Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 10 No. 2 (2026) 398 - 406 e-ISSN: 2580-0760

IoT-Based Water Quality Monitoring and Sustainable Modeling for Smart Campuses

Fachrul Kurniawan[1], Miladina Rizka Aziza[2], Novrindah Alvi Hasanah[3], Fadia Irsania Putri[4], Aji Prasetya Wibawa[5], Jehad Hammad[6], Yuhefizar[7]

[1,2,3]Departement of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia

[4,5]Departement of Electrical Engineering, Universitas Negeri Malang, Malang, Indonesia

[6]Departement of Computer Information Systems, Al-Quds Open University, Bethlehem, Palestine

[7]Departement of Information Technology, Politeknik Negeri Padang, West Sumatera, Indonesia

*fachrull@ti.uin-malang.ac.id

Abstract

This study proposes an IoT-based water quality monitoring framework integrated with a continuous sustainable modeling approach for smart campus applications. A total of 404 sensor observations were collected, including pH, turbidity, temperature, and Total Dissolved Solids (TDS). A continuous water suitability score ranging from 0 to 1 was constructed based on WHO drinking water standards, and Multiple Linear Regression was employed to model the relationship between water quality parameters and the suitability score. The main contribution of this study lies in the development of a lightweight analytical framework that combines continuous regression modeling with threshold-based classification to support real-time decision-making in resource-constrained environments. The dataset was divided into 90% training and 10% testing data. The results show that the proposed framework achieved a classification accuracy of 88.5% based on threshold mapping of regression outputs, with a misclassification rate of 11.5%. These findings demonstrate the effectiveness of integrating IoT-based monitoring with interpretable and computationally efficient analytical models for sustainable campus water management.

Keywords: IoT-based monitoring; monitoring water; smart green campus; visualization

How to Cite: F. Kurniawan, "IoT-Based Water Quality Monitoring and Suitability Modeling for Smart Campuses", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 10, no. 2, pp. 398 - 406, Apr. 2026.
*Permalink/*DOI: https://doi.org/10.29207/resti.v10i2.7360

Received: November 25, 2025
Accepted: April 12, 2026
Available Online: April 24, 2026

1. Introduction

Water quality monitoring is essential for maintaining public health and supporting environmental sustainability within campus environments [1-2]. Despite the increasing adoption of IoT-based monitoring systems, many existing implementations primarily focus on hardware integration [3], with limited incorporation of lightweight predictive analytics suitable for real-time campus-scale decision-making.

In campus environments, the quality of raw water as a drinking water source is a critical factor influencing public health and sustainability [4-5]. Smart Green Campus initiatives emphasize sustainable resource management, including monitoring water quality [6]. The term Smart Green Campus is given to campuses that effectively utilize their environment to meet the needs and improve the welfare of the community on campus [6], including at UIN Maulana Malik Ibrahim Malang where the main uses of water are for drinking.

To ensure the safety and quality of drinking water consumed by the campus community, the water must comply with the Minister of Health Regulation No. 492 of 2010, including chemical, physical, radioactive, and microbiological requirements [7]. This study aims to assess the quality of water sources around the UIN Maulana Malik Ibrahim Malang campus area and utilize the Internet of Things, including turbidity, pH, temperature, and TDS sensors, to monitor drinking water quality.

IoT devices generate continuous, heterogeneous sensor data streams that require efficient and lightweight analytical models for effective campus-scale environmental monitoring [8]. So that the data can be used as new information, it is necessary to analyze using a specific technique or method. Analyzing data can be done using machine learning or deep learning to filter data, so it is ready for use. Although numerous IoT-based water quality monitoring systems have been proposed, most existing studies primarily emphasize hardware integration or complex machine learning approaches, with limited focus on lightweight, interpretable analytical frameworks that support continuous monitoring and real-time decision-making in resource-constrained smart campus environments [9]. This gap highlights the need for an integrated approach that balances predictive capability, computational efficiency, and practical deployment.

The data collected by the IoT circuit will be analyzed using the Linear Regression method to classify water quality as clean, potable, or undrinkable [10]. The Linear Regression method was selected because several previous studies have demonstrated its strong performance in solving similar problems [11]. For instance, the study by [12] showed that the linear regression model was adequate for analyzing the AQI of Velachery based on chemical pollutants such as CO_2, $PM_{2.5}$, O_3, and SO_2.

The study by [13] highlighted that among all the reviewed techniques, Linear Regression demonstrated the greatest efficiency for predicting water quality in Iraq as evidenced by the R^2 and R values that come close to 1 for all of the observed months. Likewise, study by [14] highlighted Linear Regression was the most convenient model and proved efficient for the analysis of water quality in Shanghai as it was better than the conventional models used to calculate the AQI.

Multiple Linear Regression was selected in this study due to its balance between computational efficiency and interpretability, making it particularly suitable for deployment in resource-constrained IoT environments [9]. Furthermore, the model enables clear quantification of the contribution of each water quality parameter, which is essential for transparent and practical environmental decision-making [15]. This model also allows for easy elucidation of the contribution of each coefficient, thus making it possible to determine the influence of each of the individual parameters (e.g., pH, TDS, turbidity, and temperature) on the overall suitability of the water. Its lightweight computational structure makes it appropriate for deployment in resource-constrained IoT-based monitoring systems.

Although linear regression is primarily used for continuous prediction tasks, in this study it also serves as a foundation for threshold-based classification, enabling practical categorization of water quality levels. This approach maintains computational simplicity while supporting interpretable decision-making in IoT-based systems.

In addition, the water quality monitoring program developed in this research is compatible with the Smart Green Campus framework, which is becoming more popular. A smart green campus is a framework for educational institutions striving for greater sustainability and less environmental impact [16]. Concerning the various issues, researchers have extensively examined the use of IoT, data analytics, linear regression, and smart water quality monitoring and control systems. The integration of these systems with advanced data visualization can revolutionize the management of water systems in universities, increasing automation and sustainability of campus ecosystems .

Previous studies have shown the potential of IoT water monitoring systems; however, they have focused more on the integration of hardware, assessment of sensor accuracy, and traditional data visualization methods [17]. There are prediction models that are less resource-intensive and that could be run continuously, while still providing useful information for managing resource-limited campuses. Most other studies use rule-based classification or very complex machine learning methods while ignoring explainability and scalability to the institution [18]. This illustrates the need for a synthesis that balances predictive modeling with practical deployment.

This study introduces an analytical framework for campus-wide use, which, unlike prior IoT-based water monitoring studies that concentrate on the hardware or traditional data reporting, focuses on the framework's ability to function within informational and computational constraints. The study's main contribution is the ability to combine real-time IoT sensor data acquisition with the linear regression model, which is integer valued for real-time computing, coupled with threshold classification that is compliant with the World Health Organization (WHO) standards. The framework is specifically designed for use within limited-resource settings, enhancing the potential of the smart green campus framework for the concentration, ease of interpretation, and limited computational cost. This focus distinguishes the study from other works that center their frameworks around the use of advanced, unexplainable machine learning techniques without consideration of the implementation of machine learning techniques in educational settings.

As detailed above, prior studies have examined systems based on the Internet of Things (IoT) coupled with linear regression analytics to assess the improvements made in water quality. Several previous studies have explored IoT-based water quality monitoring systems, primarily focusing on sensor integration and data acquisition processes. For instance, existing works utilizing platforms such as Node-RED and LabVIEW demonstrate high sensor accuracy and real-time monitoring capabilities [19]. However, these studies largely emphasize system implementation rather than analytical modeling. In contrast, some studies incorporate decision-making methods such as SAW or regression-based approaches [20], yet often lack integration with continuous monitoring frameworks or real-time deployment considerations..

IoT-based water quality monitoring has also been applied in specific domains such as aquaculture and aquarium systems, demonstrating the capability of real-time monitoring and remote control [21]. However, these applications are generally domain-specific and do not emphasize scalable analytical frameworks suitable for broader environmental monitoring contexts such as smart campuses. The use of IoT technologies in monitoring and controlling water quality. in ornamental fish aquariums, is an enormous step towards giving aquarium enthusiasts the ability to support the health and comfort of the aquarium's inhabitants by assisting in the control of optimal water quality settings. The system provides aquarium owners with the ability to automate water quality control via the Telegram app to continuously monitor water conditions, and take necessary actions in real time, even if they are away from the aquarium.

These limitations indicate that existing approaches have not fully addressed the integration of IoT-based monitoring with lightweight and interpretable analytical models. Therefore, this study proposes an integrated framework that combines continuous suitability modeling using linear regression with threshold-based classification to support real-time decision-making in smart campus environments.

Another work titled 'Automated System of Water Quality Monitoring for Prawn Industry via LabView And Internet of Things' also discusses the use of the Internet of Things relative to monitoring the water quality of the prawn sector [22]. The work discusses two vital parameters of water quality of the prawn sector, which are, temperature, and pH. The water quality sensors are interfaced with a data acquisition system via LabVIEW, while the servo motor moves the probe of the pH sensor. The work makes use of ThingSpeak to display water quality data in real time and is outfitted with an alarm system which can be configured to alert users to the problem [22]. The work is accurate and can be remotely detected as a result of the data being processed in real-time.

As stated in 'History, How It Works, and the Benefits of the Internet of Things,' the essence of the Internet of Things is to interlink multiple devices to a single network in order to create an infrastructure for the seamless transfer of data [23]. The Internet of Things is a paradigm shift in technology as it allows for the automation of numerous sectors of life through a system of intelligent and collaborative sensors and tools over the internet [24]. RFID, WSN, WPAN, WBAN, HAN, NAN, M2M, CC, and DC are all components involved within IoT [23]. These components are part of the technology IoT uses for sensing.

In summary, existing studies on IoT-based water quality monitoring primarily focus on hardware implementation or domain-specific applications, while studies incorporating analytical models often rely on complex approaches with limited interpretability. There is still a lack of integrated frameworks that combine real-time IoT data acquisition with lightweight, interpretable, and continuous analytical modeling. This study addresses this gap by proposing a framework that integrates IoT-based monitoring with multiple linear regression and threshold-based classification to support efficient and practical decision-making in smart campus environments.

2. Methods

This section describes the overall research methodology employed in this study, including system design, data acquisition procedures, dataset preparation, model construction, and performance evaluation. The proposed approach integrates an Internet of Things (IoT)-based monitoring framework with statistical modeling to analyze water quality conditions [25]. The methodology is structured systematically to ensure reproducibility, measurement reliability, and analytical validity.

2.1 System Architecture and Data Acquisition

The proposed system consists of four core components: data collection, data transmission, data storage, and data retrieval, as illustrated in Figure 1. The data retrieval section is specifically designed for the collection of water quality parameters using Total Dissolved Solids (TDS), pH, turbidity, and water temperature sensors. This system is therefore capable of monitoring these parameters continuously, which suggests that the system is designed with multiple data collection sensors.

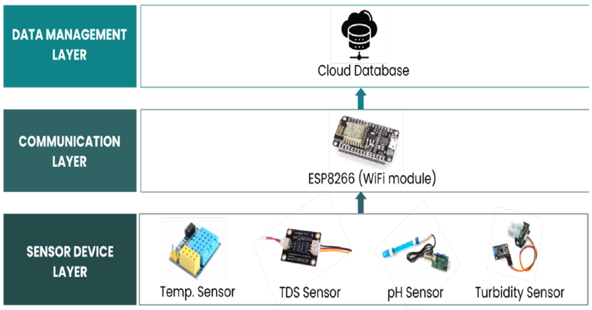

Figure 1. The block diagram of the proposed water quality measurement system

The pH sensor has a range between 0 to 14. The conductivity (TDS) sensor has a range of 0 to 1000 ppm with an accuracy of +/- 2 %. The turbidity sensor has a range of 0 to 3000 NTU with an accuracy of +/- 5 %. The temperature sensor (DS18B20) has a range of −10 °C to 85 °C with an accuracy of +/- 0.5 °C. All sensors prior to installation, were calibrated with a standard buffer solution to improve reliability and reduce systematic error.

The sensors measured every 10 seconds, and readings were transmitted to the microcontroller unit (MCU) to cross reference sensor readings. Two sensors (the pH and TDS sensors) provided an analog output which needed to be converted to a digital signal using a built-in analog to digital converter (ADC) of the microcontroller. The temperature and turbidity sensors provided data in a digital format which was used to provide an output to the MCU [26]. After processing, the data were sent to the ThingSpeak cloud server through an ESP8266 Wi-Fi module. This server is used for real time monitoring and centralized storage services [27].

2.2 Dataset Description and Preprocessing

A total of 404 observations were collected from campus drinking water sources under consistent operational conditions. Four predictor variables were evaluated for each samples: TDS (mg/L), pH, turbidity (NTU), and temperature. These variables were selected for the purposes of measuring standards for drinking water and were applicable using IoT-based sensing devices.

Before modeling, a number of processing steps were taken to improve the quality of the data and the accuracy of the analysis. Mean substitution was used on the missing data to keep the integrity of the data set and to not alter the data too much statistically [28]. Outliers related to sensor measurements were detected and removed using the Interquartile Range (IQR) method to reduce the impact of anomalous readings caused by temporary sensor disturbances [29]. To improve the comparison between the predictor variables and to improve numerical stability, the variables were normalized using Min-Max scaling methods.

2.3 Construction of Continuous Suitability Score

To enable regression modeling with a continuous dependent variable, a water suitability score ranging from 0 to 1 was constructed based on WHO drinking water standards [30]. This score quantitatively represents the degree to which each parameter satisfies the recommended thresholds.

For parameters with upper acceptable limits, namely Total Dissolved Solids (TDS) and turbidity, the suitability score was defined as Equation 1.

$$S_j = \begin{cases} 1, & X_j \le T_j \\ \frac{Tj}{Xj} & X_j > T_j \end{cases} \quad (1)$$

Xj denotes the observed value of parameter j, and Tj represents the corresponding WHO threshold. This formulation ensures that measurements within acceptable limits receive the maximum score of 1, while values exceeding the acceptable thresholds are penalized proportionally. For parameters with upper acceptable limits, such as Total Dissolved Solids (TDS) and turbidity, the suitability score was defined to assign a maximum value of 1 when the observed measurement falls within the acceptable threshold. Values exceeding the threshold are proportionally reduced based on the degree of deviation, ensuring a gradual penalty rather than a binary classification.

For pH, which has an acceptable range between 6.5 and 8.5, the suitability score was formulated based on the deviation from the midpoint of the acceptable range. This approach allows the score to decrease smoothly as the pH value moves away from the optimal range, thereby avoiding abrupt classification boundaries, the suitability score was defined as Equation 2.

$$S_{pH} = \begin{cases} 1, & 6.5 \le X \le 8.5 \\ 1 - \frac{|X - Xmid|}{\Delta}, & otherwise \end{cases} \quad (2)$$

$Xmid$ = 7.5 represents the midpoint of the acceptable range and Δ represents half of the acceptable pH range. This formulation ensures a gradual reduction in suitability as pH deviates from the optimal range, avoiding abrupt binary penalties.

The overall suitability score was calculated as the arithmetic mean of the individual parameter scores, providing a normalized representation of overall water quality on a continuous scale between 0 and 1, Equation 3.

$$Y_i = \frac{S_{TDS}+S_{pH}+S_{Turbidity}}{3} \quad (3)$$

The resulting continuous score Yi was used as the dependent variable in the linear regression model. This lightweight normalization approach provides a simplified alternative to conventional Water Quality Index (WQI) methods while maintaining interpretability and computational efficiency.

2.4 Modelling

The relationship between water quality parameters and the constructed suitability score was modeled using Multiple Linear Regression (MLR), Equation 4.

$$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{3i}+\beta_4X_{4i}+\varepsilon_i, i=1,2,...., n \quad (4)$$

Yi denotes the continuous water suitability score for observation i. The variables X_{1i}, X_{2i}, X_{3i}, and X_{4i} correspond to TDS, pH, turbidity, and temperature, respectively. The parameters $\beta0$-$\beta4$ represent regression coefficients, and εi denotes the random error term [31-32].

The regression coefficients were estimated using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared residuals between observed and predicted values [33-34]. Model assumptions—including linearity, independence of errors, homoscedasticity, and normality of residuals—were evaluated through residual diagnostics to ensure statistical validity [35].

2.5 Model Validation and Classification Procedure

The dataset was divided into 90% training data and 10% testing data. The training set was used to estimate regression parameters, while the testing set was reserved exclusively for performance evaluation to reduce overfitting bias.

Regression performance was assessed using the coefficient of determination (R^2), which measures the proportion of variance in the suitability score explained by the predictor variables.

Although the model produces continuous suitability scores, categorical water quality levels were derived for practical decision-making purposes. The classification thresholds were determined empirically based on compliance strictness and dataset distribution analysis, reflecting high, moderate, and low levels of standard compliance. The categories were defined as follows: Safe for drinking: $Y \geq 0.80$, Suitable but not safe for drinking: $0.50 \leq Y < 0.80$, Not suitable: $Y < 0.50$

Classification performance was evaluated using Accuracy, Precision, Recall, F1-score, and Confusion Matrix analysis to provide a comprehensive assessment of predictive capability.

This integrated approach enables both continuous predictive modeling and operational classification to support smart campus water monitoring applications.

3. Results and Discussions

In designing the IoT water quality monitoring system, the integration of temperature, TDS, pH, and turbidity sensors is of utmost importance. These sensors were purposefully included to enable the system to record and analyze water quality parameters with precision. Valuable water quality assessment IoT systems utilize real-time data recorded by water quality monitoring systems. More than 404 records pertaining to time and TDS, temperature, pH, and turbidity comprise the dataset. Temperature, TDS, pH, and turbidity were recorded as independent variables.

Table 1 contains real-time data of TDS, temperature, pH and turbidity sensors which measure the primary constituents of water quality. TDS measures the amount of dissolved solids in water, temperature affects chemical and biological processes within the water, pH indicates the water's acidity or alkalinity, and turbidity measures the clarity of the water. The parameters of the water quality monitoring system collected real-time data which is illustrated in the graph for effortless interpretation. The first phase of preprocessing included the collection and processing of over 400 data points. These dynamic water quality parameters are well understood and a comprehensive analysis made more data points available.

In this study, a continuous water suitability score was first estimated using linear regression. For interpretative and operational purposes, the predicted scores were subsequently mapped into three categories: "safe for drinking," "suitable but not safe," and "not suitable". This criteria stemmed from the quality of water regulations and standards outlined by the World Health Organization (WHO) as well as countries' regulations on drinking water. The assessment was based on water quality standards from WHO which include the following: pH (6.5 - 8.5), TDS (< 500 mg/L), and turbidity (< 5 NTU), and for the following parameters: pH, TDS and turbidity, ATP, and Chemicals.

The results exemplify how monitoring using IoT and algorithms based on linear regression can serve to aid campus administrators at the initial stages of the decision-making process regarding the maintenance of the filtration system and other decision points. The method provides disturbance monitoring in support of green, smart, and sustainable campus initiatives, enables data driven responses, and offers systems for prompt response to primary disturbance arising, and therefore, provides early warning systems [31]. The research contributes to the theory of scientific knowledge and practice of water resource management in institutional settings.

Table 1. Dataset Sample Acquired from Internet Of Things Device

No	Time	TDS	Temp.	pH	Turbidity	Water Quality
1	2024-08-24 26:22:05	162.4	26.13	3.05	1872.76	Not suitable
2	2024-08-14 16:32:07	266.47	25.94	3.08	2489.25	Not suitable
3	2024-08-14 16:37:07	330.27	25.75	2.78	2141.58	Not suitable
4	2024-08-14 09:47:16	325	24.81	2.78	1543.01	Not suitable
5	2024-08-15 11:28:13	88.73	27.44	8.84	5.19	Suitable, not safe for drinking
6	2024-08-15 11:33:18	125.31	27.38	8.94	4.13	Suitable, not safe for drinking
7	2024-08-15 11:38:02	139.57	27.31	2.81	3.08	Suitable, not safe for drinking
8	2024-08-15 11:43:08	143.11	27.31	7.81	3.43	Suitable, not safe for drinking
9	2024-08-15 13:05:07	104.87	27.19	6.63	0.5	Safe for drinking
10	2024-08-15 13:10:12	50.5	27.38	6.08	0.67	Safe for drinking

3.1 Data Distribution

The dataset was split into a training set (90%) and a testing set (10%), with the former being utilized to estimate the regression coefficients and the latter to assess the model's predictive performance independently. There is a regression equation that produces a continuum of outputs and these outputs are later categorized into discrete units for simplicity and ease of implementation.

This study's dataset consists of approximately 404 records, the data are at this point a representation of the trial data from the implementation of an IoT-based water quality monitoring system on campus. We note that this is a small dataset and that we are not able to claim that we are working with "big data" so we recognize that this is an initial trial designed to evaluate the system and test the integration of the sensors, we are working toward the goal of on implementing a fully functioning system that collects data continuously over long periods of time and that has the ability to add new measurement locations in a configurable manner and to incorporate external data sources. Thus, the handling of issues related to the storage, processing, and predictive model performance of large datasets will remain a primary focus of the future development work.

The results of this study are consistent with existing literature using IoT and statistical methods for environmental monitoring [32]. Numerous studies have established that the combination of sensors and analytical models can facilitate real-time decision-making regarding water management. The trends recorded in pH, TDS, turbidity, and temperature are in agreement with previous studies that demonstrated the ability of sensor-driven monitoring systems to identify changes in water quality. The consistency of these findings reiterates the reliability of our results in the context of existing literature.

Following the acquisition of real-time data via the Internet of Things architecture, three water quality categories were derived by applying predefined threshold values to the predicted continuous suitability scores, as shown in Figure 2. The term "not suitable" describes non-potable water, which is dangerous to consume, cook with, clean with, or even bathe in because it contains dangerous substances including chemicals, germs, and viruses. Water that is acceptable for everyday use but not safe for drinking is classified as "suitable but not safe for drinking." Water that is acceptable for drinking and safe for consumption is represented by the "safe for drinking" category.

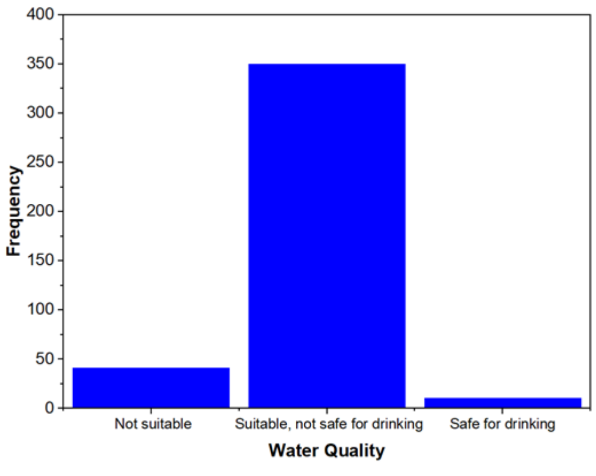

Figure. 2 Distribution graph of water quality category

3.2 Visualization

Figure 3 shows four graphic panels comparing data from several water quality parameters used in the study, namely Total Dissolved Solids (TDS), Temperature, pH, and Turbidity. Each parameter is displayed in two formats: a line graph showing changes in data over time (in sample counts), and a histogram showing the frequency distribution of the data.

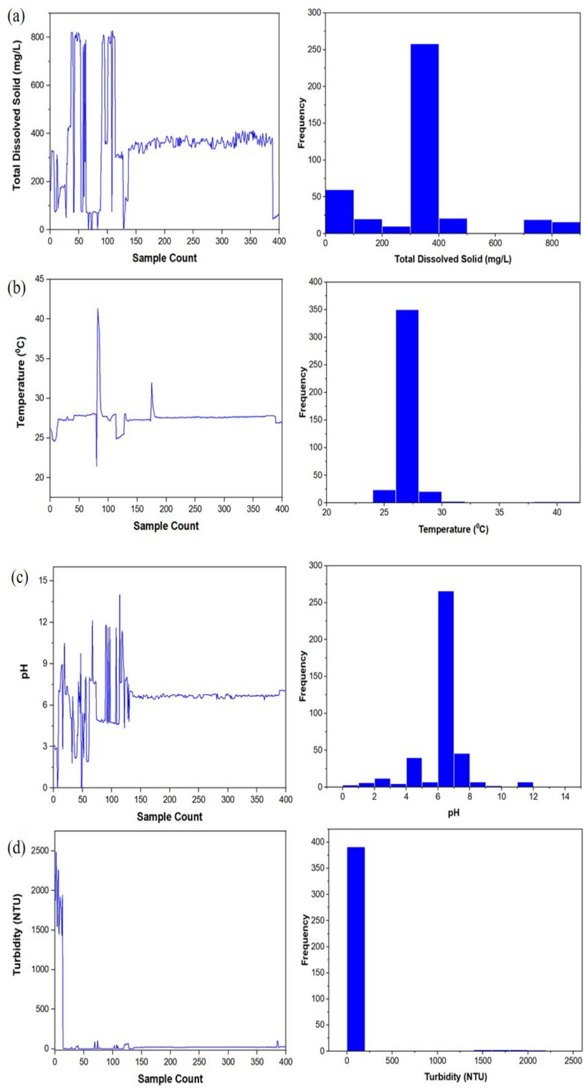

Figure 3. Visualization of water quality parameter data, including (a) Total Dissolved Solids (TDS), (b) temperature, (c) pH, and (d) turbidity.

The results of the data processing show that TDS exhibited the highest frequency within the range of 320-360 mg/L. The graph of the test results using the TDS sensor shows the left graph (line) is the change in TDS in mg/L units against the number of samples. At the beginning of the graph, TDS fluctuated sharply, then returned to stability until the end of the observation. While the right graph (histogram) shows the frequency distribution of the TDS value.

Then for the temperature test in the range of 27.5-28.75°C. The left graph (line) shows the change in temperature in degrees Celsius against the number of samples. The right graph (histogram) of temperature shows most of the temperature values.

The pH sensor shows the pH value obtained in the range of 6.5-7.0. The left graph (line) shows the change in pH value against the number of samples. Initially, there was a significant pH fluctuation, but after that, the pH was stable until the end of the observation. The right graph (histogram) shows that most of the pH values were in the range of 6.5-7.0.

The last graph is a turbidity test that shows results in the range of 0-200 NTU. The left graph (line) shows the change in turbidity value in NTU units against the number of samples. The right graph (histogram) shows the frequency distribution of data against turbidity.

The real-time data collected by the ESP8266 was transferred and safely stored in a cloud-based database [33]. The current water conditions were visualized and analyzed using graphical representations of the raw sensor data that were shown in real-time in the database.

Figure 3 illustrates the variation and distribution of water quality parameters across all observations. It can be seen that TDS and turbidity exhibit significant fluctuations, particularly in the early samples, indicating potential anomalies or sudden environmental changes. In contrast, temperature and pH values tend to be more stable, with most observations concentrated within a narrower range. The histogram plots further confirm that the majority of values fall within specific intervals, supporting the reliability of the dataset for modeling purposes.

3.3 Performance of Linear Regression

The linear regression model produces a continuous suitability score for each observation. To facilitate operational interpretation, the predicted scores were converted into three categorical levels using predefined threshold values. In this study, the regression model is primarily used to generate a continuous water suitability score, which is subsequently transformed into categorical water quality classes based on predefined thresholds. Therefore, the performance evaluation is conducted within a classification framework, where accuracy is used as the main evaluation metric. The confusion matrix is then used to evaluate the consistency between threshold-mapped predictions and reference categories.

Figure 4 illustrates the confusion matrix constructed to evaluate the classification performance after threshold-based mapping of the regression outputs. The matrix organizes predictions into three categories: not suitable; suitable, not safe for drinking; and safe for drinking, in a 3×3 format where rows represent true labels and columns represent predicted labels. Each cell captures the correspondence between predicted and actual values, including both correct classifications and misclassifications. The use of a color gradient facilitates visual interpretation, where darker shades indicate higher correct predictions and lighter shades represent misclassifications.

From the confusion matrix analysis, it can be observed that the darkest cell corresponds to the prediction of "suitable, not safe for drinking," indicating that this class was correctly predicted 101 times. However, some misclassifications are still present; for example, 2 samples from the "suitable" category were incorrectly classified as "not suitable." Furthermore, the confusion matrix was used to compute the misclassification rate and accuracy using Equations 5 and 6, resulting in values of 11.5% and 88.5%, respectively. The model achieved an accuracy of 88.5%, indicating its effectiveness in classifying water quality conditions.

Figure 4. Confusion matrix after threshold-based mapping of regression outputs.

$$Accuracy = \frac{Total\ Correct\ Prediction}{Total\ Predictions} x\ 100\% \quad (5)$$

$$Misclassification\ Rate = 1 - Accuracy \quad (6)$$

Although regression-based metrics such as the coefficient of determination (R^2) and Root Mean Square Error (RMSE) can provide additional insight into model performance, they are not the primary focus of this study since the final objective is classification. These results demonstrate that the proposed model is capable of capturing the relationship between water quality parameters and suitability levels with a relatively high degree of accuracy. The achieved performance also indicates that the threshold-based classification derived from the regression model is effective in distinguishing between acceptable and non-acceptable water conditions. This finding further suggests that the selected parameters, including pH, turbidity, temperature, and TDS, provide sufficient information for predicting water suitability in the studied environment.

Table 2 shows Threshold-Based Classification Based on Regression Output. Accuracy tells how precise each method is in determining the quality of water. Accuracy is the method's ability to diminish false positives or to determine the ratio of genuinely accurate predictions to all positive predictions. The method's ability to diminish false negatives is termed Recall; it refers to the ability of a method to capture true positives in the population of actual positives. The F1 score is the overall evaluation of predictive performance, it is the average of precision and recall measured as a harmonic mean.

Table 2. Performance of Threshold-Based Classification Derived from Regression Output

Parameter	Linear Regression
Accuracy (%)	88.5

Precision (%)	88.9
Recall (%)	88.5
F-1 Score (%)	86.4

Compared to previous studies that primarily focus on IoT-based monitoring without integrated analytical modeling, the proposed approach provides additional value through its predictive capability [34]. While some studies employ complex machine learning models with potentially higher accuracy, such approaches often require greater computational resources and lack interpretability [35]. In contrast, the proposed method achieves competitive performance while maintaining simplicity and transparency, making it more suitable for real-time implementation in resource-constrained environments.

The findings of this study have important implications for smart campus water management. The proposed framework enables continuous monitoring and real-time assessment of water quality, allowing campus administrators to make timely and data-driven decisions. In addition, the use of a lightweight and interpretable model facilitates easier deployment and maintenance compared to more complex approaches. This supports the development of sustainable and efficient water management systems within smart campus environments.

However, this study has several limitations. The model was developed based on a dataset collected from a specific campus environment, which may limit its generalizability to other contexts with different water characteristics. In addition, the use of Multiple Linear Regression assumes linear relationships between variables, which may not fully capture more complex interactions among water quality parameters. Future work may explore the integration of more diverse datasets and the application of advanced modeling techniques to further improve predictive performance.

4. Conclusions

This study presents an integrated approach for smart campus water management by combining IoT-based monitoring with continuous water suitability modeling. The proposed framework enables real-time data acquisition and provides an interpretable and normalized suitability score to assess water quality conditions.

The experimental results demonstrate that the proposed model achieves an accuracy of 88.5%, indicating its effectiveness in classifying water quality conditions based on the defined suitability thresholds. This result confirms that the selected parameters, including pH, turbidity, temperature, and TDS, are sufficient to represent water quality characteristics in the studied environment. In addition, the use of a Multiple Linear Regression model offers a balance between predictive performance and computational efficiency, making it suitable for real-time implementation in resource-constrained environments.

Despite these contributions, several limitations should be acknowledged. The model is developed based on data collected from a specific campus environment, which may limit its applicability to other contexts. In addition, the assumption of linear relationships may not fully represent complex interactions among water quality parameters. The limited number of sensors used in this study may also affect the comprehensiveness of the assessment.

Future research should focus on expanding the dataset across multiple locations and longer time periods to improve generalization. The application of more advanced machine learning techniques can be explored to enhance predictive performance. Furthermore, integrating additional parameters and developing an automated decision-support system for real-time monitoring and alert generation would further strengthen the proposed framework.

References

[1] T. Rahaman, "SMART ENVIRONMENTAL MONITORING SYSTEMS FOR AIR AND WATER," vol. 01, no. 01, pp. 1–19, 2025, doi: https://doi.org/10.63125/a08zay47.

[2] G. I. Emran, R. Barma, A. H. Khan, and M. Roy, "Reasons behind the Water Crisis and its Potential Health Outcomes," vol. X, no. Id, pp. 1–10, 2023, doi: https://doi.org/10.48550/arXiv.2403.07019.

[3] S. Abdulmalek *et al.*, "IoT-Based Healthcare-Monitoring System towards Improving Quality of Life : A Review," 2022, doi: https://doi.org/10.3390/healthcare10101993.

[4] K. Olatunde *et al.*, "Factors Affecting the Quality of Drinking Water in the United States of America : A Ten-Year Systematic Review," vol. 10, no. 1, pp. 24–34, 2022, doi: 10.12691/ajwr-10-1-4.

[5] R. Mahmood, N. Hassan, A. Chamseddine, R. Rangarajan, and R. Yassoub, "Examining Drinking Water Preferences among University Students : A Comparative Assessment," 2024, doi: https://doi.org/10.3390/ijerph21101271.

[6] Q. Liu, "Enhancing Green Smart Campus Development with Data Mining Technology," vol. 5, no. 1, pp. 523–538, 2025, doi: https://doi.org/10.53759/7669/jmc202505041.

[7] M. Hadani, I. Yana, A. S. Efendi, and Z. Rasyid, "Analysis Of Drinking Water Depot Sanitation Hygiene In The Working Area Of The Siak Hulu Public Health Community , Kampar District , In 2023," 2023, doi: https://doi.org/10.25311/miracle/Vol1.Iss1.2024.1841.

[8] Â. Sousa and P. J. S. Cardoso, "Smart Energy Monitoring for Sustainable Campuses : A Hybrid Anomaly Detection Approach Based on Prophet and Isolation Forest," pp. 1–18, 2026, doi: https://doi.org/10.3390/su18052589.

[9] M. N. Khan, "Edge AI based on Real-time Robotic Decision-Making in Resource-constrained," vol. 1, no. 2, pp. 19–26, 2025, doi: https://doi.org/10.63056/ijair.1.2.2025.92.

[10] M. A. Rahu *et al.*, "Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality," *IEEE Access*, vol. 11, no. September, pp. 101055–101086, 2023, doi: 10.1109/ACCESS.2023.3315649.

[11] M. Abu-faraj, A. Al-hyari, and Z. Alqadi, "Experimental Analysis of Methods Used to Solve Linear Regression Models," 2022, doi: 10.32604/cmc.2022.027364.

[12] A. Loganathan, P. Sumithra, and V. Deneshkumar, "Estimation of Air Quality Index Using Multiple Linear Regression," *Appl. Ecol. Environ. Sci.*, vol. 10, no. 12, pp. 717–722, 2022, doi: 10.12691/aees-10-12-3.

[13] Q. M. Ajaj, N. A. Awad, H. J. Jumaah, and H. M. Rizeei, "Air quality regression analysis over Iraq during severe dust periods using GIS and," vol. 6, pp. 300–308, 2025, doi: 10.30493/DAS.2025.478260.

[14] W. Jiang, "The Data Analysis of Shanghai Air Quality Index

Based on Linear Regression Analysis The Data Analysis of Shanghai Air Quality Index Based on Linear Regression Analysis," pp. 0–7, 2021, doi: 10.1088/1742-6596/1813/1/012031.

[15] K. H. Ahmad, K. Y. Adjei, and C. C. Ekechi, "Explainable AI for Water Quality Monitoring : A Systematic Review of Transparency , Interpretability , and Trust," *IEEE Sensors Rev.*, vol. 2, no. August, pp. 419–443, 2025, doi: 10.1109/SR.2025.3595500.

[16] P. Ivo and A. F. Chim-miki, "A Smart Campus Framework : Challenges and Opportunities for Education Based on the Sustainable Development Goals," pp. 1–17, 2022, doi: https://doi.org/10.3390/su14159640.

[17] R. Mudiyanselage, P. Niroshan, S. Bandara, and A. B. Jayasignhe, "The Integration of IoT (Internet of Things) Sensors and Location-Based Services for Water Quality Monitoring : A Systematic Literature Review," pp. 1–26, 2025, doi: https://doi.org/10.3390/s25061918.

[18] M. Nagahisarchoghaei, N. Nur, L. Cummins, and N. Nur, "An Empirical Survey on Explainable AI Technologies : Recent Trends , Use-Cases , and Categories from Technical and Application Perspectives," 2023, doi: https://doi.org/10.3390/electronics12051092

[19] A. H. Embong, L. Asbollah, and S. B. A. Hamid, "Empowering industrial automation labs with IoT : A case study on real-time monitoring and control of induction motors using Siemens PLC and Node-RED," vol. 18, no. 2, pp. 10004–10016, 2024, doi: https://doi.org/10.15282/jmes.18.2.2024.3.0790.

[20] S. S. A. Wahid, S. A. Azli, M. S. Ramli, and K. K. Hasan, "Intelligent water flow monitoring system based on internet of things for residential pipeline," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 1, pp. 20–27, 2022, doi: 10.11591/ijeecs.v27.i1.pp20-27.

[21] N. Azreen *et al.*, "Improved accuracy in IoT-Based water quality monitoring for aquaculture tanks using low-cost sensors : Asian seabass fish farming," *Heliyon*, vol. 10, no. 8, p. e29022, 2024, doi: 10.1016/j.heliyon.2024.e29022.

[22] N. S. Damanhuri *et al.*, "Automated System of Water Quality Monitoring for Prawn Industry via Labview And Internet of Things," vol. 17, no. 3, pp. 393–400, 2024, doi: 10.22266/ijies2024.0630.31.

[23] M. K. Tiwari, R. Pal, V. Chauhan, and A. Kumari, "Comprehensive review of internet of things (IoT) network protocols : Features , open challenges , and trends," vol. 5, no. 2, pp. 150–167, 2024, doi: https://doi.org/10.33545/27076571.2024.v5.i2b.111.

[24] T. Van Hoang, "Impact of Integrated Artificial Intelligence and Internet of Things Technologies on Smart City Transformation," vol. 19, no. 01, pp. 64–73, 2024, doi: https://doi.org/10.54644/jte.2024.1532.

[25] A. Bhardwaj, V. Dagar, M. Owais, K. Akarsh, and A. Rafael, "Smart IoT and Machine Learning - based Framework for Water Quality Assessment and Device Component Monitoring," *Environ. Sci. Pollut. Res.*, pp. 46018–46036, 2022, doi: 10.1007/s11356-022-19014-3.

[26] J. Rocher, J. M. Jimenez, J. Tomas, and J. Lloret, "Water BodiesLow-Cost Turbidity Sensor to Determine Eutrophication in Water Bodies," 2023, doi: https://doi.org/10.3390/s23083913.

[27] M. N. M. Aashiq *et al.*, "An IoT-based handheld environmental and air quality monitoring station," vol. 12, no. 3, pp. 1–8, 2023, doi: https://doi.org/10.21014/actaimeko.v12i3.1487.

[28] K. Maharana, S. Mondal, and B. Nemade, "A review : Data pre-processing and data augmentation techniques," vol. 3, no. April, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.

[29] S. Zafeirelli and D. Kavroudakis, "Comparison of outlier detection approaches in a Smart Cities sensor data context," vol. 17, no. 1, 2024, doi: 10.2478/ijssis-2024-0004.

[30] S. I. Abba *et al.*, "Drinking Water Resources Suitability Assessment Based on Pollution Index of Groundwater Using Improved Explainable Artificial Intelligence," 2023, doi: https://doi.org/10.3390/su152115655.

[31] I. F. Almojela *et al.*, "WatAr: An arduino-based drinking water quality monitoring system using wireless sensor network and GSM module," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2020-Novem, pp. 550–555, 2020, doi: 10.1109/TENCON50793.2020.9293896.

[32] R. P. P, A. J. L, A. K. G. B, P. G. R, S. K. S, and S. K. KN, "IoT based Smart Water Quality Monitoring and Flow Control System," *Met. Finish.*, vol. 93, no. 9, p. 85, 1995, doi: 10.1016/0026-0576(95)92051-x.

[33] M. Babiuch, P. Foltynek, and P. Smutny, "Using the ESP32 microcontroller for data processing," *Proc. 2019 20th Int. Carpathian Control Conf. ICCC 2019*, no. May 2019, 2019, doi: 10.1109/CarpathianCC.2019.8765944.

[34] A. Bassi *et al.*, "Introduction to the Internet of Things," *Enabling Things to Talk Des. IoT Solut. with IoT Archit. Ref. Model*, no. March, pp. 1–349, 2013, doi: 10.1007/978-3-642-40403-0.

[35] T. Hariyanto, M. Rahayu, F. Satria, and M. Y. Fadhlan, "Improving Temperature Sensor Accuracy in the IoT Trainer Kit by Linear Regression Method," *Proc. 2019 Int. Conf. Mechatronics, Robot. Syst. Eng. MoRSE 2019*, no. December, pp. 237–240, 2019, doi: 10.1109/MoRSE48060.2019.8998639.