

An Intelligent IoT-Based Water Quality Monitoring and STORET Index Prediction System Using Random Forest

Yohanes Adi Nugroho

Mekatronika dan Kecerdasan
Buatan

Universitas Pendidikan Indonesia
Purwakarta, Indonesia

yohanesadinugroho@upi.edu

Dewi Indriati Hadi Putri

Mekatronika dan Kecerdasan
Buatan

Universitas Pendidikan Indonesia
Purwakarta, Indonesia

dewiindri@upi.edu

Mahmudah Salwa Gianti

Mekatronika dan Kecerdasan
Buatan

Universitas Pendidikan Indonesia
Purwakarta, Indonesia

msg.salwa@upi.edu

Abstract— Hygiene sanitation water quality fluctuates due to environmental dynamics, yet conventional monitoring systems generally lack the predictive capabilities compliance with health standards (Permenkes No. 2 of 2023). This study aims to develop an intelligent Water Quality Monitoring System (WQMS) capable of predicting water quality status based on the STORET index using the Random Forest algorithm. [Methods] The proposed system integrates an ESP32-S3 microcontroller with calibrated low-cost sensors for real-time data acquisition. To ensure data integrity, regression was applied for sensor calibration, while the STORET method was utilized to determine pollution levels and water feasibility. A Random Forest regression model was then trained using these processed datasets to classify water quality status. Experimental results demonstrated high hardware precision, achieving Mean Absolute Percentage Error (MAPE) values of 5.38% for pH, 2.24% for TDS, and 0.22% for the Flow Meter. Furthermore, the Random Forest model exhibited superior predictive performance, yielding a Coefficient of Determination (R^2) of 0.9977, a Mean Absolute Error (MAE) of 0.2213, and a Root Mean Squared Error (RMSE) of 0.5144. These findings indicate that the integrated system effectively combines accurate sensing with robust predictive modeling. Consequently, the system is categorized as highly capable of providing real-time insights and early warnings, offering a significant improvement over traditional monitoring methods for public health safety.

Keywords— Random Forest Algorithm, Prediction, Water Quality Monitoring System, STORET

*Article info: Date Submitted: 2026-01-07 | Date Revised: 2026-04-14 | Date Accepted: 2026-04-16
This is an open access article under the CC BY-SA license*



I. INTRODUCTION

Clean water serves a vital role in human life and is a basic necessity that must meet specific quality standards to ensure public health. In Indonesia, the quality standards for hygiene sanitation water are strictly regulated under the Minister of Health Regulation (*Permenkes*) No. 2 of 2023. However, water sources are increasingly vulnerable to contamination due to environmental factors and anthropogenic pollution, making the maintenance of water quality a significant challenge [1]. Conventional methods for monitoring water quality often rely on manual sampling and laboratory analysis. While accurate, these methods are time-consuming, labor-intensive, and fail to provide real-time data, preventing early detection of contamination events [2]. Water utility providers and household users still lack access to real-time information regarding their water consumption and quality, leading to potential health risks [3].

Development of the Internet of Things (IoT) offers a solution to these limitations by enabling continuous, real-time monitoring of physical and chemical water parameters [4]. Previous studies have successfully implemented IoT for Water Quality Monitoring Systems (WQMS) using parameters such as pH, turbidity, and Total Dissolved Solids (TDS) [5]. However, most existing IoT systems focus primarily on data acquisition and visualization without incorporating predictive capabilities [1], [2], [4], [3]. Data monitoring alone is often insufficient for proactive decision-making; a system that can predict water quality status based on historical patterns is required to enhance safety and management

efficiency [6]. Integrates IoT technology with Machine Learning algorithms to enhance the system's capabilities [5]. Specifically, this research utilizes the Random Forest Regression algorithm to process sensor data and generate predictive insights. By combining real-time data acquisition from IoT sensors with the analytical power of Random Forest, the system is designed to not only monitor current conditions but also accurately predict water quality status (STORET), thereby maximizing the effectiveness of the WQMS in ensuring compliance with hygiene sanitation standards [7], [8].

Previous studies have successfully implemented IoT for Water Quality Monitoring Systems (WQMS) using physical and chemical water parameters. However, a critical gap remains in the current body of literature. Although previous studies have implemented IoT-based water quality monitoring systems, most of them focus solely on real-time data visualization without integrating standardized pollution indices such as STORET and predictive machine learning models. Furthermore, limited studies evaluate the performance of low-cost sensors combined with predictive analytics under hygiene sanitation regulations in Indonesia.

Therefore, this study proposes to design and build an intelligent WQMS that integrates low-cost sensors (pH, TDS, Turbidity, and Flow Meter) with an ESP32-S3 microcontroller. The specific objective is to monitor water quality parameters in real-time and apply the Random Forest algorithm to predict the water quality status using the STORET (Storage and Retrieval) method index [9]. By combining IoT with predictive analytics, this system intends to provide a comprehensive solution for monitoring hygiene sanitation water compliance, offering users actionable insights into water feasibility and consumption [5], [8].

II. METHODOLOGY

This study adopts the Research and Development (R&D) method utilizing the 4D development model, which consists of four main stages: Define, Design, Develop, and Disseminate. Figure 1 illustrates research framework, research flow begins with requirement analysis (Define), followed by the design of hardware and software architecture (Design) [10]. The development stage involves assembling the prototype, developing the firmware, and training the Random Forest model (Develop) [11]. Finally, the system undergoes functionality and performance testing [12] before the results are analyzed (Disseminate).

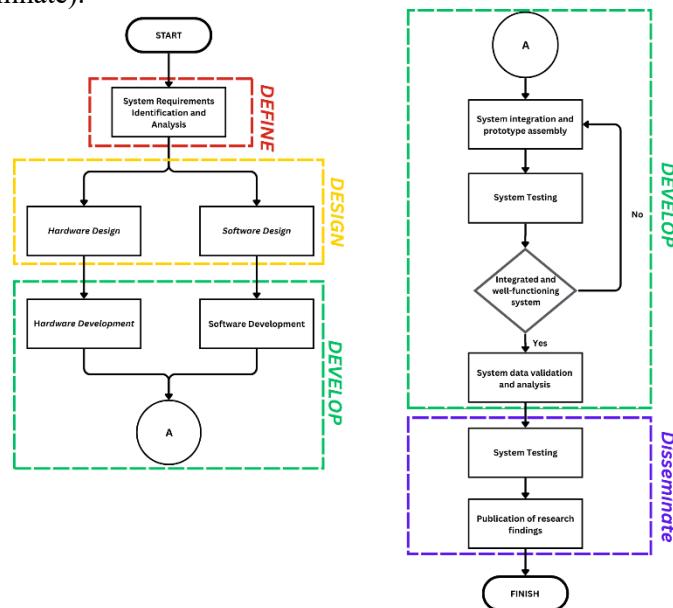


Figure 1. Research Flowchart

A. System Architecture

The system architecture is designed based on the Internet of Things (IoT) framework, comprising four layers: the Perception Layer, Networking Layer, Processing Layer, and Application Layer. The system architecture structure is illustrated by Figure 2.

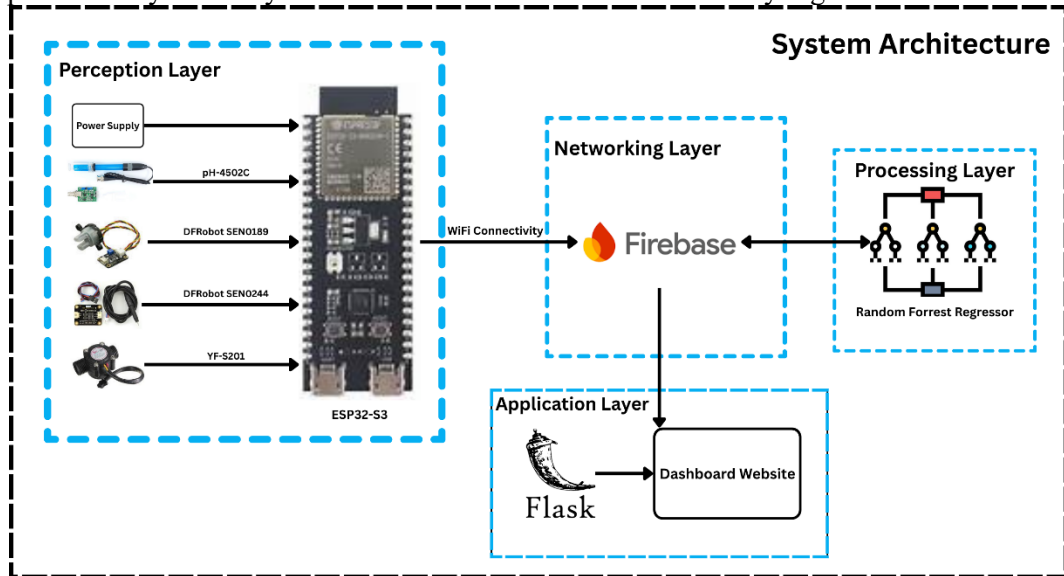


Figure 2 System Architecture

1) Perception Layer

This layer functions as the physical interface with the environment. The ESP32-S3 microcontroller is utilised for integrated AIoT capabilities. The system integrates four key sensors to acquire water quality data: pH-4502C sensor for acidity measurement; DFRobot SEN0244 analog TDS sensor for detecting total dissolved solids; DFRobot SEN0189 turbidity sensor for measuring water clarity; and YF-S201 flow sensor to monitor water usage volume. The ESP32-S3 is capable of processing raw analogue signals from the sensors utilising signal conditioning techniques. The primary function of these techniques is to ensure data accuracy and acquisition to transmission.

2) Networking Layer

This layer facilitates reliable communication between the hardware and the cloud server. The ESP32-S3 utilizes its integrated WiFi module to transmit sensor data over the TCP/IP protocol. The data is formatted as JSON (JavaScript Object Notation) objects to ensure lightweight and efficient data transfer. The system connects to the Firebase Realtime Database, which acts as a centralized cloud storage system. This setup allows for low-latency synchronization, ensuring that data uploaded by the microcontroller is immediately available for the processing and application layers.

3) Processing Layer

This layer is responsible for data management and advanced analytics. It is implemented using a Flask-based backend that retrieves data from Firebase. The core component of this layer is the Random Forest Regression machine learning model. This model takes the historical and real-time sensor data (pH, TDS, Turbidity) as input features to predict the water quality status based on the STORET index. The processing layer also handles data cleaning and feature engineering to maintain the integrity of the dataset before it is fed into the prediction model.

4) Application Layer

This layer provides the human-machine interface (HMI) for end-users. It is developed as a Web Dashboard using HTML, CSS, and JavaScript. The dashboard visualizes real-time monitoring data, historical trends via graphical charts, and the predicted STORET

water quality score. Additionally, this layer includes an alert system that triggers warnings when water quality parameters exceed the safety thresholds defined by the Ministry of Health Regulation No. 2 of 2023, allowing users to take immediate action.

B. Water Quality Monitoring System

Development of the IoT monitoring device, integrating both physical components and embedded software[13]. The core processing unit is the ESP32-S3, selected for its high-performance dual-core processor and internal WiFi capability, which is essential for handling sensor data and network tasks simultaneously. This microcontroller interfaces with a suite of sensors to capture water quality parameters. To ensure signal stability and minimize noise, all components are assembled on a custom-designed Printed Circuit Board (PCB) powered by a Hi-Link HLK-5M05 module. The electrical schematic and PCB layout are shown in Figure 3.

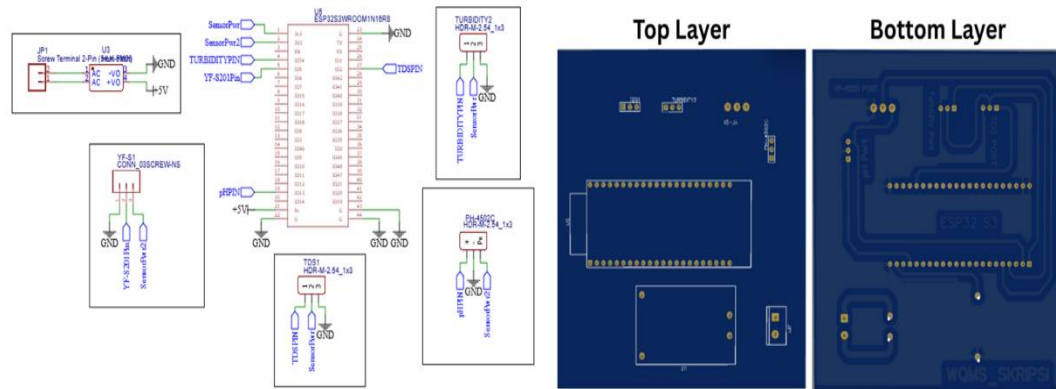


Figure 3. Electrical Schematic and PCB Layout

Firmware was developed using the Arduino IDE. To ensure data accuracy, the firmware implements signal conditioning techniques, specifically oversampling and averaging, for the analog sensors (pH, TDS, and Turbidity). Meanwhile, the flow sensor utilizes an Interrupt Service Routine (ISR) to accurately count pulses without blocking the main processor loop [14]. The acquired data is then formatted into JSON objects and transmitted via WiFi to the Firebase Realtime Database [15], ensuring seamless synchronization with the cloud server.

C. STORET Method

To classify the water feasibility, this study utilizes the STORET (Storage and Retrieval) method. Developed by the United States Environmental Protection Agency (US EPA), the STORET method is selected due to its strict orientation toward regulatory compliance [9], [16]. In this study, the water quality assessment is benchmarked against the Ministry of Health Regulation (Permenkes) No. 2 of 2023, which serves as the national standard for sanitation hygiene water quality. The STORET method operates on a scoring mechanism that evaluates compliance with established quality standards [17]. The scoring system used in this study follows the guidelines for datasets with more than 10 samples (time-series data), as detailed in Table 1.

Table 1 STORET Method Scoring System

Data Sample	Statistic Value	Physic Parameter	Chemical Parameter
≥ 10 data sample	Maximum	-2	-4
	Minimum	-2	-4
	Average	-6	-12

The final determination of water quality status is performed by summing the negative scores of all parameters. Based on the accumulated total score, the water quality is classified into four distinct categories (Class A to Class D), as presented in Table 2.

Table 2 STORET Index

Category	Score	Result
Very Good (A)	0	Meet quality standards
Good (B)	-1 s/d -10	Light contamination
Middle (C)	-11 s/d -30	Medium contamination
Bad (D)	>-31	Heavy polluted

D. Random Forest Regression Model

Random Forest was selected based on a comparative analysis with other machine learning models, including XGBoost, Support Vector Regression (SVR), Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) [7], [5]. Random Forest is the most suitable model for handling the non-linear relationships between water quality parameters (pH, TDS, Turbidity) and the STORET score [8]. The explanation step is as follows:

1) Data Preparation

Data Preparation is process that involves transforming raw data into analysis-ready datasets [18]. The dataset was constructed by combining primary data collected directly from the sensors with a supplementary dataset from Kaggle, specifically the "Water Quality and Potability" dataset [19]. This external dataset was selected due to its strong characteristic similarity with the developed system's sensor specifications, particularly containing matching parameters such as pH, TDS, and Turbidity. It is important to emphasize that this supplementary dataset serves strictly for data augmentation to enrich the training variance and improve model generalization, not to replace the primary field data. Raw dataset sample shown in Table 3, the initial dataset contained 55,264 data points, which were organized into 5,527 batches to represent time-series sequences.

Table 3. Raw Dataset

Batch	pH	TDS	Turbidity	STORET
1	7.08	207.91	2.96	-16
1	3.72	186.3	4.5	-16
1	8.1	199.1	3.06	-16
.....
5527	7.23	191.36	0.67	-2
5527	7.27	135.78	2.04	-2
5527	8.09	246.05	4.11	-2

To ensure data integrity, a rigorous data cleaning process was applied. This involved identifying and removing inconsistent or incomplete batches. As a result of this cleaning process, one invalid batch was removed, resulting in a final high-quality dataset of 5,526 batches with no missing values. This clean dataset serves as the foundation for the subsequent feature engineering and training stages.

2) *Feature Engineering*

Since water quality data is dynamic, raw sensor readings alone may not capture the full trend within a specific timeframe. To address this limitation, feature engineering was implemented by aggregating data at the batch level to extract meaningful statistical attributes [18]. For each physical-chemical parameter specifically pH, TDS, and Turbidity three key statistical features were calculated: the minimum value (min) to identify the lowest recorded point, the maximum value (max) to detect peak levels, and the mean value (mean) to represent the average condition within the batch. This process generated a total of nine distinct input features (e.g., pH_min, pH_max, pH_mean, tds_mean, etc.), which collectively provide a comprehensive representation of water quality behavior, serving as robust predictors for the STORET score target variable.

3) *Model Training and Hyperparameter Tuning*

To establish a robust predictive model, the Random Forest Regression algorithm was selected due to its ensemble nature, which effectively mitigates the risk of overfitting common in single decision trees. To find out the ideal value of the hyperparameters, the models were run on a predefined of values. This process was performed applying the function RandomizedSearchCV of Scikit-Learn and using cross-validation.

E. Performance Evaluation

The evaluation of the system begins with the assessment of sensor data acquisition performance. Ensuring the accuracy of the input data is critical before processing it with the machine learning model. The sensor readings were compared against standard instruments to calculate the Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). For the acidity parameter, the pH sensor demonstrated a strong linear correlation with the standard buffer solutions, achieving an R^2 value of 0.96. The error analysis resulted in a MAPE of 5.38% and a MAE of 0.34, indicating that the sensor readings remain within the acceptable tolerance for hygiene sanitation monitoring. The Total Dissolved Solids (TDS) sensor exhibited superior precision. The calibration process yielded an exceptionally high linearity with an R^2 of 0.99. The deviation from the standard measuring tool was minimal, as evidenced by a MAPE of 2.24% and a MAE of 1.66 ppm. This high level of accuracy confirms the

reliability of the low-cost TDS sensor for detecting dissolved particles in water. Regarding the turbidity measurement, the calibration process adopted the methodology proposed in previous studies [20], which establishes the relationship between the sensor's analog voltage output and Nephelometric Turbidity Units (NTU). By implementing this reference-based calibration technique, the system successfully converted voltage signals into meaningful turbidity values, ensuring consistent physical water quality detection.

To ensure the reliability and applicability of the predictive model [21], a comprehensive evaluation was conducted using two distinct testing approaches. First, the model's theoretical performance was assessed using the reserved 20% testing dataset, employing statistical metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2) to quantify prediction accuracy and variance. Second, to validate the system's effectiveness in a real-world operational environment, field testing was performed over a two-week period divided into seven testing cycles. This practical validation compared the model's real-time predictions against actual STORET values derived from manual sampling standards, ensuring the system's robustness against dynamic environmental fluctuations.

III. RESULTS AND DISCUSSION

The following are the results of research on the implementation of the Water Quality Monitoring System (WQMS) using the Random Forest algorithm. The presentation of the results is structured in accordance with the method used, covering sensor calibration accuracy, machine learning model performance evaluation, and real-time system validation based on the STORET index.

A. Sensor Calibration and Accuracy

This system utilizes low-cost sensors, including the pH-4502C, DFRobot SEN0244 (TDS), SEN0189 (Turbidity) and YF-S201 (Flow Meter). Despite their affordability, rigorous calibration against standard instruments demonstrated their reliability. pH sensor was calibrated using buffer solutions with pH values of 4.01, 6.86, 7.01, 8.2 dan 10.6 while the TDS and turbidity sensors were validated against a commercial TDS meter, calibration process of turbidity sensors using scientific papers to be used as a reference in calculations [22]. Additionally, the flow sensor was calibrated by comparing digital readings with manual volume measurements. Sensor calibration result shown in Table 4.

Table 4. Sensor Calibration Result

Sensor	Accuracy	MAE	MAPE
pH-4502C	94.63%	0.344	5.37%
SEN0244	97.76%	1.6 ppm	2.244%
SEN0189	98%	-	-
YF-S201	99.78%	2.89 mL	0.224%

The evaluation results indicate high precision, with Mean Absolute Percentage Error (MAPE) values of 5.37% for pH, 2.24% for TDS, and 0.22% for the Flow Meter. These low error rates confirm that these low-cost sensors are capable of providing accurate data input for the machine learning model.

B. Random Forest Model Performance

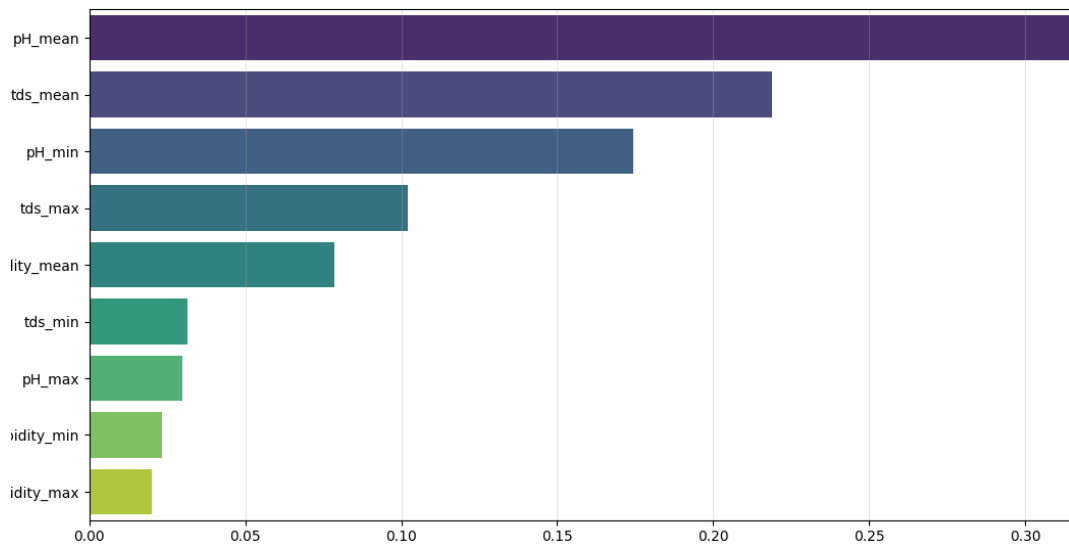
The development of the prediction model focused on optimizing the Random Forest Regression algorithm to predict the STORET water quality score. Random Forest is designed to improve predictive accuracy by constructing a multitude of decision trees during training and outputting the mean prediction of the individual tree; hence it is often known as an ensemble learning method [23], [24]. To find an optimal regression model,

the Random Forest algorithm is applied using RandomizedSearchCV for hyperparameter tuning [7]. Hyperparameter tuning result shown in Table 5.

Table 5. Hyperparameter Tuning Result

Hyperparameter	Value
n_estimator	300
max_features	sqrt
max_depth	30
min_sample_split	2
min_sample_leaf	1
bootstrap	False

With the optimized parameters established, a feature importance analysis was conducted to identify the dominant predictor variables contributing to the model's accuracy. This analysis calculates the effectiveness score of each feature across decision branches, as visualized in Table 3. The results reveal that pH_mean and tds_mean exert the strongest influence on the model's decisions, indicating that pH and TDS parameters play a vital role in determining water quality within this dataset. Although the turbidity parameter demonstrated a less significant statistical contribution in this specific model, its inclusion remains essential for the physical detection of water pollution. Feature Importance visualize shown in Figure 4.



This finding aligns with STORET method, where acidity (pH) indicators of chemical

Figure 4. Feature Importance

contamination that significantly alter the STORET score calculation. Although turbidity showed a lower statistical weight in the decision trees, its inclusion remains vital for detecting physical pollutants (suspended particles) that visually affect water hygiene.

The model's predictive performance was first evaluated using the reserved 20% testing dataset. The results indicate exceptional accuracy with minimal error margins. As shown

in the metrics at Table 6, the model achieved an (R^2) score of 0.9961, implying that 99.61% of the variability in the STORET index can be explained by the model.

Table 6. Model Evaluation on Test Dataset

Evaluation Metrics	Value
Coefficient of Determination (R^2)	0.9961
Mean Absolute Error (MAE)	0.2281
Mean Squared Error (MSE)	0.5145
Root Mean Squared Error (RMSE)	0.7173

To validate the significance of these results, the model’s performance was compared with previous related studies. While Dewi et al. [7] successfully applied Random Forest for water quality prediction, this study achieved a higher degree of precision $R^2 > 0.99$ by integrating specific signal conditioning on the hardware side and rigorous hyperparameter tuning. Furthermore, comparisons with Jayaraman [8], who reported an R^2 of 0.94 and RMSE of 0.15 for their Random Forest model, indicate that the proposed system offers a superior explanation of data variance. This improvement proves that the specific signal conditioning and rigorous hyperparameter tuning applied in this study allow the model to effectively handle the non-linear interactions between pH, TDS, and Turbidity to predict the complex STORET index.

The low MAE value of 0.2281 suggests that, on average, the model's prediction deviates by less than 0.3 points from the actual STORET score, which is a negligible margin in water quality classification. Scatter plot of actual vs predicted visualize shown in Figure 5.

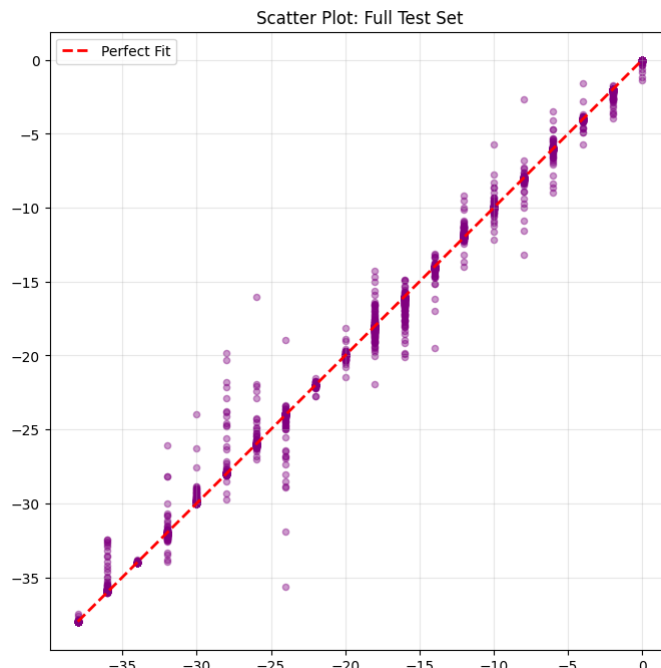


Figure 5 Model Scatter Plot

C. Deployment

To verify the system's robustness in a dynamic environment, the model was tested against real-world data collected over several days in December 2025. This phase aims to assess how well the model adapts to daily fluctuations in water quality. The validation results per day are summarized in Table 7.

Table 7 Deployment Validation Metrics

Date	MAE	MSE	RMSE	Score (R^2)
Dec 13, 2025	1.538	4.544	2.132	0.805
Dec 14, 2025	0.577	0.587	0.766	0.930
Dec 16, 2025	0.391	0.169	0.412	0.798
Dec 18, 2025	0.813	0.689	0.830	0.834
Dec 19, 2025	0.905	0.875	0.935	0.272
Dec 20, 2025	0.593	0.449	0.670	0.959
Dec 21, 2025	0.485	0.3450	0.587	0.975

Beyond statistical comparisons, the Mean Absolute Error (MAE) holds significant practical implications for the accuracy of the STORET score prediction. Since the STORET index is calculated by summing negative integer penalty scores (e.g., -2, -6, -10) based on parameter violations, the predictive model must maintain a low deviation to ensure the final score lands within the correct classification range. In the deployment, the MAE fluctuated, reaching a maximum of 1.538 on December 13, 2025. While this deviation is higher than the test phase, it remains operationally safe due to the wide intervals of the STORET classes. This deviation is minimal relative to the broad class interval, providing a substantial safety buffer.

The external validation shows that the model maintains high stability. On December 20 and 21, the model achieved near-perfect linearity with R^2 scores exceeding 0.95. A noticeable drop in correlation occurred on December 19 R^2 0.272. However, upon closer inspection, the MAE for that day remains low (0.9049), and the RMSE is below 1.0. This discrepancy suggests that while the statistical correlation R^2 appeared low due to limited variance in the actual data values for that specific day, the magnitude of the prediction error remained very small. Overall, the consistently low MAE values across all days (< 1.6) confirm that the system provides reliable predictions for hygiene sanitation monitoring purposes.

IV. CONCLUSION

This study successfully developed an IoT-based Water Quality Monitoring System (WQMS) integrated with a Random Forest Regression model to predict the STORET index. The implementation of low-cost sensors ensured reliable data acquisition precision suitable for hygiene sanitation monitoring. Regarding the predictive model, the Random Forest algorithm, optimized via RandomizedSearchCV, proved highly effective in capturing the non-linear dynamics of water quality. Crucially, deployment validation confirmed the system's robustness, where the model achieved R^2 scores exceeding 0.95 on optimal days and maintained a Mean Absolute Error (MAE) consistently below 1.6 across all daily testing cycles, demonstrating its adaptability to dynamic environmental conditions. Consequently, this system has strong potential to support local governments and households in implementing proactive water quality management aligned with national hygiene sanitation regulations (Permenkes No. 2 of 2023). For future development, the system can be enhanced by integrating biological parameters, to provide a more holistic water quality assessment compliant with stricter health standards. Additionally, expanding the dataset with diverse water samples and developing a dedicated mobile application with

push notifications would significantly improve the system's scalability and user accessibility.

REFERENCE

- [1] V. Lakshmikantha, A. Hiriyannagowda, A. Manjunath, A. Patted, J. Basavaiah, and A. A. Anthony, "IoT based smart water quality monitoring system," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 181–186, Nov. 2021, doi: 10.1016/j.gltp.2021.08.062.
- [2] S. Pasika and S. T. Gandla, "Smart water quality monitoring system with cost-effective using IoT," *Heliyon*, vol. 6, no. 7, Jul. 2020, doi: 10.1016/j.heliyon.2020.e04096.
- [3] A. Humnabadkar, B. Shivbhakta, A. Karve, and M. A. A. Kokate, "Water Quality Monitoring and Measures of Water Usage in Homes." [Online]. Available: www.ijfmr.com
- [4] R. P. Shete, A. M. Bongale, and D. Dharrao, "IoT-enabled effective real-time water quality monitoring method for aquaculture," *MethodsX*, vol. 13, Dec. 2024, doi: 10.1016/j.mex.2024.102906.
- [5] I. Essamlali, H. Nhaila, and M. El Khaili, "Advances in machine learning and IoT for water quality monitoring: A comprehensive review," Mar. 30, 2024, *Elsevier Ltd.* doi: 10.1016/j.heliyon.2024.e27920.
- [6] S. A. H. Almetwally, M. K. Hassan, and M. H. Mourad, "Real Time Internet of Things (IoT) Based Water Quality Management System," in *Procedia CIRP*, Elsevier B.V., 2020, pp. 478–485. doi: 10.1016/j.procir.2020.03.107.
- [7] D. A. Dewi, A. S. Wei, L. C. Lin, and C. D. Heng, "Water Quality Prediction using Random Forest Algorithm and Optimization," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1354–1362, Sep. 2024, doi: 10.47738/jads.v5i3.348.
- [8] P. Jayaraman, K. K. Nagarajan, P. Partheeban, and V. Krishnamurthy, "Critical review on water quality analysis using IoT and machine learning models," *International Journal of Information Management Data Insights*, vol. 4, no. 1, Apr. 2024, doi: 10.1016/j.ijime.2023.100210.
- [9] T. K. Nufutomo, Y. S. Manalu, and B. S. Muntalif, "STORET (Storage and Retrieval) method for analysis and identification of water pollutants in the Upper Citarum Watershed, West Java, Indonesia," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2023. doi: 10.1088/1755-1315/1201/1/012050.
- [10] R. Budi Indaryanti, B. Murtiyasa, and B. Soemardjoko, "4D Research and Development Model: Trends, Challenges, and Opportunities Review," 2025. doi: 10.31599/na7deq07.
- [11] Torang Siregar, "Stages of Research and Development Model Research and Development (R&D)," *DIROSAT: Journal of Education, Social Sciences & Humanities*, vol. 1, no. 4, pp. 142–158, Oct. 2023, doi: 10.58355/dirosat.v1i4.48.
- [12] M. Waruwu, "Metode Penelitian dan Pengembangan (R&D): Konsep, Jenis, Tahapan dan Kelebihan," *Jurnal Ilmiah Profesi Pendidikan*, vol. 9, no. 2, pp. 1220–1230, May 2024, doi: 10.29303/jipp.v9i2.2141.
- [13] N. Reddy, S. Gautham, P. V Adithya, and P. Kumar, "WATER QUALITY MONITORING SYSTEM BASED ON IOT," 2023. [Online]. Available: www.ijnrd.org
- [14] A. C. Tasong and R. P. Abao, "Design and development of an IoT application with visual analytics for water consumption monitoring," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 205–213. doi: 10.1016/j.procs.2019.08.159.

- [15] S. C. Olisa, C. N. Asiegbu, J. E. Olisa, B. O. Ekengwu, A. A. Shittu, and M. C. Eze, “Smart two-tank water quality and level detection system via IoT,” *Heliyon*, vol. 7, no. 8, Aug. 2021, doi: 10.1016/j.heliyon.2021.e07651.
- [16] W. Indrasari, Z. Nabilla, and S. Sunaryo, “Water quality status analysis based on a parameter of Physics and Chemistry using the STORET method in Bekasi river,” in *Journal of Physics: Conference Series*, Institute of Physics, 2023. doi: 10.1088/1742-6596/2596/1/012055.
- [17] M. Yacub et al., “Kajian Penggunaan Metode IP, STORET, dan CCME WQI dalam Menentukan Status Mutu Sungai Cikapayang, Jawa Barat,” 2022. doi: 10.26418/jtllb.v10i1.53296.
- [18] X. Wei et al., “Study on prediction model of nitrogen oxide concentration in reprocessing plant based on random forest,” *International Journal of Advanced Nuclear Reactor Design and Technology*, vol. 7, no. 2, pp. 63–69, Jun. 2025, doi: 10.1016/j.jandt.2025.04.011.
- [19] L. Tharmalingam, “Water Quality and Potability,” <https://www.kaggle.com/datasets/uom190346a/water-quality-and-potability>.
- [20] W. L. Hakim, L. Hasanah, B. Mulyanti, and A. Aminudin, “Characterization of turbidity water sensor SEN0189 on the changes of total suspended solids in the water,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1280/2/022064.
- [21] S. Dridi, “SUPERVISED LEARNING-A SYSTEMATIC LITERATURE REVIEW A PREPRINT,” Dec. 2021. doi: <https://doi.org/10.31219/osf.io/qtmcs>.
- [22] J. Tomperi, A. Isokangas, T. Tuuttila, and M. Paavola, “Functionality of turbidity measurement under changing water quality and environmental conditions,” *Environmental Technology (United Kingdom)*, vol. 43, no. 7, pp. 1093–1101, 2022, doi: 10.1080/09593330.2020.1815860.
- [23] “Random Forest Regression in Python - GeeksforGeeks.” Accessed: Jun. 28, 2025. [Online]. Available: <https://www.geeksforgeeks.org/machine-learning/random-forest-regression-in-python/>
- [24] N. Ganesh, P. Jain, A. Choudhury, P. Dutta, K. Kalita, and P. Barsocchi, “Random forest regression-based machine learning model for accurate estimation of fluid flow in curved pipes,” *Processes*, vol. 9, no. 11, Nov. 2021, doi: 10.3390/pr9112095.