Developing an artificial neural network model for predicting the growth of Chlorella sorokiniana in a photobioreactor

J B Magdaong¹,²*, A B Culaba¹,², A T Ubando¹,², J S Chang³ and W H Chen⁴

¹ Department of Mechanical Engineering, De La Salle University, Manila, Philippines
² Center for Engineering and Sustainable Development Research, De La Salle University, Manila, Philippines
³ Department of Chemical Engineering, National Cheng Kung University, Tainan, Taiwan
⁴ Department of Aeronautics and Astronautics, National Cheng Kung University, Tainan, Taiwan

*E-mail: jeremy_magdaong@dlsu.edu.ph

Abstract. Microalgae have been long considered as a potential source of biofuel. Species such as Chlorella sorokiniana can store large amounts of carbohydrates and lipids which can be used to produce biofuels. This paper demonstrates a method for developing an artificial neural network model which can predict C. sorokiniana growth in a photobioreactor. The data used for training the model came from cultivation experiments conducted at the National Cheng Kung University in Taiwan. A feedforward backpropagation ANN model with three inputs (i.e. aeration rate, biomass concentration, and nitrate concentration) and two targets (i.e. biomass concentration and nitrate concentration after 24 hours) was used for this study. Using MATLAB, multiple configurations of this ANN model were created and tested by varying the number of neurons and hidden layers and the training algorithm. Models were initially assessed in terms of their mean square error (MSE) and training performance plots. The models were then further assessed based on their simulation capabilities. After setting the initial biomass and nitrate concentration and aeration profile, the model can already predict the daily biomass and nitrate concentration of C. sorokiniana for the whole cultivation period. The final model selected has one (1) hidden layer and four (4) hidden neurons and it was trained using the Bayesian regularization backpropagation algorithm. For the final selected model, the calculated mean absolute percentage error (MAPE) for the predicted daily biomass and nitrate concentration were all below 7.59% and 3.68% respectively. Thus, the simulation results showed that the final model can accurately predict C. sorokiniana growth at varying aeration profiles. For future studies, this model can be used to determine the aeration profile that can maximize C. sorokiniana growth in a photobioreactor while minimizing aeration costs.

Keywords: Microalgae, Biofuel, Chlorella sorokiniana, Artificial neural network (ANN), Photobioreactor
1. Introduction

The application of microalgae covers several areas from food, wastewater treatment, CO₂ capture, and even biofuel production [1]. However, one of the main challenges in utilizing this resource fully is the high cost involved in cultivating it. Currently, cultivating microalgae especially for biofuel production is not yet economically competitive [2]. There have been several studies which suggest different ways to minimize its cultivating costs such as bioengineering microalgae strains, upscaling cultivation, and improving other related parameters such as aeration and lighting [3,4]. In a related study, it was shown that both parameters had significant effects on the growth of Chlorella sorokinian [5].

The cited study provided insights on how to optimize these parameters, for instance dynamically changing the aeration rate according to the microalgae’s current cultivation state. Doing this can increase the production of microalgae while reducing the aeration requirements during cultivation [6]. One possible way to do this is to monitor the microalgae culture’s current state with sensors and adjust the aeration rate accordingly. However, the downside of this method is the additional cost of acquiring and setting up the necessary equipment. A possible alternative is to develop a model such as an artificial neural network which can predict microalgae growth at different aeration rates.

Artificial neural networks or ANNs work based on the human brain’s ability to learn. Unlike conventional computing, they learn through examples and by detecting patterns and relationships in them [7]. Predetermined growth characteristics of microalgae can be then used to train an ANN model to predict its growth characteristics at different aeration profiles. The advantage of using an ANN model is that it can treat the cultivation process as a ‘black box’ which can simplify the modeling process [8].

Thus, the objective of this study is to develop a robust and accurate ANN model that can predict C. sorokiniana growth in a photobioreactor. The ANN model will be based on the data of a previous study [5]. By setting the initial biomass and nitrate concentration and aeration profile, the model will be able to predict the daily biomass and nitrate concentration of C. sorokiniana for the whole cultivation period. The ANN model created in this study may be then used in future studies to determine the optimal dynamic aeration profile which can maximize C. sorokiniana growth in a photobioreactor while minimizing aeration costs.

2. Methodology

2.1. Preparation of the ANN Training Data

The training data for the artificial neural network model were obtained from the cultivation experiments of C. sorokiniana in NCKU. Figure 1 below represents the data used, namely the daily biomass and nitrate concentration during cultivation [5].

Figure 1. C. sorokiniana’s biomass concentration (a) and nitrate concentration (b) at different aeration rates (2.5 % CO₂ concentration; continuous lighting at 150 µmol/m²-s)
The data points from all the replicates, including their average values, were added to the final training dataset totalling to 162 data points.

2.2. ANN Configuration
A feedforward backpropagation ANN was used in this study. The input variables used were the aeration rate, biomass concentration, and nitrate concentration \((i.e. \ Q_i, \ X_i, \ N_i)\) respectively of the microalgae culture at day \(i\) while the target variables used were the biomass concentration, and nitrate concentration of the same culture after 24 hours \((i.e. \ i+1)\). The outputs from the initial iteration were then used as inputs for the succeeding iteration until the last cultivation day. Figure 2 below shows the general schematic diagram for this ANN model.

![Figure 2](image_url)

**Figure 2.** The ANN model’s general schematic diagram.

The parameters in the ANN model that were varied include the number of neurons and hidden layers and the training algorithm. These variations resulted to 122 ANN configurations. The number of hidden layers was varied between one and two, as these are commonly used for majority of applications of feedforward backpropagation ANN models [9]. The training algorithm was varied between Bayesian regularization backpropagation and Levenberg-Marquardt backpropagation as they are able to achieve lower errors [10]. The number of hidden neurons were based on the number of hidden layers and the maximum amount of hidden neurons was based on [11] to minimize overfitting in the model.

2.3. Creation and Training of the ANN Models
The Neural Network Toolbox in MATLAB R2018a was used to create and train the ANN models based on the configurations mentioned earlier. Ten (10) ANN models per configuration \((i.e. \ a\ total\ of\ 1220\ ANN\ models)\) were created and trained to increase the selection pool of the final ANN model to be used for predicting \(C.\ sorokiniana\’s\ growth\ characteristics.\)

For each cycle during training, the mean squared error \(\text{(MSE)}\) between the predicted and actual experimental data was propagated back to the ANN. The backpropagation training algorithm used for the current ANN training continually adjusted the weights and biases to reduce the MSE until convergence or until the maximum training cycles set was reached.

2.4. Assessment and Selection of the ANN Model
The models were initially assessed in terms of their mean squared error \(\text{(MSE)}\) during training. The ANN model with the lowest MSE was initially selected. Before selecting the final ANN model, the training performance plots of the models were first assessed. The models were then further assessed based on their simulation capabilities. After setting the initial biomass concentration \((0.02 \ \text{g L}^{-1})\), nitrate concentration \((1.5 \ \text{g L}^{-1})\), and aeration profile, the selected model was used to predict the daily biomass
and nitrate concentration of *C. sorokiniana* for the whole cultivation period. Poor simulation results required a new ANN model to be selected.

The final ANN model was then selected heuristically using the rules mentioned in [9, 12]. Training performance plots, simulation results, and the mean absolute percentage errors (MAPE) were also assessed to verify if the selected ANN model can accurately predict the growth characteristics of *C. sorokiniana* without overfitting.

### 3. Results

#### 3.1. Initial ANN Model

The initial ANN model selected has two (2) hidden layers with twelve (12) neurons in each hidden layer. Bayesian regularization backpropagation algorithm was used as the training algorithm for this model. It was initially selected because it had the smallest MSE after testing compared to the other ANN models generated.

The initial model’s MSE during training continued to decrease even when the maximum training cycles set (i.e. default of 1000 cycles) was reached. This was a possible indicator that this ANN model was overfitting to the training data in order to reduce the MSE. Figure 3 below also shows that majority of the errors were near the zero error line, which was another indicator of an over fitted model. An over fitted ANN has poor generalization capability outside the training data.

Figure 4 below shows the simulation results of the ANN model when the aeration varied from 0.05 vvm and 0.1 vvm, which was outside the training data. In this figure, biomass concentration dropped to below zero while nitrate concentration increased. Both scenarios contradict what should happen in an actual cultivation experiment, thus, a better ANN model was required for better generalization.

---

**Figure 3.** Error histogram of the initial ANN model after training.

**Figure 4.** Simulated daily biomass and nitrate concentration at when the initial ANN model was used

#### 3.2. Final ANN Model

A new ANN model was selected heuristically by using different rules for selecting the configuration for an ANN model [9, 12]. From these rules, the ANN model selected only has one hidden layer with four hidden neurons. It was also trained using the Bayesian regularization backpropagation algorithm.

For this ANN model, the MSE decreased as the number of training cycles increased. Training already stopped after 252 cycles since MSE did not go lower anymore after this point. The difference in MSE between the training and testing results was also very small. Figure 5 below also shows that the errors follow more or less a normal distribution pattern. These observations indicate that over fitting unlikely happened for this ANN model.

Figure 6 below shows the actual and simulated biomass and nitrate concentration of *C. sorokiniana* at an aeration rate of 0.0125 vvm. The mean absolute percentage errors (MAPE) of the biomass and
nitrate concentration for the other aeration rates (e.g. 0.025, 0.05, 0.1, and 0.2 vvm) were all calculated to be all below 7.59% and 3.68% respectively which indicates high ANN model accuracy.

A constant aeration rate of 0.005 vvm and an alternating aeration rate between 0.4 vvm and 0.005 vvm were also used for additional simulations to assess how well the selected ANN model can handle aeration profiles outside the training data. Figures 7 and 8 below show that the simulated daily biomass and nitrate concentration of *C. sorokiniana* still followed the normal patterns similar to the actual results seen from Figure 1. In Figure 8, the profiles do not appear as smooth as compared to those in Figure 7. This is possibly due to the aeration rate alternating between high (0.4 vvm) and low (0.005 vvm) levels and *C. sorokiniana* is expected to grow faster at higher aeration rates. It is important to note that this ANN model was able to capture this behavior which further validates its ability to generalize its predictions outside the training data.

4. Conclusion
A method of developing an artificial neural network model for predicting *C. sorokiniana* growth in a photobioreactor in terms of its biomass and nitrate concentration profiles was demonstrated in this study. Results showed that it is possible to create a robust and accurate ANN model using predetermined growth characteristics from previous experiments and studies. Different methods were shown on how to
validate the ANN models including assessing performance plots and simulation results. The mean absolute percentage errors (MAPE) of the simulated daily biomass and nitrate concentration at all aeration rates were all below 7.59% and 3.68% respectively, thus indicating high ANN model accuracy. The ANN model also has a good ability to generalize as it was able to make reasonable predictions about the growth of *C. sorokiniana* even at aeration profiles outside the training data. For future work, this model can be used to determine the optimal aeration profile which can maximize microalgal production in a photobioreactor while minimizing aeration costs.

**Acknowledgments**
The corresponding author would like to inform the readers that this paper is a continuation of his previous work. He would like to thank his co-authors for their contribution and guidance in writing this paper. Lastly, he acknowledges that this study would not be possible without the support of DLSU, DOST – ERDT, NCKU, and CESDR.

**References**

[1] Mata T, Martins A, Caetano N 2010 Microalgae for biodiesel production and other applications: a review *Renewable and Sustainable Energy Reviews* **14**(1) 217–32

[2] Greenwell H, Laurens L, Shields R, Lovitt R, Flynn K 2009 Placing microalgae on the biofuels priority list: a review of the technological challenges *Journal of The Royal Society Interface* **7**(46) 703–26

[3] Chen H, Qiu T, Rong J, He C, Wang Q 2015 Microalgal biofuel revisited: an informatics-based analysis of developments to date and future prospects *Applied Energy* **155** 585–98

[4] Agatonovic-Kustrin S, & Beresford, R 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal Of Pharmaceutical And Biomedical Analysis*, **22**(5) 717-727

[5] Magdaong J , Ubando A, Culaba A, Chang J, Chen W 2019 Effect of aeration rate and light cycle on the growth characteristics of Chlorella sorokiniana in a photobioreactor. *IOP Conference Series: Earth and Environmental Science*, **268** 012112

[6] Malek A, Zullo L, Daoutidis P 2015 Modeling and dynamic optimization of microalgae cultivation in outdoor open ponds *Industrial & Engineering Chemistry Research* **55**(12) 3327-3337

[7] Agatonovic-Kustrin S, Beresford R 2000 Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research *Journal Of Pharmaceutical And Biomedical Analysis*, **22**(5) 717-727

[8] del Rio-Chanona E, Fiorelli F, Zhang D, Ahmed N, Jing K, Shah N 2017 An efficient model construction strategy to simulate microalgal lutein photo-production dynamic process *Biotechnology And Bioengineering* **114**(11) 2518-2527

[9] Alsmaidi M, Omar K, Mohd Noah S 2009 Back propagation algorithm: the best algorithm among the multi-layer perceptron algorithm *International Journal Of Computer Science And Network Security* **9**(4) 378-383

[10] Kayri M 2016 Predictive abilities of Bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: a comparative empirical study on social data *Mathematical And Computational Applications*, **21**(2), 20

[11] García-Camacho F, López-Rosales L, Sánchez-Mirón A, Belarbi E, Chisti Y, Molina-Grima E 2016 Artificial neural network modeling for predicting the growth of the microalga Karlodinium veneficum *Algal Research* **14** 58-64

[12] Heaton J 2008 *Introduction to neural networks with Java*. Chesterfield (MO, USA): Heaton Research