

IoT-Based Intelligent Water Quality Monitoring and Automated Purification Using Machine Learning

Vijayendra Rane
Student

Dept of Computer Science
AISSMS Institute of
Information Technology
Pune, India

Suraj Pawar
Student

Dept of Computer Science
AISSMS Institute of
Information Technology
Pune, India

Pranit Rindhe
Student

Dept of Computer Science
AISSMS Institute of
Information Technology
Pune, India

Kishor S. Wagh
Mentor

Department of Computer Science
AISSMS Institute of Information Technology
Pune, India

Bhavana Zambare
Mentor

Department of Computer Science
AISSMS Institute of Information Technology
Pune, India

Abstract—The need for continuous and uninterrupted availability of pure and clean water entails quick monitoring of water quality degradation and immediate remediation of the quality. This paper proposes a smart system for monitoring and purifying the quality of water by incorporating the Internet of Things (IoT) and machine learning techniques. An ESP32 microcontroller-based smart water quality monitoring system, facilitated by low-cost pH, turbidity, and total dissolved solids sensors, is implemented for acquiring real-time water quality information. The acquired information can be remotely monitored by the cloud platform. A machine learning model is incorporated and trained with the acquired data for classifying the water quality and predicting the water quality in advance. According to the result of the machine learning prediction model, a water purification system can be remotely implemented through a relay. The results of the experimental evaluation of the smart water quality monitoring and purification system, implemented with the real-time results of the sensors, show a high accuracy of 92.4 percent. This indicates the reliability of implementing the proposed water quality monitoring and purification system.

Index Terms—Internet of Things (IoT), Water Quality Monitoring, ESP32, Machine Learning, Predictive Analytics, Smart Water Systems

I. INTRODUCTION

Water, which is clean and safe, is one of the prime commodities required for the survival of humanity and processes related to industry and environment upkeep. As it is genuine, it is agreed that urbanization, industrial waste, and improper waste management practices add to the water pollution dilemma to the greatest extent possible. According to the World Health Organization, it is observed that bad water quality results in waterborne diseases, especially in developing nations.

Generally, the normal method for water quality monitoring involves manual sampling, and then the data has to be sent to the laboratory for analysis, which is costly, cumbersome, and cannot be used for real-time management. It lacks the ability to respond to sudden contamination.

The recent developments in IoT technology provide a cost-effective system for real-time monitoring of environmental parameters. The suggested system based on IoT technology offers a combination of sensor technology, microcontrollers, and cloud computing for continuous data acquisition, accessibility, and representation in a graphical format in real time, considering major water quality parameters, including pH, turbidity, and TDS.

However, most of the IoT-based water quality monitoring systems developed so far can serve only data visualization and notification purposes. They cannot predict the future status of water quality. Lacking the ability of prediction means that proactive and automatic control does not exist in water management.

In order to avoid these limitations, the researcher proposes an intelligent water quality monitoring system based on concepts of the Internet of Things and machine learning. The suggested system not only monitors the water quality but also predicts the time period after which the water quality will turn unfavorable. Then, based on the result, the water purification process will take place.

Unlike the traditional threshold-based IoT water monitoring systems, this proposed system predicts time in advance for unsafe water conditions and can send commands autonomously to any linked purification processes.

The major contributions of this paper are as follows:

- Design and development of a low-cost IoT-based water quality monitoring system using pH, turbidity, and TDS sensors.
- Real-time data transmission and visualization using an open-source cloud platform.
- Development of a machine learning model to predict water quality degradation over time.
- Automated decision-making and control of a water purification system based on predictive analysis.

II. RELATED WORK

Recent advancements attained in the field of IoT have made provisions for real-time monitoring of the quality of water through IoT sensing technologies. Different systems based on IoT concepts are proposed in water-quality monitoring, which are related to parameters like pH, turbidity, temperature, and TDS, showing sensor data visualization on a cloud platform citeref1, ref2.

Other options for the implementation of smart water quality monitoring systems are microcontroller-based systems, such as Arduino, ESP8266, and ESP32. It is fair to comment that these microcontroller-based systems demonstrate the capabilities of real-time data acquisition and remote monitoring very successfully. However, these systems are often dependent on set threshold limits to ascertain safety in the water. citeref3.

To avoid the above-mentioned disadvantage, research papers published recently employ machine learning methods for the assessment and prediction of water quality. Machine learning-based SVM, ANN, and RF classifier methods were utilized to categorize the quality of water based on the historical datasets citeref4, ref5. Though the accuracy of the model is fairly high, they are being trained and validated by the offline datasets rather than being integrated with a real-time environment.

Certain researchers have tried combining IoT sensing, cloud analytics, and machine learning to realize intelligent water management systems citeref6,ref7. Although these systems allow for automated alerts and decisions, most of them lack closed-loop control mechanisms that automatically trigger corrective measures such as automatic water purification. Moreover, in the context of a water quality monitoring system, limited studies have focused on time-to-unsafe conditions.

In comparison, the proposed system implements real-time IoT-based water quality monitoring with machine learning-driven prediction and automated purifier control. Thus, the system will be able to foresee unsafe water conditions in advance of threshold violations, which allows for proactive autonomous water quality management, addressing some of the major limitations identified within existing literature.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. System Architecture

The proposed real-time water quality monitoring, predictive analysis, and purification system would follow a multi-tier architecture to make sure that the system is scalable, reliable, and efficiently processes data.

The Water Quality Monitor's Sensors will continually sense the water body or tank by measuring all the significant parameters of the water such as turbidity, pH value, and TDS, etc., and collect the required data that will further be processed by the ESP32 Microcontroller and application running in the cloud. The system will use the collected and processed data with the predicted result of the machine learning model to operate the purification system.

The suggested water quality monitoring system architectural design has four layers of functionality, namely:

- **Sensing Layer** The Sensor Layer consists of different sensors to monitor water quality, including pH, turbidity, and TDS. These sensors are primarily used to determine if there is any level of contamination in the water by providing real-time voltage signals proportional to the parameters of the water being monitored. All of these sensors will continue to monitor their respective parameters on a continuous basis.
- **Processing Layer** The processing layer is constructed with an ESP32 Microcontroller. The ESP32 receives and converts analog signals from the sensors into digital values using the built-in Analog-to-Digital Converter (ADC) of the microcontroller. An additional step that occurs at the processing layer is basic data pre-processing, which includes signal stabilization and data formatting. An ESP32 was selected based on its low power consumption, high processing capabilities, and integrated Wi-Fi module.
- **Cloud and Analytics Layer** Using Wi-Fi communication, the processed data from the sensors are wirelessly transmitted to a cloud-based IoT platform where both real-time and historical data are stored with options for visualization via dashboards, graphs and charts. This layer also serves as the interface between the IoT system and the machine learning model.
- **Control Layer** The control layer of the IoT system receives the predicted results from the machine learning model and generates a control signal to turn on the water purification system if the predicted result indicates that the water quality is not safe. This capability allows the water purification system to be purifying water continuously and with less manual intervention.

B. Methodology

To ensure proper monitoring, prediction, and automated control of water quality, the proposed system adopts a systematic approach that combines real-time monitoring, cloud-based analysis, machine learning-based prediction, and automated control systems.

- **Data Acquisition**
The water quality parameters that can be acquired by the proposed system are pH, turbidity, and TDS (Total Dissolved Solids). These parameters can be acquired by using sensors that measure the quality of water in the water body or in the storage tank. The sensors produce analog signals.
- **Dataset Description** The dataset used in the research work was generated by collecting continuous real-time data from the deployed IoT system. The data collected from the sensors was uploaded to the cloud server at a specified interval. Each record in the dataset contains the pH level, the turbidity measured in NTU, the total dissolved solids measured in ppm, and the timestamp. The timestamp can be used to track the trend of water quality degradation.

IoT-Based Water Quality Monitoring System

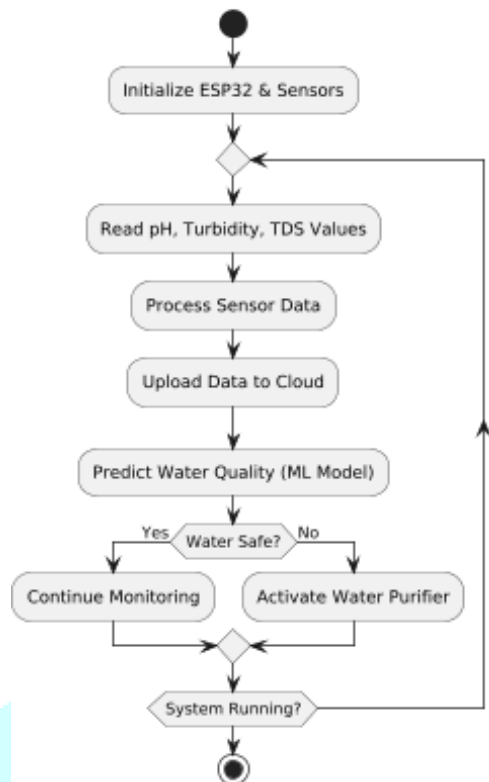


Fig. 1. Flowchart of System

The collected data pertains to the safe and unsafe water conditions.

The data set was categorized into two classes, with categories labeled Safe and Unsafe based on the limits of water quality established by the World Health Organization (WHO). The entire data set was split into an 80:20 proportion for machine learning model development and testing.

The dataset is comprised of around 2000 samples collected over a few days.

- **Data Processing and Transmission**

The ESP32 microcontroller reads the analog signals from the sensors and then converts them into digital form using the in-built analog-to-digital converter. Basic preprocessing steps such as noise removal and normalization are performed to enhance accuracy and reliability.

The preprocessed sensor data is sent to a cloud-based system using Wi-Fi connectivity at regular time intervals for real-time monitoring purposes.

- **Data Visualization and Storage**

The cloud system allows real-time graphical representation of water quality parameters using dashboards, charts, and graphical representations. All preprocessed sensor data is stored in a cloud database for historical analysis and training of machine learning models.

- **Machine Learning Prediction Based on real-time and historical data,** a machine learning model is used to predict whether the water quality is safe or potentially unsafe within a specified time period. The prediction problem involves classifying water quality status and predicting degradation trends. The model is assessed using standard evaluation criteria. Model performance is evaluated using standard metrics, as presented in Table II.

- **Automated Action and Control**

In the event that the water quality status is deemed to be in an unsafe state as per the predictions made by the system, the system is set up to automatically switch on the water purification device using a control relay. It remains in operation until a safe water quality status is attained and automatically shuts down.

IV. MATHEMATICAL MODELING

A. Sensor Data Representation

$$V_i = f(S_i) \quad (1)$$

Where S_i represents the sensor reading and V_i represents the corresponding voltage.

B. Parameter Conversion Equations

- **pH Calculation**

$$pH = a.V_{pH} + b \quad (2)$$

- **TDS Calculation**

$$TDS = k.V_{TDS} \quad (3)$$

- **Turbidity Calculation**

$$Turbidity = c.V_{tur2} + d.V_{tur} + e \quad (4)$$

where a,b,c,d,e,k are used to denote the calibration constants.

C. Feature Vector Formation

$$X = [pH, TDS, Turbidity, t] \quad (5)$$

where t denotes time index

The derived feature vector is used as input to the machine learning model described in Section VI, and its predictive performance is evaluated using the metrics presented in Section VIII.

V. PERFORMANCE METRICS AND MACHINE LEARNING MODEL

A. Classification Metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (9)$$

B. Regression Metrics

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

The classification performance of the proposed Random Forest model was evaluated using standard metrics including Accuracy, Precision, Recall, and F1-score. These metrics were computed using the confusion matrix obtained from the test dataset and are reported in Table II.

C. Problem Formulation

The problem of predicting water quality can be defined as a supervised learning issue that involves classification and regression objectives. For a set of water quality sensor readings that are obtained over a period of time, it is necessary to establish the current safety status of the water while also predicting any degradation trends.

- **Binary Classification**, wherein the instances of the data set are either classified as *Safe* or *Unsafe* depending upon the water quality standards set by the World Health Organization.

D. Feature Engineering and Data Preparation

The machine learning algorithm effectively uses the provided features for the sensor. Before the machine learning algorithm is trained, preprocessing is done on the raw sensor data to improve robustness and the capabilities of the machine learning algorithm. The preprocessing includes noise filtering and normalization to eliminate the effect of variable units of measure.

Temporal information is implicitly encoded via the feature of time index, enabling learning of trends and degradation of water quality over a set of observations. The dataset, after pre-processing, undergoes labeling according to normal water quality criteria. It is then divided into subsets as per a training and testing ratio of 80:20.

E. Model Selection

Various supervised learning algorithms have been considered, namely Logistic Regression, SVM, and Random Forest, with Random Forest selected because of robustness to noise, ability to capture non-linear relationships, and high predictive power.

F. Model Training

The training of the Random Forest Model is performed by fitting a series of decision trees constructed from a subset of samples from the initial training set, with several essential hyperparameters like the number of decision trees and the maximum number of nodes being decided using trial and error so that a balance is achieved between efficiency and complexity.

There are two different models: one for classification and one for regression. In the classification scheme, the model is

developed to predict the conditions for unsafe or safe water. In the regression scheme, the model is developed to predict the time to unsafe conditions.

G. Model Evaluation

The classification performance of the trained model is evaluated using standard metrics such as Accuracy, Precision, Recall, and F1-score. The regression model is evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to assess time prediction accuracy.

These evaluation metrics provide a comprehensive assessment of both categorical prediction reliability and temporal forecasting performance.

H. Deployment for real-time prediction

After the training and validation processes, the machine learning models are implemented in a cloud-based setup and coupled with the IoT-based monitoring system. The real-time sensor readings obtained from the ESP32-based sensing module are preprocessed and fed into the trained models for predictive purposes.

Based on the predicted status of the water quality and the time to reach the unsafe condition, the system automatically activates the water purification module via a control relay.

VI. RESULTS AND DISCUSSION

The proposed IoT-based water quality monitoring system was tested to evaluate its performance in real-time monitoring, prediction, and automatic control. The test was conducted on sensor data acquisition, cloud visualization, machine learning prediction, and system response.

A. Real-Time Monitoring Performance

All the parameters like pH, turbidity, and TDS sensors were used to continuously monitor the quality parameters for water. All the data was collected from the sensors and conveyed to the cloud through the ESP32 microcontroller. Also, the data was displayed on the cloud system with a graph for the efficient monitoring of changing parameters in the water quality. All

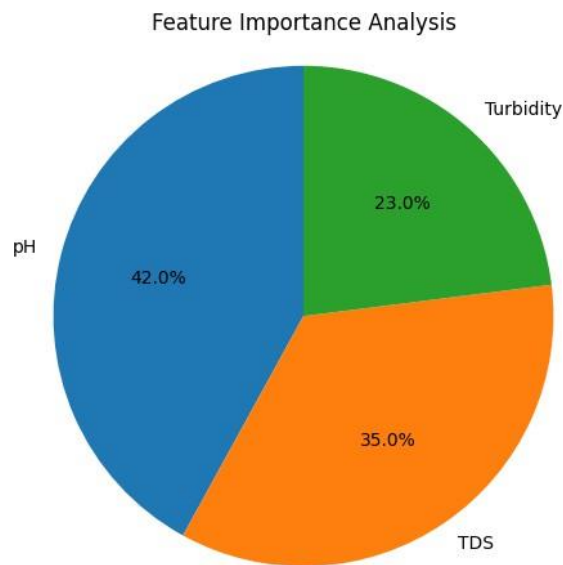


Fig. 2. Feature Importance Analysis

TABLE I
CONFUSION MATRIX FOR WATER QUALITY CLASSIFICATION

Actual / Predicted	Safe	Unsafe
Safe	930 (TP)	45 (FN)
Unsafe	35 (FP)	990 (TN)

the data was found to be conveyed efficiently with minimal loss.

Table I presents the confusion matrix of the proposed Random Forest classifier, showing effective discrimination between safe and unsafe water quality conditions.

The low false positive and false negative rates indicate the robustness of the proposed model in real-time water quality classification.

B. Machine Learning Prediction Analysis

The machine learning algorithm was trained to classify the water quality into 'Safe' and 'Unsafe' classes based on historical sensor data. Unlike other traditional methods, the model studied the degradation trends and predicted unsafe conditions in advance, enabling preventive action.

TABLE II
PERFORMANCE METRICS OF RANDOM FOREST MODEL

Metric	Value
Accuracy (%)	92.4
Precision (%)	91.1
Recall (%)	93.0
F1-Score (%)	92.0
RMSE	0.37

The results demonstrate that the Random Forest model achieved high classification accuracy and reliable prediction performance, validating its suitability for real-time water quality prediction.

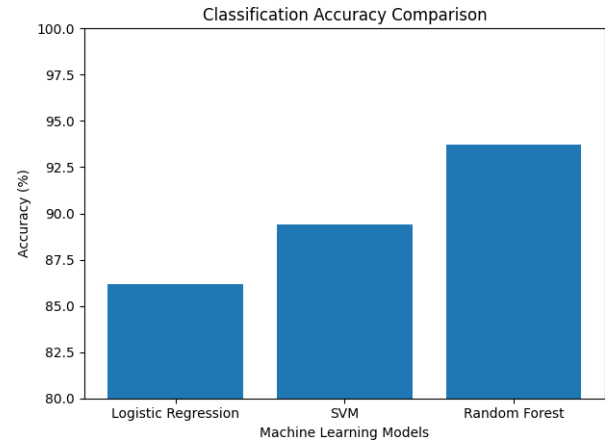


Fig. 3. Model Accuracy Comparison

C. Automated Purifier Control

If the system recognized the unsafe condition, it activated automatically the water purifier using a relay module. There were appropriate responses from the system within an acceptable time period; the system turned off upon recognizing the safe condition. Table II summarizes the standard water quality threshold values used to classify water as safe or unsafe based on pH, total dissolved solids (TDS), and turbidity. These thresholds serve as reference limits for both rule-based evaluation and machine learning-based prediction in the proposed system.

TABLE III
WATER QUALITY THRESHOLDS

Parameter	Safe Range	Unsafe Range
pH	6.5 – 8.5	< 6.5 or > 8.5
TDS (ppm)	< 500	≥ 500
Turbidity (NTU)	< 5	≥ 5

D. Comparative Discussion

The proposed system is proactive, as compared to conventional monitoring systems in its water treatment system, by predicting unsafe conditions before the threshold value is violated. Addition of machine learning in the system brought intelligence in the system.

TABLE IV
COMPARISON OF CONVENTIONAL AND PROPOSED WATER QUALITY MONITORING SYSTEMS

Feature	Conventional System	Proposed System
Real-time monitoring	Yes	Yes
Prediction capability	No	Yes
Automated purifier control	No	Yes
Preventive action	No	Yes
System intelligence	Low	High

Table II compares the conventional water quality monitoring system with the proposed intelligent system, highlighting the

added capabilities of prediction, automation, and preventive decision-making enabled through machine learning.

VII. CONCLUSION AND FUTURE SCOPE

This paper proposed an IoT-based intelligent water quality monitoring and automatic purification system that includes the integration of intelligent features like real-time monitoring, cloud-based management, prediction via the machine learning technique, and management. The continued monitoring of critical water-quality-related parameters like pH, turbidity, and TDS enables the efficient evaluation of the water's quality. It should be differentiated from the water monitoring approach, which is conventional in its use of threshold-based methods, as the proposed work uses the machine learning technique to recognize trends in the decline of water-quality-related parameters.

Besides, the evaluation of this proposed approach came through the utilization of 2000 real-time data samples regarding the sensor data. In light of the above results, the successful implementation of the proposed Random Forest-based method ensured the following metrics: a high accuracy rate of 92.4 percent, precision rate of 91.1 percent, recall rate of 93.0 percent, and F1-score rate of 92.0 percent. In this regard, the proposed method is reliable in terms of applying it to the classification of the water conditions as both safe and unsafe. Besides, the proposed regression model results ensured a high value for the Root Mean Square Error with a measure of 0.37. This confirms the effective prediction of time to unsafe water conditions.

Moreover, with the integration of an open-source cloud platform, there exists a feature for real-time visualization as well as analysis, thereby enhancing its use for implementation in a water tank, water reservoir, or a smaller water body. Additionally, with this automated purifier control mechanism, there exists a minimized response time due to the feature of instant activation of the purifying unit once the levels are deemed unsafe. Therefore, as mentioned above, this proposed system is a reliable intelligent and automated solution with regard to implementing smart water quality management.

In future, the focus may also include adding more parameters to the system, which may be based on other parameters for the quality of water. These parameters may include temperature, dissolved oxygen level, conductivity, etc. In future, a better machine learning or deep learning-based model may be used to increase the level of accuracy. Implementing the above system or models as a mobile or web application may increase the usability and make the system more apt for the purpose.

REFERENCES

- [1] S. Geetha and S. Gouthami, "Internet of Things enabled real time water quality monitoring system," *Smart Water*, vol. 2, no. 1, pp. 1–19, 2016.
- [2] M. S. U. Chowdury, T. A. Shuvro, M. A. Rahman, and S. Islam, "IoT based real-time river water quality monitoring system," *Procedia Computer Science*, vol. 155, pp. 161–168, 2019.
- [3] V. Lakshmikantha, R. K. Kodali, and B. S. Saravanan, "IoT based smart water quality monitoring system," *International Journal of Applied Engineering Research*, vol. 12, no. 18, pp. 7306–7310, 2017.

- [4] M. S. Islam, M. T. Rahman, and M. M. Hasan, "Water quality prediction using machine learning algorithms," in *Proc. IEEE Int. Conf. on Smart Cities*, 2019, pp. 1–6.
- [5] S. Bhardwaj, R. Gupta, and A. Chauhan, "Water quality assessment using machine learning techniques," *International Journal of Engineering Research & Technology*, vol. 9, no. 3, pp. 227–232, 2020.
- [6] P. P. Ray, "Internet of Things for smart water management: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4633–4649, Dec. 2018.
- [7] A. R. Al-Ali, I. Zualkernan, and F. Aloul, "A mobile GPRS-sensors array for air pollution monitoring," *IEEE Sensors Journal*, vol. 10, no. 10, pp. 1666–1671, Oct. 2010.
- [8] K. Srinivasan and R. Ranjith Kumar, "Smart water quality monitoring system using IoT," *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 2017, pp. 1–4.
- [9] S. R. K. Somani, S. P. Chavan, and M. S. Joshi, "Machine learning approaches for water quality prediction," *IEEE International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 1056–1061.
- [10] Y. Wang, J. Zhang, and L. Chen, "Real-time water quality monitoring and early warning system based on IoT and cloud computing," *IEEE Access*, vol. 7, pp. 171129–171140, 2019.

