanna, [4]B. Kalyani,

[1]Assistant Professor, Department of Electronics and Communication Engineering, Bapatla Women's Engineering College, Bapatla, A.P, India-522101

[2]Professor, Department of Electronics and Communication Engineering, Bapatla Women's Engineering College, Bapatla,
A.P, India-522101

[3,4]UG Student, Department of Electronics and Communication Engineering, Bapatla Women's Engineering College, Bapatla, A.P, India-522101

Abstract: This paper proposes a Smart Industrial Real-Time Water Quality Monitoring and Prediction System that integrates the Internet of Things (IoT) and machine learning to improve industrial water management and environmental safety. The system monitors key water parameters -Total Dissolved Solids (TDS), ammonia concentration, pH, turbidity, and temperature via dedicated sensors connected to an Arduino microcontroller, with data transmitted to the ThingSpeak cloud platform using an ESP8266 Wi-Fi module. Real-time alerts are facilitated through an on-site buzzer and a Telegram bot to notify users of abnormal conditions. For predictive analytics, the system employs machine learning algorithms such as Random Forest, Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), XG-Boost, Logistic Regression, and Decision Tree to classify water quality status based on historical data. This unified framework provides a scalable and cost-effective solution for continuous monitoring, early warning, and data-driven decision-making across industries such as manufacturing, agriculture, and wastewater treatment.

***Index Terms* -** Water Quality Monitoring, Internet of Things (IoT), Machine Learning, Real-Time Prediction, Industrial Water Management, Environmental Safety.

1.INTRODUCTION

Water is one of the most vital natural resources, essential not only for sustaining life but also for driving industrial operations and supporting agricultural activities. However, with rapid industrialization, the increasing discharge of untreated waste, and chemical runoff, water quality is deteriorating at an alarming rate. This poses a significant threat to environmental health, human well-being, and compliance with stringent regulatory frameworks.

Traditional water quality monitoring methods, which rely heavily on manual sampling and laboratory-based analysis, are inherently time-consuming, costly, and unsuitable for continuous real-time monitoring. These limitations necessitate a paradigm shift toward intelligent, automated, and scalable solutions for water management, especially in industrial settings where the risk of contamination is high and demands prompt action.

This paper presents the design and implementation of a Smart Industrial Real-Time Water Quality Monitoring and Prediction System that leverages the Internet of Things (IoT), cloud computing, and machine learning to transform the assessment and management of water quality through innovative approaches. The system integrates a suite of sensors - including Total Dissolved Solids (TDS), pH, turbidity, temperature, and ammonia sensors - connected to an Arduino microcontroller. Data collected from these sensors is transmitted in real-time via the ESP8266 Wi-Fi module to the ThingSpeak cloud platform, enabling remote monitoring and visualization.

To ensure proactive response capabilities, the system features both a local buzzer for on-site alerts and a Telegram bot that delivers immediate notifications to stakeholders upon detecting abnormal conditions. Furthermore, the collected data is analyzed using a range of machine learning algorithms, including Random Forest, Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), XG-Boost, Logistic Regression, and Decision Tree models. The predictive models categorize water quality conditions into 'normal' or 'abnormal,' enabling industries to respond proactively and make data-driven decisions. This intelligent and automated framework not only addresses the limitations of conventional methods but also offers a cost-effective, scalable, and sustainable approach to water quality management. Its applicability spans diverse industrial sectors, including manufacturing, wastewater treatment, agriculture, and food processing, marking a significant step toward smart environmental monitoring and control.

2. RESEARCH METHODOLOGY

Recent advancements in IoT and machine learning (ML) have revolutionized water quality monitoring systems.

Forhad et al. [1] proposed an IoT-based real-time water quality monitoring system for water treatment plants using PLCs and a range of sensors to collect data on parameters like pH, DO, TDS, and temperature. The system was designed for centralized monitoring, cloud storage, and historical data analysis, with high reliability and low power consumption.

Wiryasaputra et al. [2] developed a potable water quality monitoring and prediction system using Arduino, NB-IoT, and machine learning. The collected data was analyzed using classifiers including Decision Trees, Gradient Boosting, and SVM to predict water potability. The system featured real-time alerts and achieved high classification accuracy, showing the potential of combining low-cost IoT and AI.

Essamlali et al. [3] conducted a comprehensive review emphasizing the integration of ML with wireless IoT technologies for water quality monitoring. Their study highlighted how supervised and unsupervised learning models enhance real-time data interpretation and predictive capabilities across various water systems, including industrial, agricultural, and municipal settings.

Ashwini et al. [4] proposed a cost-effective system using multiple sensors and neural networks to predict contamination and alert users before water quality degrades. Such smart systems enhance public health safety by providing early warnings and are scalable for both residential and industrial applications.

3. PROPOSED METHOD

3.1. IoT Sensor Packages for Water Quality Monitoring

In the proposed system, a suite of IoT-enabled sensors is deployed to continuously monitor key water quality parameters. These sensors provide real-time data on critical indicators such as pH, turbidity, temperature, total dissolved solids (TDS), and the presence of toxic substances, forming the foundational layer of the smart monitoring architecture. Figure 1 shows the assembly of various sensors used in the experimental setup.

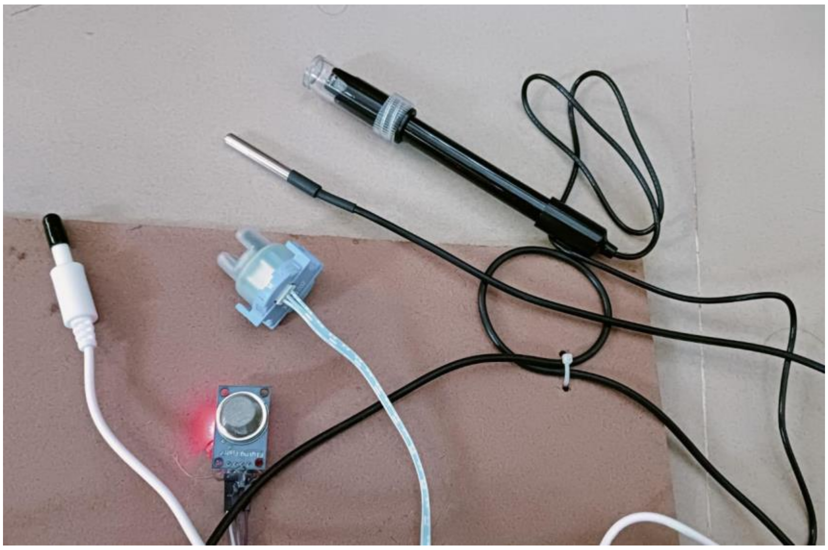


Figure 1. IoT Sensor Assembly for Real-Time Water Quality Monitoring

3.1.1 Total Dissolved Solids (TDS) Sensor

The TDS sensor estimates the concentration of dissolved solids in water by measuring its electrical conductivity, which correlates with the presence of ions from inorganic salts and contaminants such as calcium, sodium, and heavy metals. The sensor typically consists of three terminals: VCC, GND, and an analog data output connected to a microcontroller input. In this system, the TDS sensor serves as a critical indicator of water salinity and contamination. While effective for real-time applications, environmental variables, and sensor calibration can influence the accuracy. Typical error rates remain within acceptable margins for industrial monitoring.

3.1.2 pH Sensor

pH sensors are used to measure the hydrogen ion concentration, which determines the water's acidity or alkalinity on a scale from 0 to 14. A value below 7 denotes acidity, while values above 7 represent alkalinity, with 7 being neutral. The sensor consists of a reference electrode and a measuring electrode, which together produce a potential difference corresponding to the hydrogen ion concentration. The analog output is processed via the microcontroller to monitor pH changes in real-time. For industrial applications, maintaining a pH within the World Health Organization (WHO) recommended range of 6.5 to 8.5 is crucial for both environmental safety and equipment longevity.

3.1.3 Turbidity Sensor

Turbidity sensors determine water clarity by detecting the concentration of suspended particles. Utilizing the principle of light scattering, they emit an infrared beam and measure the amount of light transmitted through the sample. The intensity of scattered light increases proportionally with the level of turbidity in the water. This real-time monitoring helps identify pollutants or process anomalies such as industrial discharge or biofouling. The sensor operates with a simple analog output connected to the microcontroller and is essential for the early detection of water quality degradation.

3.1.4 Temperature Sensor

Temperature significantly impacts chemical reaction rates and biological processes in water systems. In this work, a waterproof digital temperature sensor (e.g., DS18B20) is employed, offering an operating range from −55 °C to +125 °C with a resolution of 0.5 °C. The sensor outputs a digital signal over a single data line, allowing for noise-immune long-distance transmission. Monitoring temperature is vital because higher temperatures can accelerate bacterial growth and affect sensor readings for other parameters like pH and TDS.

3.1.5 Ammonia Sensor

Ammonia's presence in water can indicate industrial waste, fertilizer runoff, or decaying biological material. The ammonia sensor detects the concentration of NH_3 through electrochemical sensing techniques, providing an analog output proportional to the gas concentration in water. Even low levels of ammonia can be harmful to aquatic life and affect the chemical balance in treatment processes. The integration of this sensor enables early warning and response mechanisms in industrial effluent monitoring.

3.1.6 ESP8266 Wi-Fi Module

ESP8266 is a cost-effective microchip that combines Wi-Fi capability with a complete TCP/IP stack and integrated microcontroller features. It facilitates wireless data transmission from the Arduino to the ThingSpeak IoT cloud platform. By operating in both Access Point and Station modes, the module ensures seamless real-time monitoring from remote locations. It operates at 2.4 GHz and supports 802.11 b/g/n standards, allowing efficient data handling even in resource-constrained networks. The module's compact form factor and low power consumption make it suitable for continuous water quality monitoring applications.

3.2. Machine Learning Models

The proposed industrial water quality monitoring system utilizes machine learning algorithms to analyze sensor data and predict the water condition (normal or abnormal) in real time, enabling timely alerts and preventive actions without relying on visual workflow representation.

3.2.1 Decision Tree (DT)

Decision Tree (DT) classifiers are used to construct hierarchical models that classify instances by sorting them based on feature values. In the proposed system, Decision Tree (DT) is used to predict the condition of industrial water by recursively splitting the sensor dataset based on entropy and information gain. The DT model is simple yet interpretable, making it suitable for rapid rule-based decisions. It demonstrates high classification accuracy with minimal computational complexity, particularly effective on small-to-moderate-sized datasets collected from IoT sensors.

3.2.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful supervised learning algorithm that constructs a hyperplane in high-dimensional space to separate classes (normal and abnormal water quality) with maximum margin. In the proposed system, SVM is employed due to its robustness to high-dimensional sensor data and its generalization capability. While SVM is sensitive to kernel selection and may experience computational challenges with unbalanced or large datasets, its precision, and ability to handle noisy measurements make it valuable for industrial water quality prediction when properly tuned.

3.2.3 Random Forest (RF)

Random Forest (RF) is an ensemble learning technique that constructs several decision trees during the training process. It then determines the final prediction by selecting the class that appears most frequently among the predictions of all individual trees. The model enhances predictive performance and reduces overfitting, making it ideal for real-world industrial datasets with missing or noisy sensor values. RF's capability to process high-dimensional features without normalization is particularly advantageous in water quality monitoring, where sensor outputs vary across scales and units.

3.2.4 K-Nearest Neighbors (KNN)

KNN is a non-parametric classification algorithm that classifies instances based on the 'k' nearest neighbors in the training dataset. In the proposed system, KNN assigns a water quality classification to a new instance by taking a majority vote from its closest neighbors in the feature space. Its simplicity and efficiency in low-complexity scenarios make it useful for quick assessments. However, its performance is influenced by the value of ‘k’, and it is sensitive to feature scaling.

3.2.5 Naive Bayes (NB)

Naive Bayes is a probabilistic classifier that utilizes Bayes' Theorem and assumes the independence of predictors to make predictions. Despite its simplicity, it is computationally efficient and performs well in real-time monitoring environments. For water quality prediction, NB offers rapid classification and is useful for datasets with categorical or discretized sensor readings. However, its assumption of feature independence can limit its performance when sensor parameters are correlated.

3.2.6 XG-Boost

XG-Boost is a gradient boosting framework that builds models sequentially to minimize classification error using a second-order Taylor approximation. It enhances model accuracy and handles missing data efficiently. In the proposed system, XG-Boost leverages historical sensor data to refine predictions and improve alert precision. Its built-in regularization reduces overfitting, making it suitable for complex industrial datasets.

3.2.7 Logistic Regression (LR)

Logistic Regression (LR) is employed to predict a binary outcome, such as normal or abnormal water quality, by analyzing a set of continuous input features.LR offers high interpretability and performs well with linearly separable classes. Though less powerful than ensemble methods, LR provides baseline performance and is useful for validating results from more complex models.

Table 1 presents the dataset extracted for experimental evaluation

Table 1: Dataset Extracted for Experimental Evaluation

TEMP	TDS	PH	TB	AMMONIA	Output	
15.2	50	6.5	10	0.1	0	
20.1	95	7.4	28	1.8	0	
11.6	140	8.3	46	4	0	
9.3	170	7.5	58	5.5	0	
21.9	190	7.4	66	6.5	0	
2.5	195	2.1	68	0	1	
50.2	225	5.3	80	0.06	1	
-7	275	10.5	100	50	1	
70.3	340	1.2	126	500	1	

4. BLOCK DIAGRAM

Figure 1 illustrates the block diagram of the proposed method for real-time monitoring of industrial water quality.

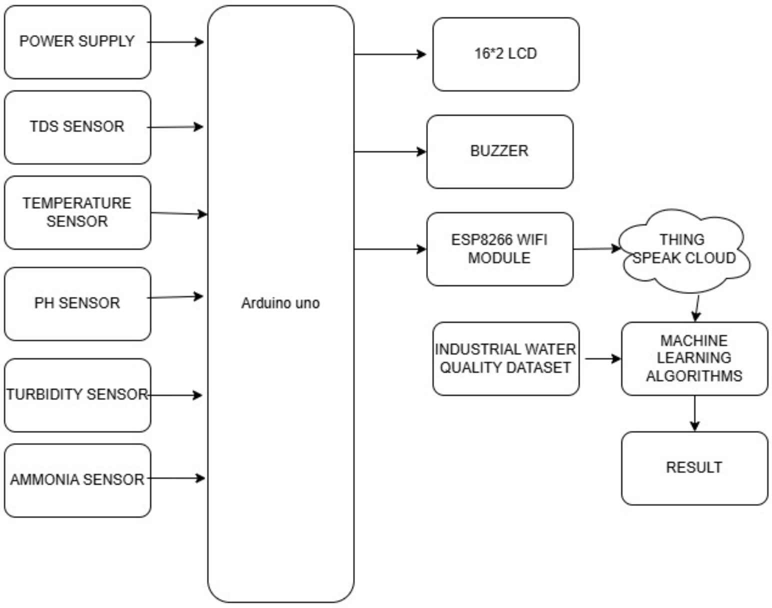


Figure 2. Block Diagram of the industrial-water quality monitoring system

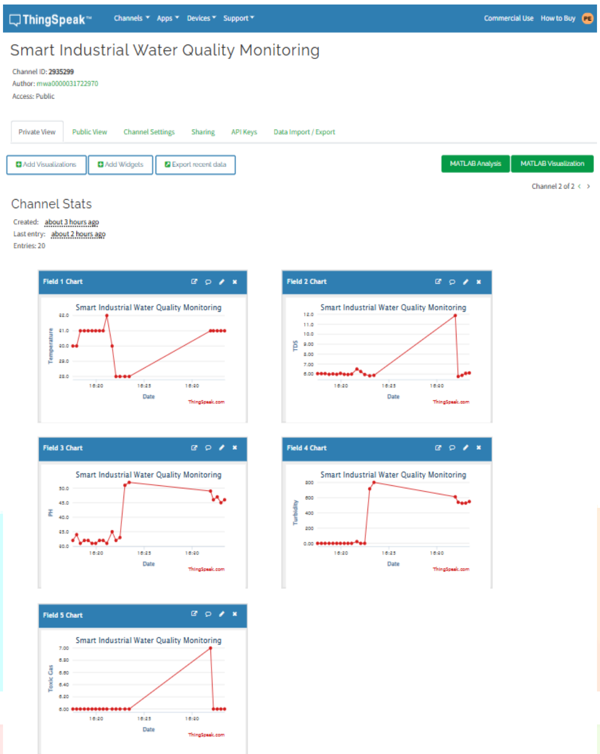
5. IMPLEMENTATION

- Multiple water quality sensors (TDS, pH, turbidity, ammonia, and temperature) are connected to an Arduino Uno, which collects real-time data from the water source and processes it for transmission.
- The ESP8266 Wi-Fi module is used to send the sensor data wirelessly to the ThingSpeak cloud platform, where the data is stored and visualized in real time through dynamic charts.
- A buzzer is activated for local alerts when abnormal values are detected, while a Telegram bot is configured to send immediate notifications to users for remote awareness and quick response.
- Historical data from ThingSpeak is used to train various machine learning models in Python, including SVM, Random Forest, KNN, Naive Bayes, Logistic Regression, XGBoost, and Decision Tree, to classify water quality conditions.
- The trained models predict incoming water quality data as normal or abnormal, enabling early detection of contamination and providing remote access to results through ThingSpeak and Telegram for decision-making.

6. RESULTS AND DISCUSSION

The proposed IoT and machine learning-based water quality monitoring system was successfully implemented and tested in real-time conditions. Sensor modules effectively captured key water parameters, including pH, turbidity, temperature, TDS, and Ammonia concentration, with minimal deviation from standard calibration values. The system reliably transmitted data to the cloud platform, enabling continuous visualization and storage. Among the applied machine learning models, Decision Tree and Random Forest classifiers demonstrated the highest prediction accuracy for water quality classification. The Telegram bot integration enabled instant alerts for abnormal readings, while the buzzer module and LCD provided immediate on-site notifications. These results validate the efficiency, accuracy, and real-time responsiveness of the system, making it a practical solution for industrial and environmental water quality monitoring. The results obtained from the experimental evaluation are illustrated in Figures 3 and 4.

Fig 3. (a) Buzzer Module Used for On-Site Alerts (b) LCD Display Module



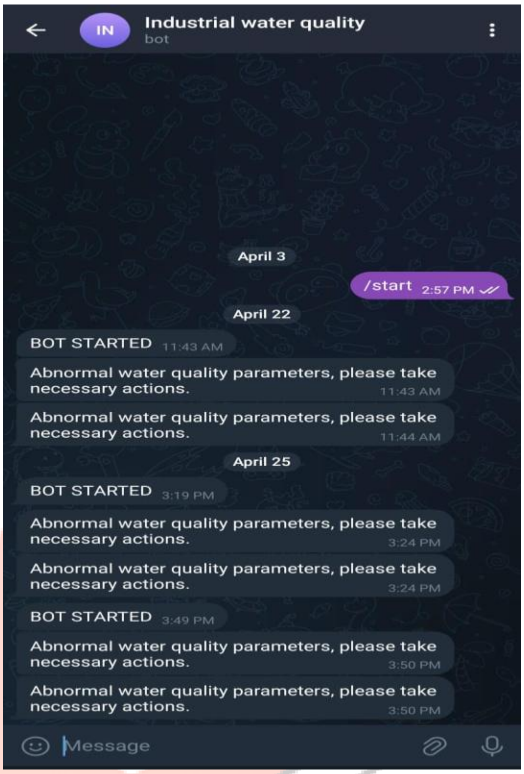


Fig 4. (a) Visualization on ThingSpeak Platform (b) Telegram Bot Alerts for Abnormal Conditions

7. CONCLUSION AND FUTURE WORK

The proposed Smart Industrial Real-Time Water Quality Monitoring and Prediction System effectively integrates IoT, cloud computing, and machine learning to provide an intelligent and automated solution for monitoring and managing industrial water quality. By utilizing sensors to collect data on key parameters and employing multiple machine learning algorithms for accurate prediction, the system ensures timely alerts and proactive intervention through Telegram notifications and buzzer alarms. This enhances environmental safety, regulatory compliance, and operational efficiency. In the future, the system can be enhanced by incorporating mobile or web-based dashboards for better user interaction, integrating automated water treatment mechanisms for immediate response, deploying solar-powered units to support remote or off-grid locations, and utilizing advanced deep learning models to achieve higher prediction accuracy. Its scalability makes it suitable for applications in smart cities, rural monitoring programs, and large-scale industrial setups.

REFERENCES

[1] H. M. Forhad *et al.*, "IoT based real-time water quality monitoring system in water treatment plants (WTPs)," *Heliyon*, vol. 10, 2024, Art. no. e40746. [Online]. Available: https://doi.org/10.1016/j.heliyon.2024.e40746.

[2] R. Wiryasaputra, C.-Y. Huang, Y.-J. Lin, and C.-T. Yang, "An IoT Real-Time Potable Water Quality Monitoring and Prediction Model Based on Cloud Computing Architecture," *Sensors*, vol. 24, no. 4, Art. no. 1180, Feb. 2024. [Online]. Available: https://doi.org/10.3390/s24041180.

[3] I. Essamlali, H. Nhaila, and M. El Khaili, "Advances in machine learning and IoT for water quality monitoring: A comprehensive review," *Heliyon*, vol. 10, no. 6, p. e27920, Mar. 2024. [Online]. Available: https://doi.org/10.1016/j.heliyon.2024.e27920.

[4] C. Ashwini, U. P. Singh, E. Pawar, and Shristi, "Water Quality Monitoring Using Machine Learning and IoT," *International Journal of Scientific & Technology Research*, vol. 8, no. 10, pp. 1046–1048, Oct. 2019. [Online]. Available: http://www.ijstr.org/final-print/oct2019/Water-Quality-Monitoring-Using-Machine-Learning-And-Iot.pdf:contentReference[oaicite:0]{index=0}.

[5] A. R. Dhruba *et al.*, "IoT-Based Water Quality Assessment System for Industrial Waste Water Healthcare Perspective," *arXiv preprint arXiv:2304.06491*, Mar. 2023. [Online]. Available: https://arxiv.org/abs/2304.06491.

[6] D. B. Krklješ *et al.*, "Multiparameter Water Quality Monitoring System for Continuous Monitoring of Fresh Waters," *arXiv preprint arXiv:2307.11630*, Jul. 2023. [Online]. Available: https://arxiv.org/abs/2307.11630.

[7] M. J. T. Kathen *et al.*, "AquaFeL-PSO: A Monitoring System for Water Resources using Autonomous Surface Vehicles based on Multimodal PSO and Federated Learning," *arXiv preprint arXiv:2211.15217*, Nov. 2022. [Online]. Available: https://arxiv.org/abs/2211.15217.

[8] S. Y. Luis *et al.*, "Deep Reinforcement Multi-Agent Learning Framework for Information Gathering with Local Gaussian Processes for Water Monitoring," *arXiv preprint arXiv:2401.04631*, Jan. 2024. [Online]. Available: https://arxiv.org/abs/2401.04631.

[9] H. S. Maghdid *et al.*, "A Novel Poisoned Water Detection Method Using Smartphone Embedded Wi-Fi Technology and Machine Learning Algorithms," *arXiv preprint arXiv:2302.07153*, Feb. 2023. [Online]. Available: https://arxiv.org/abs/2302.07153.

[10] D. P. M. Dinesh *et al.*, "IoT Based Real-Time River Water Quality Monitoring and Control System," *E3S Web of Conferences*, vol. 399, 2023, Art. no. 04013. [Online]. Available: https://doi.org/10.1051/e3sconf/202339904013.

[11] L. C. P. da Silva, J. C. de Souza, and L. A. Villas, "Smart Water Management for Smart Cities: A Review," *IEEE Latin America Transactions*, vol. 17, no. 9, pp. 1478–1484, Sept. 2019. doi: 10.1109/TLA.2019.8932337.

[12] A. Hawari, M. Haque, and M. Muhammad, "Development of an IoT-based multi-level system for real-time water quality monitoring," *Environmental Science and Pollution Research*, vol. 31, no. 5, pp. 4567–4578, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s43832-024-00092-y SpringerLink.

[13] J. V. Abrajano, K. A. Botangen, J. Nabua, J. Apanay, and C. F. Peña, "IoT-Based Water Quality Monitoring System in Philippine Off-Grid Communities," *arXiv preprint arXiv:2410.14619*, Oct. 2024. [Online]. Available: https://arxiv.org/abs/2410.14619arXiv.

[14] F. Razzano, F. Mauro, P. Di Stasio, G. Meoni, M. Esposito, G. S. Ullo, "Monitoring water contaminants in coastal areas through ML algorithms leveraging atmospherically corrected Sentinel-2 data," *arXiv preprint arXiv:2401.03792*, Jan. 2024. [Online]. Available: https://arxiv.org/abs/2401.03792.