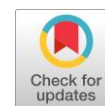


# Trophic state assessment using hybrid classification tree-artificial neural network



Ronnie Sabino Concepcion II <sup>a,1,\*</sup>, Pocholo James Mission Loresco <sup>a,2</sup>, Rhen Anjerome Rañola Bedruz <sup>a,3</sup>, Elmer Pamisa Dadios <sup>a,4</sup>, Sandy Cruz Lauguico <sup>b,5</sup>, Edwin Sybingco <sup>b,6</sup>

<sup>a</sup> Electronics and Communications Engineering Department, De La Salle University, Manila, Philippines

<sup>b</sup> Manufacturing Engineering and Management Department, De La Salle University, Manila, Philippines

<sup>1</sup> [ronnie\\_concepcionii@dlsu.edu.ph](mailto:ronnie_concepcionii@dlsu.edu.ph); <sup>2</sup> [pocholo\\_loresco@dlsu.edu.ph](mailto:pocholo_loresco@dlsu.edu.ph); <sup>3</sup> [sandy\\_lauguico@dlsu.edu.ph](mailto:sandy_lauguico@dlsu.edu.ph); <sup>4</sup> [edwin.sybingco@dlsu.edu.ph](mailto:edwin.sybingco@dlsu.edu.ph);

<sup>5</sup> [rhen\\_bedruz@dlsu.edu.ph](mailto:rhen_bedruz@dlsu.edu.ph); <sup>6</sup> [elmer.dadios@dlsu.edu.ph](mailto:elmer.dadios@dlsu.edu.ph)

\* corresponding author

## ARTICLE INFO

### Article history

Received July 21, 2019

Revised August 11, 2019

Accepted October 11, 2019

Available online March 31, 2020

### Keywords

Aquaponics

Artificial neural network

Modeling tree

Trophic state

Water quality

## ABSTRACT

The trophic state is one of the significant environmental impacts that must be monitored and controlled in any aquatic environment. This phenomenon due to nutrient imbalance in water strengthened with global warming, inhibits the natural system to progress. With eutrophication, the mass of algae in the water surface increases and results to lower dissolved oxygen in the water that is essential for fishes. Numerous limnological and physical features affect the trophic state and thus require extensive analysis to assess it. This paper proposed a model of hybrid classification tree-artificial neural network (CT-ANN) to assess the trophic state based on the selected significant features. The classification tree was used as a multidimensional reduction technique for feature selection, which eliminates eight original features. The remaining predictors having high impacts are chlorophyll-a, phosphorus and Secchi depth. The two-layer ANN with 20 artificial neurons was constructed to assess the trophic state of input features. The neural network was modeled based on the key parameters of learning time, cross-entropy, and regression coefficient. The ANN model used to assess trophic state based on 11 predictors resulted in 81.3% accuracy. The modeled hybrid classification tree-ANN based on 3 predictors resulted to 88.8% accuracy with a cross-entropy performance of 0.096495. Based on the obtained result, the modeled hybrid classification tree-ANN provides higher accuracy in assessing the trophic state of the aquaponic system.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## 1. Introduction

The Aquaponics system is composed of aquaculture and hydroponics subsystems. It is a soilless form of agriculture which recirculates water from fishpond to crop growth chambers and drains back to the fishpond. Each subsystem cultivates certain ideal dry mass, fishes for the prior subsystem, and plants for the latter one. Due to limited land space, aquaponics enables humans to grow crops in water.

Challenges in aquaponics include eutrophication of water. It is a great concern that even the creation of legislation relating to the regulation of the aquaculture industry, particularly in fish diet composition and husbandry, was implemented in some countries to mitigate this phenomenon [1]. Natural and anthropological processes can cause the changing of trophic state. Naturally, when a certain pond, lake, or body of water experiences an abrupt change in temperature and pH, and contamination of excessive dissolved nitrogen, nitrogen, and depletion of oxygen [2][3]. The biotic actions performed by bacteria

result in different nutrient loading distributed on the body of water. Anthropologically, industrial wastes, and household wastes result in water contamination leading to higher concentrations of some specific nutrients. Due to these contributions, the biogeochemistry is altered and occurs to differing trophic state.

In the study conducted by Wang *et al.* [4] in the year 2019, water quality status and eutrophication levels were assessed based on the water quality index (WQI) and trophic level index (TLI). The parametrical basis of TLI measurements are the features of Chlorophyll-a (Chl-a), total phosphorus (TP), total nitrogen (TN), Secchi depth (SD), and permanganate index (COD Mn). The comprehensive trophic state index (TSI) method was used as the basis for a trophic state assessment. Orthophosphate was resolved by limiting the nitrogen level [5]. The modified Carlson's trophic state index (TSIM) was based on water transparency, TP, and Chl-a. The physical and chemical parameters have been concluded as evaluating parameters for the trophic state [6]. Another study conducted by Wang and Qi [7] shows that unwanted natural devastations such as hurricanes and droughts are some of the most contributing factors for eutrophication. The improvement of the crop is met by controlling the fertilizer efficiency, erosion, and runoff in water. These are factors to strengthen the trophic state [8]. The consequences of increasing trophic state, particularly of eutrophication and hypertrophication, are aquatic toxification based on nutrients and harmful algal blooms (HAB). Ammonia is toxic, especially to fishes in high concentrations. Phytoplankton, periphyton, macroalgae, and macrophytes are collectively called autotrophs, which has a direct influence on increasing trophic state. Mutually, eutrophication accelerates the production of these autotrophs [9]. The relationship between climate change and trophic state was studied [10]. Harmful algal bloom productions were considered based on scaling TP, TN, and Chl-a [11]. Life cycle assessment was instrumentalized to analyze and assess the environmental impacts of the aquatic body [12]–[28]. However, the application of modeling using computational intelligence with numbers of input features is relevant to compensate for the increase of error in classification.

This study aims to create a model that provides an accurate assessment of trophic state in the aquaponic environment. Specifically, this study aims to differentiate the performance of two models, namely, artificial neural network (ANN) and the hybrid classification tree-ANN.

Trophic state scientifically expresses the nutrient productivity of the ecosystem, particularly of an aquaculture system. The visible by-products are algae and plankton development. The invisible by-products are the emission of gasses. There are four major categories for the trophic state, namely, oligotrophic, mesotrophic, eutrophic, and hypertrophic. Trophic means foods that resemble nutrients production. Oligo means few, Meso means mild, eutrophic means many foods, and hypertrophic means abundance of foods. In the view of aquaculture, food may be beneficial or harmful in extreme scales for the living species. These categories are one way to characterize ecosystem productivity scientifically.

Robert Carlson introduced a trophic state index to define the total weight of biomass for a definite time during the actual measurement. He emphasized the Carlson trophic state index (TSIC) as an object classifier for algal biomass production in any body of water. The TSIC is recommended to use for the body of water with few rooted plants.

Secchi depth represents the transparency of water based on the scale of log base 2. It specifies the concentration of particulate and dissolves materials in water. Secchi depth is mathematically defined by Eq. 1 where  $z$  is the physical depth at which the disk disappears due to shallow water and irradiation of the natural light source,  $I_0$  is the light intensity prominent on the water surface,  $I_z$  is a constant resulting to 10% of  $I_0$ ,  $k_w$  is the light attenuation coefficient,  $\alpha$  is a constant in terms of square meters per milligram and  $C$  is the particulate matter concentration in water. The relationship among TSIC, Chl-a, TP and SD is attributed in Table 1 with the trophic state as classification.

$$\left(\frac{1}{z}\right) \left(\ln \frac{I_0}{I_z}\right) = k_w + \alpha C \quad (1)$$

**Table 1.** Relationship of TSIC parameters and trophic state

TSI <sub>C</sub>	Chl-a	TP	SD	Trophic State
<30 – 40	0 – 2.6	0 – 12	4 – 8	Oligotrophic
40 – 50	2.6 – 20	12 – 24	2 – 4	Mesotrophic
50 – 70	20 – 56	24 – 96	0.5 – 2	Eutrophic
> 70	> 56	> 96	< 0.5	Hypertrophic

For oligotrophic state, no aquatic vegetation and very minimal nutrient production are expected. The mesotrophic state constitutes minimal production of algae but still on the allowable scale. Eutrophication clearly defines contamination of the body of water due to natural and anthropological contribution that makes nutrient imbalances. Hypertrophication is visible due to algae formation and died fishes that is primarily due to the low level of dissolved oxygen in the water. There are still various limnological and physical factors to be considered in assessing an aquatic system for its trophic state. It includes water temperature, pheophytin, total nitrogen, nitrite, ammonia, orthophosphate, total water alkalinity, and light intensity.

## 2. Method

The system employs two major computational intelligences, namely, decision tree and artificial neural network, to develop a hybrid technique for assessing aquaponic trophic state.

### 2.1. Decision Tree

The decision tree is a computer-based tool (CBT) used to provide decision support that is referred to as a regression tree and classification tree. A regression tree is different from the classification tree as the target variable takes continuous values in numerical figures. The other provides discrete target values. In this study, a classification tree is employed to verify the possible trophic state of certain aquaculture systems. The classification tree predicts discrete responses to data that can be true or false. Its structure was mimicked from the biological structure of plants that are composed of the root node, branches, and leaf nodes. The tree elements are considered nodes. The root node contains the predictor with the highest significance or contribution to the classification. Branches represent the conjunction of two features. The junction of two branches results in a node that signifies another predictor. A subtree is composed of branches and leaf nodes. As the tree grows, many predictors will be constituted to its structure until it ends to its leaf nodes that give the responses based on the combination of predictor data. The leaf node is also called the end node because this is the elemental node that has no children. In the graph theory of the classification tree, the graph is like the upside-down structure of a biological tree. It commences with the root node at the uppermost layer of the graph and terminates with leaf nodes at the bottommost layer. There are six ways to visually represent trees, namely, classical node-link diagram, nested set, layered icicle diagram, outline and tree views, nested parenthesis, and radial trees.

In classification tree, data is analyzed as the function  $(x, Y) = (x_1, x_2, x_3, \dots, x_n, Y)$  in which  $x$  is the column vector of predictors or features,  $x_1$  to  $x_n$  where  $n$  is the total number of columns, and  $Y$  is the response variable. Fig. 1 shows the process exhibited by the classification tree. It starts with classifying the input and output parameters as predictors and response variables, respectively. Then, it trains the selected classifiers and the validation scheme. There are three standard validation schemes for classification tree, which are cross-validation, holdout validation, and no validation. Cross-validation partitions the data space into number of folds or divisions, which gives good predictive accuracy. It is recommended to use a cross-validation scheme for small data space as it may suffer the computational cost of the system. It protects the data from overfitting. Then, the classification tree is structured out, and pruning happens. Pruning is a machine learning technique employed in search algorithms for reducing the size of decision trees. It provides improved accuracy by reducing overfitting through the removal of sections with low

significance to classify a combination of predictors. Overfitting is a statistical error that emphasizes almost complete reliance on training data. The horizon effect is one of the major problems of decision tree that happens when deciding up to what a certain number of levels a decision tree must grow. A small tree may result to unsure accuracy due to few captured sample space while a large tree may tend to overfit and poorly generalize new sample space. Thus, pruning yields a reduced size of a learning tree through two major techniques, namely, reduced error pruning and cost complexity pruning. The reduced error pruning (REP) is classified as bottom-up pruning as it is simple, speedy, and starts the reduction of sections from the leaves. The cost complexity pruning (CCP) is defined by the function  $prune(T,t)$  where  $T$  is the original tree, and  $t$  is the subtree being pruned. The removed subtree in CCP is measured by error rate function  $err(T, S)$ , where  $S$  is the overall data space. The subtree that minimizes the pruning parameter  $P$  as defined by Eq. 2 is chosen for pruning. After pruning is the training of the whole classification learner and is followed by performance measurement. In this phase, the validation predictions and accuracy are computed. Lastly, the observations from raw data space are classified.

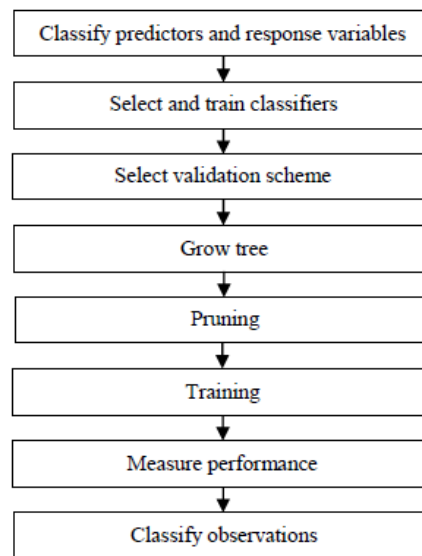


Fig. 1. Process diagram of classification tree

$$P = \frac{err(prune(T,t),S) - err(T,S)}{|leaves(T)| - |leaves(prune(T,t))|} \quad (2)$$

In performing a classification tree, there are three available general metrics to be considered, namely, Gini impurity, information gain, and variance reduction. These are the metric used to determine the best model of the system. Gini impurity determines how often the data from the global data space is labeled incorrectly. It is mathematically defined in Eq. 3 as  $I_G(p)$ ,  $J$  as the number of classes of the set of items where  $i \in \{1, 2, \dots, J\}$  and  $p_i$  is the fraction of labeled items. It multiplies the sum of all the probability of labeled items with the probability of mistakenly categorizing those labeled items.

$$I_G(p) = 1 - \sum_{i=1}^J p_i^2 \quad (3)$$

In this study, the classification model tree was preset to fine tree with the maximum number of splits of 100. The split criterion used is Gini's impurity, and the surrogate decision split is off.

## 2.2. Artificial Neural Network

Artificial neural network (ANN) is otherwise called a connectionist system as it partially mimics the biological transmittal of information and the learning paradigm of the brain. Neural networks have been trained to perform complex functions in various fields, including pattern recognition, identification, classification, speech, vision, and control systems. The ANN is basically made up of artificial neurons operating in parallel. As in nature, the connections between elements largely determine the network

function. One can train a neural network to perform a function by adjusting the values of the connection weights between elements. It is characterized by its network architecture considering the number of layers and number of neurons involved, the node characteristics considering the weights and biases, and the learning rules. The operation of a neural network is divided into two stages, namely, learning or training, and generalization, or recalling. A supervised neural network can learn by the offline or online manner of network training. Technically, training is an algorithmic procedure that repeats several times called epoch. The two categories of algorithmic training are supervised training and unsupervised training. The supervised training exhibits intervention of teaching by the user. The user provides the system with examples as inputs with corresponding outputs. The training process stops when the desired performance is accomplished. The trained system is now deployable for such applications it has been trained. The structure of ANN depicts an input, middle, and output layers that all consists of artificial neurons. Hidden neurons set in the middle layer of the neural network performs intermediate computations or processes. The output neurons are neurons that handle the network outputs.

The backpropagation algorithm is a way of feedbacking error back to the input to increase the accuracy of the system output. As shown in Eq. 4 it is the product of the weighted sum of the inputs  $x_i$  and its respective weight  $w_{ji}$ . Sigmoidal function, as mathematically defined in Eq. 5., is the most common output function that estimates values close to one for large positive real numbers and values close to zero for large negative real numbers. With sigmoidal function, there is a smooth transition from high and low outputs of artificial neurons [29].

$$A_j(x, y) = \sum_{i=0}^n x_i w_{ji} \quad (4)$$

$$O_j(x, \hat{y}) = \frac{1}{1 + e^{-A_j(x, \hat{y})}} \quad (5)$$

In this study, a feedforward backpropagation ANN of two artificial hidden layers were created with sigmoidal function as output function was employed. The mean square error (MSE) was used as the training algorithm in verifying the accuracy of the neural network.

### 2.3. Hybrid Tree-Artificial Neural Network

The hybrid classification tree and artificial neural network are the proposed scheme for assessing aquaponic trophic state as depicted by the system architecture of Fig. 2. The limnological and physical parameters are the input to the system. The limnological parameters are considerable of the water quality parameters that include chlorophyll-a, total phosphorus, water temperature, pheophytin, total nitrogen, nitrite, ammonia, orthophosphate, and total water alkalinity through calcium carbonate. The physical parameters are the Secchi depth and ultraviolet intensity. These eleven trophic state indicators are inputted to the hybrid classification tree and ANN network. The classification tree was used as a multivariable reduction technique that selects significant features that contribute to accurate classification. The ANN was used to intelligently classify the trophic state of an aquatic environment based on the erratic combinations of selected features. The ANN was trained using the pre-classified water quality parameters.

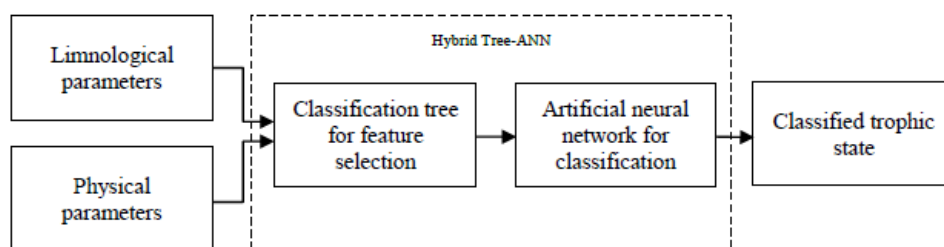


Fig. 2. Proposed system architecture for assessing the trophic state

The process diagram for hybrid classification tree-ANN is shown in Fig. 3. The classification tree involves variable identification, variable classification, and feature selection. The decision tree determines responses by following the flow of the decisions from the top down to splits dependent on the conditions met until finally reaching a leaf node that contains the response. The output of the classification tree is a diagram that shows the feature with the highest impact on the categorical output classification. Thus, instead of using the voluminous feature and data space, the ANN is set only to use selected features with the highest significance. This enhances the computational cost of the system, which includes learning time.

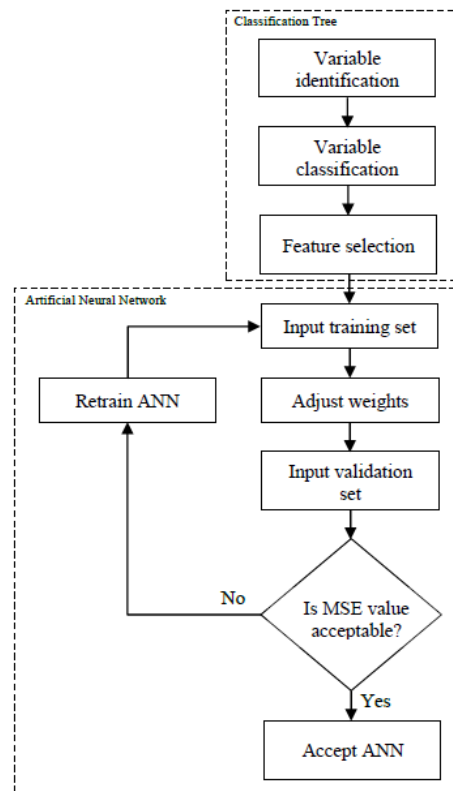


Fig. 3. Process diagram for hybrid classification tree and artificial neural network

By implementing a decision tree, specifically the classification tree, the datasets were dimensionally reduced. The selected features were used for the design of the ANN. There are eleven features used as predictors and the trophic state as the response variables, which is divided into four classes, namely, oligotrophic, mesotrophic, eutrophic, and hypertrophic. Five-fold cross-validation was implemented to protect against overfitting. The data were segmented the data into five equal-sized partitions. The subsequent decision tree diagram in classifying trophic states is shown in Fig. 4. It predicts classifications based on three predictors, phosphorus, chlorophyll-a, and Secchi depth. The classification starts at the top root node represented by a triangle ( $\Delta$ ). The first decision is whether phosphorus is smaller than 23.95. If so, follow the left branch of the root layer, and there is another decision whether chlorophyll-a is smaller than 2.545. If so, track the left branch of that tree layer, and the tree classifies the data as oligotrophic. If, however, chlorophyll-a exceeds 2.545, follow the right branch, and there is the third decision whether Secchi depth is smaller than 1.9. If not, follow the right branch, and the tree classifies the data as eutrophic. If so, it leads to the fourth level of decision that utilizes the chlorophyll-a data again and deciding whether it is smaller than 20.25. If not, follow the right branch, and the tree classifies the data as eutrophic. If so, it leads to the fifth level of decision, which utilizes the total phosphorus data and deciding whether it is less than 12.05. If so, it leads to the left branch, and the tree classifies the data as eutrophic. If not, it leads to the sixth level of decision, which utilizes chlorophyll-a data and deciding whether it is less than 2.645. If so, it leads to the left branch, and the tree classifies the data as

eutrophic. Otherwise, it is mesotrophic. Evident from the classification tree that phosphorus, chlorophyll-a, and Secchi depth are the significant tree features that must be used as input to the ANN.

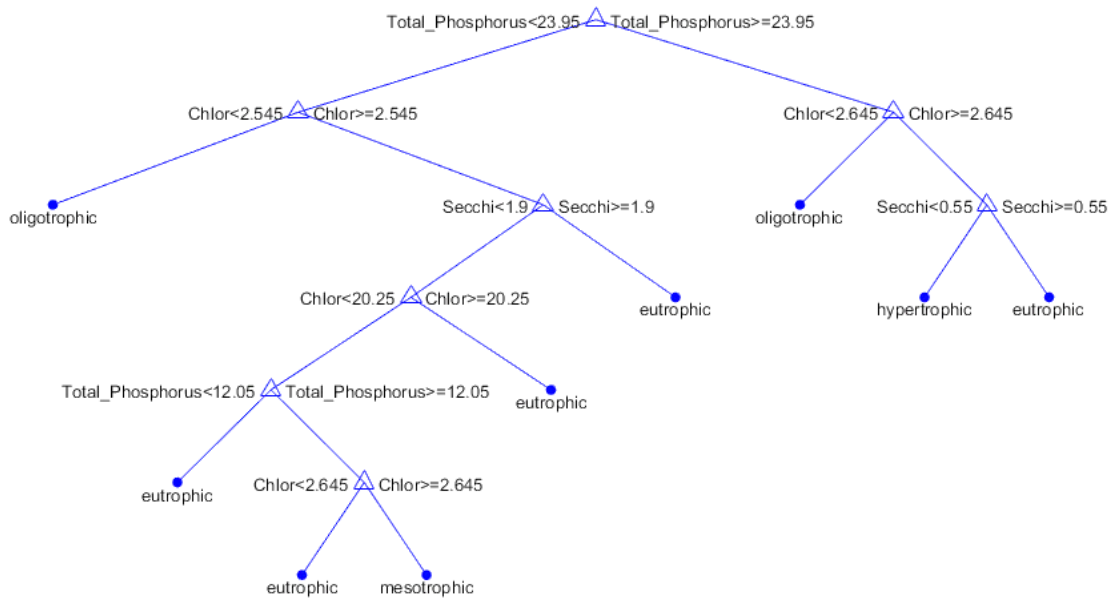


Fig. 4. Tree-classification diagram in classifying trophic states

There are two different ANN systems, as can be seen in Fig. 5 and Fig. 6 that were developed using limnological and physical parameters as input features. One system utilized all eleven features as input to its network. The other system undergone feature reduction first using classification-tree before the application of ANN. Each ANN design was subjected to similar design parameters that included a number of hidden layers, data division, training, performance, and calculation parameters. Performance evaluation of each ANN design was made using validation performance using cross-entropy, gradient and validation checks, confusion matrix, and receiver operating characteristic plot (ROC) plot.

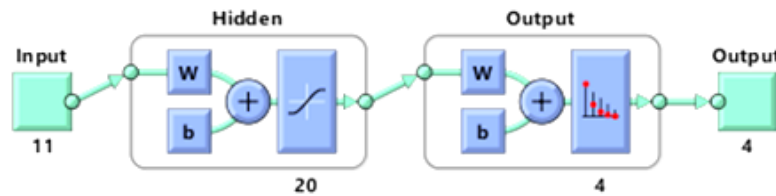


Fig. 5. ANN block diagram with eleven input features

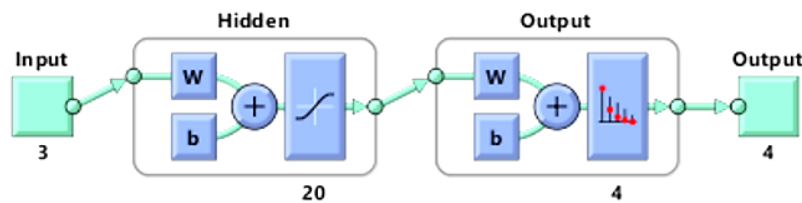


Fig. 6. ANN block diagram with reduced inputs using tree classification

The predictor vectors used in this study are shown in Table 2. The categorical output is the different trophic states: eutrophic, oligotrophic, mesotrophic, and hypertrophic. The developed ANN system is composed of a two-layer feedforward network, with a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer. Two-layer feedforward neural networks can learn any input-output relationship given enough neurons in the hidden layer. Layers that are not considered as output layers are called hidden layers. A two-layer feedforward network with sigmoid hidden-neurons and linear

output neurons can fit multi-dimensional mapping problems arbitrarily well, given consistent data and enough neurons in its hidden layer. The system consisted of a single hidden layer of 20 neurons. This value is selected for high network training performance.

**Table 2.** Input features to ANN

Input	Value Range
Chlor-a ( $\mu\text{g/L}$ )	[0.72, 97.8]
Total Phosphorus ( $\mu\text{g/L}$ )	[7.9, 103]
Secchi (m)	[0.2, 3.3]
Temp ( $^{\circ}\text{C}$ )	[4.5, 27.5]
Pheo-a ( $\mu\text{g/L}$ )	[0.01, 12.8]
Total Nitrogen ( $\mu\text{g/L}$ )	[335, 1530]
NO <sub>2</sub> -3 ( $\mu\text{g/L}$ )	[10, 22]
NH <sub>3</sub> -N ( $\mu\text{g/L}$ )	[2, 89.8]
OPO <sub>4</sub> ( $\mu\text{g/L}$ )	[0.5, 4]
UV <sub>254</sub> (abs.units)	[0.335, 254]
Total Alk (mg/L CaCO <sub>3</sub> )	[3, 13]

The predictor vectors and dependent vector were randomly divided into three sets as follows: 70% were used for training; 15% were used to validate the generalization of the network, and it stops before overfitting. The last 15% was used as a completely independent test of network generalization. Training data are presented to the network during training, and the network is adjusted according to its error. Validation data are used to measure network generalization, and to halt training when generalization stops improving. Testing data have no effect on training and so provides an independent measure of network performance and after training.

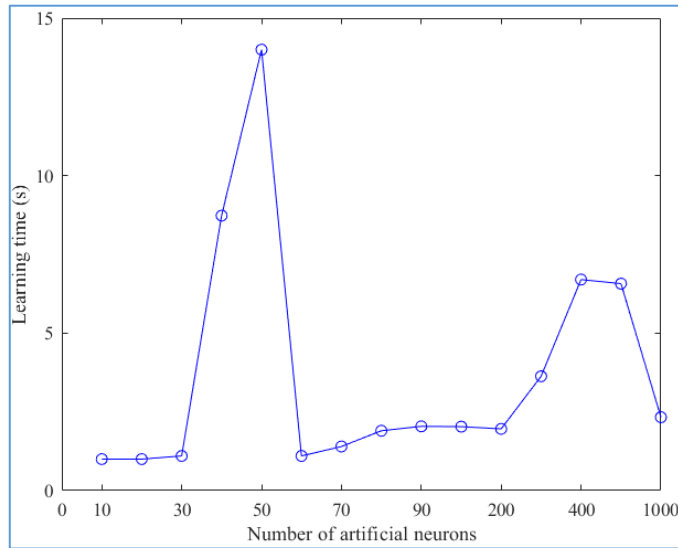
The neural network model was trained using scaled conjugate gradient backpropagation (Table 3). This kind of training uses gradient calculations, which are more memory efficient. The number of hidden neurons is varied from 0 to 1000, as shown in Fig. 7, Fig. 8, and Fig. 9. When the generalization stops improving as indicated by an increase in the cross-entropy error of the validation samples, the network training stops automatically. The key parameters that determine the best neural network are the processing time, cross-entropy (CE) value, and regression coefficient (R).

**Table 3.** ANN design parameters

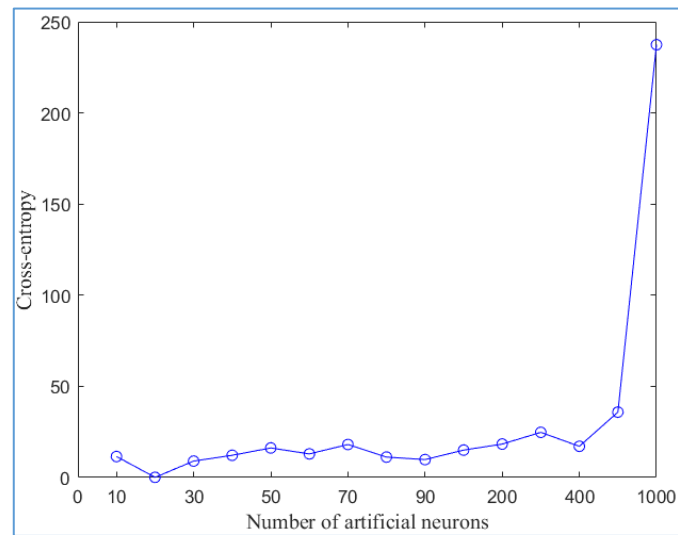
Parameters	Methods
Data Division	Random
Training	Scaled Conjugate Gradient
Performance	Cross-Entropy
Calculation	MEX

By referring to Fig. 7 to Fig. 9, it is noticeable that the lowest learning time and cross-entropy error are obtained from the hidden node size of 20. The highest regression coefficient was obtained at the hidden node size of 400. Fig. 7 depicts a considerable increase in learning or processing time as the number of hidden nodes increases. There are significant increases for hidden nodes of 30 to 60. Fig. 8 depicts a considerable increase of cross-entropy except from a hidden node size of 500 to 1000, which abruptly changed. The ideal cross-entropy is 0, and it provides a good classification of the system performance.

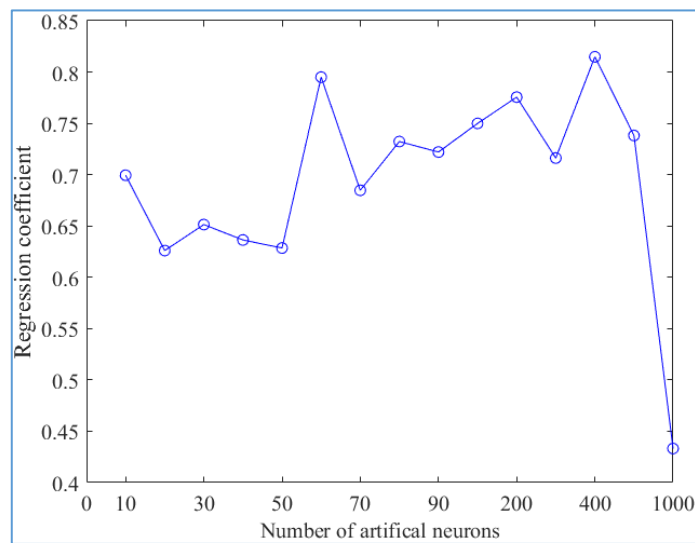




**Fig. 7.** Learning time performance for classification tree-ANN



**Fig. 8.** Cross-entropy performance for classification tree-ANN



**Fig. 9.** The regression coefficient for classification tree-ANN

### 3. Results and Discussion

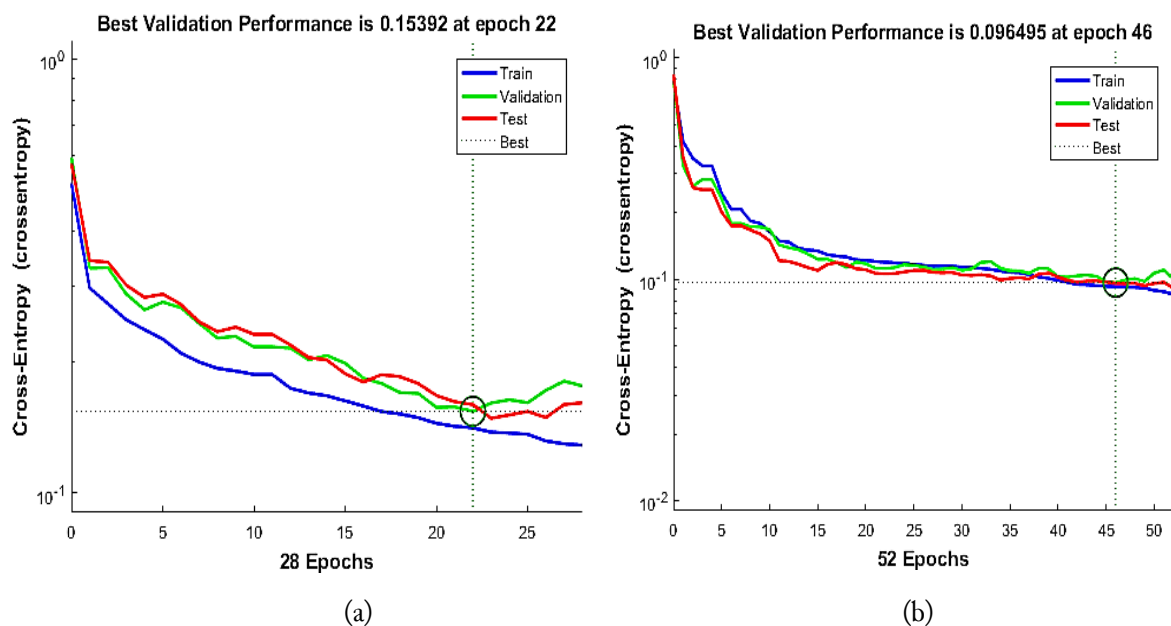
The dual binary digit of output neuron representation was implemented for trophic state classification. The '00' denotes oligotrophic, '01' for mesotrophic, '10' for eutrophic, and '11' for hypertrophic. Table 4 sampled one neuron representation per trophic state denoting actual output and artificial nodes. Through sigmoid and linear functions as estimation rules, the output neurons close to 1 are technically classified as 1.

**Table 4.** Artificial node values for trophic state classification using ANN

Trophic State	Actual Output		Artificial Nodes (N)	
	N1	N2	N1	N2
Oligotrophic	0	0	1.06E-06	1.65E-06
Mesotrophic	0	1	1.37E-07	0.9896
Eutrophic	1	0	0.9996	1.02E-05
Hypertrophic	1	1	0.9889	0.9999

MATLAB neural network fitting tool was used to simulate the expected output of the system. Fig. 10 (a) and Fig. 10 (b) show how each network's performance improved during training. As shown in Fig. 10, each training of the developed neural network stopped for each system when the validation error reached its minimum and started to increase for six iterations (validation checks), which occurred at iteration 22 for Fig. 10 (a), and 46 for Fig. 10 (b), with the best validation performance of 0.15392, and 0.096495 respectively.

The gradient is used in updating the weights and biases during iterations of testing. Fig. 11 shows the changes in gradient value with respect to validation checks. After the sixth validation check, as seen on the red diamond arranged in ladder-like form, the cross-entropy value and mean square error fails to decrease.



**Fig. 10.** Validation performance of (a) ANN (b) hybrid tree-ANN

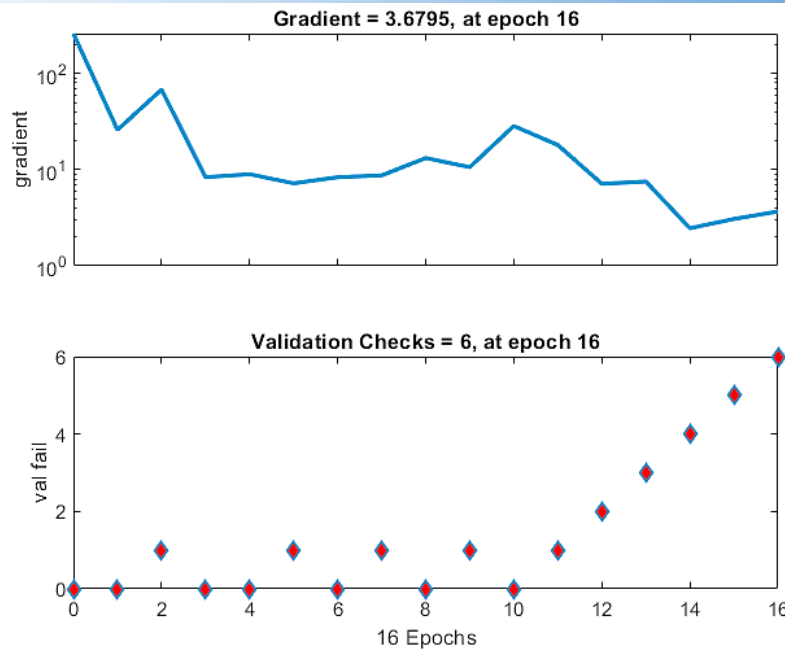


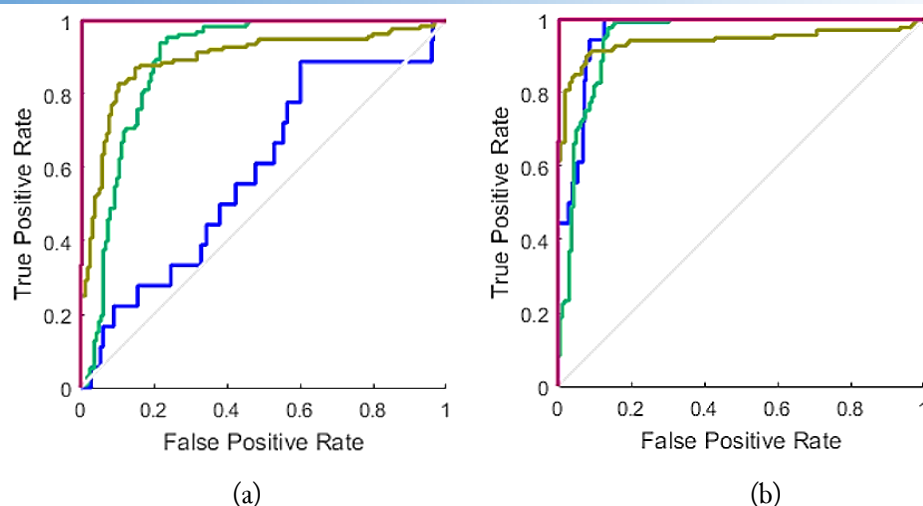
Fig. 11. Training state plot

The confusion matrix depicted in Fig. 12 describes how the developed neural network fits the feature data. It shows the percentages of correct classifications and misclassifications. It also shows how the data were distributed for each class. Correct classifications are presented as green squares on the diagonal section of the matrix. Incorrect classifications form the red squares. Fig. 12 depicts the overall confusion matrix of the developed neural network. Out of 294 attempts to classify target output, 18.7% is wrong for ANN, and 11.2% is wrong for hybrid tree-ANN. Hence, overall system accuracy is 81.3% and 88.8% for ANN and hybrid tree-ANN, respectively.



Fig. 12. Test confusion matrix of (a) ANN (b) hybrid tree-ANN

The ROC plots the true positive rate (TPR) and the false positive rate (FPR) at different classification thresholds, as shown in Fig. 13. The area under the curve (AUC) is the algorithm used to measure the two-dimensional area under the ROC curve for each threshold. Evidently, the ROC curve of the tree-ANN model lies closer to the value of 1 compared to the ANN model alone. It means that the tree-ANN model classifies accurately more than the other tested model.



**Fig. 13.** Test ROC of (a) ANN (b) hybrid tree-ANN: Class1-Oligo, class 2-Meso, class 3-Eutro, class 4-Hyper

**Table 5** summarizing the resulting classification accuracy of two presented methods. The ANN model alone used eleven features, and the hybrid tree-ANN model used the top three highest impact features. There is a difference of 7.5% between the accuracy of the two models.

**Table 5.** Summary of comparison results

Algorithm	Number of Features	Correct Classification	Incorrect Classification
ANN	11	81.3 %	18.7 %
Hybrid Tree-ANN	3	88.8 %	11.2%

#### 4. Conclusion

Assessment of trophic state is essential to maintain nutrient balance in the aquatic system, particularly in smart aquaponics in which nutrients from fishpond have a significant impact not only on aquatic species being cultivated but also on crops being grown. Two algorithms were used to model the trophic state assessment, namely, classification tree and artificial neural network. The classification tree was used as a multidimensional reduction algorithm, and ANN was used for the assessment of the trophic state. There are eleven predictors preliminary used in this study and were reduced to only three. Based on reduction, chlorophyll-a, phosphorus and Secchi depth are the significant predictors which dominantly signifies trophic state into oligotrophic, mesotrophic, eutrophic and hypertrophic. Two models were compared, ANN and hybrid classification tree-ANN, which provides higher accuracy in classification. Future work involves applying algorithms such as fuzzy logic and optimization techniques to model the ideal combinations of impacting nutrients. Consideration of newly discovered trophic state indicators is needed to gain higher accuracy. Adapting wireless sensor networks for effective real-time assessment strategy is to be considered [30].

#### Acknowledgment

The authors would like to thank the Engineering Research and Development for Technology (ERDT) of the Department of Science and Technology (DOST) of the Philippines and the Intelligent Systems Laboratory of the De La Salle University for granted supports.

#### References

- [1] Andrea Sulis, P. Buscarinu, and G. M. Sechi, "Using reservoir trophic-state indexes in optimisation modelling of water-resource systems," *Environ. Model. Softw.*, vol. 26, no. 6, pp. 731-738, 2011, doi: [10.1016/j.envsoft.2011.01.001](https://doi.org/10.1016/j.envsoft.2011.01.001).

- [2] B. Vinçon-Leite and C. Casenave, "Modelling eutrophication in lake ecosystems: A review," *Sci. Total Environ.*, vol. 651, pp. 2985–3001, Feb. 2019, doi: [10.1016/j.scitotenv.2018.09.320](https://doi.org/10.1016/j.scitotenv.2018.09.320).
- [3] J. C. Puno, E. Sybingco, E. Dadios, I. Valenzuela, and J. Cuello, "Determination of soil nutrients and pH level using image processing and artificial neural network," in *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2017, pp. 1–6, doi: [10.1109/HNICEM.2017.8269472](https://doi.org/10.1109/HNICEM.2017.8269472).
- [4] J. Wang, Z. Fu, H. Qiao, and F. Liu, "Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China," *Sci. Total Environ.*, vol. 650, pp. 1392–1402, Feb. 2019, doi: [10.1016/j.scitotenv.2018.09.137](https://doi.org/10.1016/j.scitotenv.2018.09.137).
- [5] X. Peng, S. Kang, L. Zhang, Y. Zhang, D. Hu, and W. Yao, "The Assessment of the Eutrophication Status and Analysis on the Main Controlling Factor of Erlongshan Reservoir," in *2009 International Conference on Environmental Science and Information Application Technology*, 2009, pp. 120–123, doi: [10.1109/ESIAT.2009.351](https://doi.org/10.1109/ESIAT.2009.351).
- [6] Q. Liu, H. Pei, W. Hu, and J. Xie, "Assessment of Trophic Status for Nansi Lake Using Trophic State Index and Phytoplankton Community," in *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, 2010, pp. 1–4, doi: [10.1109/ICBBE.2010.5514833](https://doi.org/10.1109/ICBBE.2010.5514833).
- [7] B. Wang and Q. Qi, "Modeling the lake eutrophication stochastic ecosystem and the research of its stability," *Math. Biosci.*, 2018, doi: [10.1016/j.mbs.2018.03.019](https://doi.org/10.1016/j.mbs.2018.03.019).
- [8] Y. Tao, J. Yu, G. Lei, B. Xue, F. Zhang, and S. Yao, "Indirect influence of eutrophication on air – water exchange fluxes, sinking fluxes, and occurrence of polycyclic aromatic hydrocarbons," *Water Res.*, vol. 122, pp. 512–525, Oct. 2017, doi: [10.1016/j.watres.2017.06.026](https://doi.org/10.1016/j.watres.2017.06.026).
- [9] G. M. Wilkinson, "Eutrophication of Freshwater and Coastal Ecosystems," 2017, pp. 145–152, doi: [10.1016/B978-0-12-409548-9.10160-5](https://doi.org/10.1016/B978-0-12-409548-9.10160-5).
- [10] K. Katsuki, K. Seto, A. Tsujimoto, H. Takata, and T. Sonoda, "Relationship between regional climate change and primary ecosystem characteristics in a lagoon undergoing anthropogenic eutrophication, Lake Mokoto, Japan," *Estuar. Coast. Shelf Sci.*, vol. 222, pp. 205–213, Jun. 2019, doi: [10.1016/j.ecss.2019.04.016](https://doi.org/10.1016/j.ecss.2019.04.016).
- [11] P. M. Glibert, "Eutrophication, harmful algae and biodiversity — Challenging paradigms in a world of complex nutrient changes," *Mar. Pollut. Bull.*, 2017, doi: [10.1016/j.marpolbul.2017.04.027](https://doi.org/10.1016/j.marpolbul.2017.04.027).
- [12] S. Payen and S. F. Ledgard, "Aquatic eutrophication indicators in LCA: Methodological challenges illustrated using a case study in New Zealand," *J. Clean. Prod.*, vol. 168, pp. 1463–1472, Dec. 2017, doi: [10.1016/j.jclepro.2017.09.064](https://doi.org/10.1016/j.jclepro.2017.09.064).
- [13] T.-K. Liu, P. Chen, and H.-Y. Chen, "Comprehensive assessment of coastal eutrophication in Taiwan and its implications for management strategy," *Mar. Pollut. Bull.*, vol. 97, no. 1–2, pp. 440–450, Aug. 2015, doi: [10.1016/j.marpolbul.2015.05.055](https://doi.org/10.1016/j.marpolbul.2015.05.055).
- [14] R. H. Bosma *et al.*, "The financial feasibility of producing fish and vegetables through aquaponics," *Aquac. Eng.*, vol. 78, pp. 146–154, Aug. 2017, doi: [10.1016/j.aquaeng.2017.07.002](https://doi.org/10.1016/j.aquaeng.2017.07.002).
- [15] C. Jaeger, P. Foucard, A. Tocqueville, S. Nahon, and J. Aubin, "Mass balanced based LCA of a common carp-lettuce aquaponics system," *Aquac. Eng.*, vol. 84, pp. 29–41, Feb. 2019, doi: [10.1016/j.aquaeng.2018.11.003](https://doi.org/10.1016/j.aquaeng.2018.11.003).
- [16] R. de C. de S. Schneider, M. de Moura Lima, M. Hoeltz, F. de Farias Neves, D. K. John, and A. de Azevedo, "Life cycle assessment of microalgae production in a raceway pond with alternative culture media," *Algal Res.*, 2018, doi: [10.1016/j.algal.2018.04.012](https://doi.org/10.1016/j.algal.2018.04.012).
- [17] M. El Wali, S. R. Golroudbary, and A. Kraslawski, "Impact of recycling improvement on the life cycle of phosphorus," *Chinese J. Chem. Eng.*, vol. 27, no. 5, pp. 1219–1229, May 2019, doi: [10.1016/j.cjche.2018.09.004](https://doi.org/10.1016/j.cjche.2018.09.004).
- [18] B. Samuel-Fitwi, S. Wuertz, J. P. Schroeder, and C. Schulz, "Sustainability assessment tools to support aquaculture development," *J. Clean. Prod.*, vol. 32, pp. 183–192, Sep. 2012, doi: [10.1016/j.jclepro.2012.03.037](https://doi.org/10.1016/j.jclepro.2012.03.037).

- [19] D. Sanjuan-Delmás *et al.*, “Environmental assessment of an integrated rooftop greenhouse for food production in cities,” *J. Clean. Prod.*, vol. 177, pp. 326–337, Mar. 2018, doi: [10.1016/j.jclepro.2017.12.147](https://doi.org/10.1016/j.jclepro.2017.12.147).
- [20] D. Romeo, E. B. Veà, and M. Thomsen, “Environmental Impacts of Urban Hydroponics in Europe: A Case Study in Lyon,” *Procedia CIRP*, vol. 69, pp. 540–545, 2018, doi: [10.1016/j.procir.2017.11.048](https://doi.org/10.1016/j.procir.2017.11.048).
- [21] H. Wu, L. Gao, Z. Yuan, and S. Wang, “Life cycle assessment of phosphorus use efficiency in crop production system of three crops in Chaohu Watershed, China,” *J. Clean. Prod.*, vol. 139, pp. 1298–1307, Dec. 2016, doi: [10.1016/j.jclepro.2016.08.137](https://doi.org/10.1016/j.jclepro.2016.08.137).
- [22] A. Cohen, S. Malone, Z. Morris, M. Weissburg, and B. Bras, “Combined Fish and Lettuce Cultivation: An Aquaponics Life Cycle Assessment,” *Procedia CIRP*, vol. 69, pp. 551–556, 2018, doi: [10.1016/j.procir.2017.11.029](https://doi.org/10.1016/j.procir.2017.11.029).
- [23] K. Abdou, J. Aubin, M. S. Romdhane, F. Le Loc’h, and F. B. R. Lasram, “Environmental assessment of seabass (*Dicentrarchus labrax*) and seabream (*Sparus aurata*) farming from a life cycle perspective: A case study of a Tunisian aquaculture farm,” *Aquaculture*, vol. 471, pp. 204–212, Mar. 2017, doi: [10.1016/j.aquaculture.2017.01.019](https://doi.org/10.1016/j.aquaculture.2017.01.019).
- [24] A. A. Forchino, H. Lourguioui, D. Brigolin, and R. Pastres, “Aquaponics and sustainability: The comparison of two different aquaponic techniques using the Life Cycle Assessment (LCA),” *Aquac. Eng.*, vol. 77, pp. 80–88, May 2017, doi: [10.1016/j.aquaeng.2017.03.002](https://doi.org/10.1016/j.aquaeng.2017.03.002).
- [25] C. Maucieri *et al.*, “Life cycle assessment of a micro aquaponic system for educational purposes built using recovered material,” *J. Clean. Prod.*, vol. 172, pp. 3119–3127, Jan. 2018, doi: [10.1016/j.jclepro.2017.11.097](https://doi.org/10.1016/j.jclepro.2017.11.097).
- [26] A. A. Forchino, V. Gennotte, S. Maiolo, D. Brigolin, C. Mélard, and R. Pastres, “Eco-designing Aquaponics: A Case Study of an Experimental Production System in Belgium,” *Procedia CIRP*, vol. 69, pp. 546–550, 2018, doi: [10.1016/j.procir.2017.11.064](https://doi.org/10.1016/j.procir.2017.11.064).
- [27] B. König, J. Janker, T. Reinhardt, M. Villarroel, and R. Junge, “Analysis of aquaponics as an emerging technological innovation system,” *J. Clean. Prod.*, vol. 180, pp. 232–243, Apr. 2018, doi: [10.1016/j.jclepro.2018.01.037](https://doi.org/10.1016/j.jclepro.2018.01.037).
- [28] A. Ascianto, E. Schimmenti, C. Cottone, and V. Borsellino, “A financial feasibility study of an aquaponic system in a Mediterranean urban context,” *Urban For. Urban Green.*, vol. 38, pp. 397–402, Feb. 2019, doi: [10.1016/j.ufug.2019.02.001](https://doi.org/10.1016/j.ufug.2019.02.001).
- [29] R. S. Concepcion and L. C. Ilagan, “Application of Hybrid Soft Computing for Classification of Reinforced Concrete Bridge Structural Health Based on Thermal-Vibration Intelligent System Parameters,” in *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)*, 2019, pp. 207–212, doi: [10.1109/CSPA.2019.8696007](https://doi.org/10.1109/CSPA.2019.8696007).
- [30] A. Testa, M. Cinque, A. Coronato, G. De Pietro, and J. C. Augusto, “Heuristic strategies for assessing wireless sensor network resiliency: an event-based formal approach,” *J. Heuristics*, vol. 21, no. 2, pp. 145–175, Apr. 2015, doi: [10.1007/s10732-014-9258-x](https://doi.org/10.1007/s10732-014-9258-x).