

# A Raw Water Quality Monitoring System using Wireless Sensor Networks

Nahshon Mokuu  
School of Engineering,  
Dedan Kimathi University of  
Technology  
Private Bag 10143  
Nyeri - Kenya

Ciira wa Maina  
Center for Data Science and  
Artificial Intelligence,  
Dedan Kimathi University of  
Technology  
Private Bag 10143  
Nyeri – Kenya

Henry Kiragu  
Electrical and Communication  
Engineering Department,  
Multimedia University of Kenya  
P.O BOX 15653 - 00503  
Nairobi – Kenya

## ABSTRACT

Water treatment can be promoted through keen consideration of raw water quality parameters (Turbidity and pH). This paper discusses the development of a real-time water quality monitoring system using wireless sensor networks. At first, we present performance experiments on *LoRa* technology connectivity for wireless sensor networks in a rural set up of Dedan Kimathi University of Technology in Kenya. The specific sensors used for the developed system included: The DFRobot gravity Arduino turbidity sensor and the DFRobot's Gravity Analog pH Sensor. The sensed data values of these parameters were relayed to a gateway by a *LoRaWAN* transceiver. The gateway then uploaded the received parameter data values to The Things Network platform which was interfaced with a Google Cloud Platform, where an *InfluxDB* Virtual Machine database stored the received data. A web-based application (Dash Plotly app) was developed and interlinked with the database for analysis and visualization of the received data in real time. The system was deployed at the Nyeri Water and Sanitation Company treatment plant based at Nyeri town, Kenya, from 4th November, 2020 to 4th January, 2021. The dataset obtained contained a total of 2,658 records, each collected after every 30 minutes. Using a subset of 291 records, extensive experiments were performed for the evaluation and assessment of machine learning anomaly detection algorithms of the Local Outlier Factor and the Robust Random Cut Forest for each of the two parameters; Turbidity and pH. From analysis results, the Local Outlier Factor algorithm outperformed its counterpart.

## General Terms

Water quality, anomalies, machine learning, data

## Keywords

Water quality monitoring; wireless sensor networks; anomaly detection; local outlier factor; robust random cut forest.

## 1. INTRODUCTION

Human activities have various effects on the environment which have adversely impacted on human health in many ways [1]. Developing countries have been the most affected by the growth of numerous slums, poor sanitation, and post-mining effects. The collective effect heightens a deteriorating environment. Environmental monitoring programs and systems have therefore been established globally to promote environmental sustainability [2]. For this research, we shall focus on water quality systems.

Freshwater management has faced severe challenges in many world economies. These challenges are due to escalating

competition for freshwater from many quarters of the ecosystem and human activities. The overexploitation of freshwater has reduced its availability for agricultural uses. As a consequence, poverty alleviation has become more difficult because agricultural development is a critical contributor to its reduction. Since water is an integral contributor to food security [3], the 2002 World Summit on Sustainable Development focused on water management and its relation to the Millennium Development Goals (MDGs) [4]. The consensus of the summit was that for water resource sustainability to be achieved, these resources should be exploited with care, bearing in mind their importance to future generations. As is, the current exploitation trends and competition for water resources fail to guarantee that the envisioned sustainability will be achieved. Therefore, all stakeholders were tasked to make rational decisions, projections, and plans to exploit and manage water resources sustainably. It was agreed that a novel universally accepted approach must be employed at all levels of society if water management goals are to be achieved [4].

In the early 2000s, new technology was integrated into water quality management (WQM) to remediate some limitations in the manual methods employed in the previous five decades. Notably, microelectronic mechanical sensors, fibre optics, laser technology, biosensors, among other sensors, revolutionized water quality analysis [5], [6]. These sensors identify various aspects of water quality in situ. Moreover, the advanced technology introduced water telemetry, which enhanced the acquisition of water quality data and accompanying monitoring procedures. Satellite technology also facilitated the acquisition of water images used for approximation of different water quality parameters [7]. Lakes, rivers, springs and seas, among other water bodies could also be monitored using visualization architectures for water quality courtesy of modern technology [5], [8].

The introduction of wireless sensor networks (WSNs) further bolstered WQM due to improved communication systems. Their ease of operation has made them increasingly popular. These networks have promoted quick capture, transmission and analysis of data relating to the environment. Application of WSNs in WQM procedures has lowered the sensing costs and increased the amount of data and sampling points analyzed at any particular moment. The WSNs also have an in-built capability to transfer data by utilizing low power techniques. This capability enhances the easy remote receipt of data from numerous data sensors and points. Therefore, the new technology is more appealing than the original manual method.

Many countries have continually faced challenges in sustaining the supply of safe drinking water. Wireless sensor networks are a potential solution to water quality management because of their effectiveness, long-term affordability, and ease of use. They provide sufficient control capabilities to the research community since monitoring and evaluation of water quality are done remotely, quickly, and devoid of human interference. However, the communication sector poses a formidable obstacle to this proposed solution. Proper communication over these WSN requires long-lasting batteries, low-cost technology, and above all, the correct coverage.

This research developed a real-time and low-cost water quality monitoring system based on wireless sensor networks. The major goals include: Determining the range of coverage and the strength of connectivity for Long-Range (*LoRa*) technology in a rural setting; accurate calibration and integration of the pH and the turbidity sensors to be used in the sensor node; Fabrication, testing and deployment of the developed sensor nodes for water quality parameter measurement; and finally, the development of a machine learning anomaly detection algorithm for the monitored parameters.

The rest of the paper is organized as follows: Section 2 describes the theoretical background of water body monitoring systems. Section 3 gives an overview of the proposed methodology, and Section 4 presents the results of the proposed methodology. Finally, the conclusion and future work are outlined in Section 5.

## 2. PREVIOUS WORK

In the last two decades, various researchers have submitted that a WSN is the most suitable method of WQM [9]. Online platforms have also been increasingly used to analyze data and automatically discern water quality problems in the past few years [10]. Research indicates that the WSNs method overcomes most of the limitations experienced in the traditional manual based in situ and traditional manual lab-based methods. Unlike the traditional approaches, the WSNs method can replace outdated and expensive equipment with low-cost sensors. They eliminate the need for the transportation of data samples to the laboratory, thereby saving time, and reducing costs in the process. The training of workers, collection of samples, data recording, and data analysis with the WSNs has proven cheaper when compared to the traditional WQM techniques [6], [10], [11]. Therefore, the WSNs method is currently more plausible than its traditional competitors.

The WSNs can be designed to track the quality of water in freshwater sites. However, various aspects that must be considered before its implementation include; sensing abilities of the nodes, the level of signal processing, network layout, and whether the sensors are likely to use acoustic or radio communication. The WSNs can be used to track the water quality parameters such as the pH, temperature, turbidity, and the level of dissolved oxygen in water [12], [13], [14], [15]. However, there seems to be insufficient research towards incorporation of machine learning algorithms in anomaly detection. On energy management schemes, methods have been proposed and WSNs have been configured to go on sleep mode when not in use [16], a method that is adopted in this research. The Local Outlier Factor [17] and the Robust Random Cut Forest [18] are among the most popular and appropriate anomaly detection algorithms for time series water quality data.

## 3. METHODOLOGY

The adopted methodology is represented by the block diagram in Figure 1.

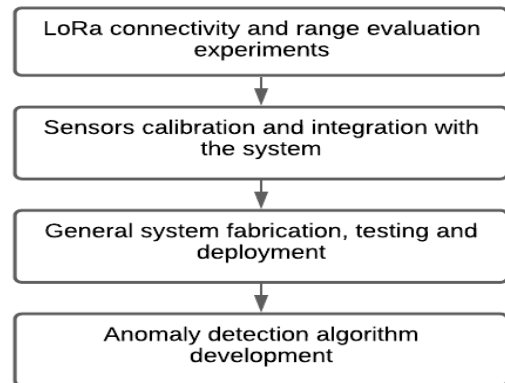


Fig 1: Methodology block diagram

### 3.1 LoRa Connectivity and Range Evaluation

An experimental evaluation of the proprietary parts of *LoRa* Technology were conducted to ascertain if it performs as advertised. The aim of this procedure was two-fold: To conduct performance experiments on *LoRa* connectivity and range evaluation for wireless sensor networks; and to present and discuss the results obtained in regards to connectivity and range evaluation for purposes of the developed WQM system.

Measurement of the outlined parameters were conducted at the Dedan Kimathi University of Technology, Central Kenya, at different times throughout the day over several days. End devices that sent payloads periodically to the base station were deployed at different locations away from the base station. These locations were 100m apart, at a 1km path range along a line of sight (LoS) from a 2.5m stand node as shown in Figure 2. For every transmitted payload, there was a measure of the Received Signal Strength (RSSI), which is used in the connectivity and range of evaluation studies hereof.

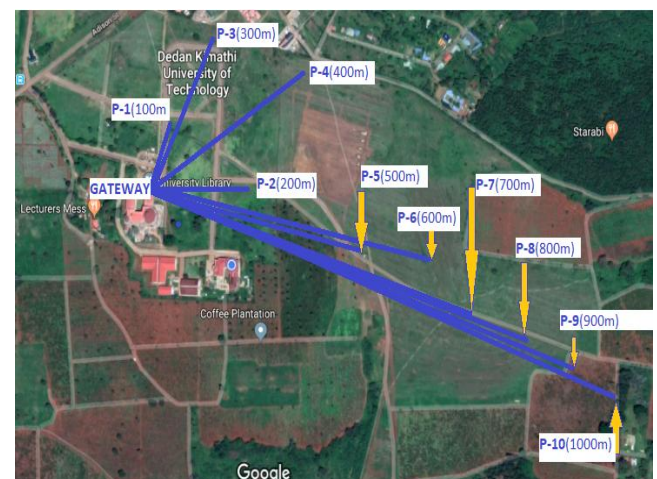


Fig 2: Test Points Geographical Locations. [Google Maps]

#### 3.1.1 Base Station

The configured and installed LoRaWAN Industry gateway at the Dedan Kimathi University of Technology was used (Figure 3). The location of the gateway is approximately 25 meters above the ground, on the roof of a centrally situated building (*The Resource Center*) at the university.

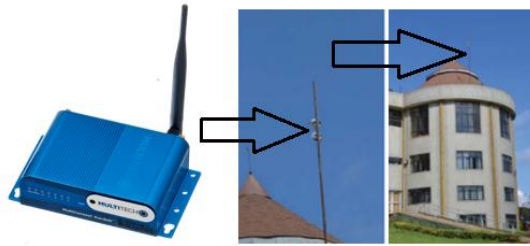


Fig 3: The LoRaWAN Industry gateway (MultiTech Conduit)

### 3.1.2 End Device

The end-device was an STM32 Nucleo board (Figure 4), equipped with a LoRaWAN Transceiver Shield. While taking the measurements, the nodes were powered by 9V batteries. The transmit power was +14 dBm at a frequency of 868MHz. For on-ground measurements, the node was attached to a stand, approximately 2.5 m height of the ground. Each device was registered on The Things Network (TTN) platform, which forwarded data to the database for storage after receiving it from the device.

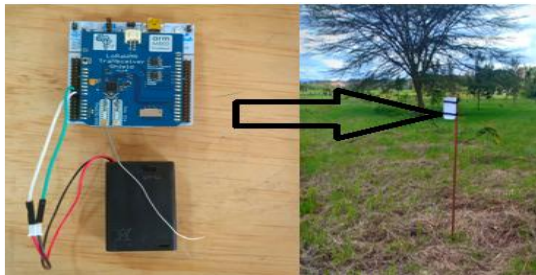


Fig 4: STM32 Nucleo board, with a LoRaWAN Transceiver Shield

## 3.2 Raw Water Quality Parameters

The developed system studied and reviewed the parameters of turbidity and pH; both which determine water treatment procedures. Equivalent sensor calibration methodologies were followed for the pH and turbidity sensors. A lab assessment (utilizing standard buffer solutions and reference instruments) was conducted at the NYEWASCO water treatment plant labs. Significant consideration was given to obtain linear responses, mitigate noise, and achieve high accuracy and quality resolution.

### 3.2.1 Turbidity (TU)

In this parameter, the level of cloudiness of water caused by microorganisms and floating particles is measured. TU is obtained by the use of the ISO 7027 approach where infrared light scatters at right angles to cross beams. TU is indicated in Nephelometric Turbidity Units (NTU). Turbid waters are susceptible to escalated growth of microbes as they provide sufficient food and shelter for pathogens.

The DFRobot gravity Arduino turbidity sensor (Figure 5), was used to detect the opaqueness levels of water. It utilizes light to sense the suspended solid particles that affect the transmission and scattering of light. The sensor operates at voltage of 5V DC and a maximum 40 mA. A temperature range of between 5°C and 90°C proves ideal for this sensor. Its response time is 500ms.

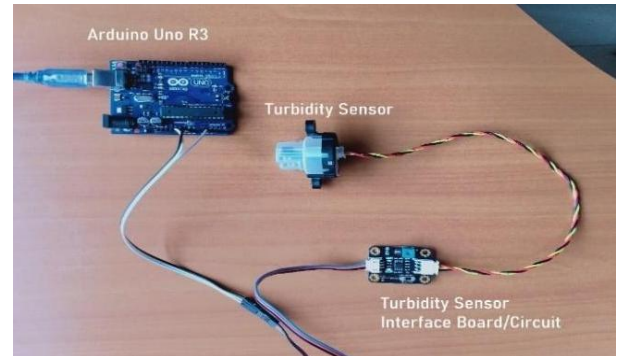


Fig 5: The DFRobot gravity Arduino turbidity sensor interfaced with Arduino UNO R3 Board for calibration

To ascertain the provided equation, a number of standard solutions (0, 20, 40, 100, 200, 1008 and 4000 NTU) available at NYEWASCO water quality laboratory were used to carry out secondary calibration. This exercise was conducted on Wednesday, August 2020 at a room temperature of 22° C. Using a turbidimeter as a primary instrument, the turbidity of these standard solutions was measured and the experiments repeated with this sensor while records were being kept.

### 3.2.2 pH

pH is the level of acidity or alkalinity of water. It is usually indicated by the use of a negative logarithm of the concentration of hydrogen ions in water. If the pH of water suddenly changes by a minimum of 0.5 pH units, there is a reason to suspect contamination.

A DFRobot's Gravity Analog pH sensor to gauge the solution pH and mirror its acidity or alkalinity was deployed for this framework (Figure 6). Its activity voltage ranges between 3.3 to 5.5V, and an accuracy of  $\pm 0.1$  at 22°C; recognition scope of 0 to 14, and activity temperature range of between 5 and 60°C. Its response time is stipulated to be one minute and a resolution of 0.01.

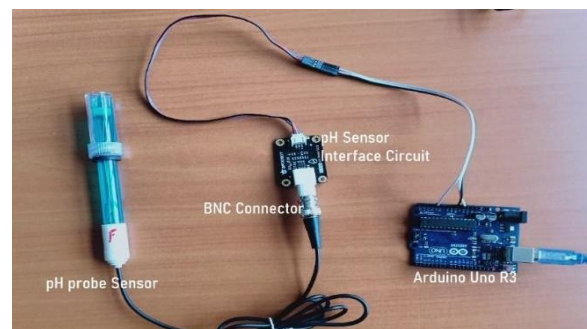


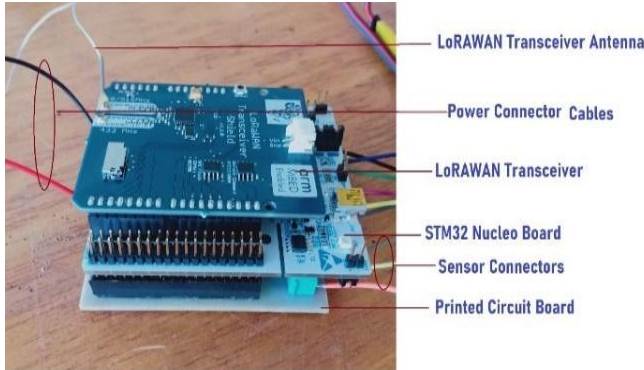
Fig 6: DFRobot's Gravity Analog pH sensor interfaced with Arduino UNO R3 Module for calibration

Thereafter, the pH of several solutions was obtained using the Kenya Bureau of Standards calibrated pH meter alongside our sensor probe. This exercise was conducted at NYEWASCO water quality laboratory on Wednesday, August 2020 at a room temperature of 21°C. This was meant for result validation purposes.

## 3.3 Sensor Node Fabrication and Deployment

Printed circuit boards (PCBs) play a very crucial role in development of micro-controlled systems. They allow circuits to be realized with a minimum number of connectors and hence optimization of the occupied space on a fabricated

PCB. For this prototype, the STM32 Nucleo F466RE microcontroller was used. The schematic and layout of the STM32 Nucleo F466RE microcontroller board were prepared using the KiCad 4.0.7 software, a suite for electronic design automation. The fabricated sensor node before packaging is shown in Figure 7.



**Fig 7: Sensor Node PCB Fabrication Design interfaced with the STM32 Nucleo board and the LORAWAN Transceiver**

The central measurement node successfully fabricated was deployed at a local water treatment company in Kenya, as shown in Figure 8, in the raw water section on the 4<sup>th</sup> of November 2020 to the 4<sup>th</sup> of January 2020 as shown in Figure 8. The device successfully collected raw water pH and turbidity every 30 minutes for 60 days.



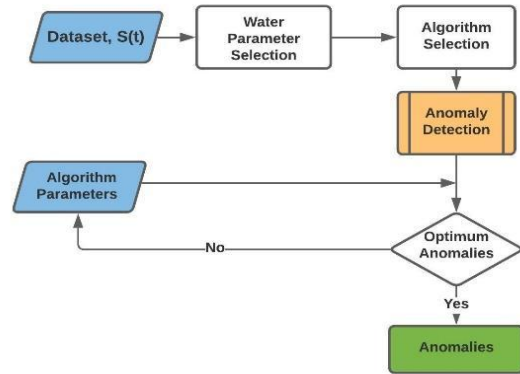
**Fig 8: The Central Measurement Node deployed**

### 3.4 System Energy Management

The general power utilization includes that of the focal measurement sensor hub, and the LoRaWAN transceiver module that sends data on water quality, and it operated at about 50mA at 5V working voltage per minute. Regular sensors expend large amounts of power. Many coordinated circuits, including the STM32 Nucleo board, did not adequately supply the power of such intensities. Interfacing these sensors and the Nucleo board power pins directly constantly could harm it. To overcome this, an H-Bridge motor control circuit utilizing the L293D Motor Driver IC to connect the sensors and the Nucleo board was deployed. This also enabled sleep mode for 58 minutes within an hour of operation of the node.

## 3.5 Data and the Anomaly Detection Algorithm Development

In this section, the anomaly detection techniques of the local outlier factor (LOF) and the robust random cut forest (RRCF) are evaluated thoroughly. A subset of the dataset, with 291 records was extracted considering a region with graphically notable anomalies, and used as the ground truth. A section of the dataset between 11<sup>th</sup> and 18<sup>th</sup> November 2020 (7 days) was used. It was then manually examined and all the outlier instances identified and later analyzed using the algorithms as depicted in Figure 9. The findings for every parameter, with each algorithm are discussed in the results section.

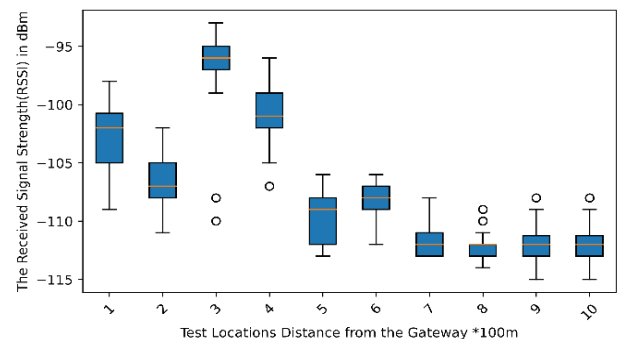


**Fig 9: Anomaly Detection Algorithms Evaluation Process**

## 4. RESULTS AND DISCUSSION

### 4.1 LoRa Connectivity and Range Evaluation

The value range of typical LoRa RSSI is -120 dBm to -30dBm. At 100m away from the gateway, a mean strength of -102.7 dBm was recorded, while at 200m, the mean signal strength was -106.5 dBm. A complete record of the computed mean RSSI values for every test location are plotted in Figure 10. The best strength was realized at test location 3 (300m), while it is notable that this value decreased (worsens) as we moved to test locations away from the gateway. Outlier RSSIs were realized in test locations 3, 4, 8, and 9 and they are plotted as individual points. The highest notable degree of dispersion (spread) and skewness in the RSSI is easily observed with test locations 1, 2, 4 and 5, whereas test location 8 depicts the contrary.



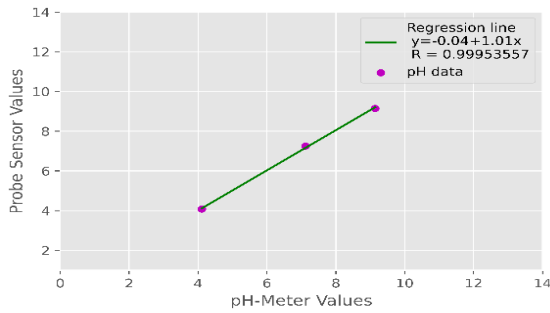
**Fig 10: The Received Strength Whisker and Box Plots for the 10 Test Locations**

### 4.2 Sensors Calibration

#### 4.2.1 The pH Sensor

For three standard solutions, values of both the pHmeter and

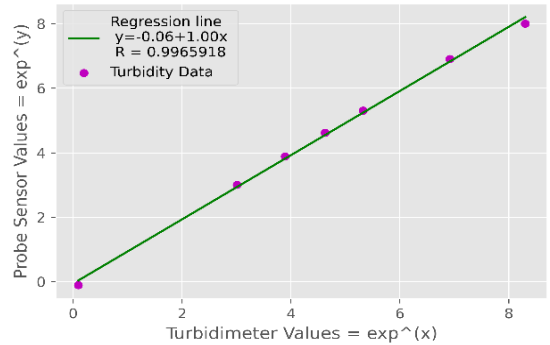
our sensor probes were recorded. A plot of the results (Figure 11) gave a correlation co-efficient approximately equal to one to ascertain proper calibration procedures.



**Fig 11: pH Standard Solutions comparison of our probe against a standard calibrated pHmeter**

#### 4.2.2 Turbidity Sensor

During validation of the measurements of the Gravity DFRobot Arduino sensors using a turbidimeter as a primary instrument, the results were obtained and graphed in Figure 12. With a correlation co-efficient of one, the probe was considered effectively calibrated.



**Fig 12: Turbidity Standard Solutions comparison of our probe against a standard calibrated turbidimeter**

### 4.3 Anomaly Detection Algorithms

#### 4.3.1 Dataset

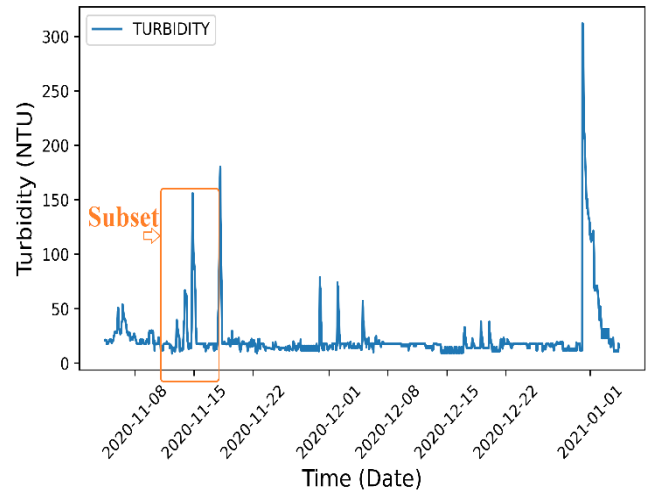
Table 1 below shows a subsection of the 2,658 records of both turbidity and pH that were collected in the period of 60 days.

**Table 1: Turbidity and pH Dataset**

	time	turbidity	pH
0	2020-11-04 11:00:31.822439+00:00	21.063435	7.34
1	2020-11-04 11:01:22.124333+00:00	20.868153	7.33
2	2020-11-04 11:01:51.663062+00:00	20.584553	7.32
3	2020-11-04 11:02:29.373718+00:00	21.185328	7.33
4	2020-11-04 11:03:45.517010+00:00	21.063435	7.32
...	...	...	...
2653	2021-01-04 07:53:20.987423+00:00	10.611506	7.35
2654	2021-01-04 08:23:37.035804+00:00	17.975997	7.35
2655	2021-01-04 08:53:53.104009+00:00	17.734662	7.34
2656	2021-01-04 09:24:09.578901+00:00	15.094176	7.36
2657	2021-01-04 09:54:25.214766+00:00	14.611506	7.36

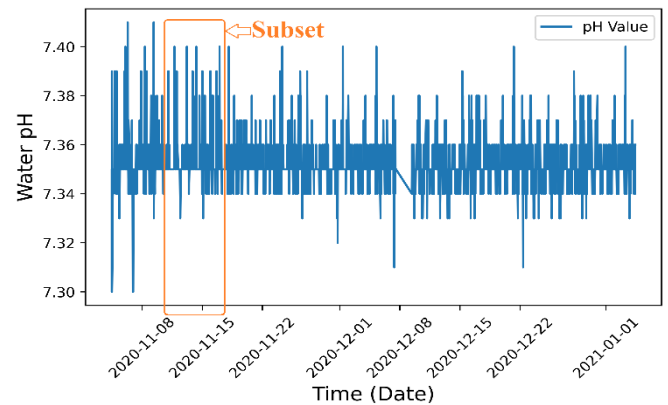
2658 rows × 3 columns

The output plot diagram for turbidity data in Figure 13 clearly shows that there are several instances which are far from the others and therefore, anomalies are evident. The subset under testing is highlighted.



**Fig 13: Turbidity Dataset for the Sixty (60) days**

Similarly, water pH data in Figure 14 below shows existence of outliers. The subset that was considered as the test data is shown in the round corner box.



**Fig 14: pH Dataset for the Sixty (60) days**

#### 4.3.2 pH

##### 4.3.2.1 Local Outlier Factor

The local outlier factor algorithm was used to detect the water pH outliers shown in Table 2. In Figure 15, the red stars diagram are the 63 instances detected as anomalies in the pH data using the number of neighbors,  $k = 100$ . The algorithm took 21 milliseconds to determine the anomalies. There were no notable false alarms as well as undetected outliers. Choosing an optimal  $k$  was essential for detection performance. For a value of  $k$  too small or very large, the error went up due to under-fitting.

**Table 2: pH outliers as detected by the LOF algorithm**

time	pH
2020-11-11 01:11:34.389403+00:00	7.39
2020-11-11 03:42:54.674932+00:00	7.36
2020-11-11 18:12:35.713854+00:00	7.40
2020-11-11 23:15:16.261228+00:00	7.36
2020-11-12 00:15:48.373825+00:00	7.36
...	...
2020-11-17 05:17:04.412038+00:00	7.37
2020-11-17 06:47:52.593118+00:00	7.34
2020-11-17 14:53:25.015789+00:00	7.36
2020-11-17 18:25:17.425846+00:00	7.36
2020-11-17 22:57:41.959856+00:00	7.34

63 rows × 2 columns

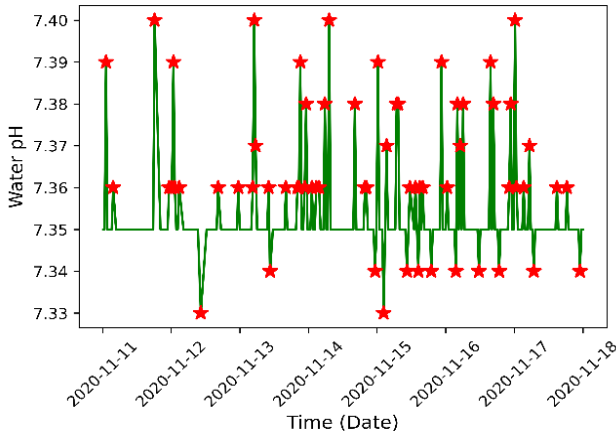


Fig 15: A plot of LOF pH outliers

#### 4.3.2.2 The Robust Random Cut Forest

The pH anomalies detected by the RRCF are shown in Table 3 and were plotted as shown in Figure 16. It was not easy find a feasible threshold to split the outliers and therefore the top 61 records having the highest outlier scores were identified. A significant number of point outliers from the beginning of the subset, that were detected by the LOF algorithm were not detected; such as the pH record 7.39 [2020-11-11 01:11:34.389403+00:00]. Also, this method takes too longer for the detection process (2.51 seconds) compared the LOF algorithm which only took milliseconds.

Table 3: pH outliers as detected by the RRCF algorithm

time	pH
2020-11-11 18:12:35.713854+00:00	7.40
2020-11-11 23:15:16.261228+00:00	7.36
2020-11-12 00:15:48.373825+00:00	7.36
2020-11-12 00:46:04.433087+00:00	7.39
2020-11-12 01:16:20.473824+00:00	7.36
...	...
2020-11-17 05:17:04.412038+00:00	7.37
2020-11-17 06:47:52.593118+00:00	7.34
2020-11-17 14:53:25.015789+00:00	7.36
2020-11-17 18:25:17.425846+00:00	7.36
2020-11-17 22:57:41.959856+00:00	7.34

61 rows × 3 columns

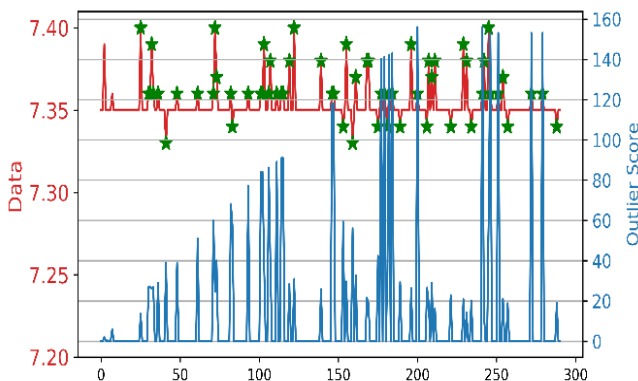


Fig 16: A plot of RRCF pH anomaly scores

### 4.3.3 Turbidity

#### 4.3.3.1 The Local Outlier Factor

For the turbidity outliers, the algorithm detected a total of 75 outliers as highlighted in Table 4. These anomalies are plotted

as shown in Figure 17. The red stars are the 75 instances detected as anomalies in the turbidity data with the value of number of neighbors  $k = 100$ . It took the LOF algorithm 38.9 ms to complete this process. Once again, finding an optimal value of  $k$  was essential for detection performance, just as it was the case for the pH data. There were no notable false alarms as well as undetected outliers.

Table 4: Turbidity outliers as detected by the LOF algorithm

time	turbidity
2020-11-12 21:57:18.752276+00:00	33.856159
2020-11-12 22:27:34.814333+00:00	35.975997
2020-11-12 22:57:50.858115+00:00	39.486692
2020-11-12 23:28:06.931813+00:00	38.856159
2020-11-12 23:58:22.991919+00:00	37.611506
...	...
2020-11-17 21:57:09.848484+00:00	50.094176
2020-11-17 22:27:25.894124+00:00	59.856159
2020-11-17 22:57:41.959856+00:00	77.975997
2020-11-17 23:27:58.035168+00:00	88.856159
2020-11-17 23:58:14.065228+00:00	100.975997

75 rows × 2 columns

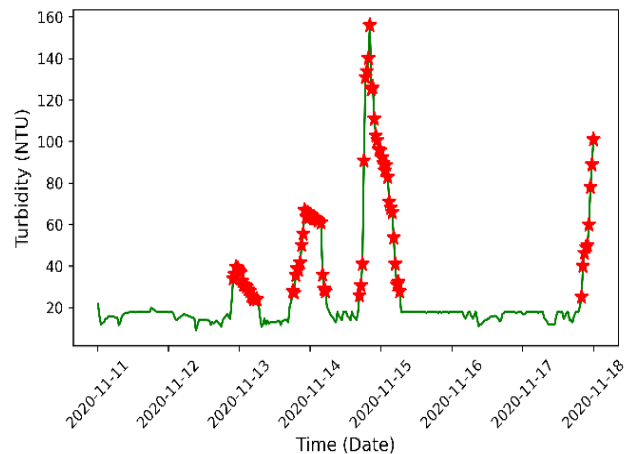


Fig 17: A plot of LOF turbidity outliers

#### 4.3.3.2 The Robust Random Cut Forest

It was not easy to find a feasible threshold to split the outliers and therefore the top 67 outlier records having the highest scores were listed in Table 5, taking 7.1 seconds. These results were plotted as shown in Figure 18. A significant number of point outliers that were detected by the LOF algorithm were not detected. False alarms are also contained in this list: For example, for turbidity record 19.856159 NTU [2020-11-11 18:12:35.713854+00:00], whose value is almost equal to the next value (only 30 minutes apart) is marked as an outlier. Also, this method took longer for the detection process compared to the LOF algorithm which only took a fraction of a second.

Table 5: Turbidity outliers as detected by the RRCF algorithm

time	turbidity
2020-11-11 18:12:35.713854+00:00	19.856159
2020-11-12 09:20:37.353014+00:00	8.856159
2020-11-12 17:55:10.311827+00:00	10.734662
2020-11-12 21:27:02.699108+00:00	22.975997
2020-11-12 21:57:18.752276+00:00	33.856159

...	...	
2020-11-17 19:56:05.608199+00:00	25.094176	
2020-11-17 20:26:21.676559+00:00	39.975997	
2020-11-17 20:56:37.735037+00:00	46.210696	
2020-11-17 22:57:41.959856+00:00	77.975997	
2020-11-17 23:58:14.065228+00:00	100.975997	
67 rows × 3 columns		

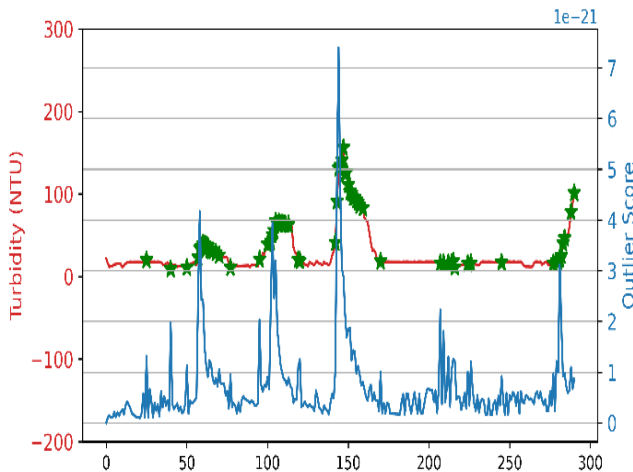


Fig 18: A plot of RRCF turbidity anomaly scores

From table 6 and table 7, the LOF algorithm successfully detects all the 63 anomalies in the time series water pH subset data as well as all the 75 anomalies in the time series turbidity data. The RRCF algorithm suffers from 19 false anomalies as well as missing 27 outliers in the turbidity subset. The case is similar for 2 undetected point anomalies in the pH subset data. Additionally, the LOF algorithm is much faster than the RRCF algorithm on detecting anomalies for both turbidity and pH data. While the LOF algorithm only took milliseconds, the RRCF algorithm consumed a number of seconds, for both the subsets.

Table 6: pH Dataset Algorithms Performance Evaluation

Algorithm	Anomalies	False Anomalies	Undetected Anomalies	Execution Time
LOF	63	0	0	21 ms
RRCF	61	0	2	2.51s

Table 7: Turbidity Dataset Algorithms Performance Evaluation

Algorithm	Anomalies	False Anomalies	Undetected Anomalies	Execution Time
LOF	75	0	0	38.9 ms
RRCF	67	19	27	7.1s

## 5. CONCLUSION

This article presents the development of a low-cost sensor nodes network that can be used to perform automated water quality monitoring on raw water in water treatment plants. First and foremost, there were LoRa technology studies on range of coverage and connectivity using the RSSI parameter of the transceiver signals in the DeKUT. The best RSSI was realized in places near the gateway (100m away), which was a mean strength of -102.7 dBm, while the least was recorded at the furthest point of testing (1 km), whose mean signal strength was -113.7 dBm. The proposed sensor node contains two water quality sensor probes that can be used to monitor water quality. These include the DFRobot gravity Arduino

turbidity sensor and the DFRobot's Gravity Analog pH sensor. The developed system is cheap, power saving, lightweight, and it can comfortably remotely transmit data using LoRa technology. Besides, this paper presents a comprehensive evaluation of two different machine learning anomaly detection algorithms on two parameters from a water sensor node at the NYEWASCO water treatment plant raw water section. The subset data that was used in algorithm evaluation had 291 records for both parameters, extracted from the 2,658 that were collected over the deployment period. The LOF algorithm emerged superior to the RRCF algorithm in contamination event detection and hence a practical water contamination detection algorithm that can trigger alarms to alert the analyzers when contamination is detected. The framework is more suitable to implement in large-scale to collect and analyze raw water quality data in water supply firms, and water authorities.

We recommend that further work can be done on LoRa technology connectivity and range evaluation to come with a wireless propagation model in a rural set up of DeKUT. This can incorporate other factors which include free space attenuation, shadowing, reflection and transmission, diffraction, among others. The developed water quality management system can be installed in multiple locations in water distribution networks to gather water quality data and classify sensor responses in practical deployment. This is because water is a vast network of interconnected bodies such as rivers, lakes, swamps, dams and other sources, and if these linked parts contain different levels of pollution, assessing water quality may be a complicated endeavor. Moreover, more water quality can be incorporated to the developed system and this can make it usable in drinking water quality parameter monitoring. Besides turbidity and water pH, other water quality parameters include: Total Dissolved Solids, Oxygen Reduction Potential, Temperature, Electrical Conductivity, Dissolved Oxygen, Free Residual Chlorine, Nitrates, to mention just but a few. Additionally, further investigation on the productivity of anomaly detection algorithms given several types of contaminants present in water should be done.

## 6. ACKNOWLEDGMENTS

The first author is grateful to the Electrical and Electronic Engineering Department members, and the Center for Data Science and Artificial Intelligence (DSAIL) in the Dedan Kimathi University of Technology (DeKUT) for their mentorship, support and inspiration. The first author acknowledges Dr. Ciira Maina, DeKUT alongside Dr. Henry Kiragu, Multimedia University of Kenya, for their valuable comments and supervision. Also, a special vote of gratitude to the Nyeri Water and Sanitation Company (NYEWASCO) water quality laboratory staff, more so Mr. Mohamed Yunis Ali (Head of the Lab), Mr. Paul Michuki (Head of Operations) and their lab technicians, who assisted and guided in sensor node deployment for data collection. Lastly, we would like to thank the Advanced RISC Machines (ARM) for hardware support.

## 7. REFERENCES

- [1] T. P. Lambrou and C. G. Panayiotou, "Collaborative Area Monitoring Using Wireless Sensor Networks with Stationary and Mobile Nodes," EURASIP Journal on Advances in Signal Processing, vol. 2009, no. 1, 2009.
- [2] K. K. Khedo, R. Perseedoss and A. Mungur, "A Wireless Sensor Network Air Pollution Monitoring System," International Journal of Wireless & Mobile Networks,

- vol. 2, no. 2, pp. 31- 45, 2020.
- [3] "Water: A Shared Responsibility – The United Nations World Water Development Report 2," *Development in Practice*, vol. 17, no. 2, pp. 309-311, 2007.
- [4] "Millennium development goals: time to reassess strategies," *BMJ*, vol. 331, no. 7525, 2005.
- [5] J. Bhardwaj, K. K. Gupta and R. Gupta, "Towards a cyber-physical era: soft computing framework based multi-sensor array for water quality monitoring," *Drinking Water Engineering and Science*, vol. 11, no. 1, pp. 9-17, 2018.
- [6] K. E. Sawaya, L. G. Olmanson, N. J. Heinert, P. L. Brezonik and M. E. Bauer, "Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery," *Remote Sensing of Environment*, vol. 88, no. 1-2, pp. 144-156, 2003.
- [7] K. S. Adu-Manu, C. Tapparelo, W. Heinzelman, F. A. Katsriku and J.-D. Abdulai, "Water Quality Monitoring Using Wireless Sensor Networks," *ACM Transactions on Sensor Networks*, vol. 13, no. 1, pp. 1-41, 2017.
- [8] G. Tuna, B. Nefzi, O. Arkoc and S. M. Potirakis, "Wireless Sensor Network-Based Water Quality Monitoring System," *Key Engineering Materials*, vol. 605, pp. 47-50, 2014.
- [9] J. Liu, X. Wei, S. Bai, X. Bai and X. Wang, "Autonomous underwater vehicles localisation in mobile underwater networks," *International Journal of Sensor Networks*, vol. 23, no. 1, p. 61, 2017.
- [10] F. Ge and Y. Wang, "Energy Efficient Networks for Monitoring Water Quality in Subterranean Rivers," *Sustainability*, vol. 8, no. 6, p. 526, 2016.
- [11] A. Alkandari, M. alnasheet, Y. Alabduljader and S. M. Moein, "Water monitoring system using Wireless Sensor Network (WSN): Case study of Kuwait beaches," 2012 Second International Conference on Digital Information Processing and Communications (ICDIPC), 2012.
- [12] J. Hall, A. D. Zaffiro, R. B. Marx, P. C. Kefauver, E. R. Krishnan, R. C. Haught and J. G. Herrmann, "On-Line water quality parameters as indicators of distribution system contamination," *Journal - American Water Works Association*, vol. 99, no. 1, pp. 66-77, 2007.
- [13] M. Allegretti, "Concept for Floating and Submersible Wireless Sensor Network for Water Basin Monitoring," *Wireless Sensor Network*, vol. 06, no. 06, pp. 104-108, 2014.
- [14] F. Gui and X. Q. Liu, "Design for Multi-Parameter Wireless Sensor Network Monitoring System Based on Zigbee," *Key Engineering Materials*, vol. 464, pp. 90-94, 2011.
- [15] Z. Rasin and M. R. Abdullah, "Water Quality Monitoring System Using Zigbee Based Wireless Sensor Network," *Int. J. Eng. Technol.*, vol. 9, no. 10, pp. 14-18, 2009.
- [16] M. Jared, K. Ngetich and M. Ciira, "Long Range Low Power Sensor Networks for Agricultural Monitoring - A Case Study in Kenya," in *2019 IST-Africa Week Conference (IST-Africa)*, Nairobi, 2019.
- [17] M. M. Breunig, H.-P. Kriegel, R. T. Ng and J. Sander, "LOF: Identifying Density-Based Local Outliers," *Association for Computing Machinery*, p. 93–104, 2000.
- [18] S. Guha, N. Mishra, G. Roy and O. Schrijvers, "Robust random cut forest based anomaly detection on streams," *International conference on machine learning*, pp. 2712-2721, 2011.