Journal of
Control & Instrumentation
https://journals.stmjournals.com/joci

ISSN: 2229-6972 (Online)
ISSN: 2347-7237 (Print)
Volume 16, Issue 2, 2025
May–August
DOI (Journal): 10.37591/JoCI

Review JoCI

Tank Water Quality Analysis Using Machine Learning

Sanika D. Gaikwad[1,*], Sakshi A. Chavan[1], Mayuri V. Jadhav[1], Rutuja K. Shinde[1], A.C. Jadhav[2]

Abstract

Tank Water quality is a critical factor for public health, agriculture, as well as industry. Continuous monitoring of tank water quality: temperature, humidity, water level, CO_2 concentration, and pH, is vital for safe usage. Using machine learning, real-time data analysis can detect anomalies, predict issues, and optimize water management, ensuring timely responses and improved safety. This intelligent approach enhances decision-making and maintains water quality effectively in various environments. We develop an IoT-based system using ESP32/Node MCU and multiple sensors to monitor water parameters in a tank. The Python Flask web application is used to retrieve the gathered data once it has been sent to the Firebase cloud platform. A machine learning algorithm, specifically a Decision Tree Classifier (DTC) and Convolution Neural Network (CNN), is employed to predict water quality based on the gathered data and showing the quality of water. If poor water quality is detected, the system provides suggestions for precautionary measures to improve water conditions. This approach offers real-time, automated water quality analysis and can be easily expanded for large-scale applications in smart water management systems.

Keywords: Tank water quality analysis, ESP32, Node MCU, humidity sensor, DHT11, ultrasonic sensor, CO_2 gas sensor, pH sensor, Firebase, Wi-Fi, Python Flask, machine learning, Decision Tree Classifier algorithm, water quality prediction, IoT sensors, data transmission, data fetching, precaution system, water monitoring, real-time analysis, Flask web

INTRODUCTION

The Tank Water Quality Analysis using Machine Learning project focuses on real-time monitoring and analysis of water quality in tanks using IoT sensors and machine learning. Sensors like DHT11, ultrasonic, CO_2, and pH sensors are used to collect essential water parameters such as temperature, humidity, water level, CO_2 concentration, and pH levels [1].

The data is sent to Firebase through an ESP32/NodeMCU and then fetched into Python Flask web app. Using a Decision Tree Classifier (DTC) algorithm, the system predicts water quality and suggests precautionary measures if the quality is found to be poor. Some sensor used in this system are: (1) IoT sensor, (2) pH sensor, (3) temperature Sensor, (4) humidity sensor, and (5) Node MCU.

*__Author for Correspondence__
Sanika D. Gaikwad
E-mail: sanikagaikwad254@gmail.com

[1]Student, Department of Computer Engineering, Shri Chhatrapati College of Engineering, Pune, Maharashtra, India
[2]Professor, Department of Computer Engineering, Shri Chhatrapati College of Engineering, Pune, Maharashtra, India

Receiving Date: April 01, 2025
Accepted Date: April 25, 2025
Published Date: May 16, 2025

Citation: Sanika D. Gaikwad, Sakshi A. Chavan, Mayuri V. Jadhav, Rutuja K. Shinde, A.C. Jadhav. Tank Water Quality Analysis Using Machine Learning. Journal of Control & Instrumentation. 2025; 16(2): 27–34p.

IoT Sensor

Sensors gather information from their surroundings and are able to identify environmental changes. Physical characteristics like temperature, pressure, wetness, and so on are measured by sensors. Water monitoring systems, biometric systems, and medical systems all require sensors [2].

pH Sensor

A pH sensor is utilized to test the water's acidity. The measurement of the pH sensor is referred to as pH. The range of the logarithmic pH scale is 0 to 14. An electrode with a pH level below 7 is considered as acidic, while a pH level above 7 is regarded as basic. A pH sensor may measure the electrode. Water is regarded as clean if its pH falls between 6.5 and 8, and impure if it falls below 6.5 or rises over 8.

Temperature Sensor

Temperature sensors measure heat and cold and converting these readings into electrical signals like in water heaters, thermometers, refrigerators, and microwaves.

Humidity Sensor

The quantity of water, or humidity, in the air is detected and measured using a device called a humidity sensor. They come in a wide variety of shapes and sizes, from tiny parts integrated into air quality systems to portable sensors. Humidity is measured via the humidity sensor.

Node MCU

The term “Node MCU”, which merges the terms “node” and “micro-controller unit”, relates to the firmware as opposed to the related development kits.

LITERATURE REVIEW

A study explores the use of IoT sensors to monitor water quality in real-time and highlighting its benefits of automated data collection for better management [1]. An investigation especially used the Decision Trees, which can be applied to predict water quality based on sensor data. The approach forms the water quality prediction process.

The approach presents a cost-effective IoT-based system for monitoring water parameters like pH and temperature, emphasizing ease of deployment, humidity and moisture. And it is provided with the required sensors [2].

Another paper presents a framework for real-time water quality monitoring using IoT devices and web-based technologies. It discusses the use of various water quality sensors (e.g., pH, temperature, and turbidity) and demonstrates how the data can be transmitted to a web application for visualization. The study emphasizes the need for real-time data monitoring to ensure the safety and quality of water, particularly in rural or remote areas [3].

A research emphasizes the public health benefits of using IoT and machine learning to monitor water quality. It discusses how machine learning models like Decision Trees can be used to predict water contamination events and prevent public health crises. The authors stress the need of prompt recommendations for remedial measures and early identification of problems with water quality [4]. Another paper investigates machine learning models for predicting water quality, focusing on Decision Tree algorithms for accurate and timely predictions [5].

METHODOLOGY

The motivation behind the Tank Water Quality Analysis project stems from the critical importance of maintaining clean and safe water for both residential and industrial purposes. Water pollution is a major worldwide problem, and unnoticed decline in water quality can have serious negative effects on equipment, the environment, and human health. The creation of this system, attempts to maintain the safety of the water for human use [6].

- *Use cases and application use cases:* Use cases represent the actions or functions that the system performs. They describe what the system does when an actor interacts with it.
- *Process of project*
  - *Send data to firebase:* ESP32/NodeMCU sends water parameters to Firebase.
  - *Predict water quality:* The system predicts the water quality based on collected data.
  - *Display result:* The system shows the water quality result to the user.

- *Analyze data:* The system processes the data to analyze the water quality.
- *Monitor water quality:* The user monitors the water quality parameters like temperature, humidity, etc.
- *Receive precaution:* The system provides recommendations if the water quality is poor.

Problem Statement

Ensuring the quality of water in storage tanks is crucial for both domestic and industrial usage. Equipment damage and health problems can result from poor water quality. Manual sampling and laboratory analysis are common components of traditional water quality testing techniques, which can be expensive, time-consuming, and impractical for real-time monitoring [7–9]. This project seeks to address this challenge by developing an automated system that uses IoT sensors to continuously monitor various water parameters, such as surface temperature, humidity, water level, CO_2 concentration, and pH levels.

The collected data will be transmitted via Wi-Fi to Firebase, from where it will be fetched into a Python Flask web app. By applying a Decision Tree Classifier machine learning algorithm, the system will predict water quality in real-time.

If the water quality is found to be substandard, the system will provide recommendations for corrective actions. This approach enables continuous monitoring and ensures timely responses to potential water quality issues, improving overall safety and efficiency.

Ensuring access to clean and safe water is a critical global challenge, particularly in settings where water is stored in tanks for extended periods. Contaminants and environmental factors such as changes in temperature, pH levels, and CO_2 concentrations can affect water quality over time, leading to serious health hazards if left undetected. The manual sample and laboratory tests that are frequently used in current water quality monitoring techniques are time-consuming, expensive, and unsuitable for ongoing monitoring [8].

There is a growing need for an *automated, real-time water quality monitoring system* that can continuously track key water parameters, predict water quality using advanced technologies, and provide timely recommendations to users when the water becomes unsafe.

Use Case Diagram

Steps to Draw a Use Case Diagram

A straightforward visual depiction of how a system communicates with users (or other systems) is called a use case diagram (Figure 1). It facilitates comprehension of the system's relationships and functional needs.

Here is an easy explanation of the key components:

Actors: Actors are entities (people, systems, or devices) that interact with the system. There are two types of actors:

1. *Primary actors:* The primary users or systems that start the conversation with the system are these. For example, a User or an ESP32/Node MCU device.
2. *Secondary actors:* These actors help or support the system but do not initiate the actions. For example, Firebase, which stores data for the system.
   i. *Identify the actors:* identify who or what will interact with the system.
   ii. *Identify the use cases:* List what actions the system will perform when interacting with the actors.
   iii. *Create the diagram:* Refer to actors with stick figures and use cases with ovals. Connect them with lines to show interactions.

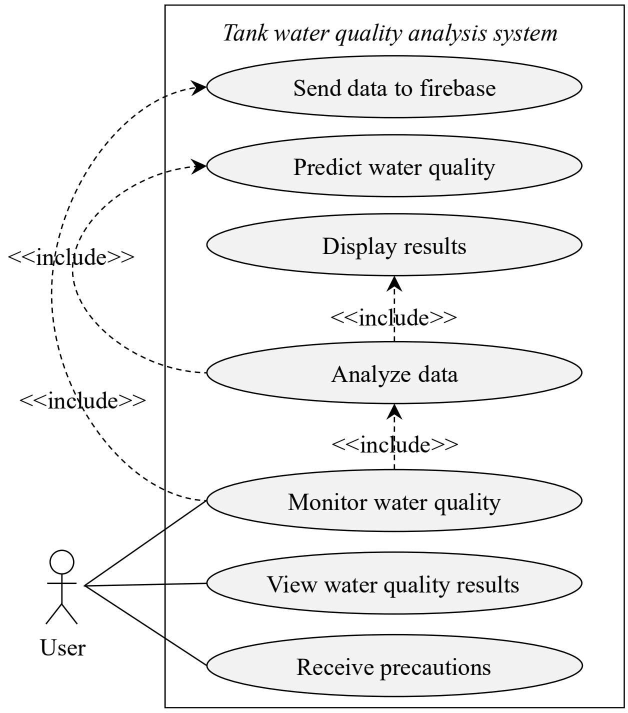

Figure 1. Use case diagram.

CASE STUDIES

Case Study 1: Urban Water Tank Quality Analysis

An urban water supply authority faced challenges with maintaining the quality of water in its large storage tanks, attributing issues to seasonal variations, industrial runoff, and aging infrastructure [9].

Objectives

- Predict water quality parameters such as turbidity, pH, and microbial levels.
- Identify potential contamination sources.
- Develop a real-time monitoring system for actionable insights.

Historical Data

- Collected from laboratory tests and maintenance logs over the last 5 years.

Data Preprocessing

- Data cleaning to handle missing values and outliers.
- Feature normalization to make model training easier.
- Category-based label encoding for machine learning.

Exploratory Data Analysis (EDA)

- Visualization of trends over time using line plots.
- Correlation analysis to identify relationships between parameters.

Case Study 2: Agricultural Water Tank Quality Assessment

An agricultural cooperative sought to improve the quality of irrigation water stored in tanks to enhance crop yield and reduce health risks related to waterborne diseases.

Objectives

- Monitor and analyze water quality parameters affecting agricultural practices.
- Leverage machine learning to predict water suitability for irrigation.

Case Study 3: Industrial Water Storage Tank Monitoring

A manufacturing facility needed to ensure the water quality of its processing tanks, as any contamination could lead to product quality issues and regulatory compliance failures.

Objectives

- Keep an eye on the quality of the water to make sure industry rules are being followed.
- Utilize machine learning to predict contamination events before they occur.

Data Preprocessing

Integrated data from various sources (sensors and logs). Techniques for analyzing time-series data were used to smooth out oscillations.

Exploratory Data Analysis (EDA)

Time series plots highlighted trends and potential seasonal patterns in chemical composition. Correlation analysis suggested relationships between manufacturing processes and water quality degradation.

Machine Learning Models

Employed algorithms such as: Gradient Boosting Machines for predicting contaminants. Time-series data may be analyzed using neural networks to identify intricate patterns. Gradient Boosting Machines outperformed other models, achieving 91% accuracy.

Model Evaluation

Model performance was validated through cross-validation techniques to assess generalization.

Results

- An automated monitoring system was established, enhancing compliance with regulations.
- Predictive alerts reduced contamination risks, leading to fewer production halts.
- Increased efficiency in monitoring processes enabled timely preventive maintenance.

ALGORITHM DTC

By applying a Decision Tree Classifier machine learning algorithm, the system will predict water quality in real-time. The system will suggest remedial measures if the water quality is determined to be below acceptable levels. This approach enables continuous monitoring and ensures timely responses to potential water quality issues, improving overall safety and efficiency.

A machine learning algorithm, specifically a Decision Tree Classifier (DTC), is employed to predict water quality based on the gathered data [10]. If poor water quality is detected, the system provides suggestions for precautionary measures to improve water conditions. This approach offers real-time, automated water quality analysis and can be easily expanded for large-scale applications in smart water management systems.

The Decision Tree Classifier (DTC) machine learning algorithm should be able to predict water quality with an acceptable level of accuracy (e.g., at least 80–90%). Predictions should correctly classify water as either "Safe" or "Unsafe" for use, based on the given sensor data.

A Decision Tree Classifier (DTC) algorithm is used to predict water quality based on sensor data (Figure 2). The machine learning component helps classify the water as either safe or unsafe, enabling real-time decision-making about water quality.

This study compares various machine learning algorithms, including Decision Tree Classifiers (DTC), for predicting water quality based on parameters like pH, turbidity, and dissolved oxygen. The accuracy and interpretability of the Decision Tree Classifier were determined to be good [11].

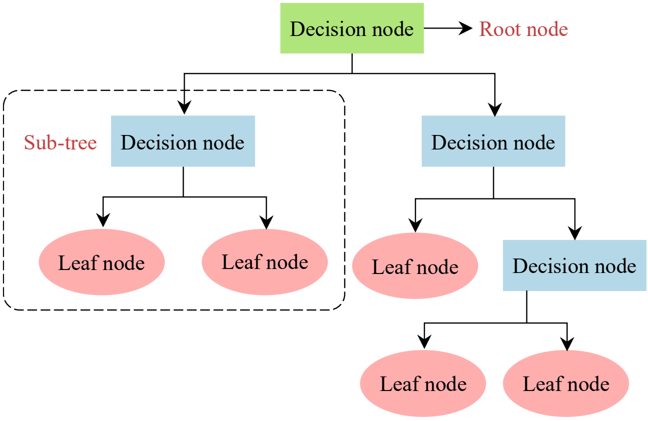

Figure 2. DTC architecture.

DTC can effectively classify water as safe or unsafe based on input parameters, supporting the proposed use of DTC in the project for predicting water quality based on real-time sensor data.

System Design

During this stage of system design, we create a system that is user-friendly, or simple enough for end users to understand (Figure 3). To comprehend the system flow, system module, and execution sequence, we create a few UML and data flow diagrams.

1. *IoT integration for real-time monitoring*
   i. The system will utilize ESP32/NodeMCU microcontrollers and a set of sensors (DHT11 for temperature and humidity, ultrasonic sensor for water level, pH sensor, CO_2 sensor) to measure water quality parameters continuously.
   ii. Sensor data will be transmitted wirelessly to Firebase, ensuring real-time data logging and availability.
2. *Cloud data storage and processing:* The project will use Firebase as the cloud platform to store, manage, and retrieve real-time water quality data. The cloud-based system will allow scalability and remote data access for users to monitor water quality from anywhere [6].
3. *Data analysis and visualization:* A Python Flask web application will be developed to fetch sensor data from Firebase and display it in a user-friendly interface. Real-time water condition updates and water quality parameter visualization will be provided via the application.
4. *Machine learning for water quality prediction*
   i. The system will implement a Decision Tree Classifier (DTC) to analyze the sensor data and predict the water quality (e.g., whether the water is safe or unsafe for consumption).
   ii. The classifier will be trained using historical water quality data to improve prediction accuracy over time.

System Architecture

The system architecture for the Tank Water Quality Analysis Using Machine Learning is designed to enable real-time monitoring, data collection, prediction, and analysis of water quality using IoT technology and machine learning. Here is a simple explanation of the components and their roles in the system:

1. *IoT sensors:* These sensors collect important water parameters:
   i. *CO_2 sensor:* Measures carbon dioxide concentration in the water.
   ii. *pH sensor:* Detects the pH level of the water, indicating its acidity or alkalinity.
   iii. *The DHT11 sensor* detects the humidity and temperature surrounding the tank.
   iv. *The ultrasonic sensor* keeps track of the tank's water level.

   These sensors are essential for collecting information about the water's present state and its surroundings.
2. *ESP32/NodeMCU*
   i. The ESP32/NodeMCU microcontroller collects data from the IoT sensors.

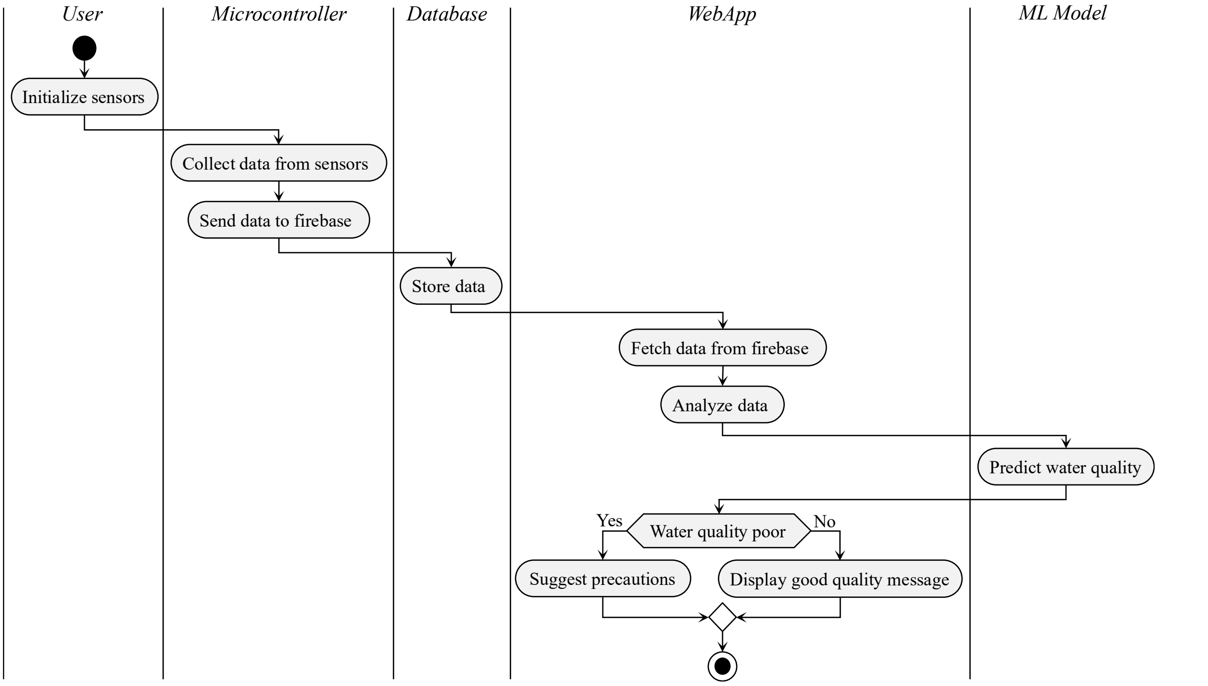

Figure 3. System design.

ii. It establishes a Wi-Fi connection and transfers the gathered data to the cloud for further analysis and storage [4].
iii. This part acts as the conduit for communication between the cloud and the sensors.
iv. Firebase: Firebase is used as the cloud storage system.
v. It stores the data received from the ESP32/NodeMCU and allows easy retrieval of the data by other components.
vi. Firebase offers remote availability of the data for monitoring in real time.

3. *Python flask web application*
   i. This is the interface that interacts with the user and performs the analysis of the data.
   ii. The web app fetches data from Firebase, analyzes the water quality, predicts whether the water is safe or unsafe, and suggests precautions.
   iii. It acts as the central processing hub that integrates the data with machine learning predictions.
4. *Machine learning model (DTC)*
   i. The Decision Tree Classifier (DTC) is the machine learning model used to analyze the water quality data.
   ii. It takes the input from the sensors (such as pH, CO_2 levels, etc.) and predicts the overall quality of the water.
   iii. This prediction helps in deciding whether the water is safe to use and what precautions or actions need to be taken [7].

User

- The web application is how the user communicates with the system.
- The user can see the real-time status of the water quality, predictions on whether the water is safe, and any recommendations or actions suggested by the system.
- This gives the user full control and insight into the state of their water storage.

CONCLUSION

The Tank Water Quality Analysis using Machine Learning system successfully meets the primary objectives of continuous, real-time monitoring of water quality parameters in a tank using an IoT-based approach. The system integrates various sensors (temperature, humidity, pH, CO_2, water level) to collect water quality data, which is then transmitted to a cloud platform (Firebase) for analysis and storage. The data is processed using a Decision Tree Classifier (DTC) model to predict water quality, and actionable suggestions are provided for improving water quality if necessary.

REFERENCES

1. Saravanan K, Anusuya E, Kumar R, Son LH. Real-time water quality monitoring using Internet of Things in SCADA. Environ Monit Assess. 2018 Sep; 190(9): 556.
2. Huang S, Xia J, Wang Y, Lei J, Wang G. Water quality prediction based on sparse dataset using enhanced machine learning. Environ Sci Ecotechnology. 2024 Jul 1; 20: 100402.
3. Solanki A, Agrawal H, Khare K. Predictive analysis of water quality parameters using deep learning. Int J Comput Appl. 2015 Sep; 125(9): 29–34.
4. Vergina SA, Kayalvizhi S, Bhavadharini R, Kalpana Devi S. A real time water quality monitoring using machine learning algorithm. Eur J Mol Clin Med. 2020; 7(8): 2035–41.
5. Khan Y, See CS. Predicting and analyzing water quality using machine learning: a comprehensive model. In 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT). 2016 Apr 29; 1–6.
6. Jibrin AM, Al-Suwaiyan M, Aldrees A, Dan'azumi S, Usman J, Abba SI, Yassin MA, Scholz M, Sammen SS. Machine learning predictive insight of water pollution and groundwater quality in the Eastern Province of Saudi Arabia. Sci Rep. 2024 Aug 28; 14(1): 20031.
7. *Hu XC*, Dai M, Sun JM, Sunderland EM. The utility of machine learning models for predicting chemical contaminants in drinking water: Promise, challenges, and opportunities. Curr Environ Health Rep. 2023 Mar; 10(1): 45–60.
8. Zhu M, Wang J, Yang X, Zhang Y, Zhang L, Ren H, Wu B, Ye L. A review of the application of machine learning in water quality evaluation. Eco-Environ Health. 2022 Jun 1; 1(2): 107–16.
9. Ahmed AN, Othman FB, Afan HA, Ibrahim RK, Fai CM, Hossain MS, Ehteram M, Elshafie A. Machine learning methods for better water quality prediction. J Hydrol. 2019 Nov 1; 578: 124084.
10. Valarmathi K, Vinoth R, Sneka T, Srinivethitha S, Venkadeswari V. Machine learning based automatic water tank cleaner. In: Kaur S, Kamboj S, Kumar M, Dagur A, Shukla DK, editors. Computational Methods in Science and Technology: Proceedings of the 4th International Conference on Computational Methods in Science & Technology (ICCMST 2024), 2–3 May 2024, Mohali, India. Volume 1. London (UK): CRC Press; 2024. p. 7. doi: 10.1201/9781003501244-26.
11. Lowe M, Qin R, Mao X. A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. Water. 2022 Apr 24; 14(9): 1384.

