Malaysian Journal of Chemical Engineering & Technology 5(2) (2022) 77–83

**Malaysian Journal of Chemical Engineering & Technology**

e-ISSN: 2682-8588

*Journal Homepage: http://myjms.mohe.gov.my/index.php/mjcet*

# Optimisation of neural network topology for predicting moisture content of spray dried coconut milk powder

[a] Nadiah Syafiqah Shaharuddin, [a] Zalizawati Abdullah*, [b] Farah Saleena Taip

[a]School of Chemical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[b]Department of Process and Food Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Malaysia

*Corresponding email: zalizawati8653@uitm.edu.my

**Abstract**

Moisture content is an important parameter to be controlled in an agricultural product to prevent rapid degradation and promote stabilisation of product and longer shelf life. There are various techniques on predicting the moisture content of spray dried powder that had been used by past researchers. In this study artificial neural network (ANN) is proposed to be used for its well-known benefits of simplicity and accuracy. The aim of this research is to evaluate the effect of hidden layer and hidden neuron in ANN in predicting moisture content of spray dried coconut milk. The effect of training algorithm, e.g., Gradient Descent (GD) back propagation and Levenberg-Marquart (LM) back propagation, and activation functions, e.g., hyperbolic tangent sigmoid (tansig) and log sigmoid (logsig) functions are also studied. Based on the result of correlation coefficient of determination ($R^2$) value of 0.9951 and root mean square error (RMSE) value of 0.0145 that was used to evaluate the ANN performance, it can be concluded that the best ANN topology is 2-10-1 with Levenberg-Marquart for learning algorithm, and tangent sigmoid as activation function.

**Article Info**

*https://doi.org/10.24191/mjcet.v5i2.15444*

*Article history:*

Received date: 15 Sept 2021
Accepted date: 28 June 2022
Published date: 31 October 2022

*Keywords:*

Moisture content
Neural network
Spray drying process
Coconut milk
Network topology

## 1.0 Introduction

Coconut milk, commonly known as 'santan' in Malaysia, is a white milky liquid squeezed from coconut flesh. In the absence of water, during the extraction process, it is a protein-water-oil emulsion with an oil content of around 35% (Seow & Gwee, 1997). Coconut milk cannot be stored for long periods of time due to low stability, especially at room temperature (Abdullah et al., 2018).

Preservation is crucial as it prevents food spoilage caused mainly by microorganisms and enzyme activities. With a low or no water content presence, the agricultural products will become inhabitable for the microorganism, thus, inhibit their activity and growth and help preserve food for a longer time (Bhat & Paliyath, 2012). Spray drying has been widely used as a preservation method to improve coconut milk shelf life in the commercial sector. Coconut milk powder has not only become more stable, and it also decreases the manufacturing cost.

Spray drying is an advanced particle drying method that is widely utilised in various industries, including food, pharmaceuticals, and chemicals. Rapid drying of liquid or slurry feed with hot gas converts the feed to dry powder. On top of that, spray drying is applied in the food industry since it improves food preservation and ease storage, handling, and transportation.

Artificial intelligence (AI) technique which is an artificial neural network (ANN) has been widely used for variety of problems. Neural network is a form of a multiprocessor computer system with simple processing elements, a high degree of interconnection, simple scalar messages, and has adaptive interaction between elements. A properly trained ANN model can accurately predict outputs for new input data sets which leads to the wide application of this approach in maximising yield and drying process identification.

Although there is an increasing trend of utilising ANN, the topology for every process differs from one another. Optimising an excellent topology and selecting the right parameter becomes an important task since it affects the smoothness of approximation and accuracy of the neural network (Ukaoha & Igodan, 2019).

The number of hidden layers and the number of neurons in the hidden layer determines the accuracy of the model. Incorrect selection of hidden layers might lead to overfitting or underfitting condition (Uzair & Jamil, 2020). Onu et al. (2020) reported that five neurons in one hidden layer in neural network model was optimum in predicting the moisture content of the dried potato. On the other hand, Sodeifian et al., (2016) found out that six neurons in one hidden layer was optimum in training the neural network model. The developed neural network model was used to predict the extraction yield of Ferulago angulate. Higher number of neurons in the first hidden layer was optimised by Taylan & Haydar (2004). They used two hidden layers in their network topology, and they found out that a neural network with seven neurons in the first hidden layer and three neurons in the second hidden layer provided highest accuracy in predicting the moisture content of porcelain powder.

Proper selection of the training algorithm is needed as it determines the prediction and generalisation ability of the model (Wilamowski, 2013). Perazzini et al. (2020) used the Levenberg-Marquardt algorithm to develop a neural network model to predict the moisture content of alumina particles in moving beds dryer. Torrecilla et al. (2005) found out that the application of gradient descent algorithm as the learning algorithm leads to a highly accurate neural network model in predicting the moisture content of alperujo from a fluidized bed dryer. On the other hand, they also found out that Polak-Ribiere conjugate gradient backpropagation was the best algorithm in training a neural network to predict the moisture content of dried olive oil mill waste (Torrecilla, et al., 2008). The selection of the training algorithm is a trade-off between the fast convergence and the high accuracy of the developed model.

In term of activation function, the selection of this function gives significant impact to the network response as it defines the prediction made by the network (Brownlee, 2019). This indicates that the selection of network parameters i.e., number of hidden layers, number of neurons, activation function and training algorithm will depend on the process itself.

In this study, the ANN technique is used to predict the moisture content of spray-dried coconut milk by optimising the topology of the network. In order to optimise the best topology for ANN, a significant number of different configurations of neural network (NN) needs to be constructed through testing and learning. Thus, this study is focusing on choosing the best number of hidden layers and neurons. The models are also optimised based on the training algorithm i.e., Levenberg-Marquardt backpropagation training algorithm (LMBP) and gradient descent back propagation (GDBP), and the activation function i.e., hyperbolic tangent sigmoid (tansig) and log sigmoid (logsig).

## 2.0 Methodology

### *2.1 Input-output data of neural network model*

The input-output data to develop the nonlinear model was obtained from the time series data of previous research (Abdullah, et al. 2020). The 826 time-series datasets were used to formulate the neural network model. Two variables were selected as the input to the model, namely input temperature ($T_i$), and output temperature ($T_{out}$). The moisture content (MC) of coconut milk powder was selected as the output of the model.

### *2.2 Data normalisation*

Data normalisation or also known as feature scaling, was an essential step and a best practice for neural network training before splitting the data set (training and test data) from existing dynamic data. This step was done to remove the difference between the data set values by scaling it within the range of 0 to 1. Scaling of values also enhanced the accuracy of subsequent numeric features of model for a better performance. The min – max normalisation where the scale value is in the range of [0,1] was used to normalise the data. The normalised data ($x'$) was derived from the original data ($x$) based on Eq. 1 (Tümer & Edebali, 2019).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad \text{(Eq. 1)}$$

### *2.3 Hidden neuron and hidden layer design*

Designing network is where the dimensions or parameters were established and optimised. The networks were designed with the Mathworks Matlab neural network toolbox. Nonlinear autoregressive exogenous (NARX) network configuration is used in network development. The designated dimensions in this study are the number of hidden layer and the number of neurons in the hidden layer as shown in Fig. 1. The number of neurons in hidden layer was optimised by trial-and-error method by starting with no hidden layer until three hidden layers, meanwhile the neurons in input and output layers were set based on the selected input-output variables. As for the neurons, the trial-and-error start from four to twelve neurons in the first hidden layer, one to six neurons in the second hidden layer and one to four neurons in the third hidden layer.

### *2.4 Learning algorithm and activation function design*

The optimised parameter for ANN study was activation function in which the input neurons were transmitted by hidden layer multiplied with by hyperbolic tangent sigmoid (tansig) and log sigmoid (logsig) function. The equations of the activation function are shown in Eq. 2 & Eq. 3.

$$tansig = \frac{2}{1+\exp(-2n)} \quad \text{(Eq. 2)}$$

$$logsig = \frac{1}{1+\exp(-n)} \quad \text{(Eq. 3)}$$

where $n$ is the input of the function.

The training algorithm used in this study was Levenberg-Marquardt (LM) back propagation algorithm to regularise the bias value. LM is the fastest training algorithm and provide memory reduction (Torrecilla et al., 2008). On the other hand, gradient descent (GD) back propagation was selected as it was the common learning algorithm used, easy calculation and easy implementation (Doshi, 2019).

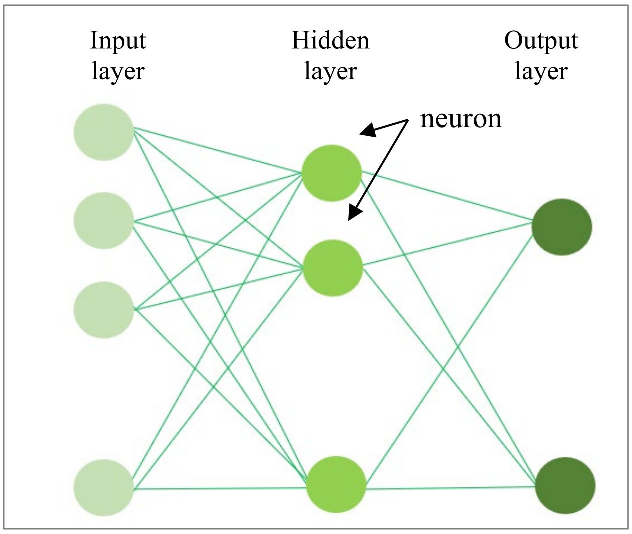



**Fig. 1**: General multilayer neural network

### *2.5 Evaluation performance of neural network*

A few performance criteria were used to evaluate the performance of neural network topology such as coefficient of determination ($R^2$), and root mean square error (RMSE). As for RMSE, it was compared between the target and output of the neural network and this criterion was the smaller the better (STB) type. Meanwhile for $R^2$ value, it measured how well the variation in the output explain by the targets. The value should be between zero (0) to one (1) and if the value was closer to one (1), it shows that the network has a great correlation between targets and output or also can be called as the larger the better (LTB) type.

## 3.0 Results and discussion

The constant parameters in this study were the number of input and output, 2 and 1, respectively. The input consisted of inlet and outlet temperature of spray dryer while the output was the moisture content of coconut milk powder.

### *3.1 Effect of hidden layers*

To study the effect of hidden layer, two types of neural networks were developed. The first model was a neural network without any hidden layer and the other one was neural network with a single hidden layer with ten hidden neurons. The transfer function used for both models was tangent sigmoid (tansig) meanwhile the training algorithm used was gradient descent (GD). The ANN models are summarised in Table 1.

The value of RMSE and $R^2$ of neural network model with no hidden layer are 0.0716 and 0.9478 respectively. The $R^2$ and RMSE values indicate the accuracy of model by comparing the predicted output of the model with the testing data sets. Low accuracy of neural network model was obtained when no hidden layer network topology was used. This is because the network acts as linear regression without presence of

**Table 1**: Effect of layer

| Variable | RMSE | $R^2$ |
|---|---|---|
| No hidden layer | 0.0716 | 0.9478 |
| 1 hidden layer | 0.0516 | 0.9719 |

**Table 2:** Effect of layers and neurons

| Single hidden layer | | | Two hidden layers | | | Three hidden layers | | |
|---|---|---|---|---|---|---|---|---|
| No. of neurons | RMSE | $R^2$ | No. of neurons in 2nd layer | RMSE | $R^2$ | No. of neurons in 3rd layer | RMSE | $R^2$ |
| 4 | 0.0941 | 0.9037 | 1 | 0.1338 | 0.8311 | 1 | 0.0969 | 0.8936 |
| 6 | 0.0799 | 0.9294 | 2 | 0.1435 | 0.7751 | 2 | 0.1158 | 0.8484 |
| 8 | 0.1562 | 0.7249 | 3 | 0.0923 | 0.9044 | 3 | 0.1105 | 0.8975 |
| 10 | 0.0516 | 0.9719 | 4 | 0.0822 | 0.9231 | 4 | 0.1382 | 0.8386 |
| 12 | 0.1296 | 0.8277 | 5 | 0.0938 | 0.9042 | | | |

hidden layer. For neural network with single hidden layer gives the result of 0.0516 for RMSE and 0.9719 for $R^2$. As for $R^2$, the closer it is to 1 represents the best fit, meanwhile for RMSE, the smaller the value represents the best performance. Therefore, these results proved the importance of hidden layer in the development of neural network model. The hidden layer has the function of transferring data and training layer to layers (Uzair & Jamil, 2020).

### *3.2 Optimising the number of layers and neurons*

To optimise the number of hidden layers and number of neurons in hidden layer, three types of NN model were developed. The developed models consisted of ANN with single hidden layer, two hidden layers and three hidden layers. In addition, the appropriate number of neurons in hidden layers were varied and the performance of each model was determined by the trial-and-error method. The results are summarised in Table 2. The transfer function and training algorithm used in this study were tangent sigmoid (tansig) and gradient descent (GD), respectively.

In single layer neural network model, the number of hidden neurons started at four and increased gradually to six, eight, ten and twelve. To avoid overfitting and much complicated NN topology, the number of hidden neurons is set not too high. Too many neurons in hidden layer allow the network to easily learn the training set. However, this network works poorly on the unseen data which leads to overfitting. It is found out that the ANN model with ten number of hidden neurons gives the best result of performance. Both criteria met the value 0.0516 for RMSE which is close to 0 and 0.9719 for $R^2$ which is close to one (1).

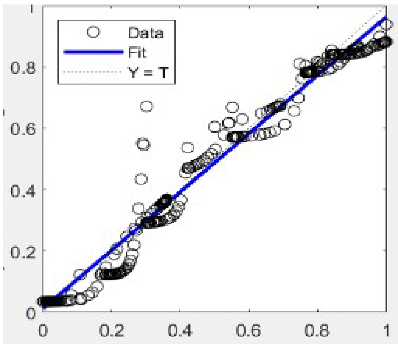
For neural network model with two hidden layers, the hidden neuron of its first hidden layer is set at ten neurons. This is due to its best performance in previous model. Therefore, only the second hidden neuron is manipulated during training. The number hidden neurons were set from one to five.

As a result, NN model with 4 hidden neurons in its second hidden layer has the best performance. The RMSE value for the best fit of 2 hidden layers is 0.0822 while the $R^2$ value is 0.9232. For NN model with three hidden layers, the number of hidden neurons set from one to four. The previous two best hidden neuron numbers i.e., ten neurons in first layer and four neurons in the second layer are maintained for the third hidden layers NN training. The result shows that NN model with three hidden neurons in the third layer has the best performance in term of RMSE and $R^2$ which are 0.1105 and 0.8975, respectively.

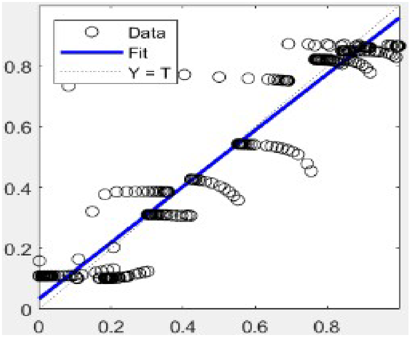
In this study, it is found out that the network model with single hidden layer with 10 hidden neurons has outperformed the two hidden layers and three hidden layers NN model. According to Perazzini (2020), single hidden layer keeps the interaction between input and output simple and guaranteed for better computational efficiency in predicting moisture content. Fig. 2 shows the correlations between the actual and NN simulated data. Strong correlation was observed between the datasets of ANN model with single hidden layer as shown in Fig. 2(a). This indicates good performance of the model in predicting the moisture content. The correlation weaker as the number of hidden layers increased. This is shown in Fig. 2(b) & 2(c) as the data point scattered in a wider band indicates weak correlation. As the number of hidden layers increased, the complexity of the model increased and exceeding the required hidden layer, thus the accuracy decrease.

### *3.3 Effect of activation function*

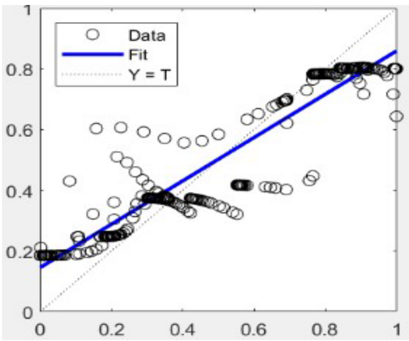
The best topology from previous result was used which is a single hidden layer NN with ten (10) hidden




(a)




(b)

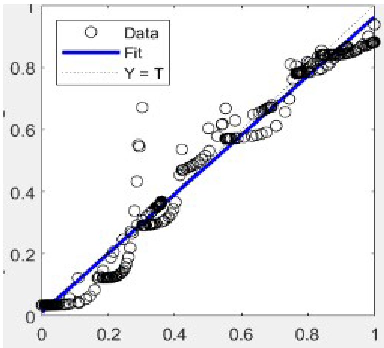



(c)

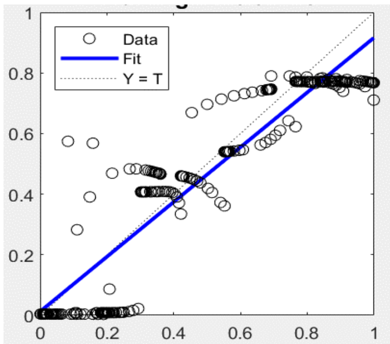
**Fig. 2**: Regression plot for training of NN with (a) single hidden layer, (b) two hidden layers and (c) three hidden layers

neurons to study the effect of activation function to the accuracy of the model. GD was used as training algorithm.

The performance of NN models which were trained by hyperbolic tangent sigmoid (tansig) and log sigmoid (logsig) is illustrated in Table 3. The performance of the NN model with tansig as the activation function is better compared to the model with logsig as it has higher accuracy based on a lower value of RMSE and higher value of $R^2$. The RMSE and $R^2$ values of the NN model with tansig are 0.0516 and 0.9719, respectively. The good performance of the NN model with hyperbolic tangent sigmoid is also supported by the good regression as shown in Fig. 3.




(a)




(b)

**Fig. 3**: Regression plot for training (a) hyperbolic tangent sigmoid and (b) log sigmoid

**Table 3**: Effect of activation function

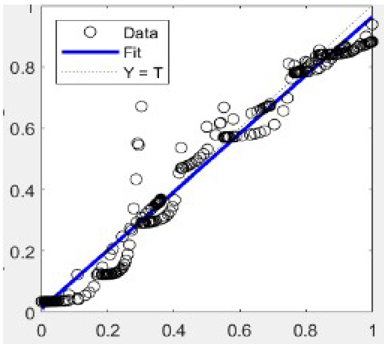
| Variable | RMSE | $R^2$ |
|---|---|---|
| Hyperbolic tangent Sigmoid | 0.0516 | 0.9719 |
| Log Sigmoid | 0.1183 | 0.8543 |

This is because its derivative is steeper and has a wider range, which helps in faster learning and grading (Kızrak, 2020).

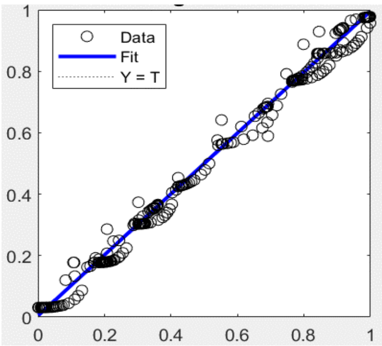
### *3.4 Effect of training algorithm*

In selecting the training algorithm of the NN model, single hidden layer NN with 10 hidden neurons was used. Activation function that gives good performance in the previous parameter study was used in this section. The performance of ANN models trained using gradient descent and Levenberg-Marquardt is summarised in Table 4.

The GD training algorithm was used in this parameter study as in the development of NN model of drying process of coconut milk and carrot as reported by Ming et al. (2020) and Nazghelichi et al. (2011), respectively. During training, the ANN model with LM has faster training speed compared to GD. On the other hand, excellent performance was obtained by LM as the R2




(a)




(b)

**Fig. 4**: Regression plot for training (a) Gradient Descent and (b) Levenberg-Marquardt

**Table 4**: Effect of training algorithm

| **Variable** | **RMSE** | $R^2$ |
|---|---|---|
| Gradient Descent | 0.0516 | 0.9720 |
| Levenberg-Marquardt | 0.0146 | 0.9951 |

value reached 0.9951 and RMSE value of 0.0146 and this is conformed to a previous study that mentioned LM is excellent at minimising the network error and regularises the bias value (Onu et al., 2020). This finding is supported by the regression plot shown in Fig. 4. Good regression was observed for model trained by LM compared to trained by GD.

## 4.0 Conclusions

Optimisation of NN parameters is important to determine the performance of the model. The parameters itself is highly reliant on the datasets of the process. The optimum configuration of neural network for predicting the moisture content of spray-dried coconut milk is NN with single hidden layer with 10 neurons in hidden layer i.e., 2-10-1 topology. The best NN was obtained when the network was trained with Levenberg-Marquart (LM) as the learning algorithm, and hyperbolic tangent sigmoid as the activation function. This configuration gives the lowest RMSE and the highest $R^2$ values are 0.0146 and 0.9951, respectively.

## References

Abdullah, Z., Taip, F. S., Mustapa Kamal, S. M., & Abdul Rahman, R. Z. (2018). Effect of sodium caseinate concentration and sonication amplitude on the stability and physical characteristics of homogenized coconut milk. *Journal of Food Processing and Preservation, 42*(11), e13773. https://doi.org/10.1111/jfpp.13773

Abdullah, Z., Taip, F. S., Kamal, S. M. M., & Rahman, R. Z. A. (2020). Nonlinear model-based inferential control of moisture content of spray dried coconut milk. *Foods, 9*(9), 1177. https://doi.org/ 10.3390/foods9091177

Bhat, R., Alias, A. K., & Paliyath, G. (Eds.). (2012). Progress in food preservation. John Wiley & Sons. https://doi.org/ 10.1002/9781119962045

Brownlee, J. (2019). (2021, January 18). *How to choose an activation function for deep learning*. https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/

Doshi, S. (2019). (2019, January 13). *Various optimization algorithms for training neural network*. https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6

Kızrak, A. (2020). *Comparison of activation functions for deep neural networks*. https://towardsdatascience.com/comparison-of-activation-functions-for-deep-neural-networks-706ac4284c8a

Ming, J. L. K., Taip, F. S., Anuar, M. S., Noor, S. B. M., & Abdullah, Z. (2020). Artificial neural network topology optimization using k-fold cross validation for spray drying of coconut milk. *IOP Conference Series: Materials Science and Engineering, 778*, 012094. https://doi.org/ 10.1088/1757-899X/778/1/012094

Nazghelichi, T., Kianmehr, M. H., & Aghbashlo, M. (2011). Prediction of carrot cubes drying kinetics during fluidized bed drying by artificial neural network. *Journal of Food Science and Technology, 48*(5), 542–550. https://doi.org/10.1007/s13197-010-0166-2

Onu, C. E., Igbokwe, P. K., Nwabanne, J. T., Nwajinka, C. O., & Ohale, P. E. (2020). Evaluation of optimization techniques in predicting optimum moisture content reduction in drying potato slices. *Artificial Intelligence in Agriculture, 4*, 39–47. https://doi.org/ 10.1016/j.aiia.2020.04.001

Perazzini, H., Perazzini, M. T. B., Meili, L., Freire, F. B., & Freire, J. T. (2020). Artificial neural networks to model kinetics and energy efficiency in fixed, fluidized and vibro-fluidized bed dryers towards process optimization. *Chemical Engineering and Processing-Process Intensification, 156*, 108089. https://doi.org/10.1016/j.cep.2020.108089

Seow, C. C., & Gwee, C. N. (2003). Coconut milk: chemistry and technology. *International Journal of Food Science & Technology, 32*(3), 189–201. https://doi.org/ 10.1046/J.1365-2621.1997.00400.X

Sodeifian, G., Sajadian, S. A., & Ardestani, N. S. (2016). Evaluation of the response surface and hybrid artificial neural network-genetic algorithm methodologies to determine extraction yield of *Ferulago angulata* through supercritical fluid. *Journal of the Taiwan Institute of Chemical Engineers, 60*, 165–173. https://doi.org/10.1016/j.jtice.2015.11.003

Taylan, O., & Haydar, A. (2004). Artificial neural network models in prediction of the moisture content of a spray drying process. *Journal of the Korean Ceramic Society, 41*(5), 353–358. https://doi.org/10.4191/KCERS.2004.41.5.353

Tümer, A. E., & Edebali, S. (2019). Modeling of trivalent chromium sorption onto commercial resins by artificial neural network. *Applied Artificial Intelligence, 33*(4), 349–360. https://doi.org/10.1080/08839514.2019.1577015

Torrecilla, J. S., Aragón, J. M., & Palancar, M. C. (2005). Modeling the drying of a high-moisture solid with an artificial neural network. *Industrial & Engineering Chemistry Research, 44*(21), 8057–8066. https://doi.org/10.1021/ie0490435

Torrecilla, J. S., Aragón, J. M., & Palancar, M. C. (2008). Optimization of an artificial neural network by selecting the training function. Application to olive oil mills waste. *Industrial & Engineering Chemistry Research, 47*(18), 7072–7080. https:/doi.org/10.1021/ie8001205

Uzair, M., & Jamil, N. (2020, November). Effects of hidden layers on the efficiency of neural networks. *IEEE 23rd International Multitopic Conference (INMIC) 2020*, 1–6. https://doi.org/10.1109/INMIC50486.2020.9318195.

Ukaoha, K. C., & Igodan, E. C. (2019). Architecture optimization model for the deep neural network. *International Journal of Intelligent Computing and Information Sciences, 19*(2), 1–16. https://doi.org/10.21608/ijicis.2019.96101

Wilamowski, B. M. (2013, June). Comparison of training algorithms and network architectures. *IEEE 17th International Conference on Intelligent Engineering Systems (INES) 2013*, 11–11. IEEE. https:/doi.org/10.1109/INES.2013.6632792.