Development of a Raspberry Pi-based Dissolved Oxygen (DO) Meter Hydrological Modelling Using Predictive Algorithms

Engr. Maria Victoria C. Padilla[1], Allysa Y. Dula[2], Aljay R. Lorenzo[3], John David C. Munda [4], Brian Noli G. Supang[5], Neil Aldrin C. Valeroso[6]
Department of Electronics Engineering, Technological University of the Philippines
Manila, Philippines

[1]mariavictoria_padilla@tup.edu.ph [3]aljay.lorenzo@tup.edu.ph

***Abstract*— This study aimed to develop a cost-effective and efficient device that could determine the value of the dissolved oxygen (DO) level through hydrological modelling of other water parameters (temperature, pH, and conductivity) using Decision Tree, Decision forest, and Multi-layer Perceptron machine learning algorithms. Using various metrics, the most efficient model was built using Random Forest algorithm, for it yielded the most reliable metrics when compared to the other two algorithms. The evaluated model has the following metrics: The Coefficient of Determination, or how well a model explains and predicts future outcomes, is 0.99. The Mean Absolute Error, or the average magnitude of the errors in a set of predictions, is 0.32. The Mean Squared Error, utilized in order to measure the performance of an estimator, is 0.36. The Root Mean Squared Error, or how concentrated the data is around the line of best fit, is 0.60. Relative to Atlas Scientific's DO Sensor, the device can predict the dissolved oxygen level of a given water pond with 2.61% error. The final device is a handheld device consisting of the sensors for the highest-ranking parameters with respect to their relationship to DO: temperature (LM35 sensor), conductivity (Atlas Scientific Conductivity K 1.0 Sensor), and pH (DFRobot Analog pH Sensor). The DO meter developed consists of various sensors. The constructed device also amounts to roughly only 24,000PhP. Thus, the device is more cost-effective, while maintaining its expected reliability.**

***Keywords*— water parameters, dissolved oxygen, Python, machine learning, hydrological modelling**

I. Introduction

Water, easily the most ubiquitous resource, has numerous parameters with different implications regarding its state. One of the most important ones is the dissolved oxygen level. Knowing the dissolved oxygen level of a fish-filled body of water is crucial; especially in the case of fish farmers, this would dictate their livelihood.

Dissolved oxygen (DO) is the most critical indicator of a body of water's health and water quality [1]. The amount of DO present in a body of water influences the growth and survival of the aquatic organisms living in it. It is highly relevant to measure the DO level of aquaculture farms to ensure its capacity to support aquatic life. The DO level, however, depends on many factors such as temperature, salinity, oxygen depletion, oxygen source, and others [2].

However, prices for DO meters prove to be very costly, especially to the everyday farmer, reaching prices of 20,000PhP up to half a million Pesos [3].

If one, however, wishes to measure the DO level cost-effectively, the trade-off would be the process being labor-intensive. DO levels are typically and traditionally measured by means of the Wrinkler method [4], where titration is used to account for DO in a given water sample. The process uses a total of five reagents (Sodium thiosulfate, Manganese sulfate,
alkali-iodide-azide, concentrated sulfuric acid, and starch solution. Aside from being labor-intensive, the method evidently requires reagents that the common aquaculture farmer cannot simply obtain.

All these has inspired the researchers to develop a low-cost DO Meter that evaluates the dissolved oxygen level of a certain body of water by developing hydrological models using certain parameters, namely: temperature, pH Level and conductivity, through a comparative study between various machine learning algorithms specifically (a) Decision Tree Regression (DTR); (b) Random Forest (RFR); and (c) Multilayer Perceptron (MLP).

II. Methodology

In conducting this study. the researches achieved their desired goal by a process illustrated in the Fig. 1.

The study mostly relied on the physical construction of a buoy for the data acquisition. From the construction of the buoy, the researchers analyzed the data and came up with the most practical and cost-effective solution for the development of the DO meter. After the development of the final device, it was tested for accuracy.

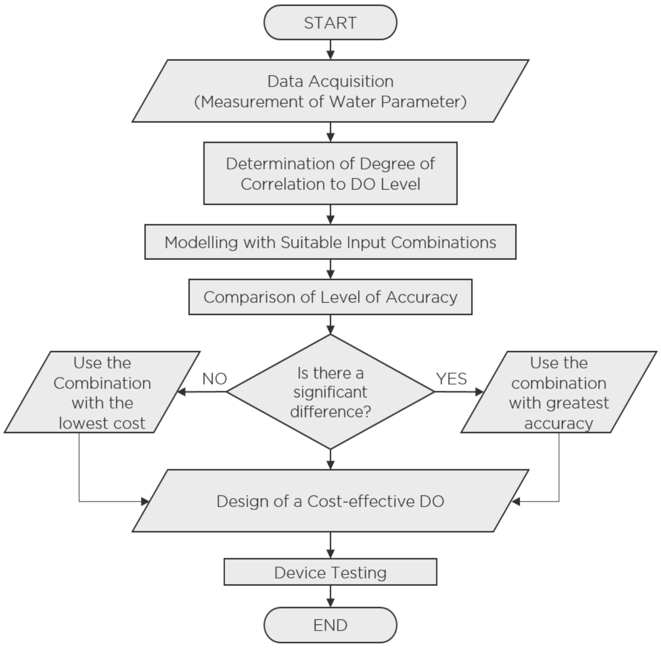

Fig. 1 Flowchart of the study

The preceding figure illustrates the main precodurs to follow for the study. It follows that the presented flow of tasks to be done must be followed to ensure that the targeted objectives would be met.

A. Data Acquisition

The core of this study is the determination of the inherent relationships that exist between the different physical parameters of water. Thus, this entails that the researchers be able to gather ample amount of data to achieve this.

Knowing this, the researchers have searched for various institutions that might support the claims of this study, both in gathering the data needed to realize the objectives of the study, and to validate the concepts and topics that the researchers have gathered to further fortify the foundations of the study. The researchers have interviewed notable personalities and scientists who are inclined and related to the study's needed knowledge.

Fortunately, after an interview with the Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development (PCAARRD), the researchers were led to the Bureau of Fisheries and Aquatic Resources' laboratory in Batangas.

In coordination with the Bureau of Fisheries and Aquatic Resources (BFAR) – Batangas, the researchers gathered three weeks' worth of dataset from a fish-breeding facility situated at Brgy. Ambulong, Batangas. To do so, the researchers have developed a floating device (buoy) equipped with multiple sensors to record various pondwater parameters. The time when these parameters were measured was also taken into consideration.

Using Arduino Mega (ATMega2560) as microcontroller, the buoy recorded measurements from the sensors and saved it in a 4GB micro SD card.

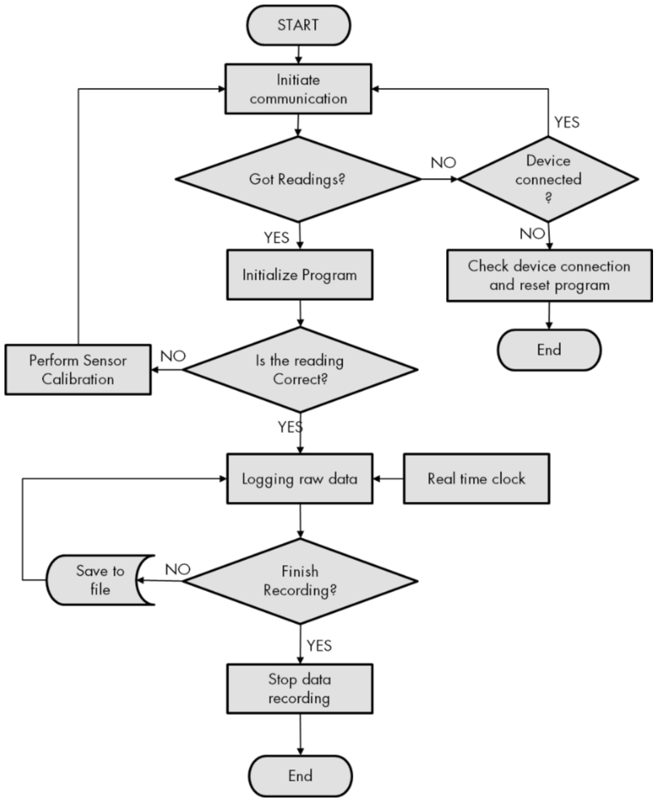

Fig. 2 Sensor node operation flowchart

Fig. 2 illustrates the flowchart of the developed sensor node. Fig. 3 shows the actual buoy; a floating device used by the researchers. Fig. 4 presents a photo of the buoy being deployed in a concreted pond at BFAR - ITSO

Fig. 3 Buoy, floating sensor design

Fig. 4 Buoy gathering parameters reading of a pond

Using Arduino Mega (ATMega2560) as microcontroller, the buoy recorded measurements from the sensors and saved it in a 4GB micro SD card.

B. Application of Various Machine Learning Algorithms to Data

Tested by other relevant literatures, the researchers utilized key algorithms to create a predictive model that can quantify levels of DO using other known physical pondwater parameters. These algorithms are: (1) Random Forest Regression, (2) Decision Tree Regression and (3) Multi-layer Perceptron.

C. Data Resampling and Parameter Selection

The dataset gathered is then filtered using various methods to rid it of errors, fragmented data, garbled sensor inputs, etc. The filtering accepts only complete data, which is only achieved when all sensors record their measured parameters, and non-erroneous values.

The filtered dataset is resampled into a minute, 5-minute, 10-minute, 30-minute and 60-minute intervals. This is done to identify which interpolation would produce the best model, as evaluated.

The sampled datasets were also utilized to quantify how relevant each considered parameter is, to the predictive model built. By using Feature Importance, the relative importance of each parameters to DO level is known.

D. Model Evaluation

Several predictive models were built based on various combination of pondwater parameters. These models were evaluated using the following criterion: Coefficient of Determination (R Square), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Higher level of R Square suggests that there is less error or unexplained variance and therefore, better prediction and more precise DO level.

MAE evaluates how huge errors affects the accuracy of the model built. MSE is kept at minimum to ensure that the predicted DO level is close to the actual DO level. Lower level of RMSE is desired to avoid large errors between the predicted and actual levels of DO.

E. Design of a Cost-Effective DO Meter

The whole process of the meter's development resulted to a handheld device shown in Fig.5.

Fig. 5 Buoy, floating sensor design

III. Experiments And Results

A. Nature of Filtered Dataset

Table 1 and Table 2 illustrate the statistical description of the filtered data set with the following parameters: Dissolved Oxygen (DO) level in mg/L; Electrical Conductivity (EC) in µS/m; Total Dissolved Solids (TDS) in ppm; Salinity (SAL) in psu; Specific Gravity (SG); Turbidity (TURB) in NTU; Temperature (Temp) in °C; pH (PH) Level.

The gathered dataset has a sample size of 1,019,189. The DO level ranges from 3.00 mg/L to 37.4 mg/L. Electrical conductivity has the highest deviation with ~133 µS/m. Total dissolved solids has a mean of ~134 ppm, and ranges from 17ppm to 450ppm. Salinity of the pondwater ranges from 0 psu to 0.4 psu. The specific gravity is constant at 1. The turbidity ranges from -2289.0 NTU to 37.4 NTU. Temperature ranges from 25.6 to 32.2. The pH of pondwater is about 6.59 to 10.8.

B. Evaluation of Parameter Relevance to DO

Using key algorithms, features were selected based on Feature Importance, a method that uses algorithms of Decision Tree Regression (DTR) and Random Forest (RFR).

For most of the evaluation made using various samples from different time intervals, the key parameters which yielded notable levels of coefficients are as follows: time, pH, temp, EC, TDS and SAL.

C. Model Evaluation

Knowing the degree of how each parameter affect DO level, various predictive models are created utilizing RFR, DTR, and MLP, with different combinations of parameters as input. These models are evaluated based on R Square, MAE, MSE, RMSE.

For the ease of discussion, evaluations using dataset from every minute, every 10 minutes and every hour time intervals are shown.

TABLE I
MODEL EVALUATION

Metrics		R Square	MAE	MSE	RMSE
Every Minute	RFR	0.992	0.322	0.360	0.600
	DTR	0.984	0.386	0.715	0.846
	MLP	0.452	3.869	23.864	4.885
Every 10 Minute	RFR	0.984	0.570	0.696	0.834
	DTR	0.967	0.695	1.456	1.207
	MLP	0.196	4.826	34.976	5.914
Every Hour	RFR	0.889	1.629	4.686	2.165
	DTR	0.656	2.424	14.487	3.806
	MLP	-129.419	69.738	5495.280	74.130

It can be inferred from Table 1 that the predictive model built using every minute dataset with top 6 parameters as input, is superior with highest R Square of 0.992, and least MAE, MSE and RMSE of values 0.322, 0.360, and 0.600 respectively. Moreover, in all evaluations, MLP performed inferior having least R square and high levels of error. Based from the table, utilizing RFR created desirable predictive models with high levels of R Square and low levels of MAE, MSE, and RMSE. The highest R Square (0.98401) is obtained by a model built using top 6 parameters as input. Similar observations were seen in every hour, RFR built a desirable predictive model with great levels of R Square and small levels of MAE, MSE, and RMSE. Likewise, this evaluation is obtained by a model built using top 6 parameters as input.

From these simulations, it can be inferred that the top performing models are built from Per Minute Sampling, using Random Forest and Decision Tree algorithms based on the following Top 6 Key Parameters: (1) Time when data was taken, (2) pH, (3) Temperature, (4) Electric Conductivity, (5) Total Dissolved Solids, and (6) Salinity.

With this, the researchers tested how well these models perform in actual setting. These models were uploaded to an R-pi microcontroller for the handheld device developed by the researchers. These models were tested its prediction schemes on the same pond where data was gathered.

Tables 2 and 3 show the parameter measurements during the testing, the predicted DO and the actual DO.

TABLE II
DEVICE EVALUATION – DECISION TREE

DECISION TREE REGRESSION ALGORITHM				
Trial No.	Time	Predicted DO	Actual DO (Atlas Scientific)	Percent Error
1	11:01	17.84	15.95	11.850
2	12:09	18.21	14.84	22.709
3	12:11	18.21	16.67	9.238
4	12:15:50	18.21	16.37	11.240
5	12:18:50	18.21	20.01	8.996
6	12:26:50	18.21	20.34	10.472
7	13:21	18.24	17.85	2.185
8	13:50:31	18.24	17.98	1.446
9	14:42:48	19.5	20.03	2.646
10	14:55:12	15.1	14.89	1.410
			AVERAGE:	**8.219**

TABLE III
DEVICE EVALUATION – RANDOM FOREST

RANDOM FOREST REGRESSION ALGORITHM				
Trial No.	Time	Predicted DO	Actual DO (Atlas Scientific)	Percent Error
1	11:10:31	15.83	16.55	4.350
2	13:36:03	19.29	19.26	0.156
3	14:51:38	19.68	20.98	6.196
4	14:56:42	19.68	19.22	2.393
5	15:12:20	21.24	21.19	0.236
6	15:21:21	21.21	22.36	5.143
7	15:23:20	21.24	21.39	0.701
8	15:25:21	18.5	18.27	1.259
9	15:28:21	21.17	21.97	3.641
10	15:29:47	18.5	18.13	2.041
			AVERAGE:	**2.612**

As can be seen in the Table 2 and 3, the average percent errors between the predicted DO level and measured level are 8.22% and 2.61%, using DTR and RFR algorithms respectively.

From this, it can be inferred that the most effective model uploaded into the device is built using Random Forest Regression algorithm, based on top 6 parameters.

In summary, the metrics of the testing done is shown in the Table 4.

TABLE IV
DEVICE EVALUATION SUMMARY

Metric	RFR	DTR
Coefficient of Determination	0.831	0.263
Mean Absolute Error	0.526	1.396
Mean Squared Error	0.460	2.901
Root Mean Squared Error	0.678	1.703

D. Statistical Analysis

To test whether the difference between these percent errors is significant or not, statistical analysis should be performed. The researchers used equal variance t-test since the sample size between observed values are equal.

The following hypotheses were established:

H_O: There is no significant difference between the percentage errors of predicted and measured DO level, using predictive model built through Random Forest Regression (RFR) algorithm and Decision Tree Regression (DTR)

H_A: There is a significant difference between the percentage errors of predicted and measured DO level, using predictive model built through Random Forest Regression (RFR) algorithm and Decision Tree Regression (DTR).

Table 5 shows the static obtained using t-test

TABLE V
T-TEST STATISTIC COEFFICIENT

Statistic	Coefficient
Mean diff.	5.607
SE	2.210
t value	2.537
df	18
two-tailed p	0.021

By performing Equal Variance t-test, with degrees of freedom *(df) = 18*, the computed two tailed p-value is 0.020638. Since $p<0.05$, the researchers reject the null hypothesis and accepts the alternative hypothesis.

Hence, there is a significant difference between the means of percentage errors of predicted and measured DO level, using predictive model built through Random Forest Regression (RFR) algorithm and Decision Tree Regression (DTR). This suggests that the predictive model built using Random Forest should be adapted.

E. Device Comparison with Atlas Scientific DO Sensor

Table 6 summarizes the comparison between DOSense, the handheld device developed, and Atlas Scientific DO Sensor.

TABLE VI
DOSENSE VS ATLAS SCIENTIFIC DO SENSOR

Device	DOsense	Atlas Scientific DO Sensor with peripherals
Nature	Multiparameter	Dissolved Oxygen
Measuring Range	DO: 3.067 – 37.43 g/mL	DO: 0 -100 g/mL
	pH: 0 – 14	
	Temperature: -50ºC – 125ºC	
	Electrical Conductivity: 5 – 200k uS/cm	
	Total Dissolved Oxygen – ppm	
	Salinity – psu	
Reading Stabilizatio n	1 to 2 minutes	1 minute
Cost	Php 23,867.00	Php 19,000.00

IV. CONCLUSIONS

Considering the study's findings, DOSense is a handheld device consisting of sensors for each of the highest-ranking parameters with respect to their relationship to DO: temperature, pH, and turbidity. In comparison with Atlas Scientific's DO Sensor, DOSense could determine the value of DO with only 2.61% error.

Although DOSense is relatively more expensive than Atlas Scientific DO Sensor with peripherals, the key difference is that DOSense is a multiparameter measuring instrument that is capable of measuring DO level, pH, temperature, electrical conductivity, total dissolved solids and salinity.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development (PCAARRD), and especially to the Bureau of Fisheries and Aquatic Resources (BFAR) of Region IV-A, for sharing both their knowledge and their resources for the fulfillment of this study. The authors also wish to extend their gratitude to the panel all the constructive recommendations that has made the study better. Finally, the authors extend their deepest gratitude to the Almighty God, that in all things, He may be praised.

REFERENCES

[1] Singh, K.P., Basant, A., Malik, A., Jain, G., “Artificial neural network modeling of the river water quality-A case study,” *Ecological Modelling 220*, pp. 888–895, 2009

[2] Wei, W., Changhui D., Xiangjun L., Jun G., “Soft-sensor Software Design of Dissolved Oxygen in Aquaculture,” 2017

[3] D. Mulkerrins, A. D. W. Dobson, and E. Colleran, “Parameters affecting biological phosphate removal from wastewaters,” Environ. Int., vol. 30, no. 2, pp. 249–259, Apr. 2004.

[4] S. W. H. Van Hulle, H. J. P. Vandeweyer, B. D. Meesschaert, P. A. Vanrolleghem, P. Dejans, and A. Dumoulin, “Engineering aspects and practical application of autotrophic nitrogen removal from nitrogen rich streams,” Chem. Eng. J., vol. 162, no. 1, pp. 1–20, Aug. 2010.

[5] Kalff, J., *Limnology: Inland Water Ecosystems*. Prentice-Hall, Upper Saddle River, NJ., 2002

[6] Willemsen, D. (2018, April 3). Lights and Buoys Sailing Issues. [Online]. Available: http://www.sailingissues.com/navcourse9.html

[7] (2018) The Atlas Scientific Website. [Online]. Available: https://www.atlas-scientific.com/_files/_datasheets/_circuit/do_EZO_datasheet.pdf

[8] (2019) The Atlas Scientific Website. [Online]. Available: https://www.atlas-scientific.com/_files/_datasheets/_circuit/ec_EZO_datasheet.pdf

[9] (2017) The DF Robot Website. [Online]. Available: https://www.dfrobot.com/wiki/index.php/Industrial_pH_electrode(SKU:FIT0348)

[10] (2017) The DF Robot Website. [Online]. Available: https://www.dfrobot.com/wiki/index.php/Waterproof_DS18B20_Digital_Temperature_Sensor_(SKU:DFR0198)