Discover Electronics

Research

An IoT-enabled intelligent water quality monitoring system for tourist safety using machine learning

Md. Ashraful Islam[1] · Sujan Chandra Roy[2] · Md. Fahad Ullah Utsho[1] · Laila Naznin[1] · Ripa Sarkar[3] · Ratna Rani Sarkar[4]

Received: 20 November 2024 / Accepted: 16 May 2025
Published online: 27 May 2025

Abstract
Recently, Bangladesh has been seen as a popular destination for visitors seeking to see its unparalleled natural beauty, including the world's longest sea beach at Cox's Bazar and the historic Silk Route city of Rajshahi. However, the availability of safe drinking water remains a critical concern, as contamination from industrial discharge, agricultural runoff, and inadequate sanitation infrastructure pose significant health risks to visitors. This study presents an intelligent Internet of Things (IoT) -based water purity monitoring system that combines real-time sensor data with machine learning (ML) for predictive analysis to guarantee safe drinking water in Bangladesh's tourist areas. The proposed system employs four key water quality sensors (pH, turbidity, TDS, and temperature) connected to an ESP32 microcontroller with Wi-Fi capabilities, enabling wireless data transmission to a centralized IoT server. We collected and analysed 3,178 water samples from high-traffic tourist regions, comparing results against WHO and Bangladesh safety standards. Our findings reveal that 51.5% of samples met safety thresholds, while 48.5% were contaminated, emphasizing the need for continuous monitoring. To enhance predictive accuracy, we evaluated five ML models: Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), Bagging Decision Tree (BDT), and Voting Classifier Decision Tree (VCDT). Among these, ANN achieved the highest accuracy (92.66%) in classifying water quality, followed by RF (84.28%) and BDT (83.02%). The system provides real-time alerts to tourists and local authorities when water quality deteriorates, enabling immediate corrective actions.

Keywords IoT · Water purity · EPS32 · ML technique · Sensors

1 Introduction

The global challenge is to provide clean and safe drinking water, an indispensable resource for all life forms on our planet and immensely important for the economy, ecology, and human well-being. However, pure drinking water is still not accessible to billions of people worldwide. According to the World Health Organization (WHO), more than 2 billion people worldwide consume water that is contaminated with fecal matter, which can lead to serious health risks, including cholera, dysentery, and other waterborne illnesses [1]. Although the United Nations (UN) Sustainable Development Goal (SDG) 6 emphasizes the importance of providing everyone with clean water and sanitation by 2030, there has been a lack of progress in developing countries due to pollution, climate change, and inadequate infrastructure [2].

Water contamination can occur from diverse sources, such as industrial waste (with heavy metals and chemicals), agricultural runoff (with pesticides and nitrates), urban sewage discharge (with pathogens and organic pollutants), and

✉ Md. Ashraful Islam, ras_ice@ru.ac.bd | [1]Department of ICE, University of Rajshahi, Rajshahi 6205, Bangladesh. [2]Department of CSE, Kishoreganj University, Kishoreganj 2300, Bangladesh. [3]Department of CSE, Uttara University, Dhaka, Bangladesh. [4]Department of Computer Science, National University, Gazipur, Bangladesh.

natural contaminants (arsenic and fluoride in groundwater). The consequences are devastating: Diarrheal diseases result in 485,000 deaths annually, particularly among children less than five years old. Additionally, prolonged exposure to pollutants such as arsenic has the potential to cause cancer, cardiovascular diseases, and developmental disorders [3, 4].

Bangladesh, known for its rivers and natural beauty, is experiencing a serious threat from water pollution caused by rapid industrialization, untreated sewage, and the excessive use of fertilizers. Arsenic and heavy metals are found in 20% of groundwater sources [5, 6]. The problem is particularly severe in tourist areas, where millions of visitors annually face higher risks because they depend on untreated natural or municipal water supplies. Contaminated water can not only harm tourists who are not familiar with the local conditions, but it also undermines Bangladesh's tourism economy, which contributes 4.4% of GDP. According to research, 62% of travelers prioritize health, safety, and clean water when picking destinations, emphasizing the need for reliable drinking water monitoring systems [7, 8]. This situation demonstrates the importance of effective water quality monitoring systems that can provide real-time data to guarantee drinking water safety, particularly in tourist regions.

Existing traditional methods of water quality assessment, such as laboratory testing, are time-consuming, costly, and often impractical for real-time monitoring [9, 10]. These methods require manual sample collection and analysis, which can delay the detection of contaminants and compromise public health. Therefore, a faster and cheaper solution is needed due to the horrifying consequences of water pollution.

Recently, the Internet of Things (IoT) and Machine Learning (ML) are two emerging technologies that have enabled cross-domain researchers to develop various systems beneficial to human life, including water quality prediction. The use of different devices in an IoT real-time package allows data to be transmitted and passed without human intervention. The IoT system provides advantages over conventional techniques due to its reduction of costs, efficient travel and data collection, adaptability over various system phases, and ability to predict optimal values for inaccessible locations [11, 12]. In contrast, ML models can analyze collected data, identify trends, and produce accurate forecasts, even with limited resources [13, 14]. Due to this, this type of system is particularly advantageous for developing countries such as Bangladesh, where resources are limited and efficient water management is crucial.

Therefore, this paper presents an efficient, intelligent system for accessing water quality that simplifies traditional WQ monitoring methods by integrating IoT and ML approaches. The main goal of this proposal is to ensure safe drinking water access for tourists. The effectiveness of the proposed system is evaluated through a microcontroller (MCU) and four sensors: total dissolved solids (TDS), turbidity, pH, and temperature sensors. Data is collected from several tourist locations in Bangladesh and transmitted directly to a server PC for storage. The data is then used to prepare the dataset for predicting WQ. At first, the collected data is statistically analyzed to ensure the water quality level and then processed through the framework's application of machine learning models to predict safe drinking water in real time. Secondly, several ML models are used in order to evaluate the data, which include Decision Tree (DT) [15], Random Forest (RF) [16], Artificial Neural Network (ANN) [17], Bagging Decision Tree (BDT) [18], and Voting Classifier Decision Tree (VCDT) [19]. These models ensure that the assessment of WQ is highly accurate and precise. The results demonstrate that ANN has the best prediction accuracy for WQ with a higher accuracy (92.66%) than other models after extensive experimentation. Thus, the use of advanced IoT and ML techniques in this framework provides actionable insights that allow decision-makers and policymakers to take proactive measures for efficient water quality management.

The remainder of this paper is organized as follows: Sect. 2 describes the related work, while Sect. 3 describes the methodology used in this study. Section 4 presents the proposed IoT-based water purity identification system with hardware components and IoT server configuration. Section5 and 6 discuss the results and discussion of the water quality analysis and their implications for public health and environmental sustainability. Finally, Sect. 6 concludes the paper with a summary of the findings and recommendations for future research.

2 Literature review

Several studies have explored the use of IoT for water quality monitoring, but many of these systems have significant limitations. For instance, [20] proposed a GSM-based system for monitoring pH, temperature, and conductivity, but it was limited by its inability to store large amounts of data for further analysis. Similarly, [21] developed an automated water quality monitoring system that only measured two parameters, which is insufficient for a comprehensive assessment of water safety. In [22], the authors survey the tools and techniques used in existing water quality monitoring systems. Other systems, such as those proposed by [23] and [24], relied on proprietary applications like BLYNK, which restricted their usability to specific users and limited their scalability.

Moreover, many existing systems focus on monitoring water quality in controlled environments, such as laboratories or industrial settings, and do not address the specific challenges of monitoring water in natural or tourist areas, where water quality can vary significantly due to environmental factors. For example, [25] proposed a cloud-based water quality monitoring system using IoT sensors. However, it did not account for the dynamic nature of water quality in natural sources like rivers or lakes. Similarly, [26] highlighted the challenges of IoT-based water quality monitoring in domestic applications but did not provide a solution tailored to the needs of tourists or rural communities.

A comprehensive review article has been conducted, meticulously analyzing the literature on water quality and presenting many conclusions regarding the challenges, issues, and research gaps that have emerged over the past five years (2018–2022) [27]. In this context, an Internet of Things (IoT)-based water quality system has been proposed, featuring an efficient prediction method that utilizes machine learning techniques to forecast water quality on a large scale, thereby supporting informed decision-making in smart water quality monitoring systems within smart cities [28]. Additionally, a hybrid machine learning and embedded IoT water quality monitoring system has been introduced [29], and a review focusing on IoT-based solutions for monitoring domestic water quality [30]. Furthermore, an IoT-enabled multi-level system has been developed to monitor industrial wastewater quality, tracking parameters such as temperature, turbidity, total dissolved solids (TDS), and pH levels. This system highlights alarming pollution levels and underscores the necessity for targeted interventions to safeguard public health and ecosystems [31]. An overview of IoT applications in aquaculture concerning water quality monitoring has also been demonstrated [32]. Moreover, an automated water quality monitoring system that integrates cloud technology and machine learning algorithms has been introduced, encompassing various sensors such as pH, temperature, turbidity, and conductivity sensors [33]. A cost-effective smart water quality monitoring system utilizing IoT has been proposed [34], incorporating multiple sensors to measure essential parameters like pH, turbidity, water level, temperature, and humidity. The system's Microcontroller Unit (MCU) interfaces with these sensors, with further data processing taking place on a Personal Computer (PC). The collected data is then transmitted to the cloud using the IoT-based ThinkSpeak application for continuous water quality monitoring. Additionally, an innovative real-time water quality monitoring system specifically designed for rural areas has been introduced, focusing on assessing water resource quality parameters [35]. This solution is solar-powered, waterproof, portable, and IoT-enabled, utilizing Long Range Wide Area Network (LoRaWAN) technology. A smart system aimed at ensuring clean water quality through efficient wastewater treatment has also been proposed [36], complemented by the introduction of the Smart Wastewater Intelligent Management System (SWIMS) [37], which monitors and controls both inlet and outlet flows alongside water quality, functioning as a cyber-physical system (CPS) rooted in an Environmental Internet of Things (EIoT) platform. Lastly, a solution employing Principal Components Analysis (PCA) for fault detection in a simulated wastewater treatment plant (WWTP) has been presented [38].

The integration of IoT technologies into water quality monitoring has been widely studied, but most existing systems focus on specific applications, such as industrial wastewater monitoring [39] or aquaculture [40]. While these systems have demonstrated the potential of IoT for water quality monitoring, they often lack the flexibility and scalability needed for broader applications, such as monitoring water quality in tourist areas. For example, [41] proposed an IoT-based system for monitoring water quality in rivers, but it was limited by its reliance on expensive sensors and complex data processing algorithms. Similarly, [42] developed a machine learning-based system for predicting water portability, but it required extensive computational resources, making it unsuitable for real-time monitoring in remote areas.

Until now, there have been no studies on predicting the real-time WQ in the tourist area by integrating an edge-computing enabled ML architecture, which would provide valuable insights for WQ management authorities. To address these gaps, this paper proposes a scalable, intelligent IoT-based water purity identification system tailored for tourist areas in Bangladesh. The system integrates four critical parameters—pH, TDS, turbidity, and temperature—using an ESP32 microcontroller with Wi-Fi capabilities. By combining IoT and ML for comprehensive parameter analysis, this system advances IoT applications in water quality monitoring, offering a practical solution to Bangladesh's urgent public health challenges.

3 Methodology

This paper proposes a real-time IoT-based water quality identification system that integrates sensors to collect and analyse water purity data. The methodology comprises three core components: water purity evaluation parameters, data collection protocols, and water purity classification criteria.

3.1 Water quality parameters

Four primary parameters are used to evaluate water purity, aligned with WHO and Bangladesh standards [1, 43, 44]. The first parameter to evaluate the water purity level is pH, which measures the acidic or alkaline (basic) level of the water. It is used to quantify the concentration of hydrogen ions in a solution. The pH level of drinking water is a critical parameter that influences its quality and safety [40]. pH value of water is varied from 0 to 14, with the value of 7 considered neutral. According to Bangladesh standards and WHO guidelines, the standard pH level ranges from 6.5–8.5 [43, 44]. Acidic water (pH < 6.5) can have a sour taste, and alkaline water (pH > 8.5) may taste bitter or soapy and can result in pollution in water.

The second parameter is Turbidity. It is a crucial factor in evaluating potable water quality [41]. It mainly refers to the relative clarity of water, which is affected by the presence of microscopic particles such as clay, silt, and other delicate undissolved matter. These particles scatter light, making the water appear cloudy or murky. Turbidity is measured in Nephelometric Turbidity Units (NTU) using specialized optical equipment like nephelometers or turbidity meters. The standards in Bangladesh and WHO recommendations specify that the turbidity level should not exceed 5 NTU [1, 44]. High turbidity levels can harbor pathogens, bacteria, and other contaminants, posing health risks.

The next evaluating parameter is Total Dissolved Solids (TDS). It quantifies the total amount of inorganic and organic substances that are dissolved in water, encompassing minerals, salts, metals, and other impurities. Comprehending TDS is crucial to guarantee the safety and taste of drinking water [42]. TDS is measured in parts per million (ppm) or milligrams per liter (mg/L). According to both Bangladeshi norms and WHO recommendations, the acceptable TDS value is a maximum of 500 ppm [1, 44]. While TDS itself is not necessarily harmful, elevated levels can indicate the presence of harmful contaminants such as heavy metals, nitrates, and sulfates. High TDS levels can lead to health issues such as kidney stones, gastrointestinal problems, and mineral imbalances.

The last and fourth evaluating parameter is Temperature. The Temperature of drinking water is a critical factor that affects both the quality and safety of the water. As the Temperature of water increases, the ability of minerals and salts to dissolve increases, resulting in higher levels of TDS. Elevated temperatures further stimulate the proliferation of micro-organisms, which might potentially lead to an increase in cloudiness and the likelihood of pathogen infection [41]. In addition, variations in Temperature influence the solubility of gases such as carbon dioxide, which in turn affects the pH of water by modifying its acidity or alkalinity [40]. Elevated water temperature can expedite chemical reactions and increase the rate of corrosion in distribution systems, thus impacting the overall quality, scent, and safety of drinking water [42].

3.2 Data collection

Figure 1 illustrates the study area map of Bangladesh, highlighting key tourist regions where water samples were collected. The map showcases diverse geographical features, ranging from urban centers to coastal and riverine environments, underscoring the complexity of water quality assessment across these areas. Traditional methods that involve manual sample collection and laboratory analysis are both time-consuming and limited in scalability. The results can take 48–72 h, making them ineligible for mobile tourist populations. To overcome these challenges, we propose an IoT-based water quality monitoring system that automates data collection and real-time analysis. Our methodology involves:

- Strategic sensor deployment across five high-traffic tourist regions (Cox's Bazar and Chittagong, Dhaka and Comilla, Rajshahi, Bogra and Rangpur, Kustia and Mujibnagar), covering natural sources (rivers, tube wells), municipal supplies (tap water), and commercial sources (bottled water, hotel/restaurant reservoirs).
- High-resolution sampling: Over 3,178 water samples [45] were analyzed to calibrate and validate the system, ensuring robustness across heterogeneous water sources.
- Continuous multi-parameter monitoring: The system tracks critical indicators—temperature, pH, turbidity, and TDS—with wireless data transmission to a centralized hub for instantaneous purity assessment.

This approach eliminates manual bottlenecks, providing a scalable framework for real-time water safety evaluation—a critical advancement for public health and tourism sustainability in developing regions.

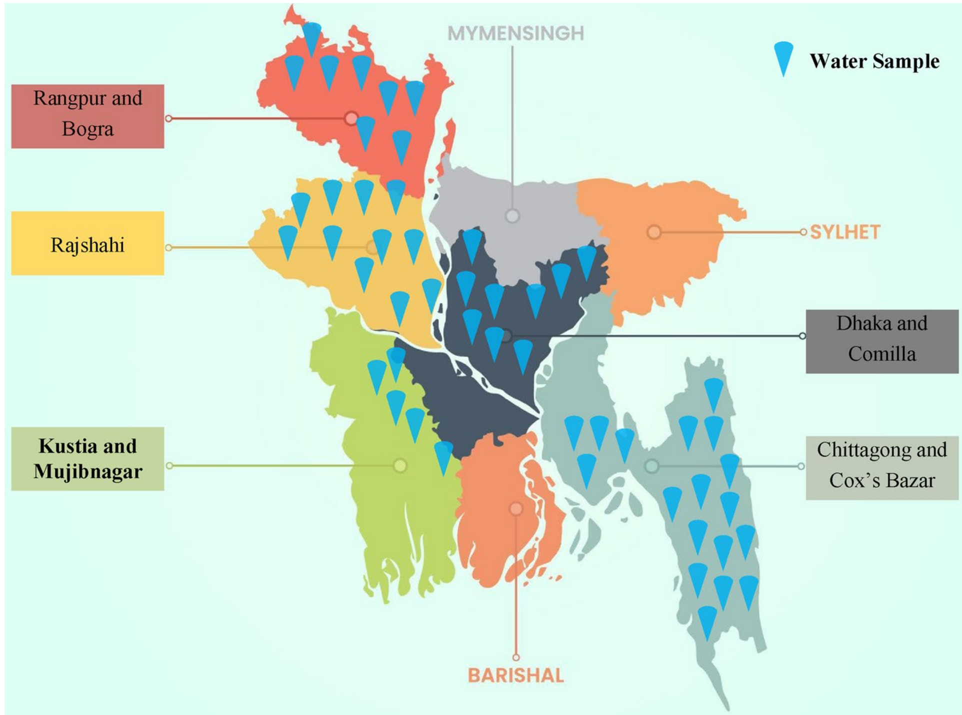


Fig. 1 Geographical locations of the study area

3.3 Water Purity Classification

Water samples are classified into three categories based on parameter thresholds (Table 1):

- **Pure**: All parameters within safe ranges; safe for drinking.
- **Polluted**: Parameters exceed safe limits; unsafe for consumption.

4 Proposed IoT-based water purity identification system

Figure 2 presents the system architecture of the proposed Intelligent IoT-enabled water purity monitoring framework while Table 2 provides an overview of the necessary equipment. The system employs four essential sensors such as TDS, temperature, pH, and turbidity—interfaced with an ESP32 MCU. The MCU digitizes the analog sensor outputs and transmits the data to an IoT server in real time. To ensure data reliability, the server performs continuous acquisition while filtering anomalous readings and outliers. Sensor data is aggregated every four seconds and subjected to statistical pre-processing before being passed to a machine learning model.

4.1 IoT server configuration

Figure 3 depicts the configuration of the IoT server responsible for wireless data acquisition from the sensor network described previously. Implemented on a personal computer, the IoT server facilitates the reception, storage, and processing of sensor data in real time. The system captures readings at one-second intervals, automatically recording the information into a CSV file stored locally on the server. Upon collecting the required number of data points, data acquisition is halted, and the dataset is forwarded to a statistical model for classification. The model outputs a categorical assessment, where a

Table 1 Water purity classification criteria

Category	pH	TDS (ppm)	Turbidity (NTU)
Pure water	6.5–8.5	50–500	Turbidity < 5
Polluted water	pH > 8.5	TDS > 500	Turbidity > 5

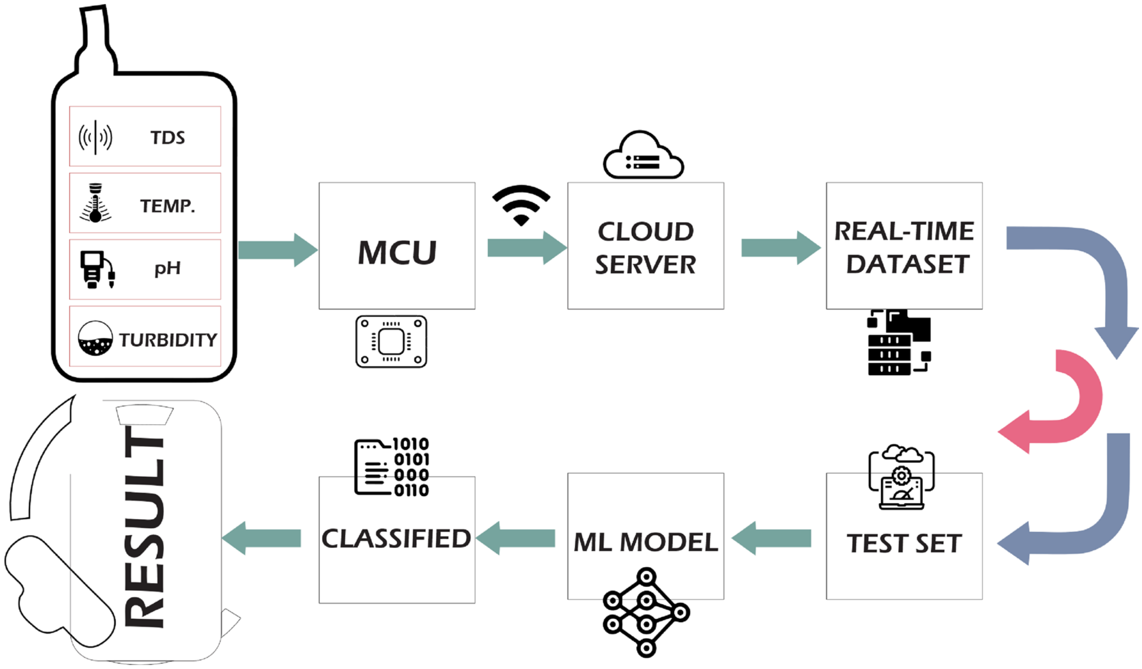


Fig. 2 IoT and ML implementation system for water purity identification

Table 2 Equipment for the proposed system

Name	Amount	Model No
Ph Sensor	1	RBD-0523
TDS Sensor	1	PRD-00003442
Turbidity Sensor	1	SEN-00205
Temperature Sensor	1	DS18B20
Control Unit	1	ESP32
PC	1	Pentium IV
Connectors	Few	–
Board	3	–
Plastic Bottle (for water sample collection)	100	–
Name	Amount	Model No

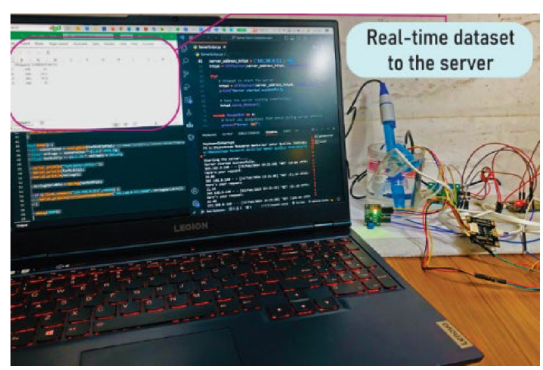


Fig. 3 Real-time data transfer to the IoT server

"NO" label indicates that the water is contaminated and unfit for consumption, while a "YES" label may correspond to either fully pure or moderately pure water. The use of a dedicated local IoT server enhances data security by maintaining all sensitive information within the internal network. Furthermore, local hosting significantly reduces latency, improves real-time processing capabilities, and ensures a smooth and efficient system operation.

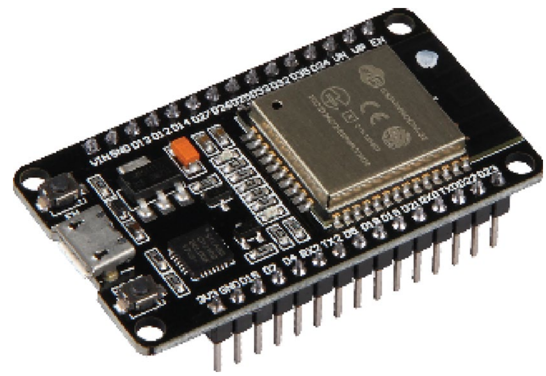


Fig. 4 Multifunctional control unit (EPS32)

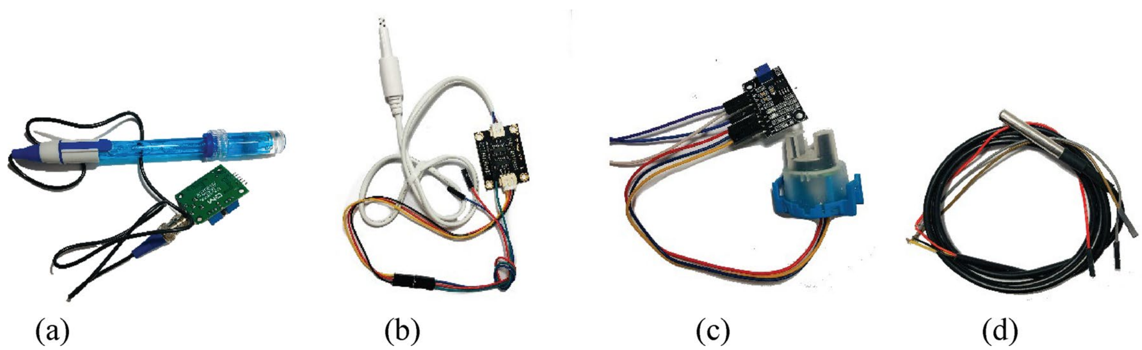


Fig. 5 Different sensors for data collection; **a** Analog pH sensor, **b** Turbidity sensor, **c** TDS sensor, **d** Temperature sensor

4.2 Hardware equipment of the proposed system

4.2.1 Control unit (EPS32)

Figure 4 illustrates the ESP32 control unit, a versatile and cost-effective microcontroller developed by Espressif Systems. Widely adopted in IoT and embedded system applications, the ESP32 is equipped with integrated Wi-Fi (IEEE 802.11 b/g/n) and Bluetooth 4.2 (including both Classic and BLE), offering robust wireless communication capabilities. At its core, the ESP32 features a dual-core Tensilica Xtensa LX6 processor operating at up to 240 MHz, delivering significant computational performance for real-time processing tasks. The device includes 520 KB of SRAM and typically supports up to 4 MB of onboard flash memory, with some variants allowing for external SPI flash expansion. These features collectively make the ESP32 an ideal platform for real-time data acquisition and processing in IoT-based water monitoring systems [46].

4.2.2 Sensors for data collection

The proposed system features four advanced sensors dedicated to assessing water quality:

- **Analog pH Sensor:** The analog pH sensor is engineered to quantitatively assess the pH level of a water sample, effectively determining its acidity or alkalinity. As depicted in Fig. 5a, this pH sensor is pivotal for accurate readings. Calibration is performed using a pH 7 buffer solution, followed by a rinsing process with distilled water. For enhanced precision, a secondary calibration is conducted with either a pH 4 or pH 9.21 buffer solution. This sensor is extensively utilized in both drinking water quality evaluations and environmental monitoring [47–49].
- **Turbidity Sensor**: The turbidity sensor gauges water clarity by examining light transmittance and scattering, which fluctuate in relation to the concentration of total suspended solids (TSS). Figure 5b showcases the turbidity sensor implemented in this system. Increased levels of TSS lead to higher turbidity, which signifies a decline in

water quality. This sensor finds common applications in:—Monitoring drinking water quality—Assessing rivers and streams—Wastewater treatment processes—Researching sediment transport [47, 50].

- **Analog TDS Sensor:** The Total Dissolved Solids (TDS) sensor plays a crucial role in measuring the concentration of dissolved ions in water, providing essential insights into its purity and overall quality. As depicted in Fig. 5c, this sensor is widely applied in various settings, including residential water testing, hydroponic systems, and industrial water monitoring. To ensure optimal accuracy, the TDS sensor is calibrated using a standard buffer solution with a conductivity of 1413 µS/cm [47, 49]. The advantages of utilizing this sensor are manifold: it offers a cost-effective solution for water quality assessment, features user-friendly operation for individuals at all skill levels, and guarantees reliable real-time monitoring of TDS levels [46]. By integrating the TDS sensor into water quality management practices, users can make informed decisions that enhance the safety and efficiency of water usage.
- **Temperature Sensor (DS18B20):** The DS18B20 temperature sensor is designed for accurate water temperature measurement over a wide range, with an accuracy of ± 0.5 °C. Figure 5d illustrates this sensor. It operates within a measurement range of –55 °C to + 125 °C and is widely used in environmental monitoring and industrial water quality management [46].

4.3 ML implementation

The objective of predictive modeling for classification is to approximate a mapping function (type of classifier) from input variables (data) to discrete output variables (classification). Before employing ML models to analyze the data, the data was prepared as input for the model by dividing them into training and testing sets to train and evaluate its performance. In addition, the dataset was cleaned by eliminating any errors and replacing empty cells with the median of the dataset's input variables. Figure 6 depicts the proposed methodology for the current study. The following subsections describe each component of the proposed system's framework and its implementation.

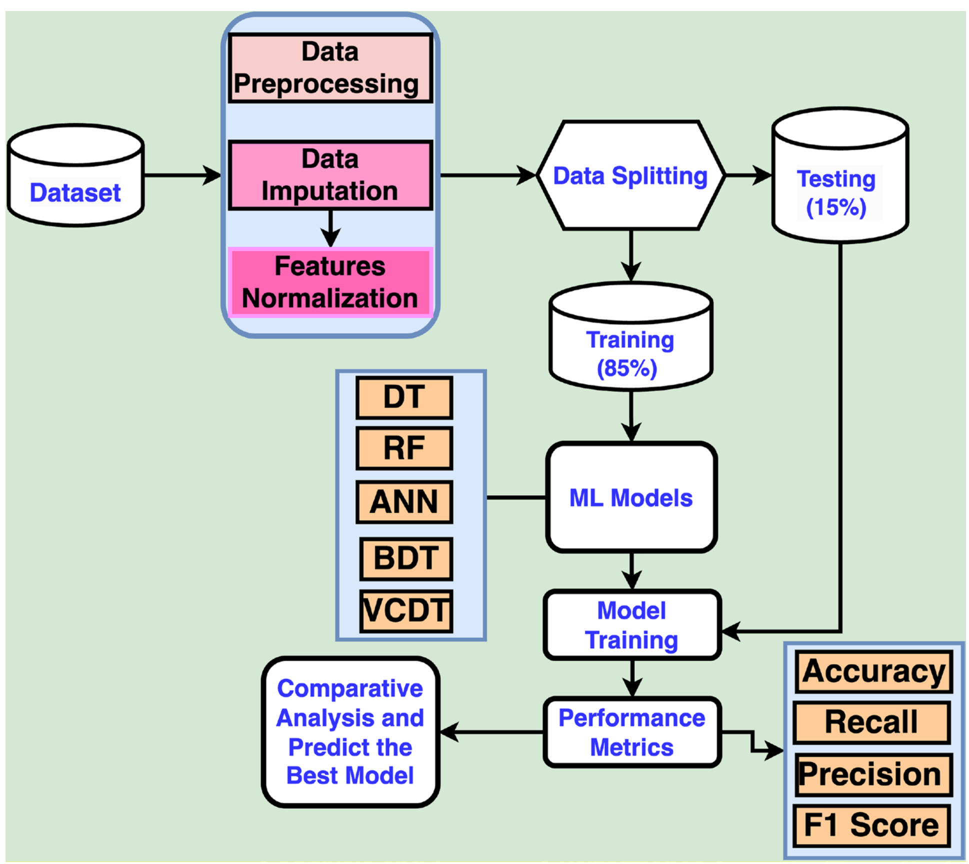


Fig. 6 Machine learning working procedure

4.3.1 Dataset description and processing

The drinking water quality data was gathered from different tourist destinations, including Rajshahi, Dhaka, and Chittagong Division of Bangladesh, and a total of 3,178 samples [45] were collected and examined for each parameter. In addition, the data was preprocessed to enhance its quality. Data preprocessing involves all the necessary steps to avoid irrelevant entries, which includes cleaning (i.e., removing outliers) and labeling the data. To address missing data that constituted less than 3% of the records, forward and backward filling, a widely accepted method for maintaining temporal continuity, was applied. Afterwards, RobustScaler was applied to standardize the data, as it is particularly effective for datasets with outliers. These preprocessing steps provide the foundation for an effective model implementation, enhancing the overall accuracy and robustness of the system. Proper use of these techniques contributes significantly to the model's ability to learn efficiently from the data.

The exploratory stage of our analysis is heavily reliant on graphical visualization to uncover insights and patterns in the CSV dataset. To begin with, heatmap is employed to display Spearman correlation coefficients between different water quality parameters, which indicate the strength and direction of monotonic relationships, as depicted in Fig. 7. In the heatmap, the color intensity of WQ parameters in both rows and columns represents the correlation strength between the intersecting parameters.

4.3.2 Data splitting

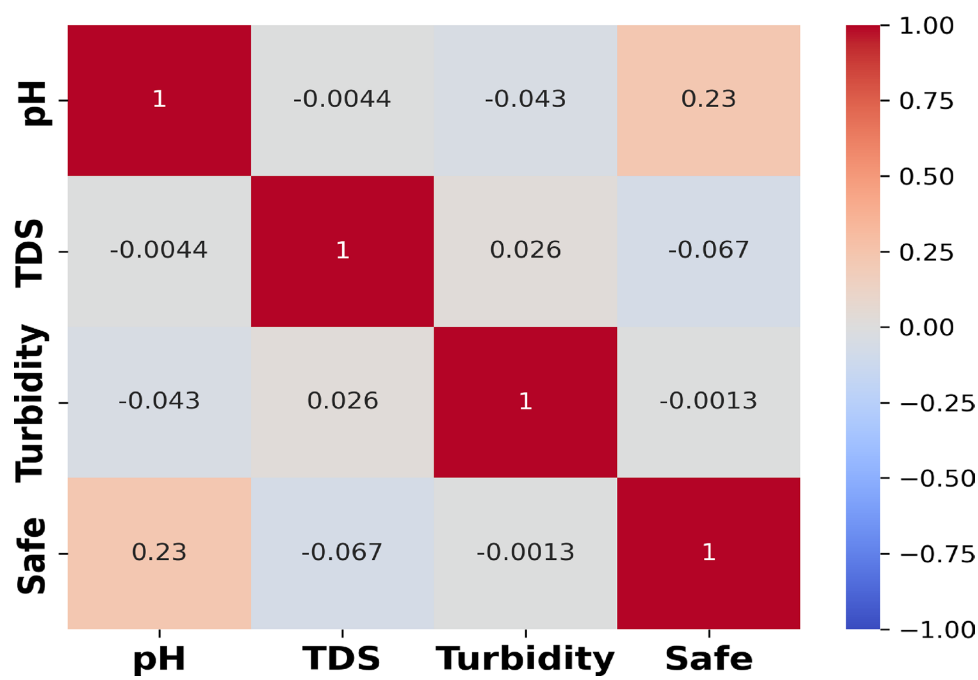
The dataset was separated into two groups: a training set and a test set. 85% of the dataset was allocated to the training set, while the other 15% was used for the test set. In order to predict or select an alternative, the independent and dependent parameters are correlated by the ML model. The test data is then used to evaluate the effectiveness of the machine learning technique.

4.3.3 ML models

The development process for the water quality prediction model involves learning the patterns in the water quality dataset and constructing classification models using ML models such as DT, RF, ANN, BDT, and VCDT.

- **DT:** DT is a non-parametric supervised learning algorithm that is made up of a hierarchical tree structure with a root node, branches, and leaf nodes. It relies on entropy to determine the root variable, which is then orientated towards

Fig. 7 Pearson correlation matrix analysis



other attribute values. DTs are known for their simplicity, high performance accuracy (even with fewer than a thousand data points), and the ability to categorize all testing samples. A decision tree approach can be used to derive decision rules by partitioning the sample dataset using the highest entropy samples [15].

- **RF**: In RF, base models are employed on different subsets of data and decisions are made on the basis of all models. The base model in RF is typically a decision tree, which carries all the benefits of a decision tree with the added efficiency of using multiple models [16].
- **ANN**: ANN models can be used as robust machine learning algorithms for time series prediction in various engineering applications, making them a powerful computational method for developing various real-world medical applications. In an ANN model, there are three layers: one is the input layer, one is the hidden layer, and one is the output layer. Each layer has weight and bias parameters that are used to manage neurons. An activation function transfers data from the hidden layer to the output layer [17]. This study employed the ANN algorithm to determine water quality. Three key layers are involved in ANNs: input, hidden, and output. Four hidden layers were utilized to transfer input training data from the input to the output through the relay function. However, the output layer consisted of two classes.
- **BDT**: BDT is a classifier that utilizes decision trees as baseline learners and applies multiple base classifiers to random subsets of data, resulting in an averaged prediction that greatly aids in variance [18].
- **VCDT**: The VCDT is a machine learning model that trains on an ensemble of various models and predicts an output (class) based on the most probable class to select. The goal is to develop a single model that is trained by these models and predicts output based on the combined majority of voting for each output class, rather than creating separate models for individual teams. There are two types of voters available: hard and soft. In hard voting, the outcome is determined by the majority of votes, but in soft voting, it is determined by the average of the votes. The reason for selecting ensemble learning of people is because it has a low error rate and low overfitting. To enhance the model, we have employed the growing and pruning method in ensemble learning. The process of growing involves adding models to the ensemble to enhance accuracy and pruning members from a fully defined ensemble to decrease the model or computational complexity of the ensemble with little or no impact on its performance [19].

4.3.4 Performance metrics

The evaluation model for MLR techniques for water quality prediction includes four elements: TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). TP is the process of correctly identifying a positive sample, while TN is the process of correctly predicting a negative sample. FP happens when a negative sample is mistakenly assumed to be positive. FN happens if a positive sample is mistakenly assumed to be negative (or when a true positive is missed by the model). The model's performance was evaluated using the Eqs. (1) to Eq. (4), which correspond to accuracy, F1-score, recall, and precision, respectively [51, 52].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$precision = TP/TP + FP \tag{2}$$

$$Recall = Tp/TP + FN \tag{3}$$

$$F1Score = 2 * precision * Recall/(Precision + Recall) \tag{4}$$

5 Results

5.1 Experimental setup

Figure 8 presents a detailed visual representation of the physical experimental setup, showcasing the system in action. This setup has been meticulously designed to continuously monitor and assess various water purity parameters in real time. We utilized Visual Studio to develop these simulations with the Python programming language. The hardware setup consisted of an Intel Core i7-11 th generation processor and 16 GB of RAM.

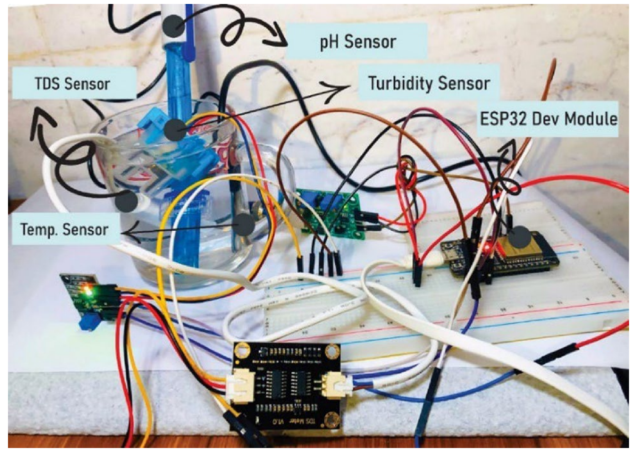


Fig. 8 Experimental setup of the IoT-based water purity identification system

5.2 Experimental result based on statistical analysis

The statistical result for water purity levels at tourist places in Bangladesh is shown in Fig. 9, with 0 representing safe water and 1 representing unsafe water. This figure ensures the water purity level based on the combined values of pH, TDS, Turbidity, and temperature of each water sample. It is noticeable from this figure that the study categorizes 51.5% of samples as pure and 48.5% as polluted. The 48.5% of polluted samples necessitate urgent remediation, particularly in areas with high tourist traffic. The IoT-based system effectively detects contamination in real-time but encounters several barriers to widespread adoption:

- **Cost:** The initial investment required for deployment (including sensors, microcontrollers, and IoT servers) may be prohibitive for resource-constrained regions. Although ESP32 microcontrollers are relatively affordable, scaling the solution across multiple locations can significantly increase costs.
- **Maintenance:** Regular calibration and technical maintenance of sensors present challenges in remote areas, where expertise may be lacking.
- **Accessibility:** The reliance on internet connectivity can restrict functionality in regions with poor service coverage. Additionally, setting up local IoT servers for data storage may overwhelm infrastructure that is already inadequate.

5.3 ML results

In order to predict water quality, this study utilizes a minimal set of low-cost sensor measurements, specifically temperature, turbidity, pH, and TDS. The Artificial Neural Network (ANN) is the best classification algorithm, with a

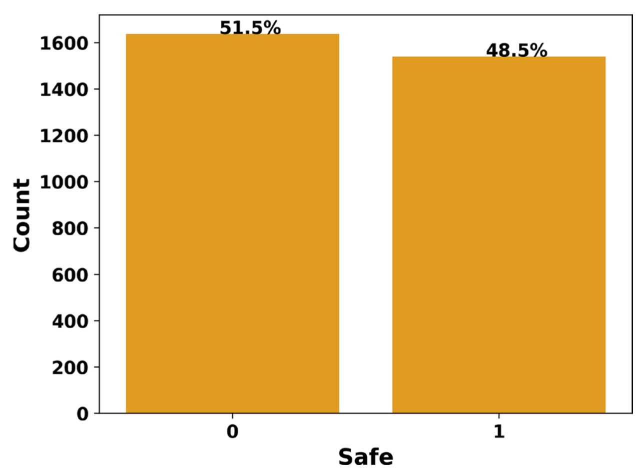


Fig. 9 Overall statistical result

performance score of 92.66%, precision of 93.33%, recall of 92.87%, and an F1-score of 92.65% than other algorithms as demonstrated in Table 3.

ANN's superior performance is likely due to its ability to capture complex, non-linear relationships between water quality parameters, which traditional tree-based methods such as RF, BDT, and DT may not be able to model effectively. Moreover, the multi-layer architecture of ANN allows for better feature abstraction, while ensemble methods like RF and BDT, although robust, may not be as generalizing with limited input features. The Voting Classifier, which combines multiple models, may also underperform if its base learners (e.g., DT) are less accurate. Thus, ANN's adaptability and superior pattern recognition make it particularly effective for water quality prediction tasks.

Table 4 shows the general structure of a confusion matrix, where'0'indicates that the water is safe and'1'indicates that it is not safe. Table 5 displays confusion matrices that are produced by applying the dataset to each of the algorithms presented in Sect. 3.

6 Conclusion

This study presents the design and implementation of a real-time Intelligent IoT-based water purity identification system aimed at enhancing drinking water safety in tourist regions of Bangladesh. By integrating low-cost, high-precision sensors with an ESP32 microcontroller, the system effectively monitors key water quality parameters (pH, TDS, turbidity, and temperature) and transmits data wirelessly to a local IoT server for real-time processing and classification. Analysis of 3,178 water samples revealed that 51.5% of the sources were fully compliant with WHO and national drinking water standards, while 48.5% were identified as polluted and unsafe for consumption. These findings underscore the critical need for continuous monitoring, particularly in high-traffic tourist areas where water contamination poses significant public health risks. The system not only provides reliable, real-time data collection and remote monitoring but also facilitates timely interventions through automated classification and alert mechanisms. Its integration with machine learning algorithms further enhances predictive accuracy, with the Artificial Neural Network (ANN) model demonstrating superior performance across multiple evaluation metrics.

Future enhancements should focus on expanding the range of monitored parameters such as biological oxygen demand (BOD) and heavy metal content, incorporating predictive analytics to forecast contamination trends, and addressing deployment challenges related to cost, sensor maintenance, and connectivity in remote areas. In summary, this research contributes a robust, intelligent framework for IoT-enabled water quality monitoring, offering practical

Table 3 Performance metrics for the machine learning algorithms utilized in this study's dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
ANN	92.66	93.33	92.87	92.65
Random Forest	84.28	85.19	84.53	84.23
Bagging Decision Tree	83.02	84.01	83.29	82.96
Voting Classifier	83.23	84.49	83.53	83.15
Decision Tree	75.26	75.41	75.37	75.26

Table 4 Structure of the confusion matrix

Case	Predicted (0)	Predicted (1)
Actual (0)	TP	FP
Actual (1)	FN	TN

Table 5 Confusion matrix of machine learning model

Model	TP	FP	FN	TN
ANN	212	34	1	230
Random Forest	188	58	17	214
Voting Classifier	182	64	16	215
Bagging DT	184	62	19	212
Decision Tree	177	69	49	182

applications for public health protection and sustainable water management in developing regions. Continued advancements in IoT and machine learning technologies hold the potential to further improve system accuracy, accessibility, and impact.

Author contributions Md. Ashraful Islam, Sujan Chandra Roy, and Md. Fahad Ullah Utsho contributed to idea generation and implementation, assisted in literature review and article writing, and proofread and verified the content of the article. Laila Naznin, Ripa Sarkar, and Ratna Rani Sarkar assisted in evaluating the results and analyzing the findings.

Funding This research did not receive any dedicated financial support from any funding agency.

Data availability The data collected for this study is available "Dataset. [Online]. Available: https://github.com/Sujan-Roy/Dataset (Accessed: April, 2025)". The dataset includes raw sensor readings for the 3,178 water samples collected from tourist areas in Bangladesh.

Code availability The analysis file is available from the corresponding author by request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. World Health Organization. www.who.int. Accessed Apr 2024.
2. UN SDG 6 Synthesis Report on Water and Sanitation, www.unwater.org/sites/default/files/2023-08/UN-Water_SDG6_SynthesisReport_2023.pdf. Accessed Apr 2025.
3. Lin L, et al. Effects of water pollution on human health and disease heterogeneity: a review. Front Environ Sci. 2022;10:880246.
4. Hasan MK, et al. Water pollution in Bangladesh and its impact on public health. Heliyon. 2019. https://doi.org/10.1016/j.heliyon.2019.e02145.
5. Sarkar A, et al. The groundwater arsenic contamination in the Bengal Basin-a review in brief. Chemosphere. 2022;299: 134369.
6. Jaffar A, et al. The influence of physico-chemical parameters to determine water quality: a review. JEESR. 2020;17:116–21. https://doi.org/10.24191/jeesr.v17i1.016.
7. "World Travel & Tourism Council. Bangladesh Economic Impact Report," https://wttc.org/research/economic-impact. Accessed Apr 2025.
8. Robina-Ramírez R, et al. The trinomial health, safety and happiness promote rural tourism. BMC Public Health. 2023;23(1):1177.
9. Das B, Jain PC. Real-time water quality monitoring system using Internet of Things. In2017 International conference on computer, communications and electronics (Comptelix). 2017; 78–82. https://doi.org/10.1109/COMPTELIX.2017.8003942.
10. He D, Zhang LX. The water quality monitoring system based on WSN. 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet). 2012; 3661–3664. https://doi.org/10.1109/CECNet.2012.6201666.
11. Zimba S, Chembe C. Model for water quality monitoring based on internet of things. Zambia ICT J. 2022;6(1):44–51. https://doi.org/10.33260/zictjournal.v6i1.133.
12. Promput S, et al. Design and analysis performance of IoT-based water quality monitoring system using LoRa technology. TEM Journal. 2023;12(1):29–35. https://doi.org/10.18421/TEM121-04.
13. Uddin MG, et al. Robust machine learning algorithms for predicting coastal water quality index. J Environ Manage. 2022;321: 115923.
14. Nasir N, et al. Water quality classification using machine learning algorithms. J Water Process Eng. 2022;48: 102920.
15. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2(3):18–22.
16. Quinlan JR. Decision trees and decision-making. IEEE Trans Syst Man Cybern. 1990;20(2):339–46.
17. Maier HR, Dandy GC. The use of artificial neural networks for the prediction of water quality parameters. Water Resour Res. 1996;32(4):1013–22.
18. Breiman L. Bagging predictors. Mach Learn. 1996;24:123–40.

19. Akiba Y, et al. Turning majority voting classifiers into a single decision tree. In Proceedings Tenth IEEE International Conference on Tools with Artificial Intelligence (Cat. No. 98CH36294). 1998;224–230.
20. Gabhane MJP, et al. Smart Homes System Using Internet-of-Things: Issues, Solutions and Recent Research Directions. International Research Journal of Engineering and Technology (IRJET). 2017; 5(5).
21. Berawi MA. The fourth industrial revolution: managing technology development for competitiveness. Int J Technol. 2018;9(1):1–4. https://doi.org/10.14716/ijtech.v9i1.1504.
22. Kumar S, et al. Internet of Things is a revolutionary approach for future technology enhancement: a review. J Big Data. 2019;6(111):1–21. https://doi.org/10.1186/s40537-019-0268-2.
23. Mukta M, et al. IoT based Smart Water Quality Monitoring System. 4th International Conference on Computer and Communication Systems (ICCCS). 2019;669–673. IEEE. https://doi.org/10.1109/CCOMS.2019.8821742.
24. Pasika S, Gandla ST. Smart water quality monitoring system with cost-effective using IoT. Heliyon. 2020;6(7):2405–8440. https://doi.org/10.1016/j.heliyon.2020.e04096.
25. Jan FJK, et al. Cloud-based smart water quality monitoring system using IoT sensors and machine learning. IJATCSE. 2020. https://doi.org/10.30534/ijatcse/2020/141932020.
26. Jan F, et al. IoT based smart water quality monitoring: recent techniques trends and challenges for domestic applications. MDPI. 2021. https://doi.org/10.3390/w13131729.
27. Zulkifli CZ, et al. IoT-based water monitoring systems: a systematic review. Water. 2022;14(22):3621.
28. Hemdan EED, et al. An efficient IoT based smart water quality monitoring system. Multimed Tools Appl. 2023;82(19):28827–51.
29. Adeleke IA, et al. A hybrid machine learning and embedded IoT-based water quality monitoring system. IoT. 2023;22:100774.
30. Jan F, et al. Iot based smart water quality monitoring: recent techniques, trends and challenges for domestic applications. Water. 2021;13(13):1729.
31. Nishan RK, et al. Development of an IoT-based multi-level system for real-time water quality monitoring in industrial wastewater. Discover Water. 2024;4(1):43.
32. Prapti DR, et al. Internet of Things (IoT)-based aquaculture: an overview of IoT application on water quality monitoring. Rev Aquac. 2022;14(2):979–92.
33. AlZubi AA. IoT-based automated water pollution treatment using machine learning classifiers. Environ Technol. 2024;45(12):2299–307.
34. Pasika S, Gandla ST. Smart water quality monitoring system with cost-effective using IoT. Heliyon. 2020;6(7):e04096.
35. Jabbar WA, et al. Development of LoRaWAN-based IoT system for water quality monitoring in rural areas. Expert Syst Appl. 2024;242: 122862.
36. Cicceri G, et al. A novel architecture for the smart management of wastewater treatment plants. IEEE International Conference on Smart Computing (SMARTCOMP).b2021;392–394. https://doi.org/10.1109/SMARTCOMP52413.2021.00080.
37. Cicceri G, et al. SWIMS: the smart wastewater intelligent management system. IEEE International Conference on Smart Computing (SMARTCOMP).2021; 228–233.
38. Garcia-Alvarez D, et al. Fault detection and diagnosis using multivariate statistical techniques in a wastewater treatment plant. IFAC Proc. 2009;42(11):952–7.
39. Mahajana P, Shahane P. An IoT based system for water quality monitoring. ICICNIS.2020.
40. Dewangan SK, et al. Investigating the impact of pH levels on water quality: an experimental approach. Int J Res Appl Sci Eng Technol. 2023;11(9):756–9. https://doi.org/10.22214/ijraset.2023.55733.
41. "OF TURBIDITY : Information for regulators and water suppliers," (2017). [Online]. https://api.semanticscholar.org/CorpusID:21643299. Accessed Apr 2024
42. Munjal S, et al. Effect of TDS on human immune system. Int J Innov Res Growth. 2019. https://doi.org/10.26671/IJIRG.2018.2.8.101.
43. Water Quality Parameters Bangladesh Standards & WHO Guide Lines, https://dphe.gov.bd/site/page/15fa0d7b-11f1-45c0-a684-10a543376873/Water-Quality-Parameters. Accessed Oct 2024
44. W. Organization, "Guidelines for drinking-water quality, Vol. 1, 3rd edition incorporating 1st and 2nd addenda". Accessed Oct 2024.
45. "Dataset". [Online]. https://github.com/Sujan-Roy/Dataset. Accessed Apr 2025.
46. Farouk MIHZ, et al. Towards online surface water quality monitoring technology: A review. Environ Res. 2023;238: 117147. https://doi.org/10.1016/j.envres.2023.117147.
47. "IEEE Standard for Information Technology--Telecommunications and Information Exchange between Systems—Local and Metropolitan Area Networks--Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Redline," IEEE Std 802.11–2020 (Revision of IEEE Std 802.11–2016) – Redline: 1–7524, (2021). Accessed Apr 2024.
48. Adityas Y, et al. Water quality monitoring system with parameter of ph, temperature, turbidity, and salinity based on internet of things. Jurnal Informatika dan Sains (JISA). 2021;4(2):138–43. https://doi.org/10.31326/jisa.v4i2.965.
49. Junaidi, et al. Measurement of physical parameters of water quality in real-time based on arduino. J Phys Conf Ser. 2021;1751(1):012067. https://doi.org/10.1088/1742-6596/1751/1/012067.
50. Samuelsson O, et al. To calibrate or not to calibrate, that is the question. Water Res. 2023;229: 119338. https://doi.org/10.1016/j.watres.2022.119338.
51. Sokolova M, et al. Beyond accuracy, Fscore and ROC: A family of discriminant measures for performance evaluation. Joint Conf Artif Intell. 2006;1015–1021.
52. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-Score, with implication for evaluation. European conference on information retrieval.2005; 345–359.