



PREDICTIVE WATER MONITORING: A DIGITAL TWIN APPROACH USING DECISION TREES

Pavada Santosh, R. Umamaheswari, B. Maha Lakshmi, Md Faizan Bin Haque, Palakonda Dakshayani,
Nukapeyyi Tanuja and Kanaka Vishnu Vardhan

Department of Electronics and Communication Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, India
E-Mail: santoshpavada@vignaniit.edu.in

ABSTRACT

This work presents a compact IoT-based Digital Twin architecture implemented for water quality monitoring. It is designed to overcome the increasing challenge of pollution and the need for real-time analysis in urban environments. Unlike conventional IoT water-quality systems that basically focus on data collection and cloud display, the proposed architecture introduces key innovations: a multi-sensor model consisting of TDS, ultrasonic, DHT11, and water flow measurements to generate a richer and more accurate representation of physical water conditions. A Digital Twin layer that mirrors real-world water behaviour in real time, enabling early abnormality detection rather than passive monitoring, and an embedded machine-learning pipeline, where a Decision Tree classifier trained on empirically labelled data will predict water quality levels and identify abnormalities through a Python-based interface.

Keywords: IoT-based monitoring, digital twin, water quality monitoring, machine learning, decision tree classifier, smart water management, cloud-based visualization.

Manuscript Received 6 March 2026; Revised 20 April 2026; Published 10 July 2026

1. INTRODUCTION

Water quality monitoring in an efficient and reliable manner has become a crucial requirement due to growth in urban population and increased usage of water. Traditional methods are low, intermediate, and not capable of providing the continuous data that is required for implementing a smart city. Recent IoT advancements such as wireless sensing, machine learning, and digital twin technologies provide a good base for architecture and real-time monitoring systems. [9] [14] [10].

According to recent studies, each sensing technology contributes uniquely to water-management systems. Ultrasonic sensors have proven effective for non-contact water-level measurement and provide a low-cost solution for tank and reservoir monitoring [1]. IoT-based platforms perform TDS measurement using cloud-connected sensors to capture real-time dissolved solids measured in ppm and file alerts to the system [2]. AI-driven IoT systems further improve monitoring speed, accuracy, and parameter calculation [9] [12] [16]. Multiple-input IoT water monitoring systems which integrates pH, turbidity, temperature, and conductivity have also shown improved accuracy and remote accessibility [4] [8] [5], and long-distance wireless networks such as LoRa enable scalable, energy-efficient requirements [13] [12].

Along with the advancements in IoT, AI and Digital Twin technologies are increasing as powerful tools for real-world simulation, prediction and analysis, and water monitoring optimization [14] [10] [15]. Technologies like Deep Learning have additionally enhanced the real-time water quality categorization accuracy in dynamic environments [10].

Microcontrollers that are connected to the cloud example ESP8266, enable the continuous flow of data to achieve measurement and leakage information [3], while IoT-based monitoring in treatment plants shows reliable real-time performance in operating areas [11]. Limitations such as sensor intensity, calibration, dependence on networks, and demand of full scale digital twin's points that the additional research is still required [15]. "To provide scientific clarity, the Digital Twin is a virtual representation of a physical system continuously updated using real-time sensor data to support monitoring, prediction, and decision-making".

The earliest efforts focused on cost-effective, contactless solutions for physical data. These low-cost systems relied on ultrasonic sensors to measure water levels in tanks and reservoirs [1]. These solutions were simple also being at a very low cost, but their accuracy was compromised by factors in the environment such as wave travelling and surface disturbances. IoT-based work described cloud connectivity for single-parameter tracking such as Total Dissolved Salts [2]. It was effective for basic remote supervision but failed to provide a complete analysis of overall water quality.

Subsequently, research in the field saw the development of more complex multi-parameter IoT models that, at the same time, measured physical and chemical indicators like pH, turbidity, conductivity, temperature, and TDS [4, 8, 5]. The systems have reported great advancements in terms of data quality. However, they also bring very important system problems. At present, most use fixed threshold-based alarm systems which did not incorporate Machine Learning for prediction-based classification or abnormality analysis. Additionally, they require



frequent sensor usage and are reported to degrade over time in field use. Many large-scale solutions that rely on Wi-Fi-based microcontrollers (such as the ESP8266) for data transmission have connectivity problems [3]. Although they were successful in identifying flow variations, their dependence on this technology made the monitoring systems unreliable and highly vulnerable to real-world malfunctions.

The majority of our DT investigations produced conceptual frameworks that lacked the integration necessary for real-time functioning. They did not integrate effectively with real-time sensor networks or see through to a successful field deployment. Advanced communication techniques like LoRa and MQTT were also introduced, which increased range and decreased power consumption, but they weren't put forward in conjunction with predictive analytic models. All things considered, the current study lacks an inexpensive framework that incorporates multivariable sensing, machine learning (ML) for forecasting capabilities, decision-making and real-time operation (DT), and coverage under connection failures.

2. RESEARCH GAPS AND EXISTING CHALLENGES

Why AI Models Are Difficult to Adopt (The "Black Box" Problem): The primary obstacle to the implementation of sophisticated AI systems in water safety is trust. For instance, Deep Learning [9, 10] models function as a black box when they detect contamination. They give responses, but the operators are unable to comprehend the underlying reasoning or the parameters that led to that conclusion. They lack obvious explanation, and it is challenging to accept and implement systems that lack clear rationale [14]. Concept drift is a phenomenon where AI models that are solely data-driven may abruptly "break" due to unforeseen changes in the environment [15]. These models can rapidly lose their capacity to make accurate predictions if a flood or other typical polluting event

happens. We require a quick model that maintains stability under extreme circumstances.

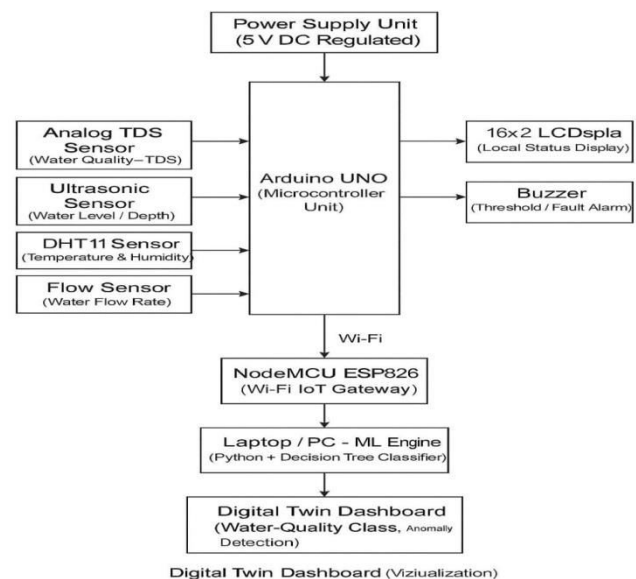
Many systems excel at a single job, yet they are unable to provide comprehensive, reliable, and timely responses. There are several IoT initiatives that mostly concentrate on basic data groupings. They have successfully collected a number of variables [4, 8, 5], but this lacks the integrated intelligence (the ML algorithm) required to provide complicated data and predictive analysis. Instead of clever variation, they are based on straightforward, fixed alerts. For processing and alerting, the data is sent straight to the cloud. Wi-Fi provides the basis for this procedure, which causes delays and implies that if the connection breaks, the entire system may fail [3]. When a system is unable to operate correctly, reliability is severely degraded. Many systems function as a "Digital Shadow" where data travels solely in one direction (from the physical sensor to the virtual display) [6, 7]. They don't have the necessary low-latency, quick intervention capabilities. The system should be able to rapidly initiate an on-site alert in the case of a critical incident, without waiting for the data to go to the cloud and back. To gain a solid grasp of water quality, several intricate studies using multiple sensors are conducted. The variables of pH, turbidity, temperature, and conductivity were simultaneously produced and utilized by systems that track and upload to cloud services [4, 8, 5]. Despite being able to assess a wide range of factors, these systems still had problems with scalability and a predictive intelligence gap. Because of its potential in modeling, predictive control, and system optimization, the Digital Twin concept [16] has developed into a viable instrument for smart water management [6, 7]. The majority of DT research is still mostly conceptual or review-based, concentrating on modeling frameworks without attaining deep, real-time integration with field data and inexpensive, deployable sensor networks [6, 7].

**Table-1.** Challenges and implemented solutions in the proposed architecture.

Challenge Addressed	Innovation and Solution
ML Black-Box & Interpretability (Gap 1)	This system needs a Decision Tree classifier, commonly referred to as a "white box," which is a very comprehensible model. It offers transparent, verifiable decision logic, which is a black box and increases operational trust in safety-type forecasts.
Computational Overhead & Complexity (Gap 1)	Decision trees are lightweight implementations of sophisticated DL that boost computing performance, reduce overhead, and improve storage economy.
Architectural Incompleteness (Gap 2)	We introduce a unified intelligent framework integrating multi-sensor IoT, cloud storage, a Decision Tree ML pipeline, and a Digital Twin visualization layer. It provides a very important, complete end-to-end outline that was missing in previous work.
Synchronization Latency & Reliability (Gap 2)	The Arduino Uno introduces a two-level solution that reduces cloud latency by managing data cleansing and quick buzzer or LCD notifications. Through local storage and retransmission during network problems, the NodeMCU further assures data reliability.

3. METHODOLOGY

All sensor readings in this system first go through the Arduino UNO, which refreshes the 16x2 LCD display after filtering the data and looking for any dangerous levels. The Arduino instantly sounds a buzzer to notify anyone in the vicinity whenever it detects an anomalous situation, such as elevated TDS or erratic water flow. The verified data is then transferred to the cloud via the NodeMCU ESP8266, enabling services such as ThingSpeak to save each reading and present it as live graphs. These graphs make it simple to identify typical trends and identify unusual behavior in water level, temperature, TDS, humidity, or flow. Later, labeled datasets for machine learning are created using the cloud data that has been stored. A Decision Tree model is trained to categorize water quality into Good, Moderate, and Poor categories following preprocessing. In order to forecast water quality and identify anomalies, the trained model operates on a local computing unit and analyzes incoming data in real time. A digital twin that replicates the real system behavior displays all of this data, including sensor readings, machine learning predictions, and alarms.

**Figure-1.** Block diagram representation of proposed method.

3.1 Component Analysis

A. Ultrasonic Sensor

When it comes to employing flight distance measurement to monitor the water level, the ultrasonic sensor is quite important. It measures the water's height inside the tank or storage space continuously and identifies significant level changes. Threshold limits guarantee safe operation; levels above 90% suggest a potential overflow, while levels below 10% indicate a shortage alarm. In addition to preventing overfilling or drying out of the system, these restrictions aid in identifying unusual swings.

Time of Flight Distance Calculation:

$$D(\text{cm}) = (v \times t)/2$$

D = Distance to water surface



v = Speed of sound in air ($\approx 343\text{m/sat}20^\circ\text{C}$) $v = 331 + 0.6T$
 t = Measured round-trip time

B. TDS Sensor

This sensor is responsible for measuring water purity by measuring the concentration of dissolved solids. It converts conductivity to TDS values, giving real-time monitoring of contamination or mineral buildup. Below is the voltage-to-EC conversion equation.

$$E_{\text{raw}} = \frac{\text{Voltage Reading Output}}{V_{\text{ref}}} \times \text{Calibration factor}$$

Temperature Compensation is given by

$$E_{25} = \frac{E_{\text{raw}}}{1 + \alpha(T - 25)}$$

α = Temperature coefficient ($\approx 0.02\text{per}^\circ\text{C}$)

T = Water temperature ($^\circ\text{C}$)

EC-to-TDS Conversion can be done by

$$TDS = EC_{25} \times k$$

k = total dissolved solids conversion factor range (0.5-0.7).

C. DHT11 Sensor

The DHT11 sensor monitors all of the environmental conditions around the water system by measuring temperature and humidity. It helps ensure that pumps, electronics, and storage containers are working within safe surroundings. Thresholds are defined to detect abnormal situations: temperature above 40°C indicates overheating, while humidity above 75% suggests leakage risks. These alerts help prevent hardware damage and environmental conditions that could affect water quality. Real-time data is transmitted to the cloud, allowing the Digital Twin to visualize working conditions instantly. The sensor's data supports ML-based abnormality detection and maintenance in the system.

D. Flow Sensor

It allows for precise tracking of water supply and consumption by measuring the speed at which water moves through pipes. It produces velocity-proportional pulses that aid in real-time variation or irregularity detection. A burst pipe or excessive consumption may be indicated by flow exceeding 20 L/min, whereas flow below 1-2 L/min signals leakage or obstruction. This restriction guarantees safe distribution and prompt abnormal condition identification. The Digital Twin is synchronized with the sensor's continuous data stream, enabling real-time viewing and ML-based predictive maintenance. This increases smart water-monitoring systems' dependability and efficiency.

The pulse output equation is as follows:

$$F = k \times Q$$

F = pulse frequency units (Hz)

Q = flow rate units (L/min)

4. RESULTS AND SIMULATIONS

The results of the work are simulated for the demonstration of the outline of the integrated hardware sensors' behavior, such as TDS, temperature, humidity, ultrasonic distance, and flow rate, along with the LCD interface and buzzer-alert mechanism.

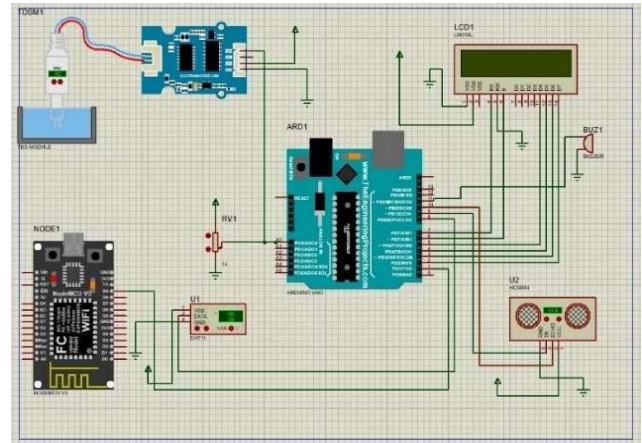


Figure-2. Proteus simulation.

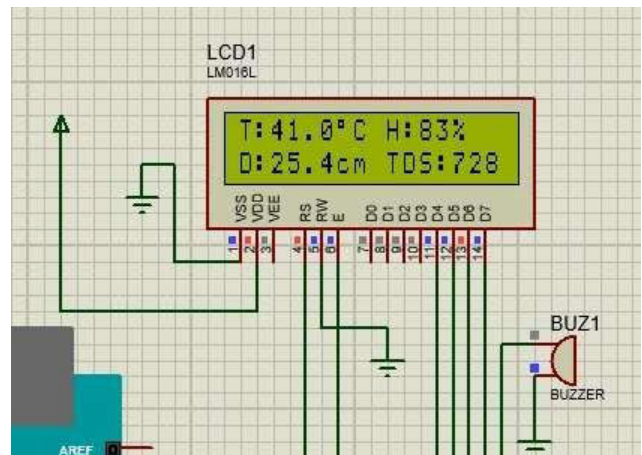


Figure-3. LCD Display of sensor readings.

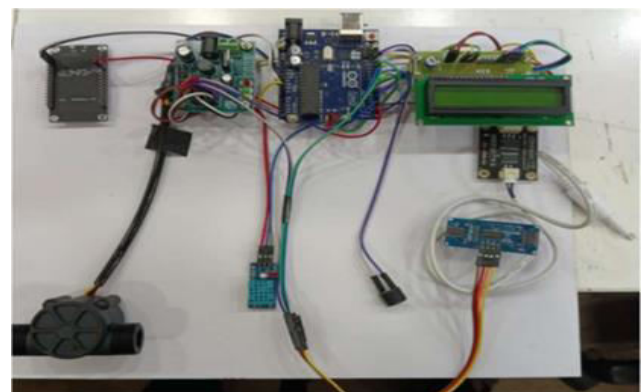


Figure-4. Hardware component connection.



A. Real Time IoT Cloud Outputs

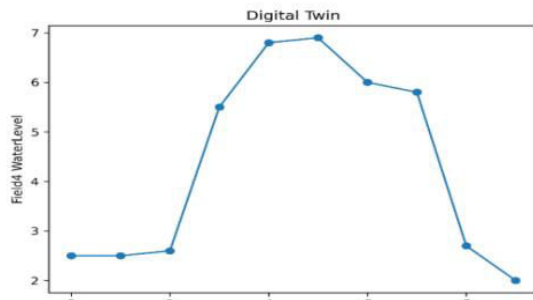


Figure-5. Flow Water level values.

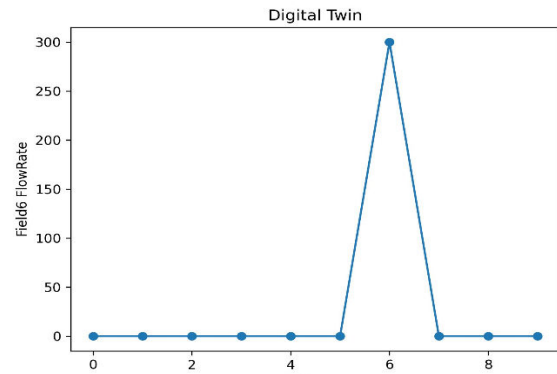


Figure-7. TDS values.

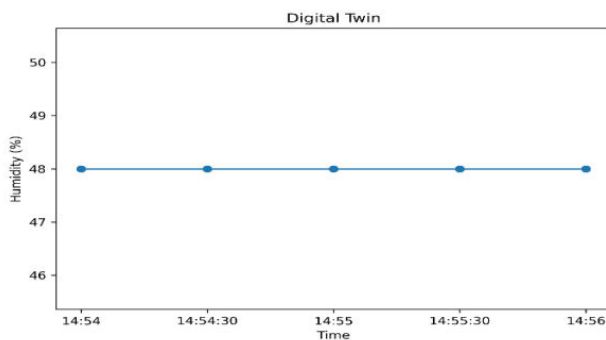


Figure-6. Humidity values.

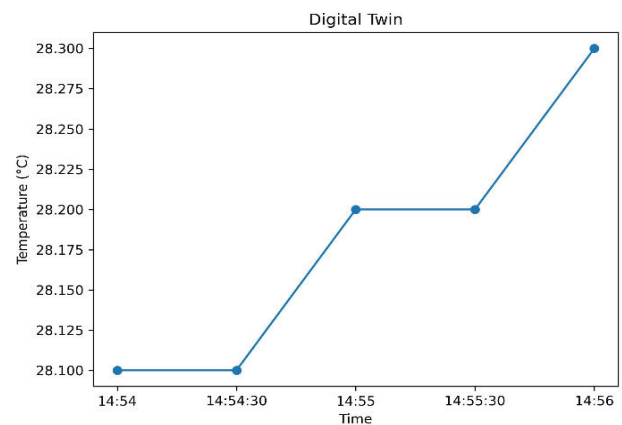


Figure-8. Temperature values.

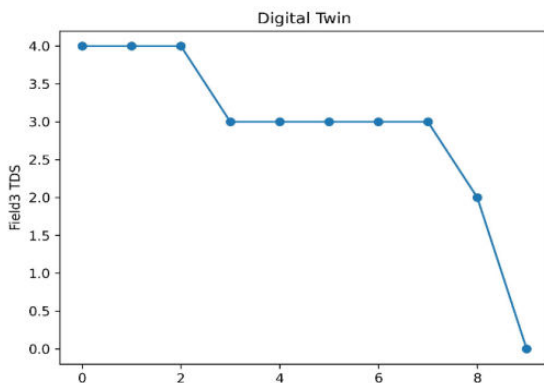


Figure-11. Flow Rate values.

All of the sensors that were continuously producing measurements for temperature, humidity, water level, flow rate, and TDS were uploaded and visualized using an IoT cloud platform called ThingSpeak.

Table-2. Threshold limiting values for each sensor used.

Sensor	Normal Range	Low Threshold	High Threshold	Critical Condition	Sensor
Flow Sensor	Q mean	$Q_{low} = Q_{mean} - n\sigma$	$Q_{high} = Q_{mean} + n\sigma$	Leakage / Burst	Flow Sensor
TDS Sensor	300-600 ppm	<300 ppm	>900 ppm	>1200 ppm	TDS Sensor
Ultrasonic Sensor	15-85% level	<15%	>85%	Over flow>95%	Ultrasonic Sensor
DHT11 Temp	20-35°C	<18°C	>40°C	>50°C	DHT11 Temp
DHT11 Humidity	30-70%	<20%	>75%	>85%	DHT11 Humidity



B. Dataset Origin Statement

This dataset is used for training the Decision Tree classifier and was generated from a mixed source. A total of 30 samples were collected, consisting of both real-time sensor readings from the implemented hardware setup and simulated data points created to cover all boundary and abnormal operating conditions. The real-time part of the dataset was captured using the sensor modules. This mixed dataset allowed the model to learn patterns across normal, transition, and extreme

cases, improving its robustness and accuracy during live prediction.

CALCULATION OF RISK%

$$\text{Risk} = \frac{\text{TDS} - 300}{600} \times 100$$

With limits,

If $\text{TDS} \leq 300$, Risk = 0%

If $\text{TDS} \geq 900$, Risk = 100%

Table-3. Dataset of 15 cases and respective output values.

Cases	Temp C	Humidity %	Level %	TDS ppm	Flow (L/min)	Class	Risk %
1	25.8	48	72.1	180	6.2	Good	0.0
2	28.4	55	65.4	260	5.9	Good	0.0
3	32.1	62	54.8	340	7.1	Moderate	6.7
4	30.5	70	41.2	410	8.5	Moderate	18.3
5	35.2	68	36.7	455	9.7	Moderate	25.8
6	33.9	73	48.6	510	6.8	Moderate	35.0
7	37.4	76	33.1	580	10.4	Moderate	46.7
8	34.7	65	27.4	620	7.3	Poor	53.3
9	36.9	78	22.8	690	9.1	Poor	65.0
10	39.1	81	19.7	730	11.0	Poor	71.7
11	31.5	58	56.2	810	8.0	Poor	85.0
12	28.9	63	44.1	865	12.4	Poor	94.2
13	33.0	72	39.6	900	9.9	Poor	100.0
14	29.7	60	52.3	295	6.5	Good	0.0
15	26.4	50	78.9	315	5.4	Moderate	2.5

Table-4. Data set composition for machine learning model.

Parameter	Description
Total data set size	15 samples
Real-time samples	10 samples collected from actual sensor readings
Simulated samples	5 samples generated to cover Boundary and abnormal conditions
Data source type	Mixed (real+ simulated)
Sensor modalities used	TDS, ultrasonic, temperature, DHT11, flow rate
Labeling criteria	Based on BISIS 10500:2012 Limits for water quality
Output classes	Good, Moderate, Poor
Purpose of simulated data	To ensure class balance and include rare abnormal/ critical conditions.
Collection duration	Real-Time data recorded over multiple varying operating conditions.



Related Work and Proposed Model

Table-5. Sensors and physical variables.

Work	TDS	Temperature	Humidity	Water Level	Flow Rate
[1] Ultrasonic Water Level Monitoring (Gupta <i>et al.</i> , 2018)	-	-	-	Yes	-
[2] Smart Water Monitoring Using TDS+ IoT (Bhope <i>et al.</i> , 2019)	Yes Yes	(limited) Yes	- -	- -	- -
[8] Real-Time Water Quality IoT System (Sugiharto <i>et al.</i> , 2023) Proposed Work	Yes	Yes	Yes	Yes	Yes

Table-6. Parameter comparison.

Work(2023–2025)	Cloud/IoT	Alerts	Quality Class/WQI	ML/ AI	DigitalTwin	Risk/Output
[9] ML+IoT Review (2024)	Yes	-	Yes(ML-basedWQI)	Yes	-	-
[10] Deep Learning WQ Prediction(2024)	Yes	-	Yes	Yes(DL models)	-	-
[12] AI-Supported WSN Sensors(2024)	Yes	Yes	Yes	Yes	-	-
[15] Hybrid Sensors		Yes			Yes (Digital Twin modeling)	-
+Digital Twin	Yes		Yes	Yes		
(2025)					Yes(Real-Time Digital Twin)	Yes(Risk%+ abnormality score)
Proposed Work	Yes	(multi- parameter alerts)	Yes (Good/Moderate/Poor)	Yes(Deci(Decision Tree ML)		

5. CONCLUSIONS

The implemented system combines multiple sensor modules like TDS, DHT, water level, and flow rate into a single architecture that is capable of monitoring and evaluating the conditions of the water and evaluating the quality in real time. The results of the simulation are shown in LCD readings; waveforms are obtained from the IoT platform, which provides a reliable response that divides the categories into “Good”, “Moderate,” and “Poor” whenever TDS levels or temperature levels cross critical values. By doing so, this work fills the practical gap between basic sensing and intelligent interpretation, which provides not only data but actionable insight. Considering overall parameters, the integration of hardware sensing, visualization, and machine learning prediction shows a dependable approach for continuous water quality observation. Hence, this work provides a solid foundation for expanding toward larger-scale implementation, which automates control systems, or even better predictive analytics in future implementation.

REFERENCES

- [1] Gupta and R. Kumar. 2018. Ultrasonic-Based Water Level Monitoring, International Journal of Engineering Science. Demonstrates a low-cost, non-contact ultrasonic method for accurate water-level measurement in tanks and reservoirs.
- [2] S. Bhope *et al.* 2019. Smart Water Quality Monitoring Using TDS and IoT, IJERT. Proposes an IoT-based system that tracks TDS in real time with cloud logging and alert mechanisms.
- [3] N. Mishra. 2000. Cloud-Based Water Flow Monitoring Using ESP8266, IJRSET. Implements an ESP8266 IoT node for continuous cloud- based water-flow monitoring and leakage detection.
- [4] R. Singh and P. Sharma. 2020. IoT-Based Water Quality Monitoring Systems, IEEE Sensors. Presents a multi-sensor IoT system measuring pH, turbidity, temperature, and conductivity with cloud support.
- [5] A. J. Ramadhan, A. Ali and H. Kareem. 2020. Smart Water- Quality Monitoring System Based on Enabled Real-Time IoT, Journal of Engineering Science and Technology. Develops a real-time IoT monitoring system for pH, turbidity, ORP, conductivity, and TDS with remote accessibility.



- [6] S. J. Orrell. 2021. Digital Twins in Water Management: A Review, Water Research Journal. Reviews Digital Twin applications in water systems for simulation, prediction, and improved infrastructure management
- [7] P. Teja. 2021. Digital Twins for Smart Infrastructure, Elsevier Smart Systems. Discusses Digital Twin applications for real-time monitoring, analytics, and predictive maintenance in smart infrastructure.
- [8] W. H. Sugiharto, H. Susanto *et al.* 2023. Real-Time Water Quality Assessment via IoT. Ingénierie des Systèmes d'Information. 28(4): 403-410.
- [9] M. Hasan. 2024. Water Quality Monitoring Using Machine Learning and IoT: A Review. Civil and Natural Resources Engineering Journal. 8(2).
- [10] A. Toumi. 2024. Harnessing Deep Learning for Real-Time Water Quality Prediction. Water. 16(23): 3380.
- [11] M. Forhad. 2204. IoT-Based Real-Time Water Monitoring in Treatment Plants. Heliyon.
- [12] Y. Singh. 2024. Smart Water Quality Monitoring with WSN and AI Sensors. Sensors. 24(9): 2871.
- [13] Y. Singh and T. Walingo. 2024. Smart Water Quality Monitoring with IoT Wireless Sensor Networks. Sensors. 24(9): 2871.
- [14] M. Frincu. 2025. Artificial Intelligence in Water Quality Monitoring: A Review of Water Quality Assessment Applications. Water Quality Research Journal. 60(1).
- [15] 2025. Water Quality Monitoring Using Hybrid Physical-Soft Sensors for River Digital Twins: A Comprehensive Review. Water Science and Technology.