A Machine Learning Based Smart Contact-less pH Sensing and Classification

Maryam Saberi, Steven D. Gardner, and Mohammad R. Haider
Department of Electrical and Computer Engineering
The University of Alabama at Birmingham (UAB)
Birmingham, AL, USA
{saberi, stevendg, mrhaider}@uab.edu

***Abstract*—With the ever increasing world population, there is a critical need for healthy food resources. Fish are the most environmentally-friendly animal protein to produce, efficiently converting feed into meat while generating a fraction of the greenhouse gasses of livestock production. Therefore, fish farming is one of the most important fields for a sustainable future. Since there is no way for fishes in fish farming pools to migrate to another water in case of any problem with the water, a key factor in this industry is to maintain the water quality in standard condition. Out of different key factors, pH is an important one to assess the water quality. In this study a portable, cheap, non contact, reusable, and machine learning based pH sensing system is introduced. This helps farmers to make sure about the pH quality of their pools without spending significant amounts of money on measurement equipment. This work introduces a sensitive, non-invasive and reflection-based optical sensor along with an Autoencoder-ESN framework for pH sensing. Using Autoencoder guarantees at least 5 percent better classification in comparison with simple Echo State Networks. Long lifetime of the sensor along with high sensitivity of the machine learning algorithm makes this system valuable for local farmers.**

***Index Terms*—Autoencoder, ESN, pH sensor, water quality**

I. Introduction

Feeding the increasing global population is a daunting challenge that forces us to look for a way to sustain food resources, with seafood being a critical source of animal protein globally. Fish farming is one of the most important food producing methods to improve due to relatively low innovation, minimal effort arrangement ventures, and the long collecting cycles. Water quality is the most imperative factor in fish creation as it influences feed effectiveness, development rates and in general fish health. As the fish are not ready to move or migrate when the water quality turns out to be excessively poor, this is an essential issue for fish farmers to monitor. Temperature, water level, dissolved oxygen, and pH are some of the key factors in water quality.

Different kinds of pH sensors are used in research works such as chemiresistors [1], potentiometric sensors, capacitive sensors, carbon nanotube-based sensors, surface acoustic wave-based sensors, complementary metal-oxide semiconductor (CMOS)-based sensors [2], and optical sensors, with examples such as [3]–[6].

Among the various types of pH sensors, fiber-optic sensors have become popular in recent years due to their minimal size, immunity to electromagnetic interference, and remote sensing capability. These sensors are divided into two categories: (1) Absorption-Based pH Sensors and (2) Fluorescence-Based pH Sensor [2]. In this study, we use a light-weight, contact-less, cheap, and user friendly optical sensor which records the data by measuring the absorbed light by the sample. Despite all developments in optical methodologies, there are still some deficiencies such as high power consumption, low sensitivity, and pH range limitation detection. To overcome these issues, a machine learning (ML) based portable pH sensing system is introduced in this study.

Machine learning is utilized to evaluate the accuracy and sensitivity of the sensor. Broadly put, ML is a set of methods that can be used to learn and detect patterns from input data for further decision making in prognostics or predicting future events. Different ML platforms have been used in literature for spectral classification, such as support vector machine (SVM) [7], Back Propagation (BP) NNs and Genetic Algorithm-Back Propagation (GA-BP) NNs [8], K-nearest neighbor regression(KNN), single layer perceptron (SLP), decision tree (DT) [9], and multi-layer perceptron (MLP) [10]. In this study, SVM is used for classification due to its generalization, low risk of over-fitting and efficient execution time.

For the tests that follow, features from the generated data need to be extracted to ensure promising ML accuracy. To reach this goal, Autoencoders (AE) are used for feature extraction. Chouikhi et al, compares an Autoencoder-Echo State Network (AE-ESN) with a random reservoir layer and Autoencoder-Extreme Learning Machine (AE-ELM) for classification purposes. The results of this study illustrates better classification accuracy by using autoencoder ESN [11]. Training of the AE is based on gradient-based methods such as the back-propagation algorithm which is a first order gradient algorithm for parameter settings and used in neural networks to optimize weight values. To avoid the well-established problem of vanishing and exploding gradients, a non-gradient-based recurrent ESN-RAE is being used [11]. The Echo State Network is a recurrent neural network (RNN) with a reservoir hidden layer. The ESN is considered here as a new recurrent representational space of the original data patterns.

The aim of this study is to introduce a new non-invasive portable pH sensing method, which is affordable for local farms. This method utilizes mixture of various ESN architectures with the AE in an ESN-RAE framework to extract efficient data representations from the original one. Lastly, results are compared with the simple ESN method. The structure of this paper is as follows. The ESN, autoencoder and machime learning methodology are given section II. In section III, sensor and system architecture are introduced. Section IV defines the data acquisition process. Finally, experimental results and evaluation benchmark are introduced in sections V, with a conclusion in section VI.

II. Machine Learning and Classification

1) Autoencoder: The Autoencoder is an unsupervised artificial neural network which learns how to ignore the noise and disturbances in data and find high-dimensional patterns. It is used to pre-processes data before classification. It has two parts, encoding and decoding. The coding part efficiently compresses and encodes data.

After encoding the input data, the decoder layer reconstruct the data back from the reduced encoded representation. Resulting hidden representation h_i is decoded back to the original input space. A standard AE has an input, output and one hidden layer. The dynamics of the AE are computed by following equations:

$$x = f_e(w_i * u + b_e) \quad (1)$$

$$y = f_d(w_o * x + b_d) \quad (2)$$

where b_e and b_d denote bias vectors and f_e and f_d represent activation functions of the encoder and decoder, respectively. u, x and y are the input, hidden, and output states. It is recommended to consider $W_i = W_o^T$ to ensure the autoencoding property is conserved.

2) Echo State Network: The echo state network is a recurrent neural network. Its hidden layer is a random sparse recurrent layer called a reservoir. Output units interpret the reservoir by linearly combining the desired output signal from the rich variety of excited reservoir signals.

To build an ESN model, a random dynamical reservoir is created by utilizing any neuron type that has non-linearity, separating property, and echo state property. Then input and output layer neurons are connected to the reservoir and the training data is applied to the input nodes to drive the dynamic reservoir. In the final step, weights from the reservoir to the output layer are calculated by linear regression. The ESN can then be used for prediction [12].

The ESN has a non-linear behavior and only the weights of output layer are needed to be trained. A simple ESN architecture mainly consists of three parts: input layer (K input units), reservoir (N reservoir units) and output layer (L output units). W^i represent the connection weight matrix of the input layer to the reservoir and W_O is the connection weight matrix of the reservoir to the output layer. $u(n)$, $x(n)$ and $y(n)$ represent the network input at time n, the state vector of network reservoir and the output of network, respectively. The input signal is fed into the ESN and the state generated by the reservoir is collected to form a state matrix. The state vector generated by the network reservoir at the moment can be calculated according to the following equation:

$$x(n+1) = f(w^i * u(n+1) + w * x(n)) \quad (3)$$

Also, the output equation is as follows:

$$y(n+1) = f^o(w^o * x(n+1)) \quad (4)$$

where f and f^o are the activation functions of the reservoir and the output layer, respectively. Once the initial parameters are set, the training can be launched and the reservoir states are gathered in a matrix H. Each column in H correspond to the hidden activation states for one pattern. The output weights are computed according to the following equation, where y_t represents the desired outputs.

$$R = pinv(H) * y_t \quad (5)$$

Having both the desired output and network output helps to compare them to qualify the reservoir. A key factor of ESN is the dynamic reservoir, so should be well chosen. The number of neurons and connectivity rate between neurons in the reservoir are important to the ESN performance. Choosing optimal parameters to design a reservoir is very crucial in order to maximize the networks accuracy while minimizing its complexity.

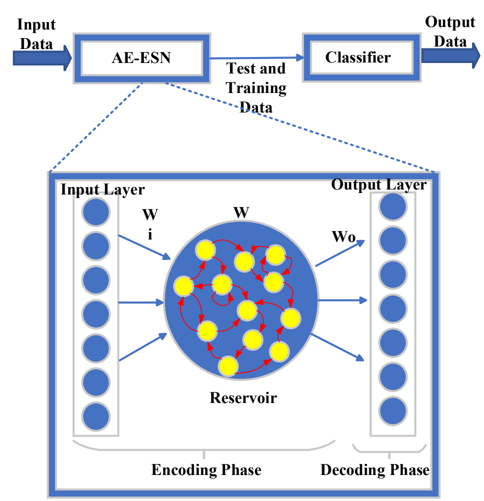

Fig. 1. Block diagram of the Echo state network autoencoder.

3) Autoencoder ESN: Figure 1 shows the main idea of this research, which is the mixture of ESN and Autoencoder with a reservoir hidden layer to achieve more dynamic and richer data from the original one. Recorded data from the spectroscope sensor are pre-processed in MATLAB and fed to the Autoencoder as input signal, and the AE output is considered as the input of ESN reservoir. Finally, corresponding reservoir states are fed to the SVM classifier. Two important factors in reservoir design are reservoir size N and inner connectivity $beta$. Table I demonstrate the values of important factors used in this work.

TABLE I
RESERVOIR PARAMETERS FOR THE ESN

Reservoir Size	Spectral Radius	Activation Function	Train Size	Test Size
300	0.1	tanh	30	30

The equations of the autoencoder-ESN are:

$$x(n+1) = f_e(w_i * u(n+1) + w * x(n) + b_e) \tag{6}$$

$$y(n+1) = f_d(w_o * x(n+1) + b_d) \tag{7}$$

III. EXPERIMENTAL SETUP

In this study, we use an 11-channel spectral sensor (AS7341) and use AS7341 EVAL KIT to evaluate and demonstrate different use cases of the pH sensor. This is a clean, contact-less, reusable and accurate sensor which senses the absorbed light spectral of the sample. Figure 3 shows the sensor.

Setup of this test is shown in Figure 2 and is build in a dark room to avoid the effect of external light contamination on the spectral signatures. As previously mentioned, the sensor has the ability to measure the absorbed light in different wavelengths (410 nm to 670 nm) but after analyzing them separately, the highest frequency band (670 nm) was selected for the rest of processing procedures due to its greater sensitivity to pH changes in the samples.

As it is demonstrated in Figure 2, different segments of the set up are sensor, sample container and an external LED. During the experiment, LED light passes through the sample container and sensor starts recording the absorbed light in different wavelengths. Sample container is a 35 milliliter bright container filled with water and indicators. Different colors of LED including green, blue, red, yellow and white were used as the source light, and after processing all data sets, red LED made more clarified spectral response. Therefore red LED was chosen in the test setup.

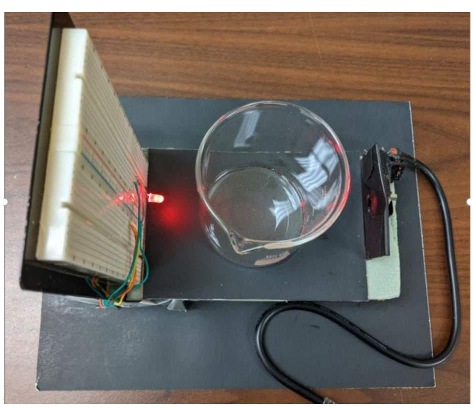

Fig. 2. Experimental setup of pH measurement from liquid samples.

IV. DATA ACQUISITION

To simulate different situations of the water in farms, three indicators with different pH values are chosen and added to the water. By adding 10 micro-liters of orange juice, lemon juice and Bleach in each step to 35 milliliter of water, we will change and discuss the effect of pH on the spectral reflection from the liquid in different wavelengths. The experiment is performed in five steps for each indicator. In each step, 10 micro-liters of indicator is added to the sample and spectral data is recorded for one minute. After recording all datasets, they are separated to two different parts for training and testing. Comparing absorbed spectral of samples with different pH shows that absorbance decreases as the acidic indicator increases. This observation is more obvious in higher wave-lengths, explaining why a higher wavelength is used for the rest of the classifications.

Fig. 3. Photograph of the spectral kit, AS7341 Sensor.

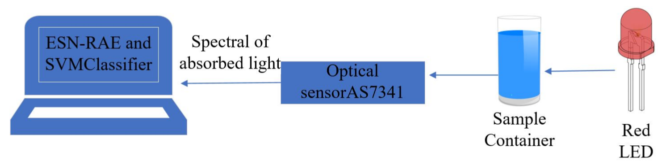

Fig. 4. Schematic of the system

V. RESULTS AND DISCUSSION

As mentioned, the strength of this system in classifying acidic and basic environments, and the ability to classify between different amounts of impurity in sample solution is explored in this study. Figure 5 presents the recorded data

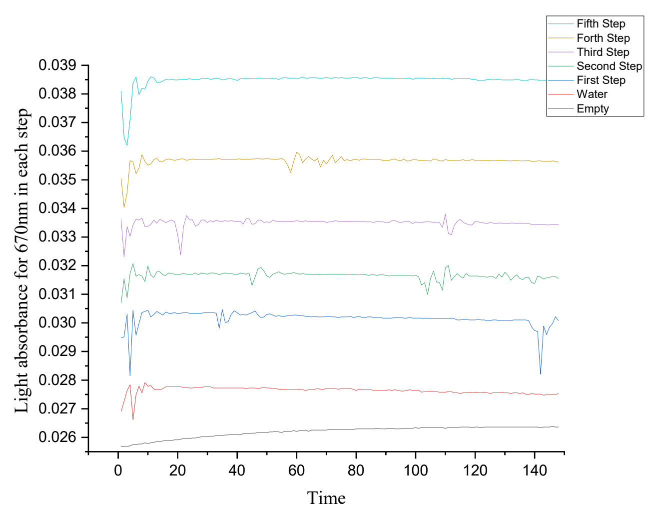

Fig. 5. Spectral output of the sensor for different conditions in 670 nm light.

for lemon juice indicator step-by-step. After pre-processing, recorded data are encoded by the AE, resulting in richer and more various forms in comparison to the original data. The new representation obtained as the output of reservoir activation states and fed into a Support Vector Machine (SVM) classifier addresses the general problem of learning to discriminate between positive and negative members of given n-dimensional vectors. The main idea of SVM classification is to a transform the original input set to a high dimensional feature space by using kernel function. Due to less chance of over fitting, better computational complexity and less execution time, SVM is used as a classification tool in this work. Final output of the classifier is compared to already existing ESN approach without AE on the database. Results of experiments are shown in Table II. As the table shows, adding AE before the reservoir network improves the final classification about 5 percent in comparison with the simple ESN. The Fisher Scientific Test Tube pH Pen and different pH buffer solutions were used to both evaluate sample's pH and made sure that the sensor could detect different ranges of pH.

TABLE II
ACCURACY PERFORMANCE OF ESN AND RAE-ESN BASED MACHINE LEARNING

Indicator	Classification precision of ESN	Classification precision of ESN-RAE
Lemon juice	0.86	0.90-0.95
Orange juice	0.75	0.85-0.95
Bleach	0.80	0.85

To evaluate the authenticity of the data, the data recording process was repeated by using industrial buffer solutions with three different pH values of 4, 7, and 10 (see Fig. 6). Comparing results with those with natural indicators, confirms the trustworthy of our data presentation.

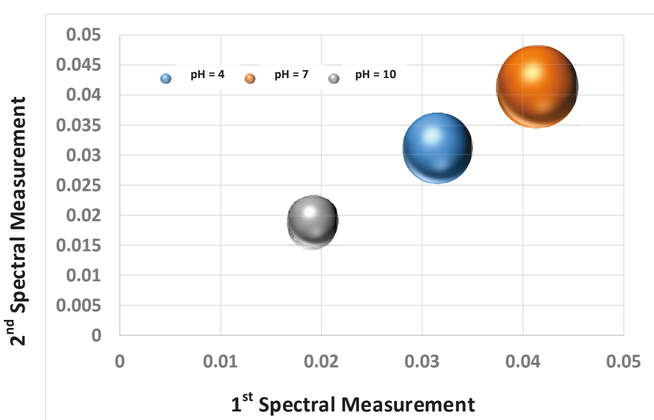

Fig. 6. Spectral response of different pH buffer solutions.

VI. CONCLUSION

In this paper, a smart contact-less pH sensing and classification algorithm is introduced, which works based on machine learning methods. The sensor makes it possible for fish farmers to measure pH value of their pools without entering the water, helping to prolong sensor lifetime and avoid water contamination. The implemented sensor is also portable and affordable, helping them to have healthy pools without spending lots of money. Furthermore, the classification accuracy of the presented method is noticeably higher than many other existed methods. As mentioned, Echo State Network is used as an Autoencoder for data representations. The high non-linearity of ESN and the strength of Autoencoder are combined to provide a new representation of the original data. The new encoded data was then fed into a SVM classifier. Experimental results show that the classification accuracy is improved by 5% compared to the simple ESN used in other works.

VII. ACKNOWLEDGMENT

This research was partially supported by National Science Foundation (Award Nos. ECCS-1813949 and CNS-1645863). However, any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] D. Jeon, W. J. Yoo, J. K. Seo, S. H. Shin, K.-T. Han, S. G. Kim, J.-Y. Park, and B. Lee, "Fiber-optic ph sensor based on sol-gel film immobilized with neutral red," *Optical Review*, vol. 20, no. 2, pp. 209–213, 2013.

[2] M. Cao, S. K. Mahto, and O. Yadid-Pecht, "Real-time optical ph sensor with cmos contact imaging and microfluidics," *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3611–3619, 2016.

[3] M. R. R. Khan, A. V. Watekar, and S.-W. Kang, "Fiber-optic biosensor to detect ph and glucose," *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1528–1538, 2018.

[4] Y. Zheng, L. H. Chen, X. Dong, J. Yang, H. Y. Long, P. L. So, and C. C. Chan, "Miniature ph optical fiber sensor based on fabry–perot interferometer," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, no. 2, pp. 331–335, 2016.

[5] T. H. Nguyen, T. Venugopalan, T. Sun, and K. T. V. Grattan, "Intrinsic fiber optic ph sensor for measurement of ph values in the range of 0.5–6," *IEEE Sensors Journal*, vol. 16, no. 4, pp. 881–887, 2016.

[6] A. Gomez, J. Gutiérrez, and R. Muñoz, "Design of an optical portable system for detection of ph," in *2018 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE)*. IEEE, 2018, Conference Proceedings, pp. 1–6.

[7] M. Hussain, M. Lv, J. Xu, X. Dong, T. Wang, Z. Wang, W. Wang, N. He, Z. Li, and B. Liu, "Rapid identification of pathogens based on mie light scattering and machine learning approach," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 2019, Conference Proceedings, pp. 1–5.

[8] J. Fan, Y. Li, I. Fryc, C. Qian, X. Fan, and G. Zhang, "Machine-learning assisted prediction of spectral power distribution for full-spectrum white light-emitting diode," *IEEE Photonics Journal*, vol. 12, no. 1, pp. 1–18, 2019.

[9] Q. Ye, P. Ye, Z. Guo, X. Dong, and M. Wang, "A corona recognition method based on visible light color and machine learning," *IEEE Transactions on Plasma Science*, vol. 48, no. 1, pp. 31–35, 2019.

[10] B. Batinić, M. Arbanas, J. Bajić, S. Dedijer, V. Rajs, N. Laković, and N. Kulundžić, "Using machine learning for improvement of reflected spectrum estimations of colorimetric probe," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[11] N. Chouikhi, B. Ammar, and A. M. Alimi, "Genesis of basic and multi-layer echo state network recurrent autoencoders for efficient data representations," *arXiv preprint arXiv:1804.08996*, 2018.

[12] O. A. Adeleke, "Echo-state networks for network traffic prediction," in *2019 IEEE 10th annual information technology, electronics and mobile communication conference (IEMCON)*. IEEE, 2019, Conference Proceedings, pp. 0202–0206.