ASHESI

ASHESI UNIVERSITY

SMART PORTABLE LOW-COST WATER QUALITY CHECKER

THESIS

B.Sc. Electrical/Electronic Engineering

Nana Yaa Pokuah-Debrah

2019/2020

ASHESI UNIVERSITY

Smart Portable Low-Cost Water Quality Multi-Sensor

CAPSTONE PROJECT

Capstone Project submitted to the Department of Engineering, Ashesi University in partial fulfilment of the requirements for the award of Bachelor of Science degree in Electrical/Electronic Engineering.

Nana Yaa Pokuah-Debrah

Supervisor: Dr. Elena Rosca

2019/2020

DECLARATION

I hereby declare that this capstone is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:

……………………………………………………………………………………………

Candidate's Name:

Nana Yaa Pokuah-Debrah

Date:

29th May 2020

I hereby declare that preparation and presentation of this capstone were supervised in accordance with the guidelines on supervision of capstone laid down by Ashesi University College.

Supervisor's Signature:

……………………………………………………………………………………………

Supervisor's Name:

……………………………………………………………………………………………

Date:

……………………………………………………………………………………………

Acknowledgements

I would like to say a big thank you to my supervisor, Dr. Elena Rosca for much appreciated support and encouragement during this project. Also, I would like to extend my sincere gratitude to Nicholas Tali for the technical support given to make this project a success.

Abstract

Water quality checking is an essentiality in the manufacture of drinking water which is required and done in all water manufacturing and distribution companies. However, it is required that the quality of water be checked at the distribution end as well, as the quality is reduced due to plumbing contamination, leakages, and prolonged storage to mention a few. The checking of water quality in the urban home is not common in Ghana and those who check through water quality specialists get the results after a number of days during which the sample could be exposed to more contamination. The work done in this project employs different technologies in building a low-cost sensor device which is easy to use by the lay man through a web application.

Table of Contents

Chapter 1

1.1 Problem Definition

Water tanks are a vital resource in many Ghanaian urban households and local businesses which have many uses in the fulfilling of daily activities. Though the quality of water is measured at water treatment plants, the same cannot be said for the consumer end—urban home tanks— after it has been distributed. Due to cross-contamination in the distribution system, unsafe storage tanks, and the periodic length of storage, the quality of water in household tanks has been depreciated. With some factors such as plumbing contamination due to long underground transportation medium, leakages, microbial infections from insects just to name a few being the root cause of reducing the quality of water used in households, it is essential that urban household dwellers routinely check the quality of water in their tanks and take the needed actions to prevent consuming unsafe water. As there has been a shortage of water supply in the country especially, in the capital due to an increase in population, many Ghanaian urban households rely on commercial water supply [1]. The intermittent supply of water is not only limited to Ghana but to other West African Countries such as Nigeria whose population is at least twice that of Ghana [2]. The commercial use for water as has become popular is not reliable and may contain bacteria harmful to the health. In general, water quality tests in home tanks should be performed on a timely basis due to leaching of plumbing material into the water, defecation by birds, and leakages [3].

1.2 Motivation for Project Topic

The main motive for this project is to provide a water quality sensor that is affordable as well as portable to provide an opportunity for consumers to undertake their own water quality risk assessments in the distribution end of water supply systems as contamination in the distribution system, leakages and prolonged storage tends to make the water unsafe for consumption.

1.3 What Has Been Done?

Many systems have been built to tackle this issue, many of which have been built outside of Ghana, specifically the United States. Commercially available sensors out there related to this work include the Compartment Bag Test (CBT) by AguaGenX, LLC which measures the presence of *E. coli* in water, Professional Plus Multiparameter Water Quality Meter a Xylem Brand by Global Water and the Hone Forest quality Tester by Hone Forest. Although CBT is known for requiring no electricity and being able to work at different temperatures, it is limited to just one parameter and quite bulky for the user as it contains ten compartment bags. The quality tester by Hone Forest can measure parameters for any type of water and is prolific for its accuracy and reliability, however, it lacks the automatic analysis of the measured parameters, and thus the user would have to manually use charts to explain the readings. Sensorex is also known for producing smart sensor devices with digital communication nevertheless, just as the CBT each of these devices measure just a parameter. The commercial system which is closely related to this project is the Xylem brand quality meter which measures multi-water quality parameters, a total of 13 parameters, very reliable as it is accurate, portable, and has an in-built analysis tool. It is however sophisticated and expensive.

Also, academic research and prototyping have been developed in this area many targeted at distribution mains and the transportation network. The design of smart water quality monitoring systems using IoT has been developed by many academic researchers which provide the measurement of data as well as analysis, however, these are bulky and designed for industrial use.

1.4 Proposed Solution

A proposed solution to this problem is to design and build a low-cost portable multi-sensor system that would be able to measure basic as well as derived water quality parameters such as pH, conductivity, turbidity, salinity, and microbiological presence. These parameters would be measured upon request and transmitted via Wi-Fi module ESP-01 to a database and web-based application software where the analysis would be carried and the resulting output—percentage quality of water—displayed to the user.

1.4 Expected Outcomes of the Project Work

The aim of this project is to be able to assess the quality of household water stored in tanks, verify that the quality of water is reduced at the distribution end of the supply system due to contamination and prolonged storage, and to give the opportunity to contribute to the risk assessments of water at the distribution end by consumers and building a multi-sensor contrary to conventional water quality sensors, that will be able to measure different water quality parameters with a single system that requires little to no technical skills. This would be done by a system that will measure three basic water quality parameters from which the water quality index of a water sample would be generated. The water quality index generated will then inform the user of the status of the water sample.

1.4 Objectives of the Project Work

- This project is aimed at being able to efficiently and accurately measure at least four of the water quality parameters such as pH, temperature, turbidity, total dissolved salts (TDS), dissolved oxygen concentration, salinity, and the presence of bacteria.
- Develop a machine-learning algorithm to predict the quality of water.
- To develop a user interface for the user through an application software to routinely check the quality of water in their home tanks.

Chapter 2: Literature Review

2.1 Background

In designing a water quality multi-sensor to curb the problem of reduced water quality in-home tanks, there is the need to assess and get acquainted with the sensors that have already been designed and built to gain more insight into the project topic and how to design a better system. Water quality multi-sensors, both those of commercial use and those of research and development, have contributed immensely to measuring and attaining the quality of different sources of water. As this project mainly entails the measuring of the parameters used to determine water quality and analyzing these results, the scope of this chapter would involve reviewing some of the prominent work closely related to this project topic with a focus primarily on the methodology of measuring the parameters and the analysis used in the determining the water quality.

This project aims to design a water quality multi-sensor for urban home users to allow them to perform their own water quality risk assessment as well as to establish a record of water quality in their households from year to year. Thus, the target stakeholders of this project are domestic, which hence, helps establish the criteria in choosing the literary works which would help in gaining insight into this project topic. Though many works have been done in measuring water quality, the criteria used in selecting the literature sources are the already existing commercial sensors and research development sensors used mainly in a domestic setting rather than an industrial setting and that which is less bulky.

2.2 Commercial Multi-Sensors

As mentioned in chapter 1, there are many commercial sensors in the market that have been developed to check the quality of water. Some of these can work in a wide range of water networks from, home tanks, to small water bodies, such as lakes, wells, rivers, and so on. The compartment bag test (CBT) and hone forest water quality tester although are structurally relative to this work as being portable, their shortcomings are adjudicated for by its single parameter measurement and manual analysis by the user. Contrary to CBT and hone forest, Xylem brand water quality meter which is closely related to this work is known for its accuracy and multi-parameter measurement capabilities. It is however bulky and very costly thus relatively expensive for the average Ghanaian. Aside from the commercial meters in the market, there have been research developed sensors that have as just potential as the ones found in the market.

2.3 Academic Research Developed Sensors

Aside from the commercial meters in the market, there have been research developed sensors that have as just potential as the ones found in the market. Chalchisa et al. [4] analyzes water samples collected in a small community in Ethiopia and determines the water quality through statistical means and correlations. The data analysis was by the use of tables and statistical methods to establish a correlation between the measured parameters of which a relationship was found between pH and bacterial counts [4]. The study determined that the water quality in storage tanks after distribution is truly reduced due to leakages. Similar to this work, one of the focus of this project is to determine the extent of the reduction of water quality in-home tanks, however, creating an opportunity for urban home dwellers to perform their own water quality risk assessment without having any expertise. Another aim is to eliminate the time wasted in transporting samples to well-equipped laboratories, of which the samples could be further contaminated, as well as the analysis being prone to human error.

The usual found in any water distribution company, employ the use of technologies in monitoring the quality of water in real-time during the distribution process. Contrary to Wong [1], Revathi [5] develops a real-time water quality monitoring system in the distribution network using embedded systems and IoT to monitor the quality of water. In his system, he makes use of fuzzy logic to determine and predict the extent of contamination while also deploying wireless communication networks, specifically, ZigBee communication module. Similarly, Srivastavae et al. [6] also develops a system using embedded systems and IoT to monitor the quality of water in distribution systems [6]. The system unlike other water monitoring systems provides real-time monitoring of the water quality in distribution networks and mains [5][6]. It thus eliminates the need to take regular samples from water tanks for evaluation and eliminates the time elapsed between when the sample is taken and when it is analyzed for the prediction of the water quality. As related to this project, the design aim is to design a multi-sensor for measuring several parameters and analyzing the data measured, which eliminates the use of charts for the manual determination of water quality which may be more subjected to inaccuracy due to human error. Considering this system is real-time based, one downside may be the communication module used as Zigbee is a low data rate communication module. The type of communication module should be critically examined to pick the right choice. [7] explores communication protocols suitable for measuring water quality using IOT technology which is a fast and adopting technique in the monitoring of water quality.

2.3 Standard Parameters to Measure

Since the water in home tanks are sometimes consumed—used for drinking and cooking—the parameters used in this project are selected according to the World Health Organization's (WHO) guidelines according to the World Health Organization's (WHO) guidelines for drinking water and that of Environmental Protection Agency's(EPA) interpretation and standard parameters of quality water. The principal physiochemical parameters which should be measured according to WHO, includes pH, turbidity, chlorine residual, and microbiological presence. Also, conductivity and salinity according to EPA and WHO, may be measured to determine the water quality. The recommended or mandatory limit values suite as drinking water for each parameter is listed in table 2.1.

Parameter	**WHO Limits**
pH	6.5 - 8.5
Turbidity	<5 FTU/JTU or <0.5 NTU for drinking water
Chlorine Residual	-
Total Dissolved Solids	1000 mg/L
Conductivity	2500 µS/cm
Salinity	-

Table 2.1: Standard permissible limits for drinking water parameters according to WHO&EPA

<u>Equations of Analysis:</u>

1. **pH:** The analysis conversion of electrical signals measured by the pH sensor to pH reading is given by the Nernst equation as shown below.

   $E = Eo-(k \times pH)$

   $E = Eo - [2.3 \left(\frac{RT}{nF}\right) \times pH]$; where $k = 2.3 \left(\frac{RT}{nF}\right)$

   This is an equation of a straight line of E against pH with a gradient (Nernst factor) of -k and E-intercept Eo. Where,

   E is the output voltage,
   Eo is a constant,

R is the gas constant,

F is the Faraday constant,

T is the temperature in Kelvin = 25°C   and

n is the ionic charge.

2. **Turbidity:** [11] discusses the relationship between voltage and turbidity reading in NTU shown accordingly in figure 2.1. Turbidity is measured in Nephelometric turbidity units which measures the absorbance, of light passing through a sample as expressed below.

   Total Dissolved Solid = 3.4216 x Turbidity Value in NTU

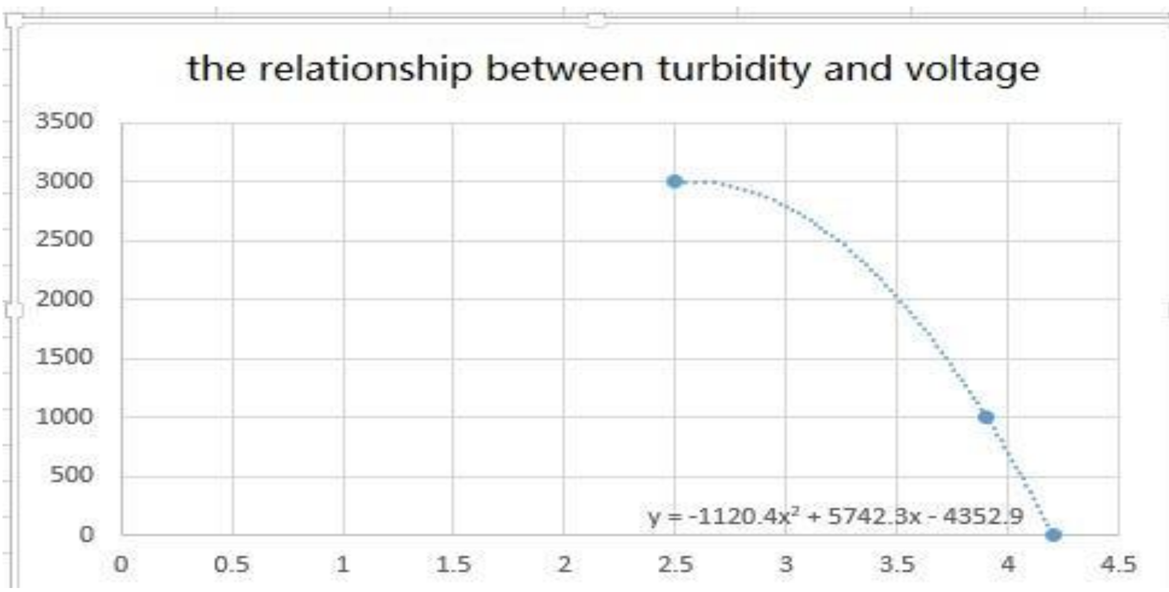


Figure 2.1: Relationship between voltage and turbidity   Source: Adapted from Turbidity_sensor_SKU__SEN0189-DFRobot." https://wiki.dfrobot.com/Turbidity_sensor_SKU__SEN0189

3. **Water Quality Index:** The Water Quality index is a weighted average sum of all the measured water quality parameters. It is calculated by multiplying the q-value by the weight factor of each parameter and summing them. The q-value is calculated by finding the percentage change between the observed and standard values. The weight factor for

each parameter to be used in this project is summarized in table 2.2 WQI is generally calculated using the formula below:

$$\text{WQI} = \frac{\sum q_{value} \times w\_factor}{\sum w\_factor}$$

$$q_{\text{value}} = \frac{100(Vo - Vid)}{Sv - Vid}$$

where, w_factor = weight factor

V_o = Observed/Measured Value

V_{id} = ideal value

S_v = Standard Value

Parameter	Weight
pH	0.11
Turbidity	0.08
Conductivity	1
Total Solids	0.07

Table 2.2: Weight factor for each parameter

Chapter 3: Design

3.1 Introduction

The purpose of this project is to design a Multi-sensor system for the measurement of various water quality parameters including pH, turbidity, and conductivity as well as analyzing the data collected from the sensors to determine the water quality of the test sample. For the design of this system, some factors must be analyzed before its implementation. Design decisions that have been developed were thoroughly considered based on each phase of the project.

3.2 Product Description

The end-product is targeted at having a probe that can measure multi-parameters, send them via some communication module to a controller which is connected to a database cloud, and display the result—percentage/level of water quality—through a web application. The overall design of the system is shown below in figure 3.1. From the system design, the project can be categorized into three phases as described below.

- Phase I, Multi-Probe Sensor: In this phase, circuitry design would be developed for each parameter measured and the designs for each would be integrated into one design with a single power source. Firstly, the design for each would be developed built and tested before the integration.
- Phase II, Algorithm Development: Here, the measured data would be collected and sent to the controller, where an algorithm would be developed to analyze the data for output results. The microcontroller after all its task is also connected to a server to save past data for future predictions.

- <u>Phase III, HMI development</u>: In this phase, a web-based application is developed to create an interface between the user and the system.

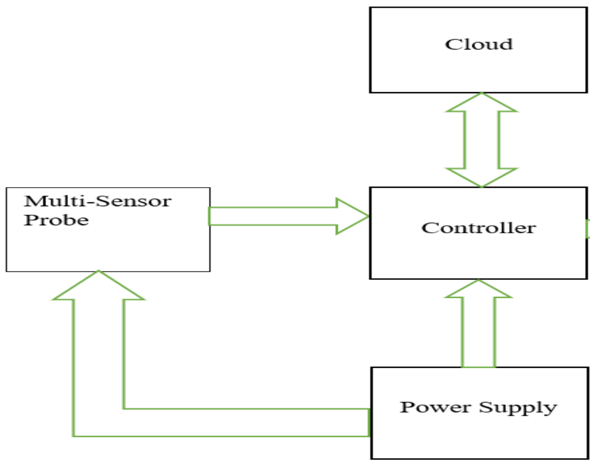


Figure 3.1: Block diagram of system Design

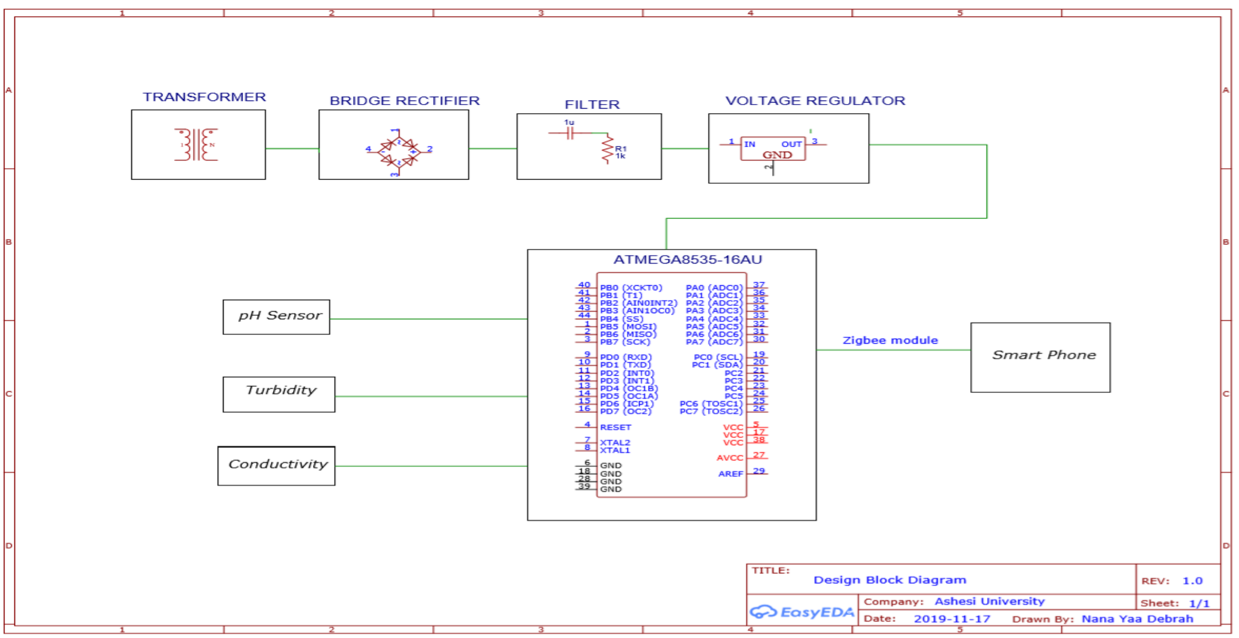


Figure 3.2: Detailed block diagram of design showing power supply.

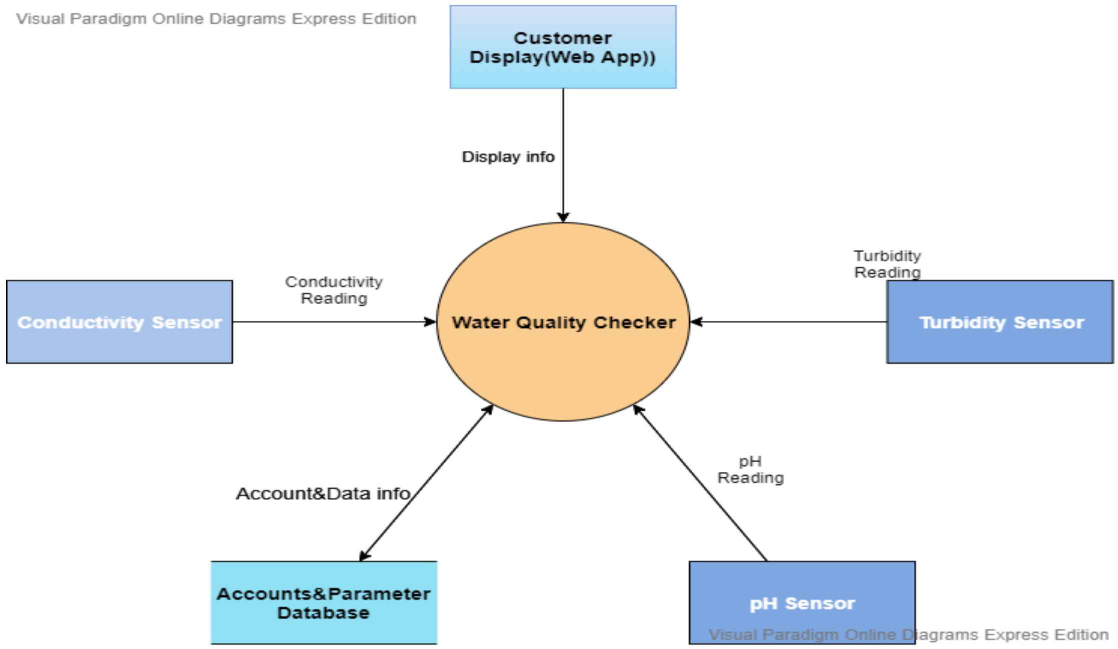


Figure 3.3: Context Diagram of System

The system consists of three main components, the sensoring device(input), the controller, and a web application. The sensoring device contains various circuitry for measuring the three basic parameters namely, pH, turbidity, and conductivity, from which the parameters, salinity, and microbiological presence would be derived. The outputs from these circuits serve as input to the microcontroller where the analysis of the parameters would be made using various statistical and mathematical tools to predict the water quality. The resulting output from the controller is sent to a web-application software for the perusal of the user. The output data is also stored on a cloud/server, to serve as a reference for future predictions.

3.3 Design Requirement

Considering the scope of this project and the target stakeholders—dwellers of urban households—design requirements have been developed to give insight and assist in the development of the various circuitry and design in each phase. The requirements for this system are grouped into user, functional, and system requirements as described below.

User/Usability Requirement:

For user requirements, the user should be able to:

- The user should be able to reset after reading and analyses of water quality.

Usability requirements here refers to how re-usable can the system be after some time of use and how long it can last from its aesthetic structure. Thus, the system should be able to

- Recalibrate after 2-3 months of use to maintain accuracy and coherence in parameter measurements.
- The probe should be able to last for at least two years.

Functional Requirements

- The device must be able to work off-grid for some time (At least an hour).
- Should be able to reset after 5 minutes of stable reading.
- Should be able to remind the user to check the quality of water periodically.

System (Technical) Requirements

The system requirements are developed to provide convenience and ease of use with little to know, technical knowledge for its operation. Considering the fact that the system is associated with health risk—likelihood of gastro-intestinal diseases and skin diseases—the parameter measurements must be as accurate as possible as well as the algorithm for the analysis, having an error of at least 0.99. The table below describes the functional/technical requirements of the system and its justification.

System Requirements	Justification
1. Responsiveness	Should be able to give an output within the shortest possible time. Should also be responsive to user input. Example, if a reset button is pressed.
2. Sensitivity	Should be able to pick up signals of elements/particulates which may be present.
3. Selectivity	Should be able to acquire elements of interests among several.
4. Accuracy	The accuracy of the system should be about 0.99.
5. Power Consumption	The system should be able to work with less power.

Table 3.1: Table of the technical requirements of the system

3.4 Design Decisions

In relation to the requirements and block diagram of the system, certain factors are considered to help make some decisions. The decisions considered are solely considered due to the availability of several options in completing each phase. The design decision for each factor as shown below is made as a result of evaluation using Pugh Chart tables. The various design decisions to be made for each phase are as follows.

3.4.1 PHASE I

1. <u>Parameters to measure</u>: There are many parameters to measure in order to determine the water quality such as pH, conductivity, turbidity, etc. The parameters essential for determining the water quality may vary depending on the water use—drinking, industrial use, aquatic life. However, considering this project is focused mainly on safeguarding the public health of urban dwellers, the parameters chosen are based on EPA's Parameters of

Water Quality; Interpretation and Standards published in 2001. Thus, the parameters most essential to this project whose sensor is manufacturable within the scope of this project is that of turbidity, pH, conductivity of which other parameters such as salinity and microbial presence can be derived.

2. Methods of Measuring Parameters: There exist several methods for measuring each parameter, designated to contribute the extent of the water quality. One major aim is to cut down on cost, while simultaneously optimizing system performance. A Pugh chart is created for each parameter, to select the best methods suitable for the scope of the project. The criteria chosen with which the different methods would be compared against a standard are the technical requirements, responsiveness, accuracy, sensitivity, material availability and manufacturability. The tables below represent the selection method for each parameter to be measured.

A. Turbidity

	Baseline	Weight	A	B
Criteria	Turbidity Meter		Turbidity Tube	Nephelometric Sensors
Responsiveness	0	1	-1	0
Accuracy	0	5	-1	0
Sensitivity	0	4	-1	0
Material Availability	0	2	+1	+1
Manufacturability	0	3	+1	-1
		Total	-5	-1

Table 3.2: Pugh chart for the different methods of measuring turbidity.

B. pH

	Baseline	Weight	A	B	C	D
Criteria	Hydrogen Electrode		Quinhydrone-Electrode Method	Antimony-Electrode Method	Semiconductor sensor methods	Glass-Electrode Method
1Response time	0	1	-1	0	+1	+1
2Accuracy	0	5	0	-1	+1	+1
3Selectivity	0	4	+1	0	+1	0
4Material Availability	0	2	-1	-1	0	+1
5Manufacturability	0	3	-1	-1	0	+1
		Total	-2	-10	10	11

Table 3.3: Pugh Chart for the different methods of measuring pH

From the table, the glass electrode method emerges as the best suited for this project, having a total score of 11.

C. Conductivity

	Baseline	Weight	A
Criteria	2-AC bipolar Method		Electronic Induction Method
1Response Time	0	1	-1
2Accuracy	0	5	0
3 Selectivity	0	4	0
4 Material Availability	0	2	+1
5 Manufacturability	0	3	+1
		Total	4

Table 3.4: Pugh Chart for various methods of measuring conductivity

D. Microbiological Presence: Measuring turbidity can give a measure of the extent of microbial particles present in the water sample.

E. Salinity: Determination of the salinity of the water, can be derived from measuring the electrical conductivity of the water**.**

3.4.2 PHASE II

For phase II, the design decision to make is, firstly, which communication module to use in transferring the measured data to the controller and also, which controller chip is most suitable. The criteria, used in selecting the most suited module include low power consumption, frequency, data rate, and availability. For the communication modules, different types such as ZigBee, Bluetooth, and Lora WAN are compared against Wi-Fi technology.

	Baseline	Weight	A	B	C
Criteria	Wi-Fi		ZigBee	Lora WAN	Bluetooth
Low Power Consumption	0	4	+1	+1	+1
Frequency	0	1	0	-1	-1
Data rate	0	2	-1	-1	-1
Availability	0	3	0	0	0
Communication Range	0	5	+1	+1	+1
		Total	7	6	6

Table 3.5: Pugh chart for selecting communication protocol.

Zigbee communication with a total score of 7, is best suited for this project having a good communication range and lower power consumption than the Wi-Fi technology. However, Wi-Fi technology, specifically esp-01 was used in this project due to its availability and the ability to implement a full TCP/IP protocol stack, unlike LoRaWan and Zigbee which require a gateway for connecting to the internet.

3.5 Schematic Diagram for Sensoring Parameters

<u>pH Schematic</u>

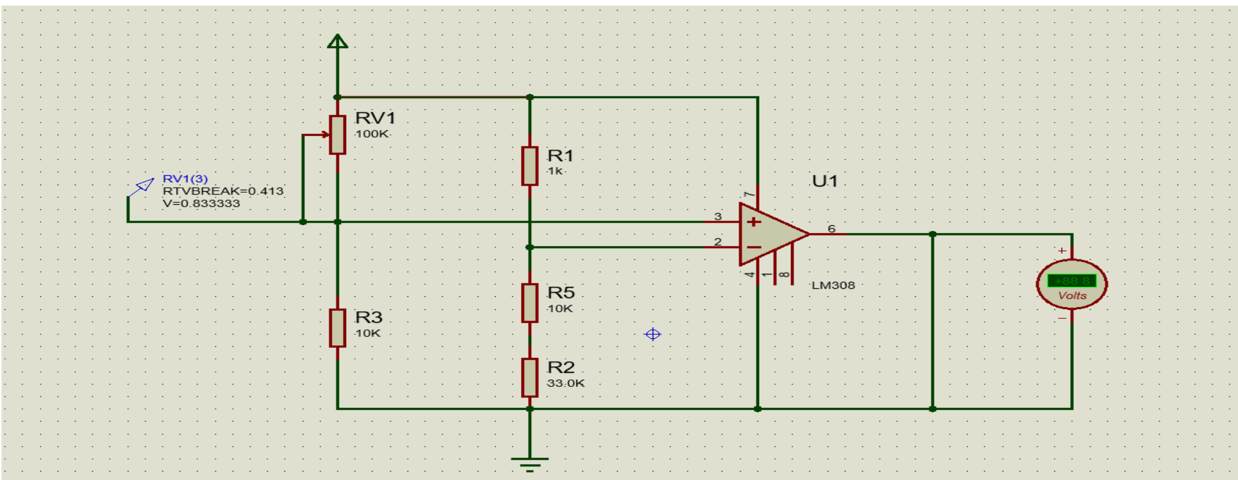


Figure 3.3: Circuit Diagram for pH measurement

The circuit consists of a voltage probe RV (13), an operational amplifier, U1 of LM308, and voltmeter to read the output. A voltage divider is built at the inverting input (-ve), to serve as a reference voltage for the system. Hydrogen ions generate a minimum of 403mV when in water, thus the voltage from the probe in solution also creates a voltage divider at the non-inverting input(+ve). If the voltage at the non-inverting input is greater than the voltage at the inverting input, the output is high otherwise it is low. The output voltage serves as an input to the controller unit where it is converted to a pH reading.

Turbidity Schematic

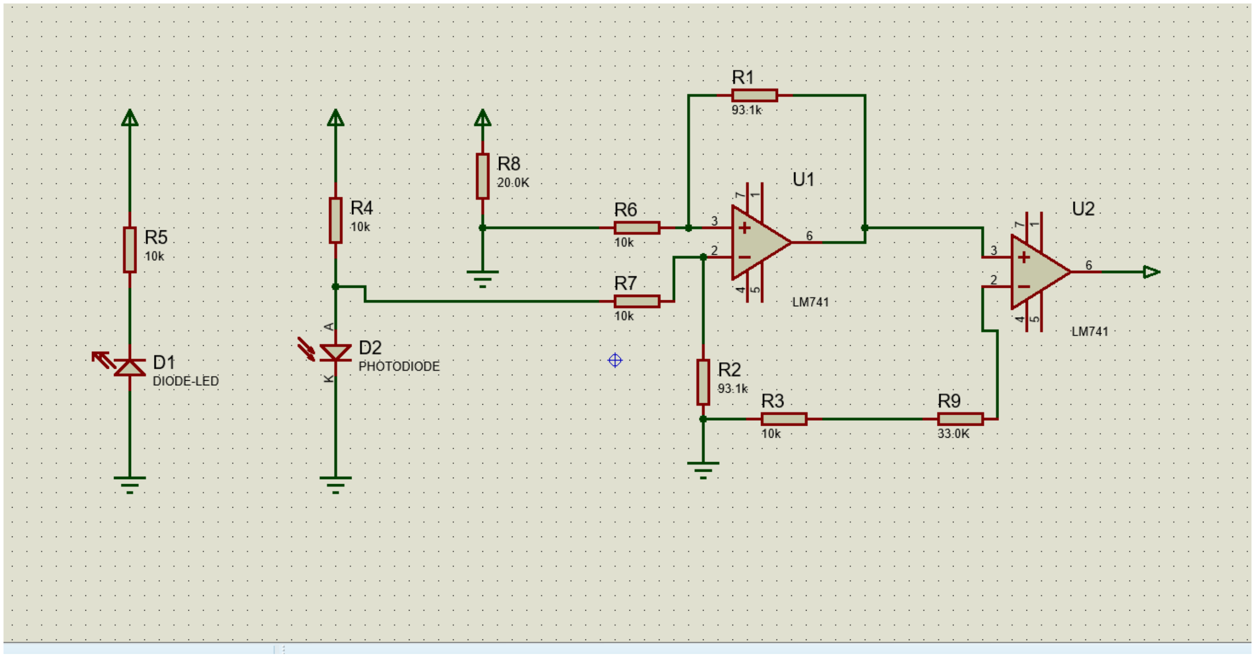


Figure 3.4: Circuit Diagram for Turbidity measurement

The circuit consists of an Led which can be replaced by a tungsten lamp, photodiode, two operational amplifiers U1 and U2 of LM741. The sample is placed between D1 and D2 where the diffraction of light through the sample is measure by the photodiode by converting the light into current. The signal from the photodiode is compared to that of the voltage at the noninverting input(+ve). The output from U1 is further amplified by U2. The output from U2 also serves as an input to the controller.

Conductivity Schematic

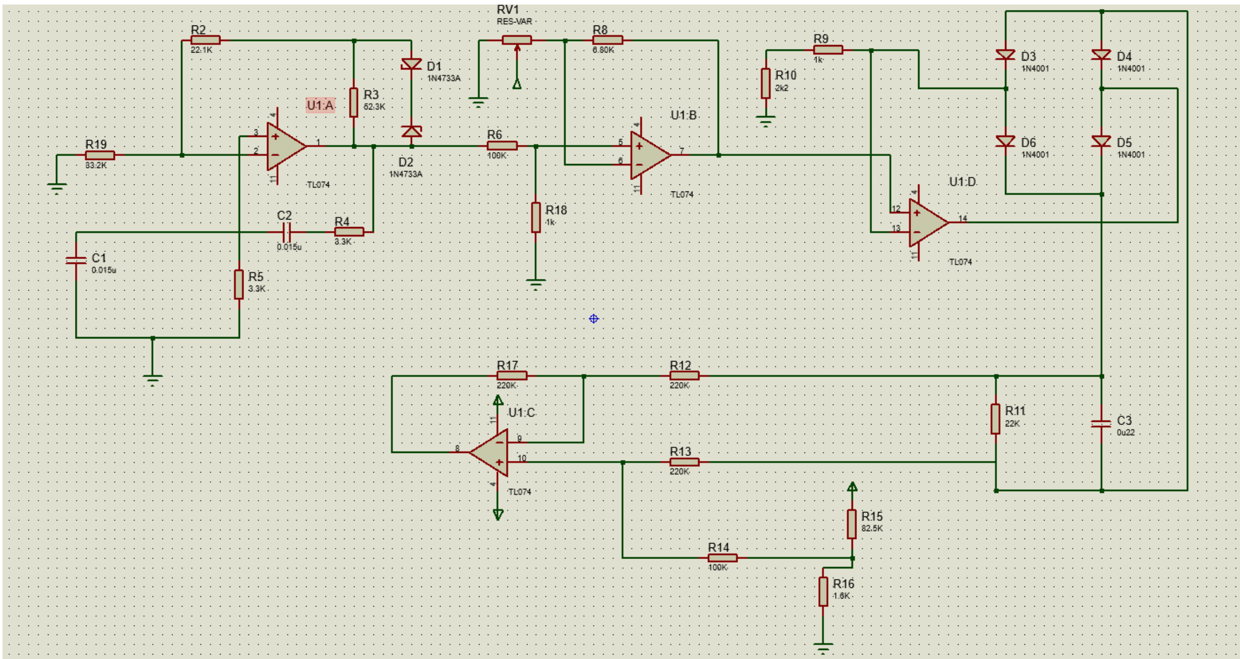


Figure 3.5: Circuit Diagram for Conductivity measurement

The schematic above is borrowed from a research on the implementation of a hydroponic system [8]. It consists of a potentiometer at RV1 where the probe would be connected.

Integrated Circuits

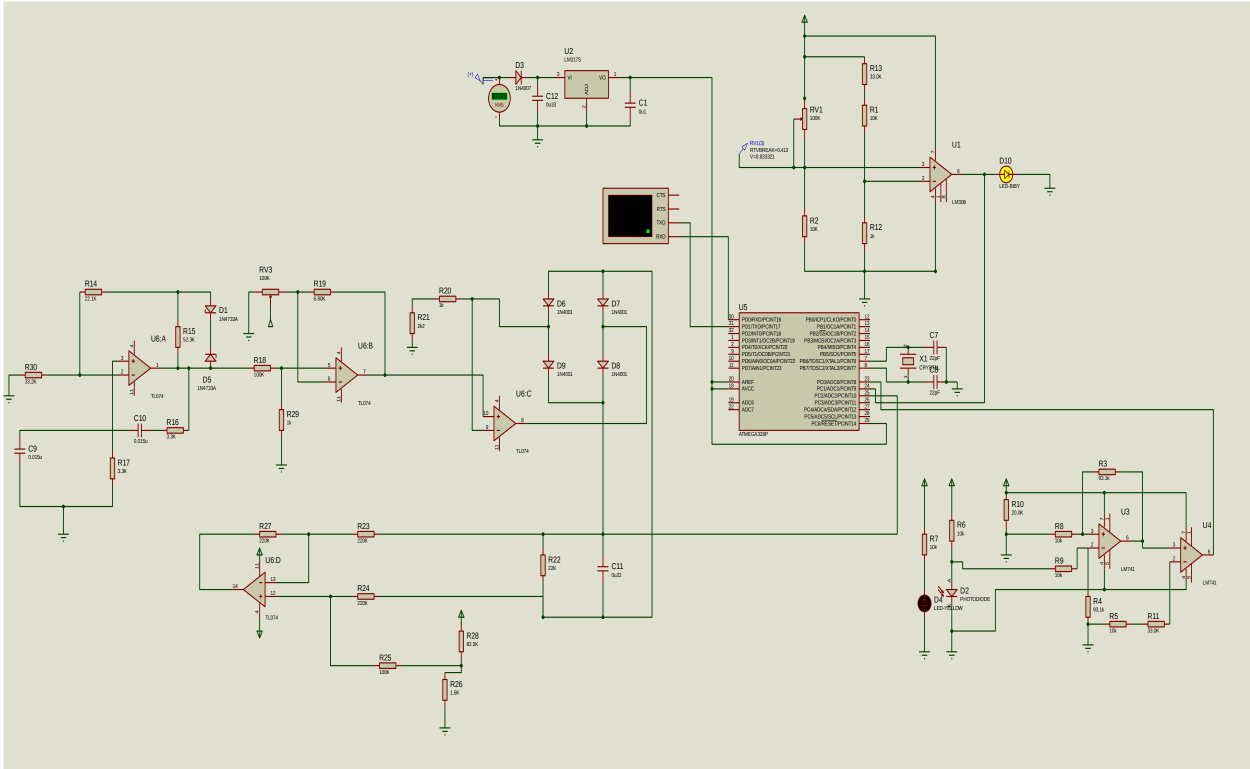


Figure 3.6: Integrated circuits of sensors with Atmega328P.

In figure 3.6, the circuits for turbidity sensor, conductivity, and pH sensor has been connected to the Atmega328P as analog inputs. A power circuit consisting of a voltage regulator—LM317s—capacitors, a diode, and battery terminals supply a voltage of 5V to the Atmega chip.

Proteus Design

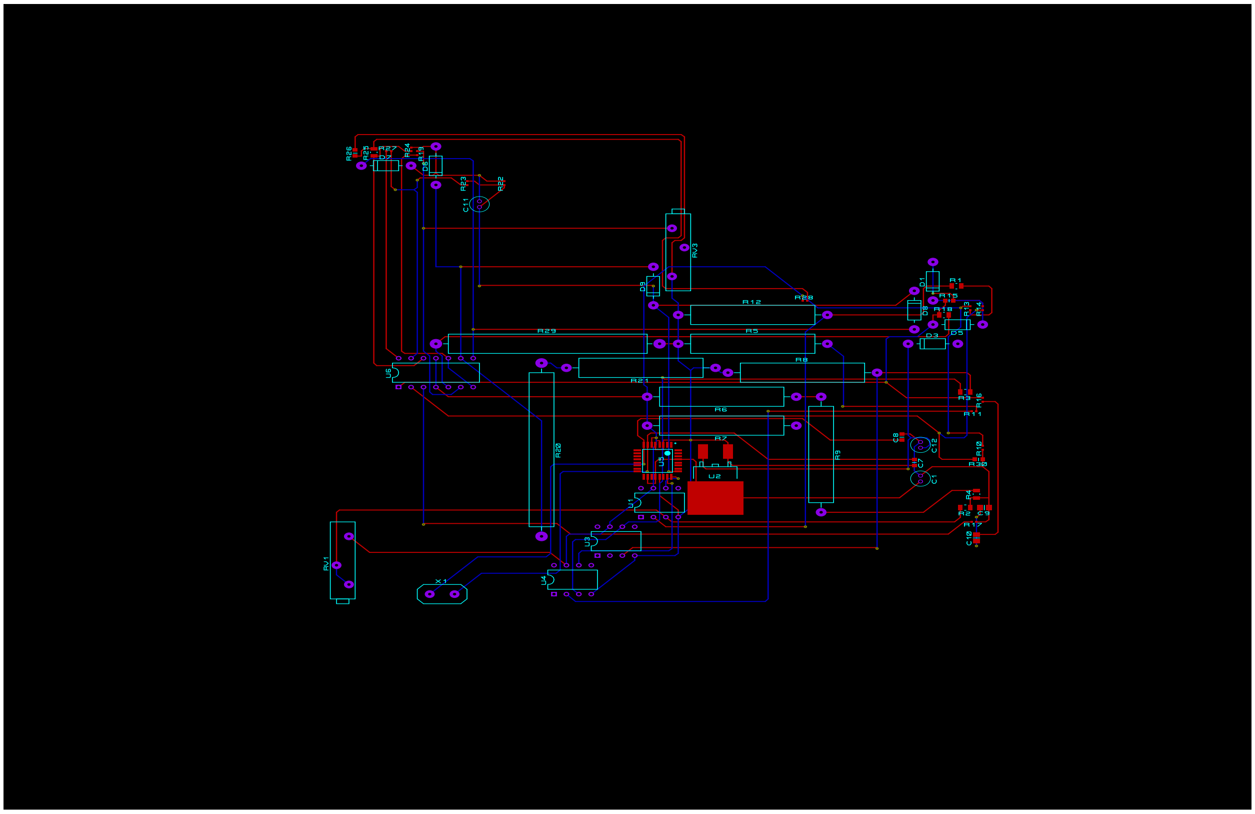


Figure 3.7: Proteus design of the integrated circuit.

The integrated sensor circuits are simulated and the printed circuit design generated as shown in figure 3.7

Chapter 4: Design Implementation/Methodology

The first phase is first implemented by the building and development of the sensor circuits and the collection of data to the microcontroller. Due to the absence of some critical components needed in the building of the conductivity and pH circuit, the chapter will detail the implementation of the system with only the turbidity sensor since the components required for that was readily available. The first phase of the project is implemented in Proteus simulation software where the code for measuring the parameter, turbidity, is uploaded onto the atmega328P in Proteus to simulate it as shown in figure 4.1 below.

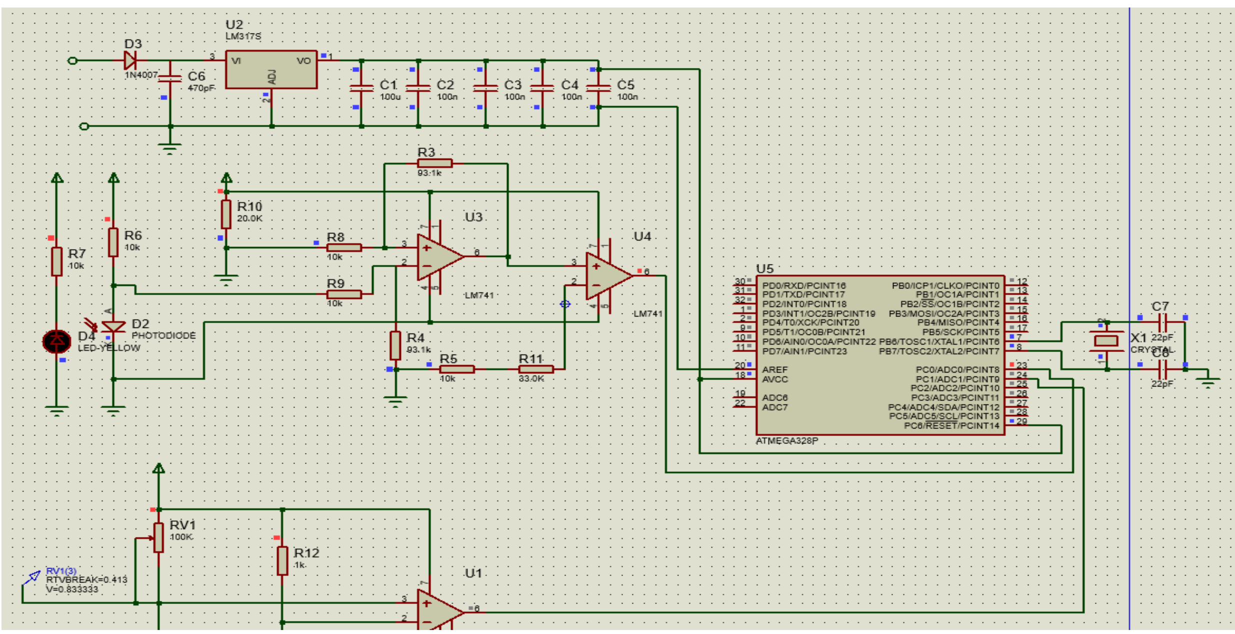


Figure 4.1: Simulated proteus design showing active low and high connections.

4.1 Building of Sensor Circuits

The turbidity sensor is built as shown in its schematic in figure 4.3. The output of the sensor is connected to the first analog pin of the ESP8266 and the output displayed on the serial monitor as shown in the figures below. The data is then preprocessed from the code uploaded to the ESP where the turbidity reading and the water quality index is predicted based on the water quality metric below in table 4.1. The measured data is sent to a web app(see appendix B) and simultaneously to a database on the local server. The nature of the database and API interface is discussed further in section 4.5.

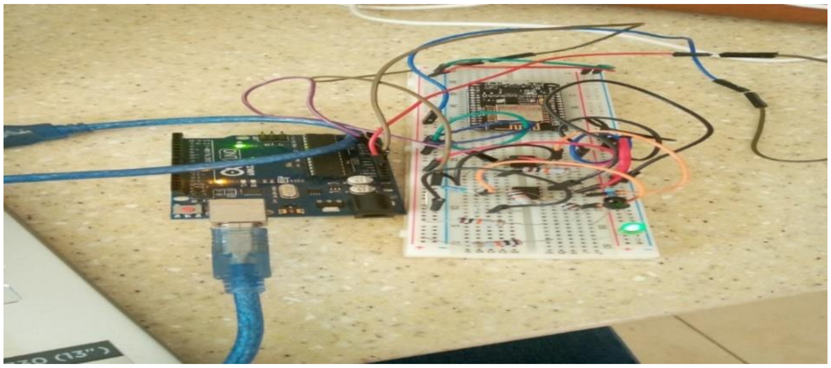


Figure 4.2: Implemented circuit of turbidity sensor without test sample.

WQI Range	Status of Water
0-25	Very Bad
25-50	Bad
50 - 70	Good
70 - 90	Very Good
90 - 100	Excellent
Total	**100**

Table 4.1: Water Quality Status based on the WQI ranges

In order to make future predictions, the data set—turbidity and WQI readings among other relevant information—are imported into **jupyter notebook**, an open-source application where different machine learning algorithms are employed in training and testing the data set acquired from measuring the turbidity of different test samples.

4.2 Setup of System of Sensor for Testing

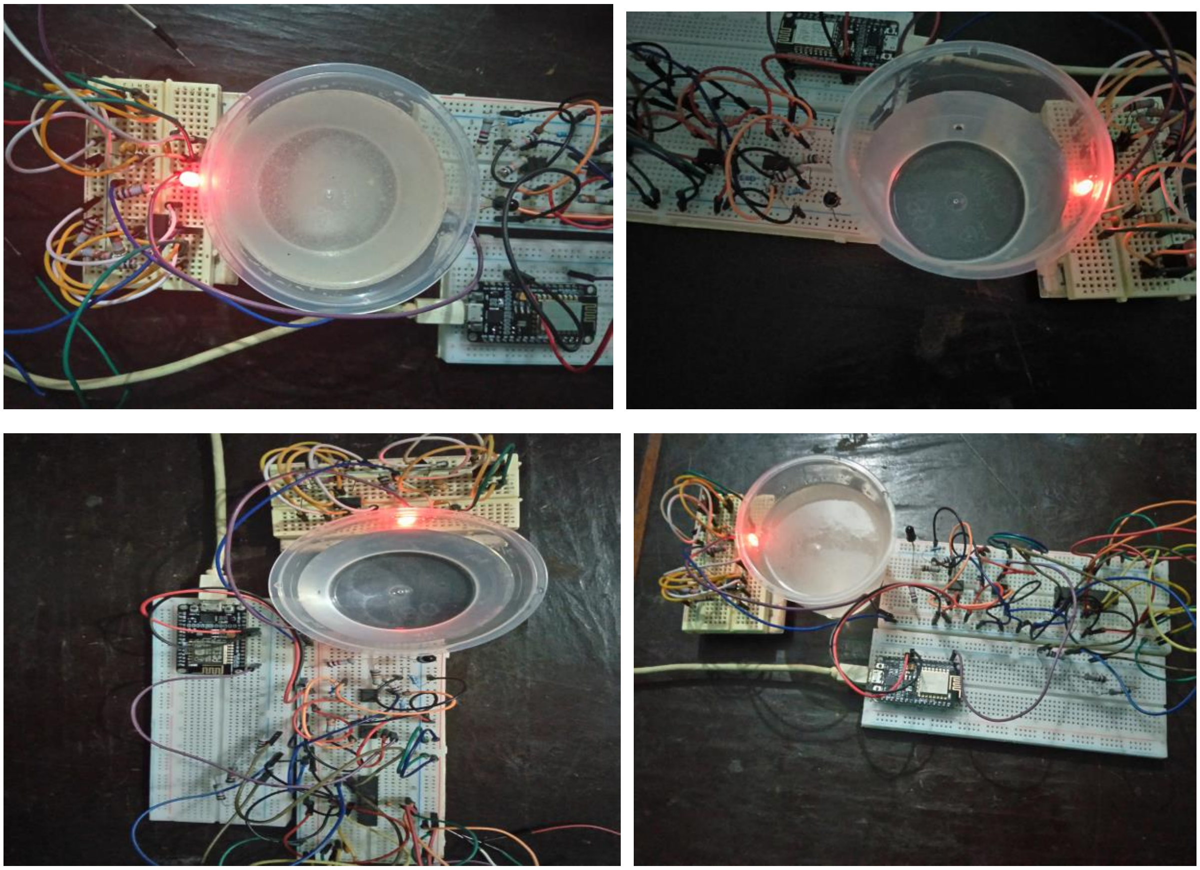


Figure 4.3: Set up of system hardware with turbidity sensor

a. *Pure Water Sample* b. *2.42 M salt solution* c. *4.86 M salt solution*

Figure 4.4: The three samples used in measuring the turbidity

Three samples were used for testing the system: **pure/clean** water, **2.42 M** salt solution, and **4.86** M salt solution shown in figure 4.4. The built turbidity sensor was then set up to test the three different test samples: pure water, salt water, and visibly cloudy water (4.86M salt solution) as shown in figure 4.3. Two observations were made for each of the test samples. An interval of 1 minute elapsed before the second reading was made. The measured data is then sent from the microcontroller to the database using the esp8266. The data gathered from the sensor is analyzed and discussed in chapter 5.

4.4 Edge Analytics of Data

Edge computing is done on the measured sensor data in the atmega328P which is programmed using Arduino. The code developed in the chip is used to take the average of ten voltage readings from the sensor, convert the voltage readings to NTU values, as well as determine the corresponding water quality index from the NTU values using Eqn(1) in chapter 2. This ensures that the measured data is pre-processed and thus relevant data to the system is transmitted to the cloud for further analysis. The next step after performing edge analytics is to send the data to a database which is described in section 4.5.

4.5 Database and API Development

Admin	**Edit User info** **Delete User Account**
User	**Sign-up and Login** **Check Water Quality** **View Past Predictions** **View Purification Suggestions**

Table 4.2: Use case table for users

The system may have two possible types of users; a user with administration rights who can edit user information and delete user accounts should the user be inactive. The user on the other hand can through the web-based application, sign-up and login onto the system, and have the three options of either: *checking the water quality, viewing past predictions, and view purification suggestions.(Refer to Appendix B for a snapshot of the website)*

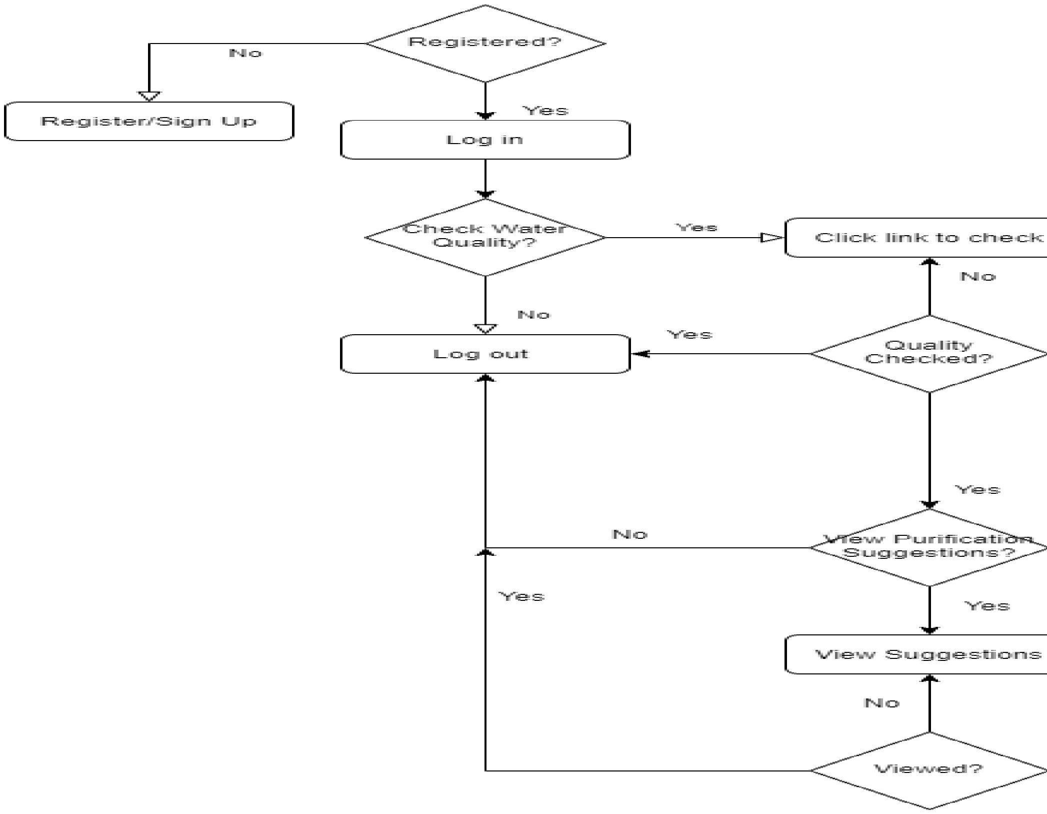


Figure 4.5: Flow chart for the system

In an attempt to access the web-app by the user, the system checks if the user has been registered. If "Yes", the user can log in, where the system sends a prompt to check the quality of a test sample; the user is made to register if otherwise. If "Yes", the system checks the water quality and displays the result to the user (individual parameter values and percent quality), the user can logout if otherwise. After checking the quality, the system further prompts the user for either of the options, *View Past Predictions* and *View Purification Suggestions.*

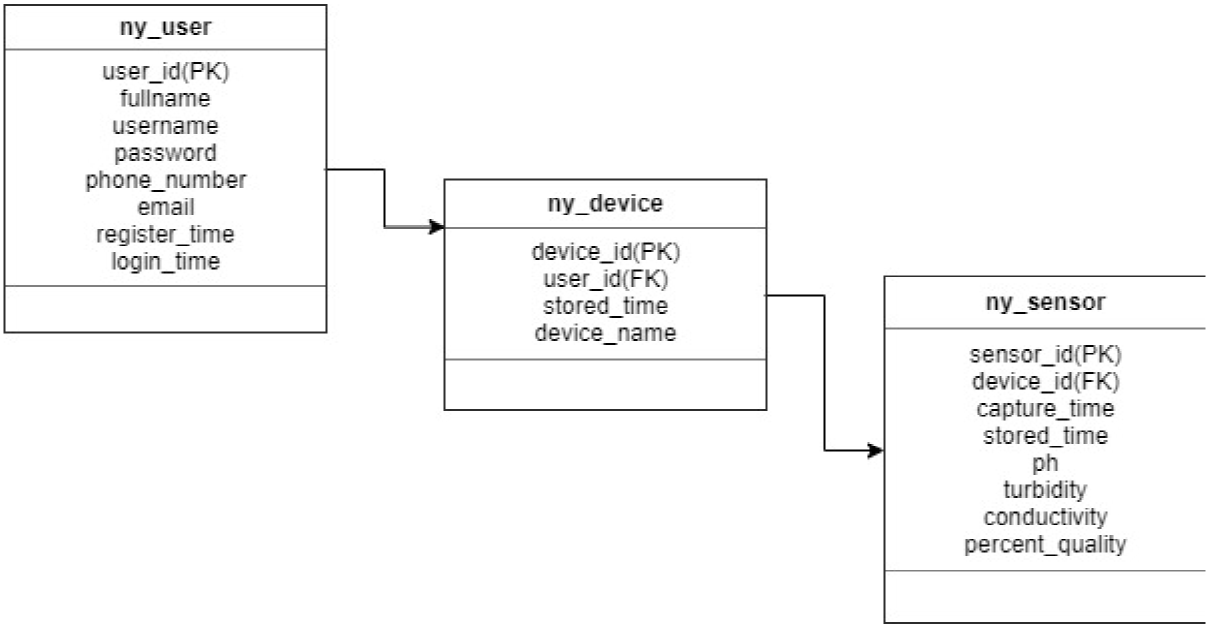


Figure 4.6: Database Schema of system

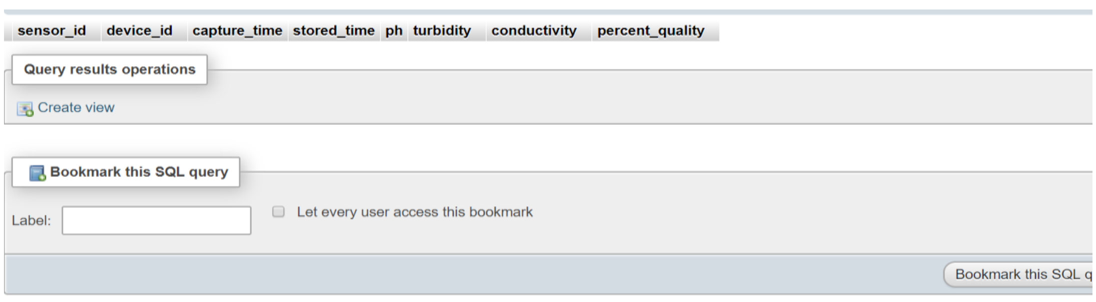


Figure 4.7: Creation of database nycapstone on the local server

A database **nycapstone** is created in the local server where 3 tables **ny_device, ny_sensor, and ny_user** is created as well. From figure 4.6, the table ny_user consists of 8 columns which contain basic information about the user, including user_id which serves as a primary key in the table, full name, username, and password. The table ny_device contains the columns device_id, user_id, and device_name for identifying each device made by the manufacturer. The table ny_device is linked to ny_user through the column user_id which serves as a foreign key. The database also contains the table ny_sensor with columns such as turbidity, pH, and conductivity for storing the measured values as well as their time of storage. The table ny_sensor is also linked to ny_device through the column, device_id which serves as a foreign key.

←T→	sensor_id 1	device_id	capture_time	stored_time	ph	turbidity	conductivity	percent_quality	Sample
Edit Copy Delete	1550	NULL	NULL	2020-05-09 09:08:37		4.32	0	86.34	DoubledConc
Edit Copy Delete	1551	NULL	NULL	2020-05-09 09:08:37		4.32	0	86.36	DoubledConc
Edit Copy Delete	1552	NULL	NULL	2020-05-09 09:08:45		4.32	0	86.34	DoubledConc
Edit Copy Delete	1553	NULL	NULL	2020-05-09 09:08:49		4.32	0	86.37	DoubledConc
Edit Copy Delete	1554	NULL	NULL	2020-05-09 09:08:49		4.32	0	86.42	DoubledConc
Edit Copy Delete	1555	NULL	NULL	2020-05-09 09:08:57		4.32	0	86.33	DoubledConc
Edit Copy Delete	1556	NULL	NULL	2020-05-09 09:09:00		4.32	0	86.34	DoubledConc
Edit Copy Delete	1557	NULL	NULL	2020-05-09 09:09:01		4.32	0	86.41	DoubledConc
Edit Copy Delete	1558	NULL	NULL	2020-05-09 09:09:07		4.32	0	86.38	DoubledConc
Edit Copy Delete	1559	NULL	NULL	2020-05-09 09:09:13		4.32	0	86.35	DoubledConc

Figure 4.8: Turbidity and WQI values entering the database nycapstone on the local server.

After the sensor circuit measures the turbidity of the water sample, the data is sent to the atmega328P for the necessary analysis, and this data is then sent to the local server as shown in figure 4.8.

4.6 Data Processing and Machine Learning Algorithms

After the data from the sensor device has been uploaded to the database, it is imported into jupyter notebooks where machine learning methods are performed on the data. The imported data set is trained and tested using machine learning algorithms. In this work, the machine learning algorithms employed are the Random Forest Regressor and Decision tree regressor. These two were chosen for this work because of their abilities to perform both regression and classification methods. Machine learning is employed in this work because of the vast number of data being processed in the system and also it is a necessary tool in providing accurate and fast predictions of the system being developed.

4.7 Prediction Models

a. Decision Tree Regressor: The decision tree regressor model is a model that uses decision trees to model and predict a data set. It can perform both regression and classification computations. The number of trees used here is usually limited and cannot be used for complex systems.

b. Random Forest Regressor and Classifier: Random Forest Regressor model like the decision tree performs both regression and classification computation for fitting the data set with its model. It is a supervised data algorithm that creates many decision trees and from which random models are made and the best of the random is chosen as the best fit for the data set. Due to it employing many decision trees it less biased than other models and can handle large amounts of data.

4.7 Model Evaluation/ Performance Metrics

a. Mean Absolute Value: This is a regression error metric used measuring how far apart a predicted output is from the actual measured data.

b. Root Mean Square Error: This error metric measures the fitness of the regression models to the measured data set.

Chapter 5: Results Discussion/Analysis

In this chapter, the results from implementing the machine learning algorithms—Decision Tree Regressor, Decision Tree Classifier, and Random Forest—are discussed and analyzed, the models evaluated, and how they fit the overall system. Also, to further evaluate the robustness and precision of the system, the data is evaluated using statistical analysis implemented in R Studio, specifically a student t-test and a scatter plot on the three samples as mentioned in chapter 4.

5.1 Machine Learning Algorithms

In this section, the results from the implementation of the various machine learning algorithms described in chapter 4, is explored and analyzed to evaluate the system. To display and observe the output of the prediction, the last five data entries into the database were used as seen in figure 1 below. These last five entries are data belonging to the doubled concentrated salt solution. The predicted values from the models are compared to the actual measured WQI value in Appendix. The results from performing the Random forest Regressor and Classifier and Decision Tree Regressor are summarized below with their respective errors.

Prediction of models

```
      weekday  month  turbidity
1241        5      5       4.32
1242        5      5       4.32
1243        5      5       4.32
1244        5      5       4.32
1245        5      5       4.32
The predictions are
[6.95740365 6.95740365 6.95740365 6.95740365 6.95740365]
```

(a) *Decision tree Regressor (Even split)*

```
      weekday  month  turbidity
1241        5      5       4.32
1242        5      5       4.32
1243        5      5       4.32
1244        5      5       4.32
1245        5      5       4.32
The predictions are
[8.47418478 8.47418478 8.47418478 8.47418478 8.47418478]
```

(b) Decision Tree Regressor (Uneven Split)

```
      weekday  month  turbidity
1241        5      5       4.32
1242        5      5       4.32
1243        5      5       4.32
1244        5      5       4.32
1245        5      5       4.32
The predictions are
[14.10905659 14.10905659 14.10905659 14.10905659 14.10905659]
```

(c) Random Forest Regressor

```
      weekday  month  turbidity
1241        5      5       4.32
1242        5      5       4.32
1243        5      5       4.32
1244        5      5       4.32
1245        5      5       4.32
The predictions are
[0 0 0 0 0]
```

(d) Random Forest Classifier

Figure 5.1: Output prediction from using the Machine Learning models

In predicting the model using the decision tree regressor, the data set was split and used to train the data set. First, the decision tree regression was used on the data set where the data set is split equally in half; one split is used to train the data to fit the model and the other, used to test the model for the prediction of the water quality index. Next, the decision tree regression is used where the data is split unevenly between those that would be used to train the data to fit the model and that which would be used to test the data.

Evaluation of Machine Learning Models

The various errors described in chapter 4 are essential in evaluating which of the employed model best fits the system based on the error produced by each. The error gives an idea of by how much the predicted value diverges from the observed/measured water quality index. The table below summarizes the relative errors obtained from predicting the water quality index of samples.

ML Algorithm	MAE	RMSE
Decision Tree Regressor (Even Split)	20.2709	34.7338
Decision Tree Regressor (Uneven Split)	23.9707	38.9411
Random Forest Regressor	20.2892	**34.7293**
Random Forest Classifier	**17.9326**	36.7184

Table 5.1: Mean errors obtained from the computation of the ML algorithms

Discussion

From table 5.1, comparing the mean absolute errors and root mean square of the various models, the Random Forest Classifier computed the least mean absolute error and the Random Forest Regressor computed the least root mean square error. For MAE, the lower the value, the better the model is at predicting. Thus, we can overly conclude that the Random Forest is a good prediction model in this regard to the other models used as it employs more decision trees in its implementation. From appendix C, the measured WQI value is between the ranges of 86. 30 and

86. 40 yet none of the models predict a value close to the measured value. Why is this so? From appendix C, a simple turbidity and WQI histogram show how far apart the data values are and the outliers between the two extremes. Due to the divergences of the data, the ML models may have built a model for these two extremes in each case that, they tend to choose the best of the models out of their decision trees. Random forest is known for its simplicity and many decision trees and thus, although its prediction of WQI values was not close to the measured, it did have the least mean absolute error.

5.2 System Precision Evaluation Based on Statistical Analysis

In this chapter, various statistical analysis—t-test, scatter plot— is made to analyze the precision of the system for the data obtained in each of the three sample cases described in chapter 4. The paired-student t-test, scatter plots and the analysis of variance, tests the analysis of the measured data obtained in each instance. As discussed in chapter 4, the measurement of turbidity in each test sample was observed twice, in order to check for precision—that is that the same data value is observed twice. The sole aim of the statistical analysis done in this section is to validate the precision of the hardware system (turbidity sensor circuit).

5.2.1 Analysis of Method

The analysis of the data is in two parts: (1) performing a scatter plot and paired t-test between the two observations of each test sample data and (2) Performing an analysis of variance (ANOVA) test on the different test sample data. The scatter plot has a 45˚ line with intercept=0 and slope= 1 as shown in figure 5.1, is used to fairly describe the change that occurred between the data in each instance A 99% confidence was used and a null hypothesis of Ho: Mean difference is equal to 0, was used in all the two-case analysis. Precision is checked by the rejection or acceptance of the null hypothesis through the p-value obtained after performing the t-test. The analogy below is used in checking for precision.

Case	**Hypothesis Status**	**Meaning**
P<0.05	Reject null hypothesis	Significant difference between the two means exist
P>0.05	Accept null hypothesis	No significant difference exists between the two means

Table 5.2: Analogy of the how precision is determined statistically for the paired t-test

5.2 Paired T-test

5.2.1 Testing of Pure Water Sample

Statistical Parameter	**Value**
p-value	0.783
Confidence level interval	-0.5956526–0.7096526
Mean of differences	0.057

Table 5.3: Statistical Values obtained after performing t-test on pure water sample values

Since a p-value of 0.783 was obtained, there is no significant difference between their means.

5.2.2 Testing of Salt-Water Sample

Statistical Parameter	**Value**
p-value	0.2512
Confidence level interval	-0.1716385—0.3796385
Mean of differences	0.104

Table 5.4: Statistical Values obtained after performing t-test on salt-water sample values

5.2.3 Testing of visibly cloudy water

Statistical Parameter	**Value**
p-value	0.8783
Confidence level interval	-0.2747786—0.3027786
Mean of differences	0.014

Table 5.5: Statistical Values obtained after performing t-test on cloudy water sample values

5.3 Using Analysis of Variance (ANOVA) Test and Tukey Test

An ANOVA test is conducted to test the null hypothesis of *Ho: mean of sample measurements is the same for all test samples*—ie pure water, salt water, and visibly cloudy water samples. The ANOVA test aims to establish if there are significant differences between the means of the measured samples. The Tukey test then was used to further distinguish among which of the three samples do these differences occur which is summarized in table 5.5. Table 5.4 below summarizes the statistical result of the ANOVA test carried it in R.

Statistical Parameter	Value
p-value	0.0372
F-Value	3.729

Table 5.6: Statistical results from performing ANOVA test on test sample data

Paired Measurement	P-Value
P_sample and C_sample	0.0332824
S_sample and C_sample	0.1731336
S_sample and P_sample	0.6966519

Table 5.7: Statistical results from performing Tukey test on test sample data. Check Appendix A.

5.4 Discussion of Results

In section 5.2, a scatter plot and paired t-test is done for each of the test samples. The p-values for all three samples, pure water, salt water, and cloudy—0.783, 0.25, and 0.87 respectively—are all above the p-value of 0.05. Thus, according to table 5.1, the null hypothesis should be accepted in each case and this indicates that there is no statistical difference between the means of the two observations of each test sample. There is thus consistency in the measurements made by the turbidity sensor as it gives fairly the same readings when the measurements of a sample were taken twice. The ANOVA test done in section 5.3 determines if there is a statistical difference between the values measured in each of the test samples. From table 5.4, the results from the ANOVA test gave a p-value of 0.0372 which is less than 0.05 and thus the null hypothesis is rejected. This means that there are significant differences between the measured turbidity readings of each test sample. The result of the Tukey test displayed in table 5.5 shows where these differences take place among the different samples. A p-value of 0.0332824 for the test between pure water sample and cloudy water sample depicts that there is a significant difference between the values as should be expected. Since there are many particles in the cloudy water solution, the light rays from the light source were scattered and thus just a fragment of the incident light was detected by the photodiode. However, for the test between salt-water &cloudy water and salty water and pure water with p-values of 0.17 and 0.69 respectively, the null hypothesis is accepted, depicting that there is not a statistical difference between their means. The particles in the salt-water sample are slightly less than that of the cloudy water and thus it should be expected that the means should be different, however, the Tukey test gave a result otherwise. This is accounted for by the type of light source used in this project. EPA standards require a tungsten lamp—of wavelength 860nm—as the light source for a turbidity meter. Thus, essentially, since an led of wavelength 360nm was used, the scatter of light rays in the pure water sample and saltwater were fairly the same as well as that of the saltwater and cloudy-water, thereby causing the Tukey test to detect no significant difference amongst them. The limitations and future works of this project are discussed in the following sections.

Chapter 6: Conclusion

The design and implementation aim of this work was to build three sensors that met international standards, use them in measuring the parameters, pH, turbidity, and conductivity. These measured parameters are sent to a web app and to a database for storage. The work done in this project, through the use of different technologies has been able to build a turbidity sensor, whereby the data from this sensor is sent to an atmega328P for analysis before being sent to a the web app and database via esp-01 Wi-Fi module. Due to the outbreak of the pandemic in China last year, some major components needed in building the other two sensors was unavailable. Thus, the project was carried out with only the implementation of the turbidity sensor. The outbreak of the pandemic in Ghana during the building stage caused some design implementation activities incapable of being done due to the absence of needed equipment needed from the Ashesi fabrication lab.

6.1 Limitations

Some limitations which were encountered during this project include unavailability of critical components, improvision made for some of the unavailable components, and some of these materials not meeting WHO and EPA standards. They are discussed below.

- The first limitation which reduced the scope of this project is the unavailability of some of the components needed to build the pH and conductivity sensor and thus reducing the number of measured parameters to one. A silver chloride reference electrode and silver lead wire both needed in the construction of the pH and conductivity was unavailable as the order from China was unable to be delivered due to their ports being shut down as a result of the COVID-19 pandemic. This electrode and wire are essential to the measurement of these parameters because they can detect the water elements needed to generate a voltage. Also, the light source was improvised and did not possess the right wavelength required by water quality measurement standards.

- Unavailability of lab technologies, such as that used in the making of printed circuit boards caused the PCB design of the system circuit to not be printed.
- The line of sight between the light source and the photodetector caused there to be inaccuracy in the measured readings.
- The percentage quality displayed to the user was only based on the turbidity readings which is insufficient data to come to such a conclusion.
- The accuracy of the system was not checked since a standard turbidity was unavailable.

6.2 Future Works

The future development of this project could see the incorporation of the pH and conductivity sensor to be able to display the correct percentage quality of a test sample. The further development of this project would create the complete package of a water quality checker for the urban user who would not have to wait three days for the results of the test. The system can be further developed by:

- Adding and building more water quality sensors such as chlorine residual, microbial presence to make the system more robust.
- Build a smart-phone app in addition to the web app
- Add the feature in the app, whereby purification suggestions can be made based on the results displayed from a test.
- Expand the power source of the device to include solar power.

- Employ more machine learning algorithms, to improve the accuracy and precision of the system.

References

[1] T, Wong (2014). [online] Web.mit.edu. Available at:

http://web.mit.edu/watsan/Docs/Student%20Theses/Ghana/2014/Thesis_FINAL_Teng_Ke_Wong_5-25-14.pdf [Accessed 13 Oct. 2019].

[2] '(PDF) Domestic Water Quality Assessment: Microalgal and Cyanobacterial Contamination of Stored Water in Plastic Tanks in Zaria, Nigeria'. [Online]. Available: https://www.researchgate.net/publication/257138907_Domestic_Water_Quality_Assessment_Microalgal_and_Cyanobacterial_Contamination_of_Stored_Water_in_Plastic_Tanks_in_Zaria_Nigeria?enrichId=rgreq-77defeca65745acca8fadaa188bc3a73-XXX&enrichSource=Y292ZXJQYWdlOzI1NzEzODkwNztBUzoxMDQ1NjczNDE1ODQzOTdAMTQwMTk0MjIwNTk2OA%3D%3D&el=1_x_3&_esc=publicationCoverPdf. [Accessed: 13-Oct-2019].

[3] Agyemang-Mensah, K. (2015). *National Drinking Water Quality Management Framework for Ghana "Prevention is better than cure"*. [online] Gwcl.com.gh. Available at: https://www.gwcl.com.gh/national_drinking_water_quality__management_framework.pdf [Accessed 13 Oct. 2019].

[4] D. Chalchisa, M. Megersa, and A. Beyene, 'Assessment of the quality of drinking water in storage tanks and its implication on the safety of urban water supply in developing countries', *Environ Syst Res*, vol. 6, no. 1, p. 12, Apr. 2017.

[5] '(2) (PDF) Design of smart sensors for real time drinking water quality monitoring and contamination detection in water distributed mains'. [Online]. Available: https://www.researchgate.net/publication/324052217_Design_of_smart_sensors_for_real_time_drinking_water_quality_monitoring_and_contamination_detection_in_water_distributed_mains. [Accessed: 06-Nov-2019].

[6] S. Srivastava, S. Vaddadi, and S. Sadistap, 'Smartphone-based System for water quality analysis', *Appl Water Sci*, vol. 8, no. 5, p. 130, Aug. 2018.

[7] ‘(PDF) Smart Water Quality Monitoring System Using Iot Technology’. [Online]. Available: https://www.researchgate.net/publication/330637347_Smart_Water_Quality_Monitoring_System_Using_Iot_Technology. [Accessed: 23-Sep-2019].

[8] (PDF) First advances on the development of a hydroponic system for cherry tomato culture. (n.d.). Retrieved December 22, 2019, from ResearchGate website: https://www.researchgate.net/publication/261082725_First_advances_on_the_development_of_a_hydroponic_system_for_cherry_tomato_culture

[9] L. Zhang and X. Gu, ‘Study on residual chlorine detecting system based on dual-wavelength optical path’, 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), Jan. 2010.

[10] *(PDF) A review of emerging trends on water quality measurement sensors*. Available from: https://www.researchgate.net/publication/281924339_A_review_of_emerging_trends_on_water_quality_measurement_sensors [accessed Dec 22 2019].

[11] “Turbidity_sensor_SKU__SEN0189-DFRobot.” https://wiki.dfrobot.com/Turbidity_sensor_SKU__SEN0189 (accessed May 08, 2020).

Appendix

Appendix A: Important Hardware Components Used

Atmega 328p

LM 741

ESP8266 WiFi Module

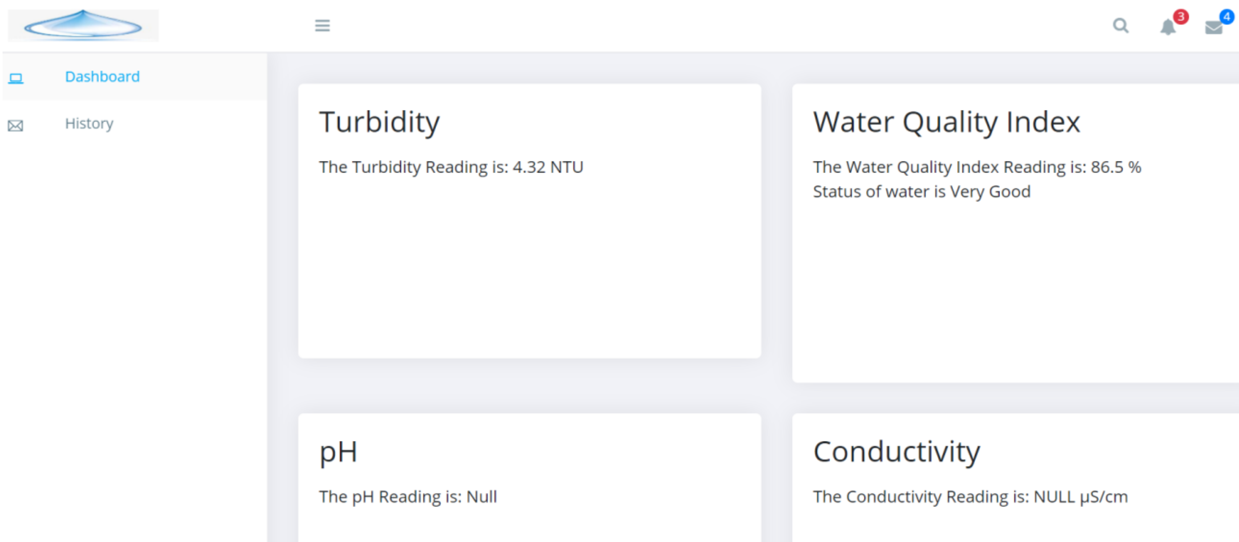
Labeling For Test Samples used in Statistical Test

P_sample and C_sample = Pure Water and 4.86 M sample

S_sample and C_sample = 2.42 M Salt and 4.86 M Salt Solution Sample

S_sample and P_sample = 2.42 M Salt Solution and Pure Water Sample

Appendix B: Web App and Database



Measured turbidity value and Water Quality Index data transmitted to web app

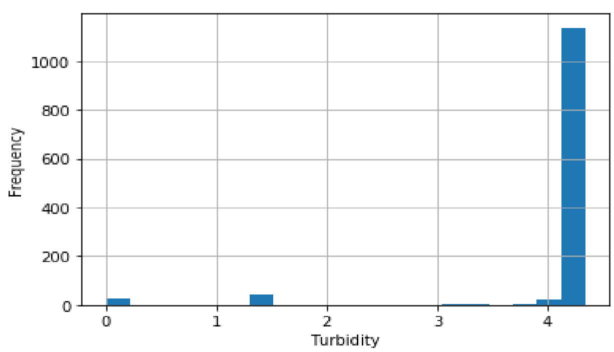
	sensor_id 1	device_id	capture_time	stored_time	ph	turbidity	conductivity	percent_quality	Sample
Edit Copy Delete	1550	*NULL*	*NULL*	2020-05-09 09:08:37		4.32	0	86.34	DoubledConc
Edit Copy Delete	1551	*NULL*	*NULL*	2020-05-09 09:08:37		4.32	0	86.36	DoubledConc
Edit Copy Delete	1552	*NULL*	*NULL*	2020-05-09 09:08:45		4.32	0	86.34	DoubledConc
Edit Copy Delete	1553	*NULL*	*NULL*	2020-05-09 09:08:49		4.32	0	86.37	DoubledConc
Edit Copy Delete	1554	*NULL*	*NULL*	2020-05-09 09:08:49		4.32	0	86.42	DoubledConc
Edit Copy Delete	1555	*NULL*	*NULL*	2020-05-09 09:08:57		4.32	0	86.33	DoubledConc
Edit Copy Delete	1556	*NULL*	*NULL*	2020-05-09 09:09:00		4.32	0	86.34	DoubledConc
Edit Copy Delete	1557	*NULL*	*NULL*	2020-05-09 09:09:01		4.32	0	86.41	DoubledConc
Edit Copy Delete	1558	*NULL*	*NULL*	2020-05-09 09:09:07		4.32	0	86.38	DoubledConc
Edit Copy Delete	1559	*NULL*	*NULL*	2020-05-09 09:09:13		4.32	0	86.35	DoubledConc

Entries of turbidity Readings into the Database

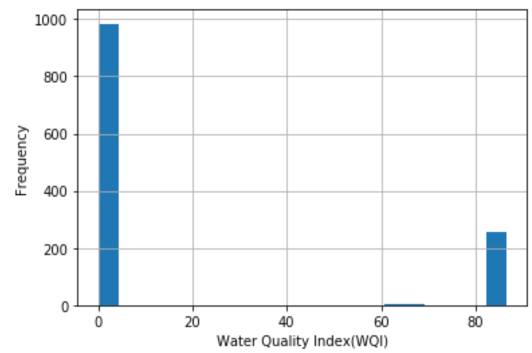
Appendix C: Graphs of Data

Total Data Entries = 1246 => (1246 rows x 10 columns)

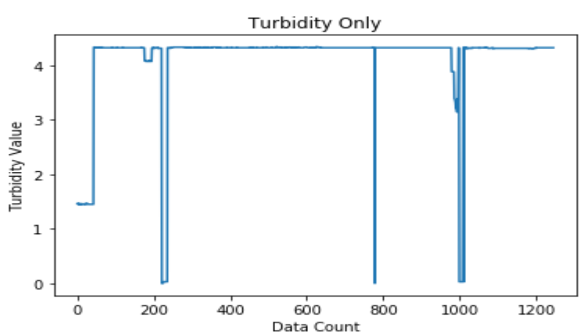
Graph from Data Set Performed in Jupyter



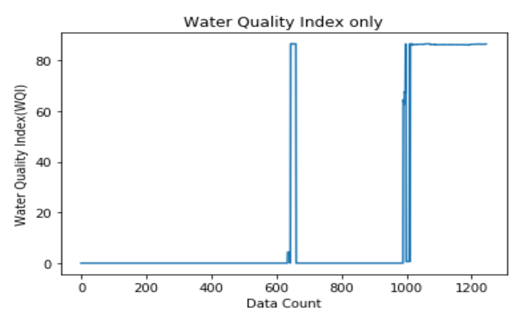
A. Histogram showing Turbidity Readings



B. Histogram showing WQI Readings



C. Line Graph of Turbidity Readings



D. Line graph of WQI Readings

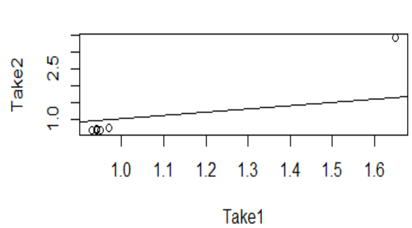
sensor_id	turbidity	conductivity	percent_quality	
count	1246.000000	1181.000000	1246.0	325.000000
mean	909.889246	4.099873	0.0	69.029846
std	415.735414	0.828972	0.0	34.001908
min	38.000000	0.000000	0.0	0.000000
25%	628.250000	4.320000	0.0	86.100000
50%	939.500000	4.320000	0.0	86.200000
75%	1258.750000	4.320000	0.0	86.330000
max	1570.000000	4.340000	0.0	86.520000

Description of data set

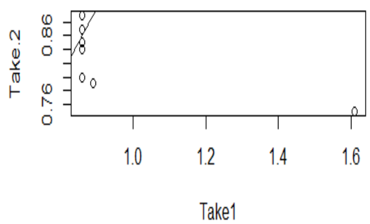
	sensor_id	device_id	capture_time	stored_time	ph	turbidity	conductivity	percent_quality	Sample	weekday	month	sampleID
1241	1566	None	None	2020-05-09 09:09:38		4.32	0.0	86.43	DoubledConc	5	5	2
1242	1567	None	None	2020-05-09 09:09:45		4.32	0.0	86.37	DoubledConc	5	5	2
1243	1568	None	None	2020-05-09 09:09:48		4.32	0.0	86.40	DoubledConc	5	5	2
1244	1569	None	None	2020-05-09 09:09:50		4.32	0.0	86.43	DoubledConc	5	5	2
1245	1570	None	None	2020-05-09 09:09:57		4.32	0.0	86.42	DoubledConc	5	5	2

Last 5 Entires of data set

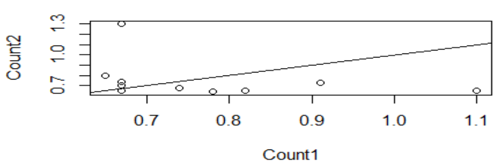
Scatter Plot for Data Set Performed in R



A. Scatter Plot for Pure Water



B. Scatter Plot for 2.43 M salt solution



C. Scatter plot for 4.46 M salt Solution

Appendix D: Calculation for Salt concentration

mass of salt = 17.06 gram (g)

volume of water = 0.00012cm^3 = 120 ml

Molecular weight = 58.5g/mol

Concentration of 17.06 gram of salt solution= 2.43 M

Concentration of 34.12gram of salt solution= 4.86 M