

---

# **Automated *in-situ* sensors & Spatiotemporal Monitoring & Machine Learning prediction of Harmful Algal Blooms**

by

**Okomo Jacob Okello**

***Case Study Lake Victoria***

*Project report submitted to the department of Geomatic Engineering and geospatial Information Systems degree of Bachelor of Science in Geospatial Information system GIS.*



**Department of Geomatic  
Engineering and Geospatial  
Information Systems (GEGIS)**

---



---

## Table of contents

Table of contents .....	III
List of figures .....	<b>Error! Bookmark not defined.</b>
List of tables.....	V
Acronyms and abbreviations .....	<b>Error! Bookmark not defined.</b>
1 Introduction .....	- 1 -
1.2 Motivation and problem statement .....	- 1 -
1.3 Research identification .....	<b>Error! Bookmark not defined.</b>
1.3.1 Research objectives.....	- 3 -
1.4 Study outline .....	- 3 -
3 Materials and methods.....	- 3 -
3.1 Study area .....	<b>Error! Bookmark not defined.</b>
3.2 Data.....	- 7 -
3.3 Methodology .....	- 7 -
References .....	A





**List of tables**

Data and sources ..... **Error! Bookmark not defined.**

Materials and Methods ..... **Error! Bookmark not defined.**

---

# 1 Introduction

## 1.1 Background

Algae, in limited concentration, are ecologically friendly however when an unanticipated bloom comes to pass, can have severe impacts on human health, aquatic ecosystems. This is the fact that the said form unsightly views and nuisance in points of impact and with cyanotoxins, initiated by the cyanobacteria being particularly problematic as they can be toxic and scum-forming, posing a risk to the ecosystem and to public health and as well detrimental to the economy.

The geoscientific preparedness to monitor and predict algal and cyanobacteria blooms of great material value to provide pre-warning to society and enable management processes to be activated in advance to limit the disastrous and catastrophic impact.

Currently, there is ongoing collaborative efforts between SERVIR Africa and the Regional Centre for Mapping of Resources for Development (RCMRD). The research is currently mainly geared for estimating algal concentrations in Lake Victoria by assessing and monitoring the Lake Victoria water quality parameters such as chlorophyll concentration, Lake Surface Temperature, and turbidity for the lake using satellite data from the Moderate Resolution Imaging Spectrometer (MODIS) sensor on the Aqua satellite.

The advent and uptake of high resolution in-lake automated water quality sensing technology together with new satellite platforms now enables a step-change in data availability that could be used for monitoring and forecasting of cyanobacteria (and algal) blooms in Lake Victoria.

## 1.2 Motivation and problem statement

The rapidly escalating demographics a long Lake Victoria riparian reserves has negatively impacted water quality through deposits of agricultural and industrial runoff and sewer refuse. **Harmful algal blooms** (HABs) particularly, are occasionally associated with the gram-negative Cyanobacteria hence the name CyanoHABs or 'red tides' are a global phenomenon and recent evidence indicates that their frequency and intensity are increasing (Shumway, 1990, Smayda, 1990, Hallagraeff, 1993, Burkholder, 1998). They are a serious threat to **human** health, aquaculture, fisheries, and ecosystem health.

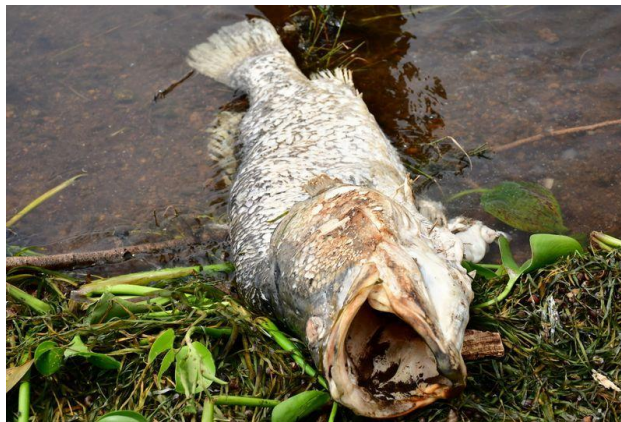
With the growth of industries along the Lake Victoria regions, there has been reported enrichment of nutrients, increasing the amount of plant and algae growth in the Lake. Lake Victoria has been reported to face eutrophication challenges, resulting in an increase of bloom-forming cyanobacteria (CynoHABs), (Martin K., et al. 2019).

A good number of cyanobacteria species rich algae blooms can create toxins that affect the nervous and chronic system ranging from liver, and skin just to highlight, causing injury to humans and their companion animals, such as pets who drink water

from them or play with them and to extreme ends might be fatal, unfortunately. HABs can also harm freshwater ecosystems by fouling beaches, producing nuisance, unsightly views, taste and pathetic odor in drinking water, and photosynthetically bringing down the potential amount of ambient light required by submerged aquatic vegetation for their survival.



*Fig 1: Sample Lake Victoria Shores and Mass Fish negatively impacted by the Harmful Blooms*



*Fig 2: A Big Fish suffocate to death by HAB*

---

The main aim of this research is to monitor and predict the occurrence of Harmful Algal Blooms (HABs) and Cyanobacteria in Lake Victoria. This is achievable through the following specific objectives:

### **1.2.1 Research objectives**

1. To monitor Harmful Algal Blooms and Cyanotoxins from Satellite RS Images data in L. Victoria.
2. To Predict occurrence of cyanobacterial and Harmful algal blooms in the Selected Lake.
3. To associate Automated Internet of Things (IoT) *in situ* sensors, machine learning Applicable in near real-time to enhance the accuracy and speed for *in situ* data analysis.
4. To develop a reporting system to alert on a short-term foreseen bloom.

## **2 Materials and methods**

### **2.1 Study Area**

Lake Victoria, with a surface area of about 68,800 KM<sup>2</sup> and an average depth of 40m at a maximum depth of 79m ranks the second largest fresh water lake in the world after Lake Superior and the Largest in Africa. Lying between 3° S to 0° 30` N latitude and 31° 40` E to 34° 50` E is distributed among these three East African countries viz Tanzania 51%, Uganda 43% and Kenya the remaining 6% (africangreatlakes.org).

That in place, the lake is privileged to serve as economical home of about 40 million residents (Dorothy et. Al 2020) in that riparian reserves. These millions of individuals solely bank on the lake for all aspects of their daily economic livelihood ranging from, fishing, agriculture, and industrial applications just to barely highlight but a few. That in place, it's ecological monitoring should be of great geoscientific interest.

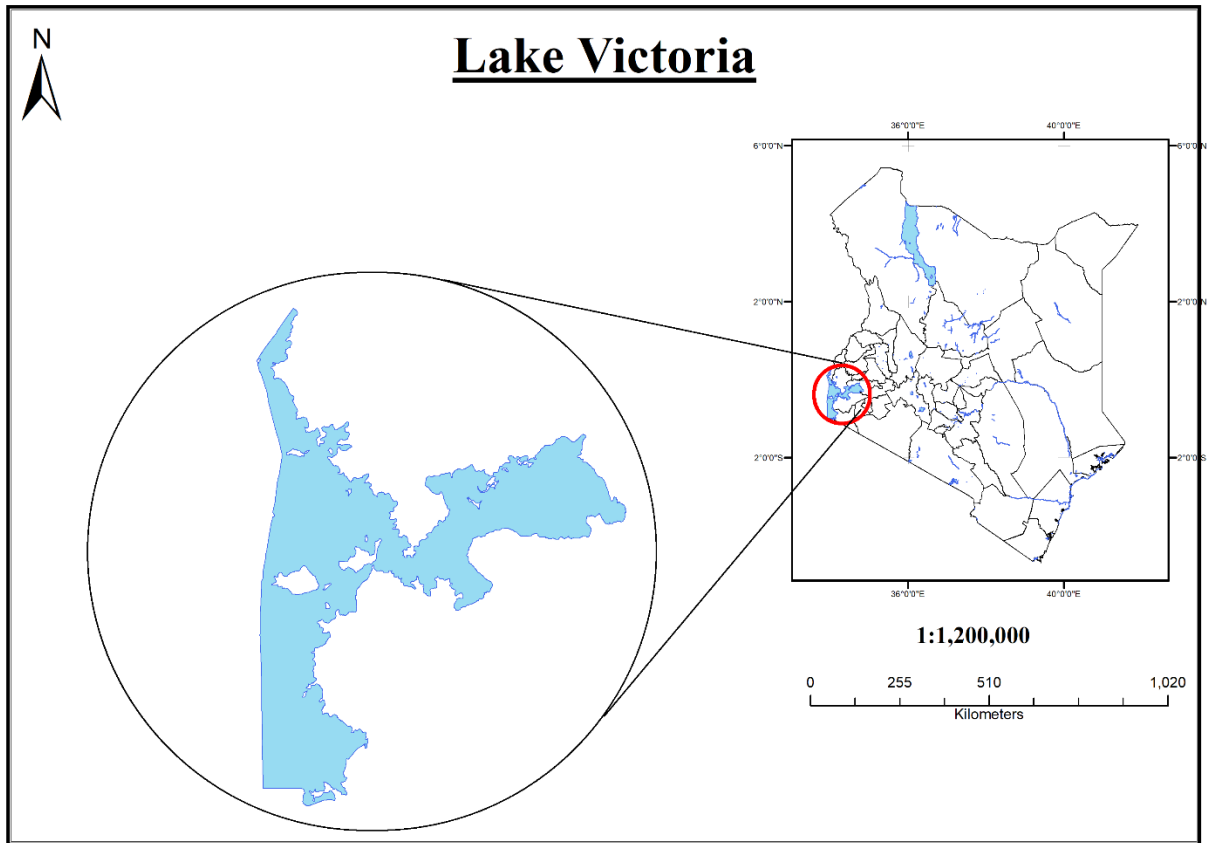
Being located in Equatorial regions of the globe, the lake has an alternating climatic condition varying from tropical rain forest with rainfall over the lake for a better portion of the year to a semi dry climate with sporadically discontinuous droughts over some locations.

This provides ambient temperatures varying between 12-26°C it therefore provided optimum host conditions for the growth and development of the Cyanobacteria in this scope.

Figure 1 shows the location and extent of the study area as discussed above.







*Fig 3 Shows the extent of the study area*



## 2.2 Data and Data sources

Data Type	Source	Access
Shapefiles	Database of Global Administrative Areas- GADM	Free
MODIS Aqua	Google Earth Engine Cloud Repo	Free
Landsat 8 OLI	Google Earth Engine Cloud Repo	Free
- Meteorology Data -In-Situ	-Lake Victoria Water Quality Mgt -My Automated <i>In-situ</i> System	-Uncertain -Within my control

### Tools and Materials

Tool/ Material	Role	Availability
Google Earth Engine	Geocomputation & Processing	Freely Available
QGIS, R & Python	Further Analysis, Maps & Prediction	Free
Microcontroller, Sensors & Gateway	In-situ data collection & Monitoring	Local Purchase



---

## 2.4 Methodology

Landsat and MODIS observations were adopted as fine- and coarse- resolution satellite data for at least Seven (7) reasons:

- (1) Freedom of Accessibility: The two NASA owned satellite missions provide unlimited free-ride to their Earth Observation data access and usage irrespective of your research budget.
- (2) Previous studies and research rate them a better score as they have been widely used to delineate the concentration and spatial distributions of CyanoHABs in inland waters (Hu et al., 2010; Vincent et al., 2004), and these scholarly works are comparatively widely documented;
- (3) Having actively stayed in orbit for about one and two decades respectively, the two sensors archive long-term and continuous observations, thereby benefiting the long-term time series analysis of CyanoHABs in Lake Victoria;
- (4) They have similar band configurations, orbital parameters (Sun Synchronous), near-nadir viewing-solar geometries, and imaging time (both cross the equator at nearly 10:00 +/- 15 mins), thereby providing a solid basis for the Spatio Temporal Image Fusion processing (Gao et al., 2006).
- (5) MODIS images can easily be geometrically matched with Landsat images by applying optimal offsets (Gevaert and García-Haro, 2015),
- (6) The Short-Wave InfraRed (SWIR) band of MODIS Aqua at 1230–1250-nm is the first choice for algae index derivation such as in Floating Algae Index (FAI) (Hu, 2009), based on MODIS.
- (7) In Landsat 8, there exists the provision of three significant spectral bands 1(Coastal), 3(Green, plant vigor), and 5(NIR, shorelines) which produce the most promising results for accurately estimating chl-a concentrations in lakes regions.

### **b. Near real-time Automated *in situ* Lake Surface temperature data.**

The internet of things, or IoT, is a network of interconnected computing devices, mechanical and digital machinery, items, animals, or people having unique identifiers (UIDs) and the ability to transfer data without requiring human-to-human or human-to-computer interaction. (2019, Alexander S.)

---

In this scope, the word IoT has been used to refer to a man-made system that has been assigned an Internet Protocol (IP) address, loaded with sensors and is able to transfer data over a network between the system and a remote computer.

### **Devices and Sensors Used**

At least the following components and sensors have been proposed for achievability of collection of the In-Situ data:

#### 1. Microcontroller (MCU):

This can be lightly defined as an integrated circuit that contains a brain or rather the microprocessor that has a memory and integrated circuits and that has the potential to control the functionalities of an embedded system or electronic device (such as our Water Quality Monitoring system). There exist a vast variety of Commercial Off The Shelf-(COTS) MCUs e.g

- ATmega (Arduinos)
- ESP32
- BCM
- Single Board Computers (SBC) e.g Raspberry Pi
- ARM Cortex e.g STM32



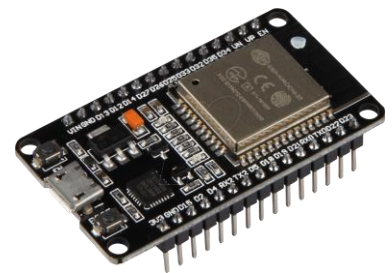
ATmega, Arduino



Raspberry pi



STM32, ARM Cortex



ESP32, LOLIN

*Fig 5, 6 & 7 showing some of the common microcontrollers*

---

Based on my previous experience with ATmega, ESP32 and Raspberry Pi microcontrollers, I will choose Raspberry pi over the rest for a few reasons highlighted below:

### 1. Powerfulness:

**Comparatively**, Raspberry Pi rates highest over other MCUs. It performs relatively similar to a full-blown computer by being able to simultaneously run **multiple tasks** at ago. The additional capability to operate as a Server gives it more potential score and rating.

### 2. Networking:

Coming with a **built in Ethernet port**, its internet strength and connectivity can be easily boosted by a local area network. Unlike for other MCUs that need Wi-Fi shields, of course with proper coding to handle the shields, Pi comes with that pre-configured. This in turn saves the developer a lot of burden in this scope.

### 3. Little to no electronics prerequisites needed

Little embedded programming languages knowledge and its components is needed here. Unlike for Arduino and other ARM Cortex MCUs, you definitively need a very good and deep electronic background, and need to know about embedded programming languages.

### 4. My good knowledge in Linux OS and related

The overarching fact that Raspbian OS runs on similar environment and understands shell commands similar as in Linux Operating system, makes me fall for it for this IoT task. I've used Linux for 3 years (at the time of compilation of this report) and a walk around it will just be another rosy experience.

## 2. Sensors

At least the following sensors have been integrated and used as part of the in-situ data collection system.

#### *a) Global Positioning System (GPS) Sensor.*

Being locally available and easy to interface and tinker with the raspberry pi, among other reasons, the **Neo-6M- Ublox GPS sensor Module** was proposed for this study.

This sensor will solely play the role of collection of the (x, y) coordinate location information about where the whole system will be at any given time. This is an intelligent sensor that runs offline and is capable of collecting the GPS location any



---

system that it's embedded to. This will tell us precisely where in Lake Victoria there exists abnormality in temperature rises, calling for an emergency mitigation action.



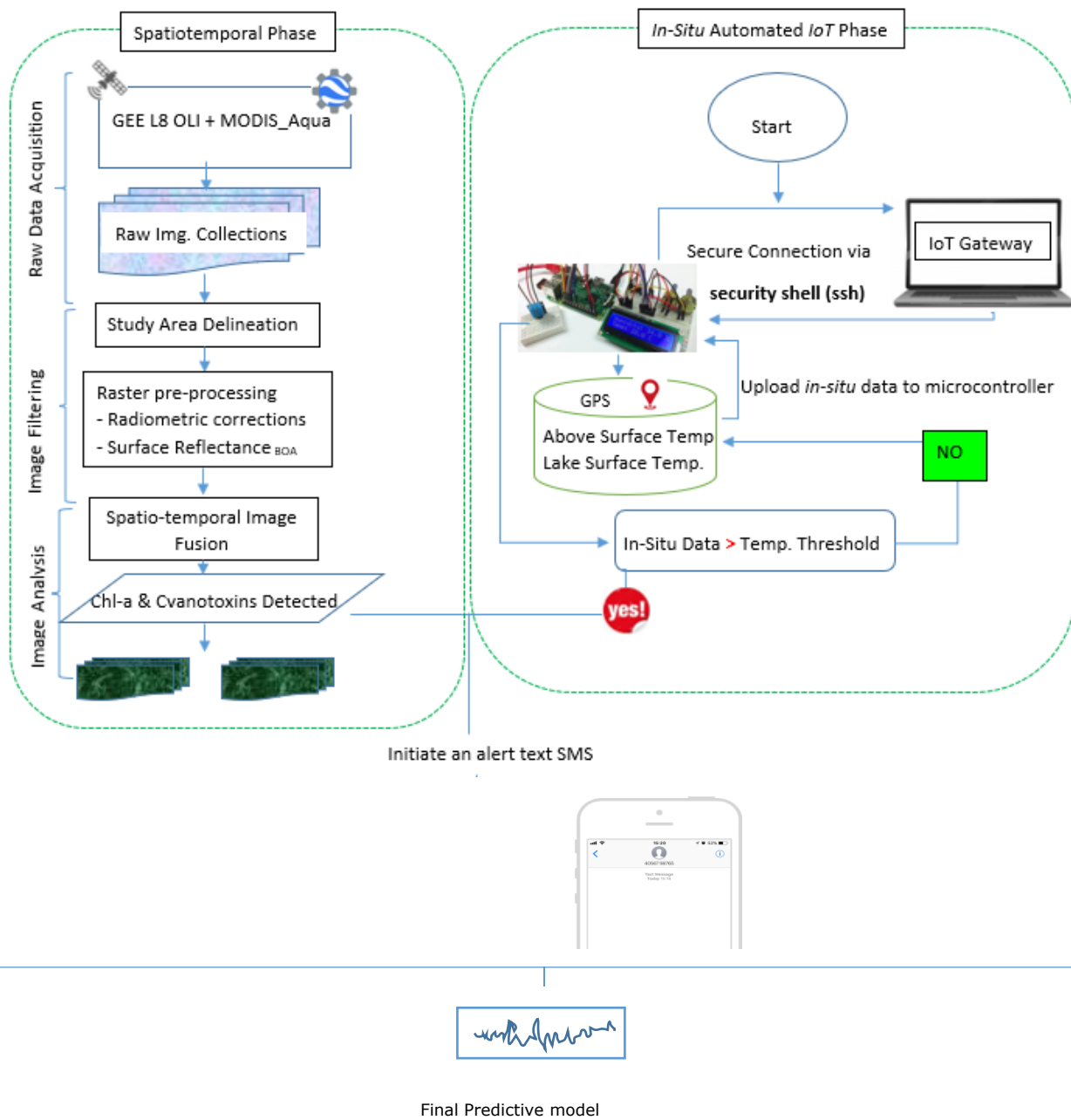
*Fig 8 The Neo-6M GPS Sensor Module.*

Lake Surface Temperature Sensor- DS18B20

b) Above Surface Temperature- DHT Temperature and Humidity Sensor

## Overall Methodology Flowchart

Herein is the overall flow of activities from data acquisition to implementation.



---

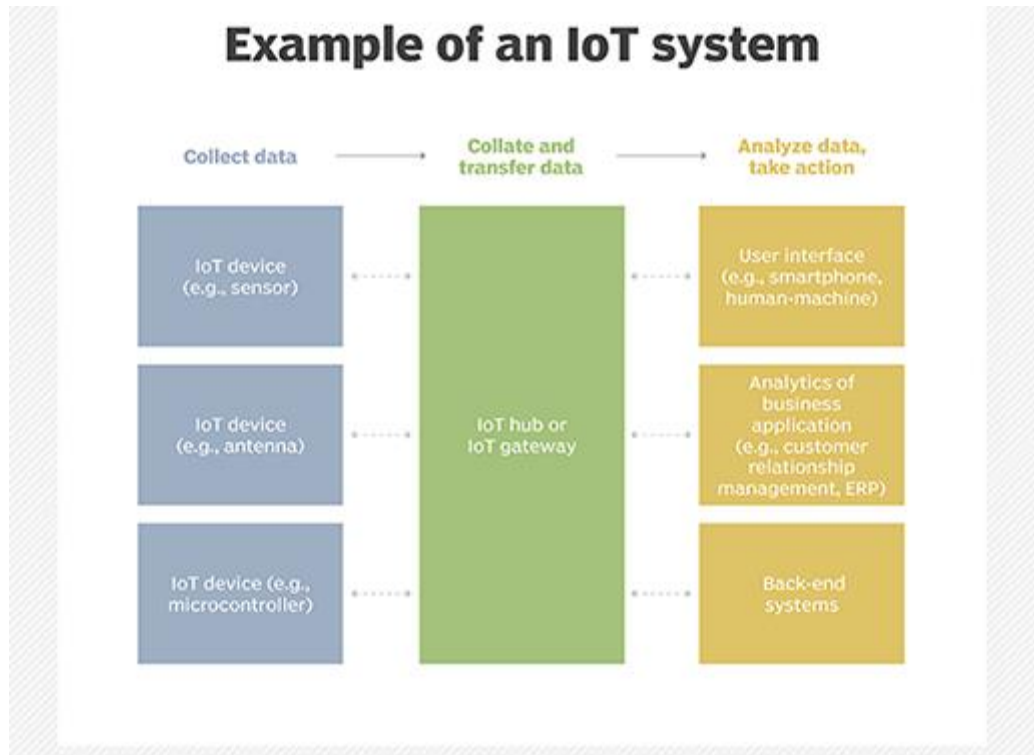
## **Working and Operation of IoT Systems**

An IoT ecosystem integrates an array of smart devices that use embedded systems technology, such as: -

1. In-built processors for example Intel,
2. An array of sensors that collect and send the data from their point of attachments.
3. Communication hardware, that disseminates the collected data.

Upon establish a secure connection with the IoT gateway device, the sensors will be able to share the sensor data they collect where data is either sent to the cloud server- common on IoT systems that run on Raspberry Pi (using POST, GET mechanisms) for analysis or the sensor data can just be locally are analysed. Our IoT devices, system and functionality are automated with little to no human intervention. However exceptionally, I do interact with the system -- for which case, I set up the whole system, code it and give it instructions to collect, package and disseminate the data.

Even though an IoT system can also incorporate Artificial Intelligence (AI), Deep Learning and Machine Learning to boost her data collection and communication processes and further make it more flexible and dynamic, I won't get into this in this project. Detailed information about that might be however availed in later versions of this project.



*Fig 9: How an IoT system works with the Gateway from collecting data to response.*

#### **2.4.1 Importance of IoT concept in this Scope**

Beyond providing ground truth data especially for validation of the remotely Sensed data, the internet of things

1. Helps the geoscientific world to work and generally live smarter at large.
2. On the same note, the IoT system provides the GIS and Remote sensing researcher with complete control over the entire monitoring and to some extent, even prediction.
3. In addition to automating the whole process, IoT essentially comes in by providing a near real-time monitoring of how the entire system operates. By

---

sending the near real time GPS position of the entire system and housekeeping data, this saves us a lot from unknown state of affairs.

4. Consumption of IoT inherently reduces the manual processes involved, thereby cutting down labour costs to be incurred. The amount, in form of money, time and labour that would otherwise be set aside for physical field collection of that Ground Truthing and validation data is saved on.

All said and done, IoT stands out as a few of the chief principal technologies of everyday GIS and Remote sensing research life. It's potential and it will continue to pick up steam as more businesses realize the potential of connected devices to keep them competitive

## **Methods**

In order to successfully achieve the previously highlighted objectives, there's that comprehensive need to employ Spatiotemporal Remote sensing data proposedly from Landsat 8 OLI (relatively fine but rare) and MODIS\_Aqua (course but frequent), automated Internet of Things (IoT) sensors and Machine Learning concepts for short-term probabilistic forecasting.

The project idea intends to aggregate Earth Observation Remote Sensing data from Google Earth Engine cloud platform repository to extract and analyse the presence of Chlorophyll-a pigment. The fact that this aid in the Long-term monitoring of the occurrence of the Harmful Algal Blooms (HABs)

The data repository of GEE is already enriched with several fine resolution satellite image data assets that have global spatial coverage and span several decades of time since 1984. These include the entire datasets collected by Landsat 8 from February 2013 to date and MODIS Aqua from 2000. The overarching idea that Google Earth Engine updates its repository on a daily basis with thousands of new image scenes from current active satellite sensors (Qiusheng W et al., 2020), makes it a near real-time image repository rightly suitable for monitoring of the somewhat frequent HAB occurrences in Lake Victoria.

The repository was also chosen due to the fact that it saves the Geoscientist the Computational anguish of Satellite Image Preprocessing and large size image downloading procedures in other Earth Observational cloud repositories like United States Geological Survey (USGS), Copernicus whatnot. Of importance to note is that GEE has an intrinsically parallel computation capability that divides massive tasks into small ones and utilizes many processors to process them individually and in parallel, hence dramatically speeding up the intensive computation required for large-scale and long-term monitoring applications

---

### Parameters to be collected from dataset for Lake Victoria:

#### 1. Ecological water quality parameters including:

- Chlorophyll-a surface concentration
- Suspended Particulate Matter (SPM)
- Lake Surface Temperature (LST)

#### 2. Sustainability indexes (evolution of land cover - land use (1990 - 2020), evolution of pollution release into the lake due to demographic pressure

#### 3. Time series (Long Short-Term Memory) of meteorological observations.

Besides that, meteorological data (for the last few years together with water quality data supplied by the Lake Victoria Basin Health Monitoring. These data will be used to identify historic occurrences of algal and cyanobacterial blooms in that specified Lake.

The result will be analysed on the fly and a short Early Warning System in the form of a text SMS will be relayed to the authorities concerned.

## **7.Expected Results**

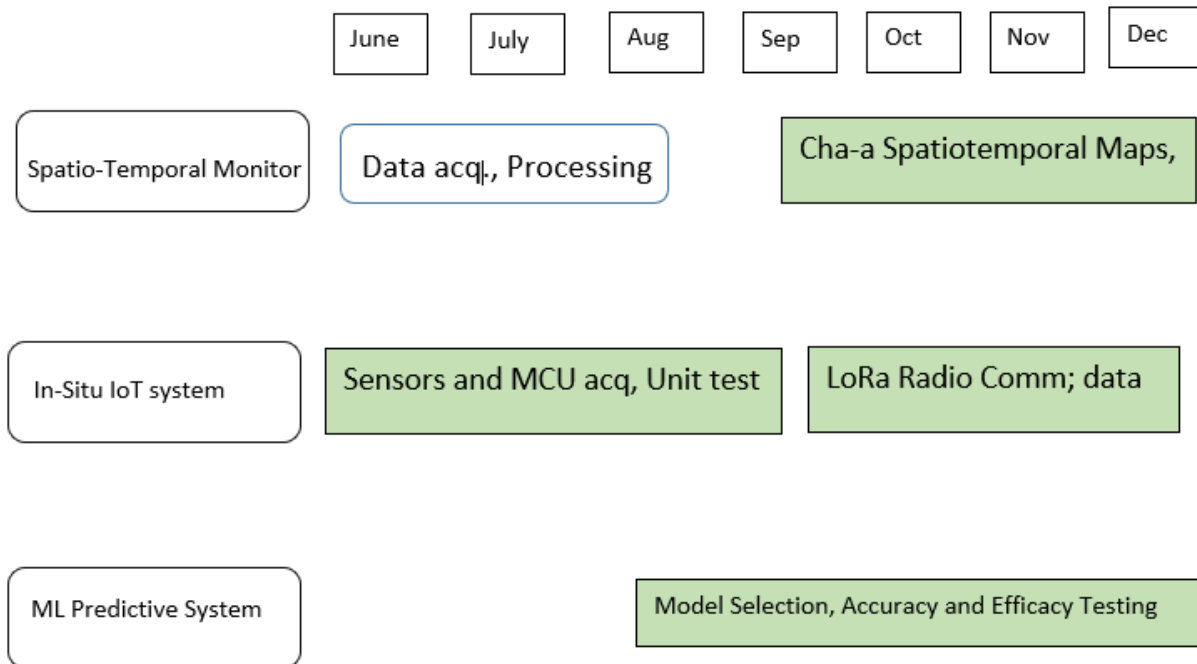
It is expected that upon successful completion of this project, there should be The system will then be usable in near real-time through the IoT utilizing incoming *in-situ* data to provide a short-term probabilistic forecast of the likelihood of a bloom, dependent on the weather conditions on previous and current trends.

- I. Chlorophyll-a Geographical Maps associating the occurrence of the Harmful Algal Blooms and Cyanobacteria.
- II. Automated system that monitors and reports geo-tagged data
- III. Reported confirmation alert Text SMS reporting in near-real time the *in-situ* status from the sensors.
- IV.** Time Series predictive model on any looming bloom.

## **Project Timeline**

The project proposal is intended to run from June 2021 and run through towards the end of the year 2021. This is within the minimum requirements of the Final year project timing of the department of Geomatic Engineering and Geospatial Information Sciences – GEGIS

Appended below is a typical sample Gantt Chart illustrating a rough overview of the project timeline. However, it is right to disclaim that this Gantt chart is highly subject to review and editing but should not however stretch beyond the end of this year 2021.



*Fig 10; The Overall Gantt Chart for Project timeline*

---

## References

Alexander S., IoT Agenda, Retrieved February, 12 2021 from <<https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT>>

Circuit Digest, Arduino versus Raspberry Pi. Retrieved on May, 10 2021 from <https://circuitdigest.com/article/arduino-vs-raspberrypi-difference-between-the-two>

Michael A., *GeoJango Maps, 5 Largest Lakes in The World*. Retrieved on May 16, 2021 <https://geojango.com/blogs/explore-your-world/largest-lakes>

National Library of Medicine, *Detection of bacteria associated with harmful algal blooms from coastal and microcosm environments using electronic microarrays*. Retrieved on May, 15 2021, <https://pubmed.ncbi.nlm.nih.gov/17298369/>

Lei W., Min X., Yang L., Hongxing L., Richard B., Molly R., Erich E., Jade Y. & Qiusheng W. *Mapping Freshwater Chlorophyll-a Concentrations at a Regional Scale Integrating Multi-Sensor Satellite Observations with Google Earth Engine*.

Landsat 8, *Archive Since 2013.*, Retrieved on May 10 2021, <https://eos.com/find-satellite/landsat-8/>

Science Inventory., *Monitoring algal blooms in drinking water reservoirs using the Landsat-8 Operational Land Imager.*, Retrieved on April, 20 2021 <https://eos.com/find-satellite/landsat-8/>

IEEE, Earthzine., *Lake Victoria's Water Quality: Past Present and Future*. Retrieved on May, 20, 2021 <https://earthzine.org/lake-victorias-water-quality-past-present-and-future/>