

# IoT-based Estimation System for *Microcystis aeruginosa* Cyanobacteria in Laguna de Bay using an Arduino-controlled Spectrophotometric Device

Lawrence D. Alejandrino <sup>1</sup>, Jessica Joy D. Jocson <sup>1</sup>, Micah Romina R. Mirarza <sup>1</sup>, Ericson D. Dimaunahan <sup>2</sup>, Ramon G Garcia <sup>1</sup>, Flordeliza L Valiente <sup>1</sup>

<sup>1</sup> School of Electrical, Electronics and Computer Engineering, Mapua University Manila, Philippines

<sup>2</sup> Department of Physics, Mapua University Manila, Philippines

**Abstract.** Laguna de Bay, the largest freshwater lake in the Philippines, provides livelihood to the fishermen and serves as a source of potable water to the locals. However, freshwater quality has degraded, whereas one of the main contributors are Cyanobacteria that produce cyanotoxins. Existing studies that uses a similar device are either too expensive or too bulky. The purpose of this study is to estimate the cyanobacteria concentration by using a low-cost 16-channel spectrophotometric device to determine the level of severity efficiently. Using Linear Regression, the dataset is modelled by the algorithm to estimate the number of cyanobacteria present on the water sample, while Support Vector Machine (SVM) algorithm for severity level classifier. This study achieved high accuracy in estimating the cyanobacteria using linear regression and classifying the level of severity by support vector machine.

## 1 Introduction

Laguna de Bay has been experiencing environmental problems, resulting in the presence of *Microcystis aeruginosa*, also called the blue-green algae. Cyanobacteria is a photosynthetic bacterium that has a blue-green colour. *Microcystis aeruginosa* from the phylum of Cyanobacteria found in Laguna de Bay is a vitally necessary factor to freshwater and marine ecosystems <sup>[1]</sup>. This phylum produces neurotoxins and hepatotoxins, which are fatal to creatures such as wild animals and birds, although, for humans, it can only cause skin irritations, fever-like symptoms, gastroenteritis, and a variety of allergic reactions <sup>[2]</sup>. Albeit the lake can still sustain fisheries, it can be exposed to contamination from pollution.

Fishers and locals are aware of the algal blooms that occur every dry season due to uncontrollable heat. Thus, they prefer to harvest early to avoid the occurrence of a fish kill. Yet, the cyanobacteria are invisible to the human eye, making it a problem for early harvest since algal blooms vary depending on the year <sup>[3]</sup>. Chlorophyll-a is a light-absorbing pigment of the cyanobacteria species, so many researchers and scientists are using spectrophotometers to detect cyanobacteria <sup>[4]</sup>. However, typical commercial spectrophotometers are bulky, power-consuming, and expensive. The portable spectrophotometer price in the market costs up to 500 USD, which is expensive. Also, manual recording, cell count, and analysis of water samples are time and energy-consuming.

A study developed a low-cost visible light spectrophotometer capable of measuring absorbance and transmittance of various materials liquid solution

within the visible range of electromagnetic spectrum <sup>[5]</sup>. Materials used are the following: Arduino Uno, Stepper Motor and Driver, breadboard, op-amp, photodiode, 9V battery, push button, LED, resistors, battery cap, switch, DVD, and test tube with a total of 21.9 USD. Another study built an LED spectrophotometer with a photodiode circuit to determine iron, manganese, and fluoride in drinking water <sup>[6]</sup>. They concluded that there is no practical difference between the LED spectrophotometers and in comparison, to the four commercial spectrophotometers they used as a basis. Multichannel spectrophotometry allows better sensitivity, corrects interferences, and measures multiple components in a solution. <sup>[7]</sup> A 16-channel spectrophotometer will produce a more sensitive output, making it easier to identify disparities that interferences could cause to the test. A study showed how to calculate the concentration of cyanobacteria by using the parameter absorbance <sup>[4]</sup>. This study also showed the correlation of concentration to cell density (cells/mL). Support vector machine is for remote sensing data classification. According to Vapnik, Support Vector Machine is a machine learning method used for small datasets and classification and regression during data analysis. It contains features such as margin maximization and kernel technology that adopted in high-dimensional feature space. High fitting accuracy and a small quantity of tuneable parameters are also some of the valuable properties of SVM <sup>[8]</sup>.

The primary purpose of this study is to develop a device that can estimate the number of cyanobacteria in Laguna de Bay by spectrophotometry and access the data remotely and determine its level of severity through machine learning. Specifically, this research aims to: (1) design a low-cost 16 channel spectrophotometer that measures the intensity of light as a function of

wavelength; (2) implement internet of things (IoT) by interfacing and accessing the ThingSpeak website or mobile application which displays the graphed data; (3) estimate cyanobacteria concentration using linear regression approach; and (4) determine the severity level by training the support vector machine.

This study proposes a fast approach in cyanobacteria estimation since the dataset is viewed and analysed remotely with the aid of the internet. Since not all cyanobacteria contaminated water is visible to the human eye, conducting a study about early detection of *M. aeruginosa* can prohibit them from inhabiting the waters, subsequently damaging the ecosystem and livelihoods dependent on the lake. The local government may establish strict environmental policies on certain parts near the lake detected by the spectrophotometer to be harmful.

This study covers the estimation of cyanobacteria species, *M. aeruginosa*, by spectrophotometry. The three severity levels observed are the minor, moderate, and massive cyanobacteria concentrations. The parameters correlated to these are absorbance and cell density. A stable internet connection is required to transmit the data to the cloud. Testing is best done within the day to preserve the fresh amount of cyanobacteria cells.

## 2 Methodology

### 2.1 Conceptual Framework

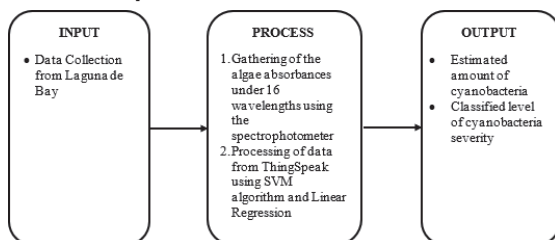


Fig. 1. Conceptual Framework

### 2.2 Prototype Design and Set-up

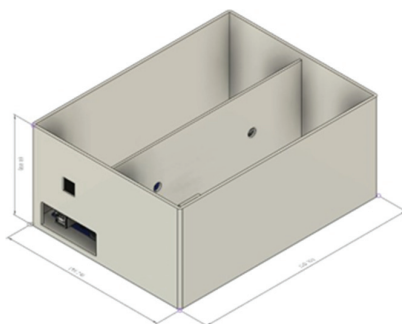


Fig. 2. Prototype Dimensions and Set-up

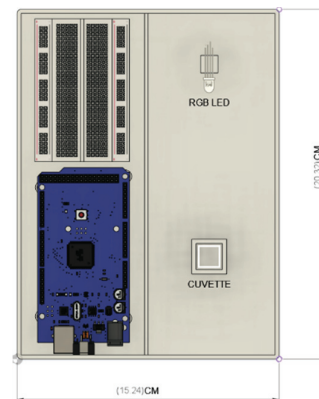


Fig. 3. Prototype Isometric View

#### 2.2.1 Prototype Enclosure

The enclosure for the prototype is made of plastic which is 3D printed to ensure that the components inside are properly positioned. The enclosure is divided into two to isolate the light from the Arduino and the RGB LED. The plastic is black in colour to minimize the reflection of light inside the enclosure that may affect the reading of the sensor.

#### 2.2.2 Microcontroller

An Arduino Mega 2560 is a microcontroller board based on the ATmega2560 (datasheet). This microcontroller will be used to collect the data from the photodiode which is connected to pin A0 of the Arduino.

#### 2.2.3 Light Source

The light source that the researchers will be using is an RGB LED. This is because a single RGB LED can produce a wide array of colors or wavelengths. This can be done by programming the RGB LED using the Arduino. Commercially available RGB LEDs which can have a wavelength range of 380 to 780 nm. Although, wavelengths at the upper end of the UV range (10-400nm) produce results that are accurate and cost-efficient since the absorption of these wavelengths is lower. One piece of RGB LED will be used for the prototype.

#### 2.2.4 Sensor

The sensor to be used is a photodiode. A photodiode reacts to visible and near infrared radiation and it acts like a miniature solar panel that converts the light energy coming from the RGB LED into voltage. In this study, it is used to measure the absorbance of the *M. aeruginosa* obtained from the Laguna de Bay. The sensor will be connected to the analog input of Arduino Mega 2560 and will be placed inside the enclosure together with the light source, cuvette, and other components. BPW34 is the model of the photodiode that will be used in this study.

### 2.2.5 Cuvette

A cuvette is a small container that is similar to test tube, but it has a square cross-sectional area that is sealed on one end. It is important to use a cuvette instead of a test tube since the shape of a test tube is round, causing refraction leading to losses in the light transmitted through the sample. The sample of water from Laguna de Bay that will be tested will be held by a cuvette when being tested in the spectrophotometer.

### 2.2.6 ESP8266 Wi-Fi Module

This Wi-Fi module is attached to the Arduino as it allows the Arduino to transmit data to the cloud using ThingSpeak.

### 2.2.7 ThingSpeak

ThingSpeak is a website which stores and analyses data from IoT devices. It is configured through a code that enables the communication between the cloud and the prototype.



Fig. 4. Sample Output from ThingSpeak

## 2.3 Data Gathering

The samples to be used for testing will be taken from Laguna de Bay at multiple sites during the afternoon, the sample collector must wear proper PPE and to avoid exposure to the cyanotoxins. The water samples collected will be transported in containers placed in a dark cooler at +4°C in order to preserve the quality of cyanobacteria [9]. In case the storage time is longer than 12 hours, quick-freezing of samples to -20°C is suggested [10].

Each sample will be placed in 3.5ml quartz cuvettes and tested in the spectrophotometer. The data obtained from the spectrophotometer is transmitted to ThingSpeak, made available for remote access and for machine learning.

Beer's Law states that the absorbance of light-absorbing material is proportional to its concentration in solution. The intensity of the light passing through the sample cell is also measured for that wavelength - given the symbol, I. If the value of I is less than that of I<sub>0</sub>, then the sample has absorbed some of the light, neglecting reflection of light off the cuvette surface.

Absorbance cannot be measured directly, thus the transmittance of a solution is measured instead. In the

diagram below, the relationship of percent transmittance and absorbance is shown. This shows that if all the light transmitted is absorbed, the percent transmittance is zero and the absorbance is infinite, and when the percent transmittance is 100, the absorbance is infinite.

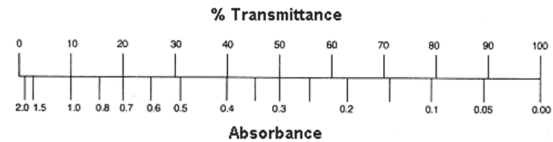


Fig 5. Relationship between % transmittance and absorbance

Transmittance is measured by the fraction of the transmitted light by the incident light. It is obtained by recording the amount of light passing through an empty cuvette while I is obtained by recording the amount of light transmitted through a cuvette with the sample.

$$T = I/I_0 \quad (1)$$

The absorbance is then calculated by the logarithm of the fraction of the incident light by the transmitted light.

$$A = \log(I_0/I) \quad (2)$$

The researchers were able to obtain 108 data observations from the spectrophotometer through the Beer-Lambert Law. These will serve as the dataset for the SVM algorithm.

## 2.4 Spectrophotometer Testing

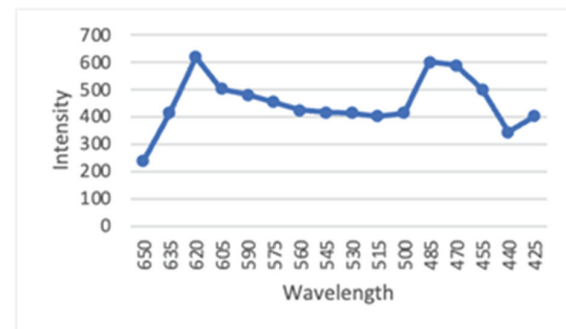


Fig. 6. Average Values of Intensity

In testing the samples that have been gathered from Laguna de Bay, the researchers need to determine the peak value of the intensity and its equivalent wavelength. After conducting some test, the average value for the intensity peaked at wavelength 620 nm. The 620 nm wavelength will be used in determining the absorbances of the cyanobacteria.

## 2.5 Calibration

Calibration is used to compare the results from a much reliable method or device to the results from the prototype. It will help the researchers determine if the prototype needs an adjustment to lessen the error. The dataset will be divided into three classifications which are Minor, Moderate, and Massive. For Minor, the

concentration should be less than 100,000 cells/ml, Moderate ranges from 100,000 to 500,000 cells/ml and Massive should be above 500,000 cells/ml. 3-point calibration will be used in the dataset in which the researchers will get a sample from each of the three categories. Only two of the samples coming from the Moderate and Massive classification will be tested in a laboratory because the sample from Minor will be zero in value.

TABLE I. SAMPLES TESTED IN PROTOTYPE

Absorbance	Concentration	Classification
0	0	Minor
0.0158	191,811 cells/ml	Moderate
0.0484	593,224 cells/ml	Massive

TABLE II. SAMPLES TESTED IN LABORATORY

Absorbance	Concentration	Classification
0	0	Minor
0.0158	191,811 cells/ml	Moderate
0.0484	593,224 cells/ml	Massive

Computing the percentage error will be the difference between the concentration of the samples tested in spectrophotometer (Actual Value) and the samples tested in laboratory divided by the samples tested in laboratory (Expected Value).

$$\%error = \frac{|Actual\ Value - Expected\ Value|}{Expected\ Value} \times 100 \quad (3)$$

For the minor classification, the percentage error would be 0%, moderate classification would be 0.95%, and massive classification would be 0.55%.

## 2.6 Estimation of Cyanobacteria Concentration

There are many ways to determine the concentration of cyanobacteria but one of them using a spectrophotometer. Using a spectrophotometer is not an ideal and an indirect method in determining the concentration of cyanobacteria but it gives you an estimate on what its exact value should be [x]. In spectrophotometry, the relationship between absorbance and concentration is linear. When the value of the absorbance increases, the value of the concentration also increases and vice versa. The formula below is the result of a linear regression that will be used to correlate the absorbance to concentration at 620 nm wavelength for all the samples [4].

$$y = (8.13 \times 10^{-8})x + (2 \times 10^{-4}) \quad (4)$$

## 2.7 Data Analysis

The support vector machine algorithm or SVM is a type of supervised machine learning algorithm. This algorithm is capable of analyzing and processing data for classification and regression analysis. SVM is the algorithm that will be used for the classification of the level of severity of *M. aeruginosa*. In this research the samples will be classified into three classes, which are minor, moderate, and massive. The data points in the 3D plane are the frequencies emitted by the *M. aeruginosa* and it is separated by the hyperplane.

From the data obtained using the spectrophotometer, the dataset is divided into three. 60% of this dataset is allotted for training, 20% for validation, and 20% for testing. SVM will be administered using MATLAB.

### 2.7.1 Validation Process

Figure 5 and 7 shows the scatter plot for the validation process and testing process. Each of these scatter plots are comprised of a random 20% from the dataset, so there are 22 observations plotted for each process. The scatter plot presents the distribution of the data observations' absorbances and their SVM classifiers. These classifiers are shown here in the colors blue, orange and yellow, whereas blue represents massive, orange represents minor, and yellow represents moderate.

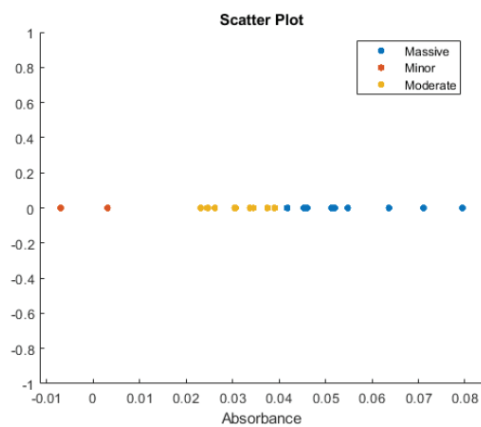


Fig. 7. Scatter Plot for Validation

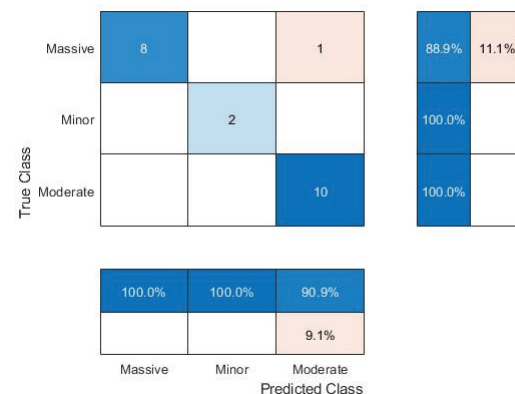


Fig. 8. Confusion Matrix for Validation

Figure 8 displays the classification and misclassification of each level using the trained support vector machine. There is a total of 22 observations for the validation test. 22 (20%) of it are randomly selected for validation test. For the massive concentration, there were a total of 9 data observations wherein 8 (88.9%) are classified correctly by the support vector machine. The remaining 1 (11.1%) is misclassified as either minor or moderate. 2 data observations are minor and all (100%) of it are correctly identified by the support vector machine. The last data observation is a total of 11 moderate concentration. In this observation, 10 (90.9%) is correctly classified and 1 (11.1%) is misclassified by the SVM. In total, the SVM validation accuracy is at 95.24%.

### 2.7.2 Testing Process

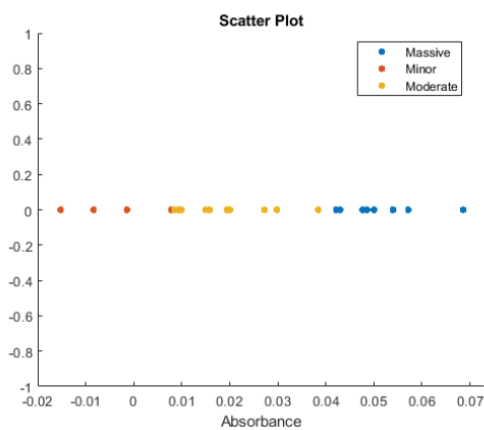


Fig. 9. Scatter Plot for Testing

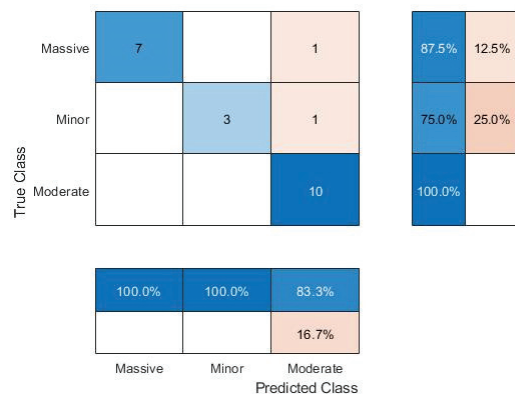


Fig. 10. Confusion Matrix for Testing

To protect against overfitting, validation and testing process are implemented and compared. 20% validation test and 20% actual test is observed to protect against overfitting. Figure Y displays the classification and misclassification by the trained support vector machine. There is a total of 108 observations for the actual test. For the massive concentration, there were a total of 8 data observations and 7 (87.5%) are correctly classified by the support vector machine. This level had

the highest prediction accuracy among the 3 levels. The remaining 1 (12.5%) is misclassified as either minor or moderate.

For the minor concentration, there were 4 data observations and 3 (75%) are classified correctly by the support vector machine. The remaining 1 (25%) of the minor data observation were misclassified as moderate or massive. In the last 12 data observations, 12 are moderate and 10 (83.3%) of it are classified correctly. The other 2 (16.7%) left were misclassified as either massive or minor.

Out of the 22 data observations, 20 were classified as correct. Thus, the SVM test accuracy is 90.91%. Table A shows that the prediction model used is accurate since the validation and test process has a high accuracy.

TABLE III. VALIDATION AND TESTING ACCURACY

Absorbance	Concentration	Classification
Validation	95.24%	4.76%
Testing	90.91%	9.09%

### 3 Conclusion

The researchers were able to develop a low-cost spectrophotometer that can measure the absorbance of cyanobacteria. Wireless communication between the prototype and the cloud was achieved. The linearity between absorbance and concentration was also proven based from the results using Linear Regression and SVM. The researchers also accomplished the classification of the samples into three categories labelled as Minor, Moderate, and Massive using Support-Vector Machine algorithm through MATLAB. Lastly, the estimation of the cyanobacteria concentration using the gathered data was also achieved.

### 4 Recommendation

For further development of the low-cost spectrophotometer, it is recommended that the case design is compatible with an in-situ set-up for real-time testing of scientists and researchers. For this kind of set-up, the suggested power source is a solar-powered battery. Last, it is recommended to look for a more stable sensor than the photodiode.

### References

1. S.H. Kim, Y. He, E.H. Lee, J.H. Kim, *Portable Fluorometer for Cyanobacteria Detection*, in *IEEE Sensors Journal* **17(8)**, 2377-2384 (2017)
2. L. Backer, W. Carmichael, B. Kirkpatrick, C. Williams, M. Irvin, Y. Zhou, Y.S. Cheng, *Recreational exposure to low concentrations of*



- microcystins during an algal bloom in a small lake. *Marine drugs* **6(2)**, 389-406 (2008)
3. C. Rosales, A. Rollon, Factors Affecting the Occurrence and Growth of Toxic Cyanobacteria *Microcystis Aeruginosa* in Laguna De Bay, Philippines. *Journal of Environmental Science and Engineering*, 1109-1116 (2011)
  4. A. Agberien, B. Ormeci, Monitoring of Cyanobacteria in Water Using Spectrophotometry and First Derivative of Absorbance, *12* (2020)
  5. E. Ashfaque, M. Alam, M. R. Islam, I.J. Faria, *Development and validation of a low-cost visible light spectrophotometer*, in 2017 4th International Conference on Advances in Electrical Engineering (2017)
  6. M.W. Prairien S.H. Frisbie, K.K. Rao, A.H. Saksri, S. Parbat, *An accurate, precise, and affordable light emitting diode spectrophotometer for drinking water and other testing with limited resources* in in *PLOS ONE* **15(1)** (2020)
  7. D.L. Witte, *Introduction to Multiple-Wavelength Spectrophotometric Measurements in Analytical and Clinical Chemistry*, in *Contemporary Topics in Analytical and Clinical Chemistry*. Springer, Boston, MA (1982)
  8. S.X. Lu, J. Meng, G.E. Cao, *Support vector machine based on a new reduced samples method*, in 2010 International Conference on Machine Learning and Cybernetics (2010)
  9. J.L. Graham, K.A. Loftin, A.C. Ziegler, M.T. Meyer, Guidelines for design and sampling for cyanobacterial toxin and taste-and-odor studies in lakes and reservoirs: U.S. Geological Survey Scientific Investigations Report 2008, 5038 (2008)
  10. World Health Organization, US Environmental Protection Agency, Commission of the European Communities. *Chapter 10: Cyanobacteria and Algae*, in *Monitoring Bathing Waters - A Practical Guide to the Design and Implementation of Assessments and Monitoring Programmes*, 19 (2000)
  11. W. Hu, Dry Weight and Cell Density of Individual Algal and Cyanobacterial Cells for Algae Research and Development, *7* (2014)