An Internet of Things-based Smart Water Meter with Machine Learning-aided Water Quality Assessment

Marvin Jagen, Joel Ariiho, Edwin Mugume and Dorothy K. Okello
Department of Electrical and Computer Engineering
Makerere University
Kampala, Uganda
jagenmarvin1@gmail.com, tajoel@icloud.com, {edwin.mugume, dorothy.okello}@mak.ac.ug

***Abstract*—This paper presents an Internet of Things (IoT)-based smart water meter with machine learning (ML)-aided water quality assessment capability. A flow rate sensor is used to measure water consumption while pH and turbidity sensors are used for water quality assessment. The collected data is sent to a remote server via a cellular network where it is used for monitoring purposes by both the utility company and customers. The system assesses the collected data against relevant thresholds and provides appropriate notifications to the service provider and the customers. The thresholds for water quality are based on the national standards for potable water while those for consumption are determined by the average monthly water consumption. This paper considers National Water and Sewerage Corporation (NWSC), the largest water utility company in Uganda, as a case study. A total of 1,760 samples collected by NWSC in the Kampala service area in 2022 were assessed using the feature selection algorithm of ML. The most dominant parameters were determined as residual chlorine, pH, turbidity, conductivity, and apparent color. In this paper, only pH and turbidity are considered.**

***Keywords*—Internet of Things (IoT), Machine Learning (ML), smart water meter, water consumption monitoring, water quality assessment.**

2023 IEEE AFRICON | 979-8-3503-3621-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/AFRICON55910.2023.10293553

I. Introduction

Water quality monitoring is the routine assessment of water supply [1], while water consumption monitoring is the measurement of water usage [2]. Both processes are crucial for maintaining public health safety and preventing water wastage. Without water quality monitoring, there is an increased risk of water-borne diseases like malaria, typhoid, and cholera [1]. Similarly, lack of water consumption monitoring results in water wastage due to leakage. Therefore, the World Health Organization (WHO) guidelines mandated water monitoring, especially in urban areas [1]. In Uganda, National Water and Sewerage Corporation (NWSC) is responsible for the supply and routine monitoring of water quality [3]. NWSC has digitized its services in the recent past but formal complaints from customers and physical billing still pose limitations [3].

An effective solution to these limitations is an Internet of Things (IoT)-based smart water meter with machine learning (ML)-aided water quality assessment system. The IoT aspect refers to a network of physical devices, sensors, and software that are interconnected and are capable of collecting and exchanging data [4], [5]. IoT facilitates the seamless integration and communication of various devices, enabling real-time monitoring and control of water-related parameters. This system measures and monitors both water consumption and quality in real-time using low-cost sensors, allowing effective monitoring by NWSC and customers. Furthermore, ML algorithms are incorporated to analyze the collected data, identifying the key parameters for water quality assessment. Thresholds for water quality assessment and water consumption monitoring are determined by the national standards for potable water [6] and average monthly consumption respectively. Finally, alert messages are incorporated to improve response time to faults and customer safety.

The main contributions of the paper are:

1) The system integrates both water quality and consumption monitoring into a single system by utilizing IoT technology. This approach provides a holistic view of water management.
2) Data-driven analysis of water quality by employing feature scaling ML algorithms to assess water quality and consumption in Uganda. This data-driven approach facilitates targeted interventions on the root causes of water pollution.

The rest of the paper is structured as follows: Section II presents a comprehensive literature review of related works. Section III describes the methodology. Section IV presents the relevant results, highlighting the system's performance. Section V concludes the paper and Section VI gives insights into the future work.

II. Literature Review

This section explores the different methodologies that have been employed to help solve problems related to both water quality and water consumption.

TABLE I
MAXIMUM PERMISSIBLE NATIONAL STANDARDS FOR POTABLE WATER (WHO)

Parameter	Maximum Permissible	Unit
pH	6.5 - 8.5	-
Turbidity	5.0	NTU
Total Dissolved Solids (TDS)	700	mg/L
Hardness (as $CaCO_3$)	500	mg/L
Chlorine Residual	0.2 - 5	mg/L
Nitrate (as NO_3)	5.0	mg/L
Fluoride	1.5	mg/L
Arsenic	0.01	mg/L
Lead	0.05	mg/L
Mercury	0.001	mg/L
Calcium	75	mg/L
Magnesium	50	mg/L
Bi-carbonate	500	mg/L
Manganese	0.2	mg/L
Chloride	250	mg/L
Iron	<0.3	mg/L
Sulphate	200	mg/L
Ammonia	0.5	mg/L
Cadmium	<0.001	mg/L
Copper	1.0	mg/L
Zinc	5.0	mg/L
E. Coli	0	CFU/100mL
Total Coliforms	0	CFU/100mL
Fecal Coliforms	0	CFU/100mL

A. Water Quality Assessment

Authors in [7] introduce a method that involves the use of specific laboratory equipment to collect water samples that are then assessed in the laboratory. In order to determine water quality, multiple water quality parameters are tested and evaluated according to national standards set by WHO [6] shown in Table 1.

Authors in [8] discuss a method in which multiple sensors including conductivity, pH, and turbidity are combined to measure the various water quality parameters. Each sensor measures its respective parameter and the data is stored in a remote server. The server then evaluates the values against thresholds set according to the national standards for potable water set by WHO [6].

In the ML-based water quality assessment method presented in [9], the process consists of three main steps: data collection and processing, model training, and model evaluation. During the data collection and processing phase, various water quality parameters are merged to form a standardized measure known as the Water Quality Index (WQI). This index serves as the basis for determining the Water Quality Class (WQC), which enables the utilization of ML models for prediction purposes. The WQI is computed using (1), and the corresponding WQC can be determined by referencing Table 2.

$$\text{WQI} = \frac{\sum W_n \times \sum Q_n}{\sum W_n}. \tag{1}$$

As indicated in (1), W_n is the unit weight, expressed as:

$$W_n = \frac{1}{\sum \frac{1}{S_n}}, \tag{2}$$

TABLE II
WATER QUALITY CLASS (WQC)

WQI Range	Class
0-25	Excellent
25-50	Good
50-70	Poor
70-90	Very Poor
90-100	Unfit for Consumption

where S_n is the standard desirable value of the given water parameter under evaluation. Q_n is the sub-index value and is expressed as:

$$Q_n = \frac{(V_n - V_o)}{(S_n - V_o)} \times 100\%, \tag{3}$$

where V_n is the mean concentration of the water quality parameter and V_o is the actual value of the parameters in pure water.

During the model training phase, feature selection classification and regression algorithms are employed to identify the predominant water quality parameters. Subsequently, in the evaluation stage, multiple Classification and Regression ML models are assessed to identify the models that exhibit the highest performance in predicting the WQC.

B. Water Consumption Monitoring

Water metering is the process of measuring the amount of water used. As such, the water meter is an essential tool for both the utility company and its customers to measure the actual water consumption. There are a number of different types of water meters such as positive displacement meters, velocity water meters, mechanical water meters, and digital water meters [10]. An IoT-based smart water meter can also be used to monitor the water usage through wireless flow sensor nodes that are capable of sending consumption data to a remote server in real time over the Internet. The consumption data can then be analyzed such that when water is excessively used, a notification is sent to the customer to actively monitor their water consumption.

III. METHODOLOGY

This section describes the entire process of developing, building, and testing the water quality assessment and consumption monitoring prototype.

A. Water Quality Assessment subsystem

The following steps were involved in developing the prototype for water quality assessment system:

- Determining the water quality parameters using ML.
- Designing the water quality system prototype.
- Implementing the water quality system prototype.

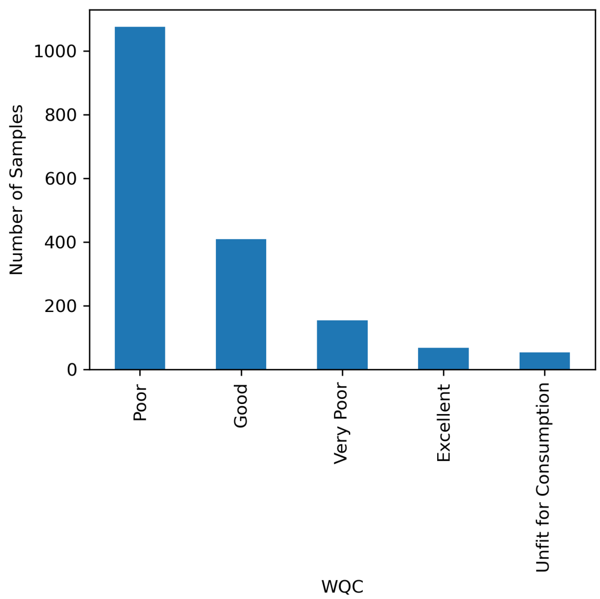

Fig. 1. WQC distribution.

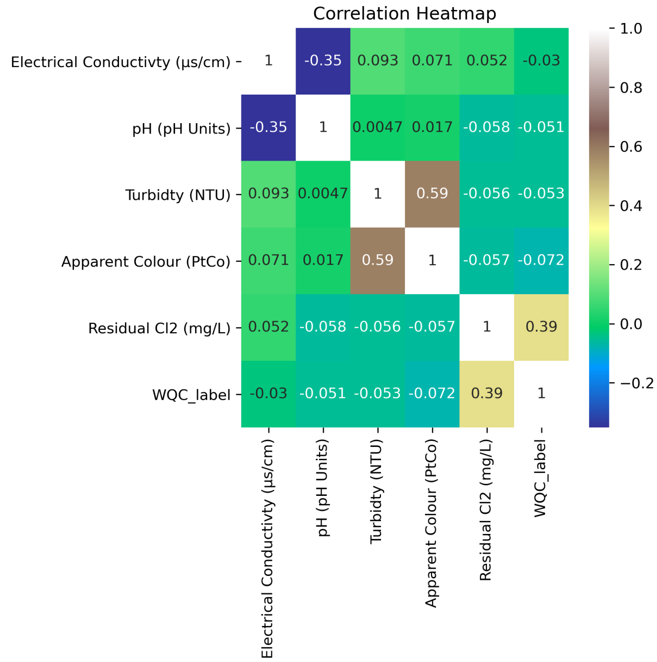

Fig. 2. Correlation of the different water quality parameters with WQC.

1) Determining the water quality parameters: The relevant water quality parameters to be measured were determined using a method of Classification ML called feature selection. The steps involved include: data processing, model training, and finally model evaluation.

a) Data processing: In this step, a dataset provided by NWSC consisting of 1,760 water samples collected from tertiary points within the Kampala service area in 2022 was processed. This involved cleaning the data and subsequently using (1) to calculate the WQI. The WQI values were then utilized to determine the WQC based on the reference provided in Table 2. The distribution of the number of samples with respect to the WQC is shown in Fig. 1.

We then conducted a correlation analysis to examine the association between various water quality parameters and their corresponding WQCs. Our findings indicated that residual chlorine exhibited the strongest correlation with the WQC, as depicted in Fig. 2. The elevated levels of residual chlorine can be attributed to its initial high concentration at the distribution point. This practice is commonly employed to ensure effective chlorine treatment reaches all customers, especially those located at the end of the distribution network.

Feature scaling was then performed to ensure the data is both normalized and fits into a single scale from 0 to 1 for all water quality parameters. This helps improve the accuracy of the ML algorithms when predicting the WQI.

Data splitting: Finally, the dataset was partitioned into an 80% training set and a 20% testing set which was used to train the ML models.

b) Model training and evaluation: The training phase involved utilizing the training set, which consisted of 80% of the collected data, to train various classification ML models.

TABLE III
PERFORMANCE OF ML MODELS

ML model	Accuracy	Precision	Recall	F1-Score
Random Forest	89.2%	85%	88%	88%
Gradient Boosting	89.20%	83%	90%	86%
Extra Trees	88.63%	86%	86%	85%
Decision Tree	86.93%	87%	87%	87%
Bagging	86.64%	88%	88%	88%
k-NN	77.27%	70%	77%	73%
Logistic Regression	38.35%	39%	56%	38%

Additionally, to assess the performance of the trained models, the remaining 20% of the data, known as the testing set, was used for evaluation. The models were evaluated based on four metrics: accuracy score, precision, recall, and F1 score, to determine their effectiveness in predicting the WQC. The results of model evaluation are shown in Table 3. Notably, the Random Forest Classifier showed the best performance, achieving an accuracy score of 89%, precision of 85%, recall of 88%, and an F1-score of 88%.

c) Feature selection: Finally, feature selection was employed to determine the key parameters based on the features with the most influence on the performance of the Random Forest Classifier. As illustrated in Fig. 3, residual chlorine is observed to be the most dominant, followed by pH, electrical conductivity and turbidity, and finally apparent colour.

In this research, pH and turbidity were selected over residual chlorine, despite residual chlorine having the highest correlation with WQI. This decision was influenced by sensor availability. While residual chlorine is an important indicator of water treatment effectiveness, pH and turbidity provide important information on the chemical and physical characteristics

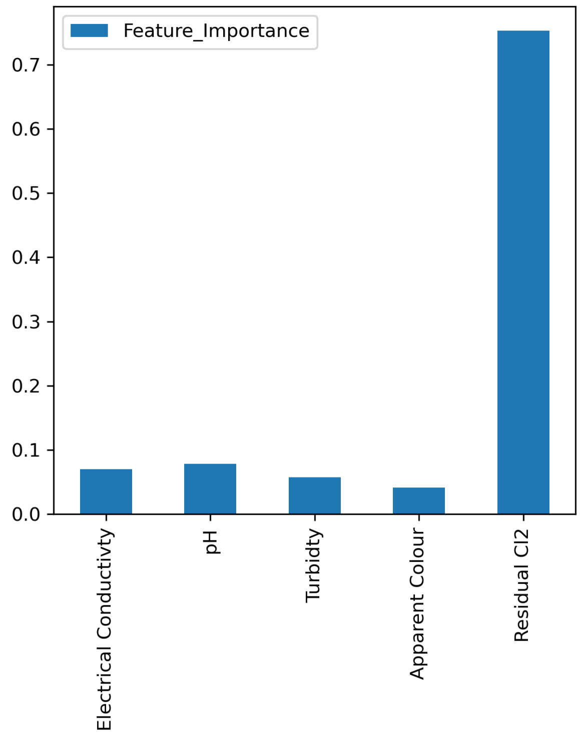

Fig. 3. Feature selection of the parameters.

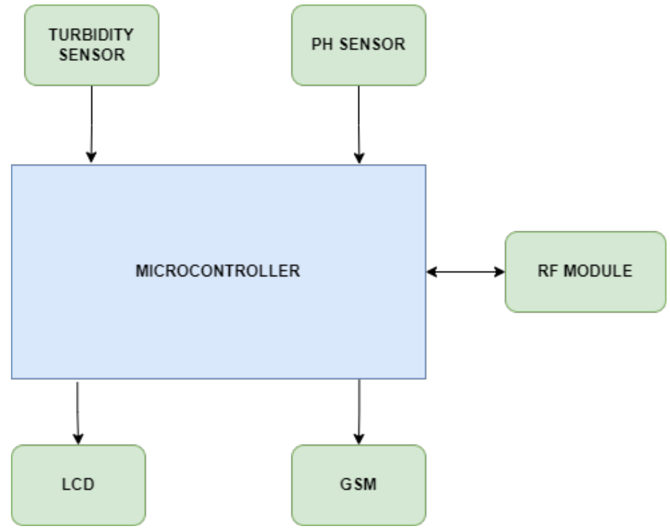

Fig. 4. Water quality system architecture diagram

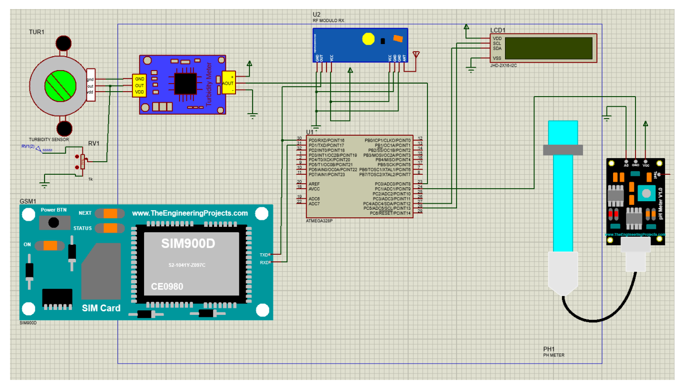

Fig. 5. Water quality system schematic diagram

of water [9] and their sensors are more widely available and easier to deploy in real-time water quality assessment systems. Additionally, pH and turbidity can be used to monitor other aspects of water quality such as the presence of suspended solids or microorganisms, which residual chlorine may not directly reflect [9]. Thus, the selection of pH and turbidity in this study offers a sufficient assessment of the water quality.

2) Designing the water quality subsystem prototype: The basic system architecture of the water quality subsystem is depicted in Fig. 4 encompassing a range of components dedicated to seamless data collection and transmission. In this comprehensive setup, the turbidity sensor and pH sensor play pivotal roles in acquiring essential water quality data. Once collected, the data is efficiently transmitted to the water consumption subsystem via a radio frequency (RF) module. Furthermore, the integration of a Global System for Mobile Communications (GSM) module is utilized to efficiently upload this critical data to a remote cloud server facilitating accessibility and data analysis. Lastly, to provide an intuitive user interface, a Liquid Crystal Display (LCD) is used to offer a visual representation of the real-time water quality information. The equivalent schematic diagram derived from the system architecture is presented in Fig. 5.

B. Water Consumption Monitoring subsystem

The water consumption monitoring subsystem was developed following a similar approach to the water quality assessment subsystem. Fig. 6 shows the basic system architecture of the subsystem. The design consists of a water flow sensor that collects consumption data, an RF module which transmits the consumption data to the water quality assessment subsystem, an SD card module which stores the data collected from all the sensors, and a Real Time Clock (RTC) module which provides the date and time of data collection. An LCD is also included to provide a visual representation of the water consumption in liters. Furthermore, the system architecture allows for scalability, enabling additional sensors to be integrated for measuring other water quality parameters. The equivalent schematic diagram as obtained from the water consumption architecture is represented by Fig. 7.

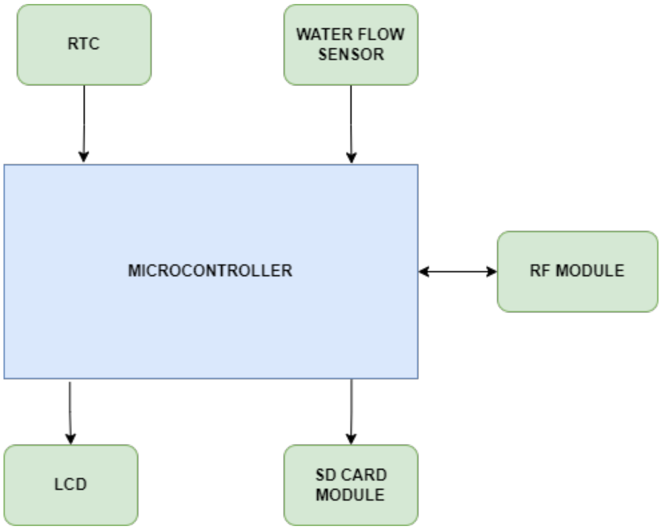

Fig. 6. Water consumption system architecture diagram

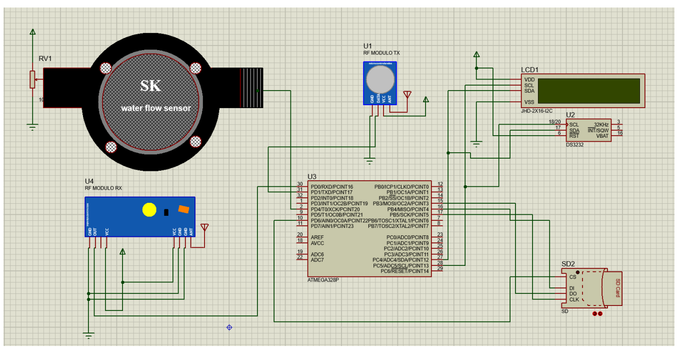

Fig. 7. Water consumption system schematic diagram

IV. Results

This section discusses the results obtained from testing the system. The tests carried out include:

- Measurement of water consumption and water quality.
- Data transmission between the two systems.
- Storage of data on the SD card.
- Uploading data and receiving email alerts from a remote ThingSpeak cloud server.

A. Measurement of water consumption and water quality

The water consumption test was conducted to test the accuracy of the water consumption subsystem by measuring the volume of water in liters (L). The results, depicted in Fig. 8, display the water consumption as 0.91 L on the LCD screen. Similarly, the water quality test aimed to evaluate the accuracy of the water quality assessment subsystem. The assessment is based on the compliance of both pH and turbidity values with national standards for potable water, where pH should range between 5.5 and 8.5, and turbidity should be between 0 and 5 [4]. Consequently, the water quality is classified as either good or bad depending on its adherence to these thresholds. Figs. 8 and 9 demonstrate the classification of water quality

Fig. 8. LCD display of water consumption.

Fig. 9. LCD display showing that water quality is "Good".

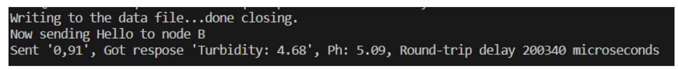

Fig. 10. RF acknowledgement message

based on these criteria, with the results conveniently displayed on an LCD screen.

B. Data transmission between the two systems

Both the water consumption monitoring subsystem and the water quality assessment subsystem are designed for seamless communication with each other. This is achieved using the NRF24L01 radio transceiver modules that can wirelessly communicate with each other over a 100m radius with line of sight. As such, when data is received by either of the systems, an acknowledgement message is returned by both modules, this message is shown in Fig. 10.

C. Storage of data on the SD card

The water consumption monitoring subsystem has an SD card that is used to store both the volume data and the quality data. This data is stored in the SD card as a text file as indicated by Fig. 11.

D. Uploading data and receiving email alerts from a remote ThingSpeak cloud server

All the data is uploaded to ThingSpeak using a GSM module located on the water quality subsystem. When the data has been successfully uploaded to ThingSpeak, it can be accessed on the ThingSpeak dashboard as observed by Fig. 12. Furthermore, the server has the capability to send email alerts to both NWSC and their customers, as depicted in Fig. 13. This feature enables faster response to faults in the water supply.

V. Conclusion

In conclusion, the development of a combined water quality assessment and consumption monitoring system brings significant advantages to water management. It enables accurate real-time monitoring of both water quality and consumption, ensuring the safety and reliability of the water supply. By integrating low-cost sensors, the system enables cost-effective

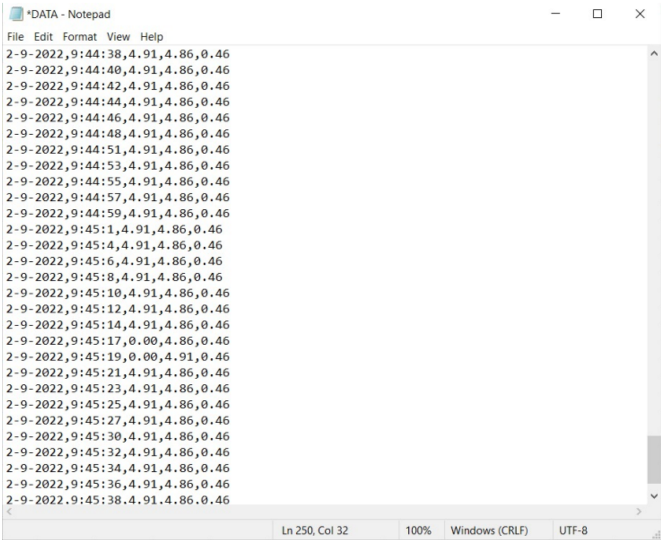

Fig. 11. Data text file saved on the SD card.

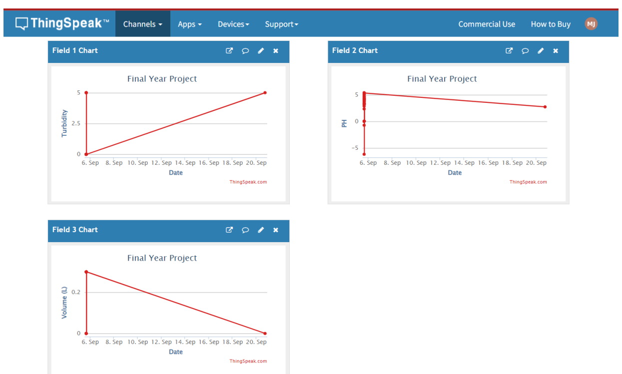

Fig. 12. Graphical data representation on ThingSpeak

measurement and monitoring of various water parameters. The use of ThingSpeak server and SD card storage facilitates remote monitoring and on-site access to data, promoting efficient management and analysis. The inclusion of email alerts enhances response times, enabling prompt actions in addressing water quality issues. Graphical data representation aids in on-site fault investigation, reducing the need for physical assessment and improving customer safety.

An essential aspect of the system is the application of feature scaling ML techniques to determine key parameters for water quality assessment. This approach allows for efficient evaluation of the water's chemical and physical characteristics and facilitates targeted investigation of water quality parameters when polluted. Our assessment showed the most dominant parameters to be residual chlorine, pH, turbidity, conductivity, and apparent color. Overall, this combined system contributes significantly to efficient water management, emphasizing the importance of integrating multiple functionalities and leveraging ML techniques for accurate assessment and monitoring.

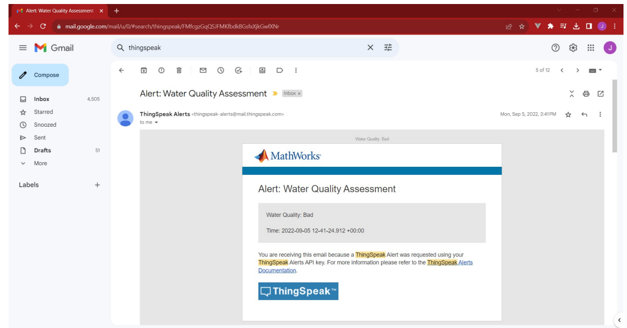

Fig. 13. Email alert from ThingSpeak

VI. Future Work

Future work will include integrating additional sensors to the system to measure other water quality parameters and provide a more comprehensive assessment. Furthermore, there is potential for scalability, enabling the system to be deployed in larger water supply networks.

Acknowledgment

We thank the management and staff of NWSC for their help and support during this study.

References

[1] K. Khatri, K. Vairavamoorthy, and M. Porto, "Challenges for urban water supply and sanitation in developing countries," in *Water for a Changing World-Developing Local Knowledge and Capacity*. CRC Press, 2008, pp. 93–112.

[2] T. Boyle, D. Giurco, P. Mukheibir, A. Liu, C. Moy, S. White, and R. Stewart, "Intelligent metering for urban water: A review," *Water*, vol. 5, no. 3, pp. 1052–1081, 2013.

[3] M. Robert Biti and A. N. Gisele, "An online web based e-billing water system: A case study national water and sewerage corporation (nwsc)," 2012.

[4] S. Li, L. D. Xu, and S. Zhao, "The internet of things: a survey," *Information systems frontiers*, vol. 17, pp. 243–259, 2015.

[5] A. Ghasempour, "Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges," *Inventions*, vol. 22, March 2019.

[6] S. Nayebare, L. Wilson, D. Carpenter, and D. Dziewulski, "A review of potable water accessibility and sustainability issues in developing countries - case study of uganda," *Reviews on environmental health*, vol. 0, 06 2014.

[7] J. Pritt and J. W. Raese, *Quality Assurance/quality Control Manual: National Water Quality Laboratory*. US Geological Survey, 1995, vol. 95, no. 443.

[8] 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, 2021.

[9] U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, R. Irfan, and J. García-Nieto, "Efficient water quality prediction using supervised machine learning," *Water*, vol. 11, no. 11, p. 2210, 2019.

[10] H. E. Mutikanga, S. K. Sharma, and K. Vairavamoorthy, "Investigating water meter performance in developing countries: A case study of kampala, uganda," *Water SA*, vol. 37, no. 4, pp. 567–574, 2011.