Artificial Neural Network Topology Optimization using K-Fold Cross Validation for Spray Drying of Coconut Milk

Jesse Lee Kar Ming, Farah Saleena Taip, Mohd Shamsul Anuar, Samsul Bahari Mohd Noor, Zalizawati Abdullah

Department of Process and Food Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400, UPM Serdang, Selangor, Malaysia.
Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400, UPM Serdang, Selangor, Malaysia.
Faculty of Chemical Engineering, Universiti Teknologi Mara, 40450 Shah Alam, Selangor Malaysia

farahsaleena@upm.edu.my (Taip, F.S.) *Corresponding Author

Abstract. In this study, the development of an optimized topology neural network model for spray drying coconut milk is investigated using K-fold cross validation technique. Performance between standalone ANN and ANN with K-fold cross validation is compared, as K-fold cross validation method is integrated into neural network to overcome the limitations of restricted dataset. With inlet temperature (140 °C-180 °C), concentration of maltodextrin and sodium caseinate (0 w/w % - 10w/w %) are established as the input parameters, while moisture content (3.64%-5.1%), outlet temperature (76.5 °C-104.5 °C) and surface free fat percentage (0.35%-34.51%) are the output parameters for the neural network. Experimental data from the spray drying process is used to develop the neural network. Selection from the best training algorithm (gradient descent backpropagation, gradient descent with momentum, resilience backpropagation, conjugate gradient backpropagation with Polak-Ribe re restarts, conjugate gradient backpropagation with Fletcher-Reeves, scaled conjugate gradient, Broyden-Fletcher-Goldfarb-Shanno backpropagation algorithm and Levenberg-Marquardt backpropagation), transfer function (tansig, logsig, purelin and satlin), number of training runs (1000-5000), number of hidden layers (1-3) and nodes (5-15) have significant effect on the performance of the ANN models based on the lowest MSE values and R² values. Overall, the optimum topology ANN model with k-fold cross validation outperformed the recorded lowest MSE value of 0.064 and highest R² value of 0.855 compared to the optimum standalone ANN model with MSE value of 0.082 and R² value of 0.832. The optimum ANN with K-fold cross validation implements the Levenberg-Marquardt training algorithm with hyperbolic tangent sigmoid transfer function using 4500 times training runs with optimal topology configuration of 3-8-2-3. Result concludes that
the developed neural network using K-fold cross validation represents the spray drying process as a highly reliable model with high degree of accuracy.

1.0 Introduction

Coconut milk is a white protein-water-oil emulsion that is extracted from grated endosperm of matured coconuts which is highly unstable and is easily separated into two separate distinct solutions - a heavy aqueous phase and a lighter cream phase (Zafisah et al., 2018). Because of this, coconut milk must undergo different food preservation treatments such as pasteurization and chemical treatment. Recently, spray drying the coconut milk into powder form has been well received by consumers and producers due to ease in transportation, longer shelf life and higher microbiological stability than its liquid form (Tallei, 2017).

Spray drying is regarded as one of the most important industrial drying system. Despite being a prevalent drying technology, the optimization process relies heavily on experimental data from pilot-scale plants and design experiences (Huang, Kumar & Mujumdar, 2006; Langrish & Fletcher, 2001). Therefore, developing an accurate mathematical model of spray drying process is a complicated procedure. It involved application of mechanical, heat and mass transfer to describe the relationship between the processing conditions and the powder properties. The complexity of spray drying process is also influenced by correlative relationship between independent variables in the system, as it leads to multiple-layer common linear phenomenon resulting in inconsistent end powder product characteristics (Avila, Rodríguez, & Velásquez, 2015). Extensive numerical simulations have to be validated through experimental works which are costly and highly time consuming (Gharsallaoui, Roudaut, Chambin, Voilley & Saurel, 2007; Patel, Chen, Jeantet & Schuck, 2010). Furthermore, none of the mathematical model can be applied over a wide range of products, conditions and spray dryer designs.

Artificial neural network (ANN) has been proposed as a simpler alternative of modelling technique for spray drying process. ANN has placed its importance in this study, where the topology configuration is the foundation towards spray drying model development and is crucial to ensure effective network function. The study emphasized on choosing the best transfer function, number of hidden nodes and hidden layers as building blocks for the feed-forward network. The importance of topology optimization has been heavily employed in the food drying industry such as convective drying, vacuum drying and thin-layer drying (Beigi & Ahmadi, 2018; Chen, Ramaswamy & Alli, 2001; Goriian, Hashjin & Khoshtaghaza, 2011). However, an ideal neural network requires large number of datasets to achieve high robustness and reliability. Development of ANN using lower number of datasets due to experiment cost and time constraint leading to poor precision and accuracy.

Multiple methods have been proposed to overcome the limitation of small datasets such as bootstrap method, multivariate regression technique and mega-trend-diffusion method (Borra & Di Ciaccio, 2010). However, K-Fold cross validation technique has proved it is viably significant in the development of ANN network. The K-fold cross validation held its advantages, as the method tends be less biased over other methods (Borra & Di Ciaccio, 2010). The importance of K-fold cross validation is to provide a better generalization capability of the neural network and fully utilize all datasets for the
neural network validation. The integration of K-fold cross validation in neural mapping have been used in various food research studies such as neural modelling of rice parboiling process (Behroozi-Khazaei & Nasirahmadi, 2017), fruit classification (Zhang, Wang, Ji & Phillips, 2014) and drying process (Erenturk & Erenturk, 2007). However, there are no significant research papers that compare and justify the potential of K-Fold cross validation method in the area of food research studies.

In this paper, the objective is to develop an artificial neural network using trained optimal topology configuration resembling spray drying process of coconut milk using K-fold cross validation technique. Comparison are made with standalone ANN by observing mean square error (MSE) and correlation of determination (R²) values.

2.0 Materials and methods

2.1 Material

Freshly grated coconut flesh was purchased from a local wet market (Selangor, Malaysia). The carrier agents used were maltodextrin (Epic Chemical Sdn. Bhd., Malaysia) and sodium casein (R&M Chemical, Malaysia).

2.2 Preparation of coconut milk powder

The data for building a feed-forward neural network is collected through experimentation. Spray drying process was carried out using a laboratory spray dryer (SD-05). In order to develop more robust spray drying neural network, a number of experiments were carried out with the independent variables ranging between 140 °C-180 °C of inlet spray drying temperature, 0 w/w %- 10w/w % of maltodextrin and 0 w/w %- 10w/w % of sodium caseinate percentage.

2.3 Moisture Content Measurement

The characteristic of moisture content represents the end quality of the powder (Sagar & Suresh Kumar, 2010). Two grams of powder is weighed in sample plate holder where the moisture content of the spray dried powder is measured using moisture analyser MX-50 (A&D Weighing, US). The measurement of moisture is done in triplicate and the average value is taken (Ferrari, Germer, Alvim, Vissotto, & de Aguirre, 2012).

2.4 Surface Free Fat Measurement

The percentage amount of free fat powder is measured using a method prescribed by Jafari et al. (2008) with minor modifications. Twenty-five millilitres of petroleum ether were added to 2 g of powder in a glass vial with a screw cap and shaken with a vortex mixer for 10 minutes. The mixture is then filtered using Whatman filter paper after the solution is left standing for 30 minutes. This is to ensure that the mixture is completely separated by density difference. The collected powder on the filter paper is then transferred to a petri dish where it is heated in a water bath as the petroleum ether evaporated from the process. The residual powder is then dried in an oven at 102 °C until constant weight is achieved. The content of the surface free fat is expressed as percentage of the powder weight.

2.5 K-Fold Cross Validation
For the development of ANN with K-fold cross validation, the datasets from the experiment runs are divided into tenfold cross validation to develop the hybrid model. Based on Figure 2, each of K fold used for testing, the remaining (K-1) folds are used for training of the ANN (Stegmayer, Milone, Garran, & Burdyn, 2013). Therefore, in the development of the network, the training and validation of the network is repeated K times and as each K turn, the network is further developed using a different set of training and validation data. The selection of K value is important as higher K value reduces variance and required high computational time (Hagan, Demuth, & Beale, 1996). Moreover, low value of K produces high variance but reduces computational time needed. For each fold cycle, the process of training and testing of the ANN are being repeated until the iteration requirements is completed. And as each iteration, the neural algorithm constantly adjusts the network’s weights in the negative descent direction to ensure that the neural network model is similar to the model of the spray drying. The process of training and testing is completed when the requirement of k-fold is fulfilled and each number of K-fold has been applied to the network. However, the K-Fold cross validation has its limitation, as the training of the neural network has to be rerun of K times. A larger computation time often lead to computationally intensive in term of time and consumed cost (Alexander, Tropsha, & Winkler, 2015). Furthermore, the length of training process also largely depends on the size of the datasets and the number of K applied.
2.6 Development of an optimal topology neural network

In this study, the method of constructing an optimal feed-forward neural network is adapted from Gnanasekharan and Floros (1995), as the authors have provided guidelines and limitations specifically in the area of food processing and neural networking. Each design step is to determine the best attributes of the topology design. Based on Figure 2, the input layer consists of three neurons, representing the value of drying air temperature (°C), concentration of sodium caseinate (w/w %) and concentration of maltodextrin (w/w %). The output layer has three neurons that are the value of surface free fat, outlet temperature and moisture content.

Using the experimental data from the spray drying coconut milk, an optimized topology ANN model was developed through MATLAB R2016 software. Each units of the input layer received input
signal $x_i$ and transferred the signal to each node in the hidden layer. At each iteration, processing and learning took place in the hidden and outer layer, in which each node is programmed to have an activation function. The weightage of each node is modified based on the back-propagation method, in which each weight modified corresponds to the negative gradient of an error measured. Weights of the bias are adjustable connection strength between neurons and signifies the inputs signal to generate outputs. The weighted sum of inputs and bias is transformed into an activation level using transfer function to produce signal $y_j$ as stated:

$$y_j = f(\text{net}_j) = f\left(\sum w_{ji}x_i + b\right) \quad (1)$$

The output of the network depends on the nature of its transfer or activation function, in which transmission occurs when weighted sum, $\sum w_{ji}x_i + b$ exceeds the threshold. Therefore, the threshold function is defined as positive output values and is defined as:

$$y_j = \begin{cases} 1 & \text{if } \text{net}_j > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

As the ANN used iterative gradient descent algorithm, the evaluation of the network prediction performance in comparison to real experiment data is based on a few criterions, which are correlation of determination, $R^2$ value and mean square error (MSE) used for this study. The formula for the criterions is given as followed:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (T_{p,exp,i} - T_{p,cal,i})^2 \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^{N}(T_{p,exp,i}-\overline{T}_{p,exp})^2 - \sum_{i=1}^{N}(T_{p,exp}-T_{p,cal,i})^2}{\sum_{i=1}^{N}(T_{p,exp}-\overline{T}_{p,exp})^2} \quad (4)$$

where $T_{p,exp,i}$ and $T_{p,cal,i}$ are the average number of experiments and calculated spray drying for its observation and $N$ is the number of runs. The singularity concept for the model is to achieve the smallest $MSE$ with the largest possible $R^2$ (Sheela & Deepa, 2013).

In this study, the performance of ANN with K-fold cross validation is compared with ANN without K-fold cross validation technique. Both models underwent same topology development; learning algorithms, training functions, number of runs and hidden layers and nodes. The hybrid of ANN with K-fold cross validation applied ten-fold cross validation. Configuration of the most efficient topology is obtained through selection of several network architectures. The optimal design is selected based on several network configurations such feed-forward training with varying number of neurons (5-15) and hidden layers (1-3), training runs (1000-5000), transfer functions and training algorithms. For this research, there are eight training algorithms and four activation functions used, each has its own distinguished speciality in achieving best neural network performance as stated in Table 1. The training of the neural network stopped when it has satisfied the stopping criteria that was set at default by MATLAB Neural Network Toolbox.
Table 1: Categorization of learning algorithm and transfer function

| Activation function               | Learning Algorithm                                      |
|-----------------------------------|---------------------------------------------------------|
| Hyperbolic tangent sigmoid (tansig) | Gradient descent backpropagation (traingd)               |
| Log-sigmoid (logsig)              | Gradient descent with momentum (traindgm)                |
| Pure-linear (purelin)             | Resilience backpropagation (trainrp)                     |
| Saturating linear (satlin)        | Conjugate Gradient backpropagation with Polak-Riebre restarts (trainscg) |
|                                   | Conjugate gradient backpropagation with Fletcher-Reeves (traincgp) |
|                                   | Scaled conjugate gradient (trainscg)                    |
|                                   | Broyden-Fletcher-Goldfard-Shanno backpropagation algorithm (trainbfg) |
|                                   | Levenberg-Marquardt backpropagation (trainlm)            |

Adapted from Gorjian et al. (2011)

3.0 Results and Discussion

3.1 Selection of Activation Function and Training Function

In this research, the inlet spray drying temperature (°C), concentration of maltodextrin and sodium caseinate (w/w%) are taken as input parameters. The outlet spray drying temperature (°C), moisture content and surface free fat are the output parameters. The selection of optimal learning algorithm, training function, number of hidden layers and hidden nodes is decided by trial and error method as proposed in this research. To obtain the most suitable training algorithm and training function, a benchmark neural network with one hidden layer of 20 neurons is tested against every combination of training functions and algorithms. The combination of training algorithm and training function with the lowest MSE is chosen.

The minimum MSE value of the training data in neural networking development is presented in Table 2. The combination of hyperbolic tangent sigmoid (tansig) with Levenberg-Marquardt training algorithm records the lowest minimum value of MSE. Table 2 shows that trainlm (Levenberg-Marquardt training algorithm) has consistently low MSE value in comparison to other training functions and the lowest MSE value with the combination of tansig activation function. In terms of training function capability, the Levenberg-Marquardt algorithm yields the minimum value between the network actual output and target output, using the stability of steepest descent method and speed of Gauss-Newton algorithm (Aronoff & Mackinney, 1996). In general, ANN with K-Fold cross validation has the lowest MSE value in comparison to the ANN model.
Table 2: MSE value of ANN with K-Fold cross validation and ANN model with different activation function and training function

| Training Function | Activation Function (MSE values) |  |  |  |  |
|-------------------|----------------------------------|---|---|---|---|
|                   | tansig | logsig | purelin | satlins |  |
| ANN with K-fold   | ANN    | ANN    | ANN     | ANN     |  |
| traingd           | 0.107  | 0.321  | 0.314   | 0.289   | 0.548  | 0.395  |
| traingdm          | 0.409  | 0.312  | 0.389   | 0.374   | 0.435  | 0.721  | 0.622  |
| trainrp           | 0.144  | 0.305  | 0.627   | 0.453   | 0.612  | 0.591  | 0.472  |
| trainseg          | 0.211  | 0.432  | 0.567   | 0.547   | 0.722  | 0.478  | 0.374  |
| traincgp          | 0.393  | 0.545  | 0.851   | 0.430   | 0.623  | 0.422  | 0.705  |
| traincfgf         | 0.368  | 0.392  | 0.578   | 0.426   | 0.394  | 0.571  | 1.257  |
| trainbfg          | 0.229  | 0.076  | 0.632   | 0.512   | 0.539  | 0.472  | 0.592  |
| trainlm           | **0.062** | 0.074 | 1.132   | 0.301   | 0.628  | 0.348  | 0.517  |

The advantage of this algorithm showed the performance function is always reduced at each iteration. Similar trends are shown in other studies on the area of drying. Gorjian, Hashjin and Khoshtaghaza (2011) have examined different neural network topologies on drying bayberry fruits and Jafari, Ganje, Dehnad and Ghanbari (2016) implemented neural network technology on fluidized bed drying of onion. Both studies have discovered that topology configuration involving Levenberg-Marquart training algorithm and hyperbolic tangent sigmoid produced best results in optimizing neural network topologies configuration, which is aligned with the current study done.

Both tansig and logsig activation functions have recorded among the lowest MSE value compared to other activation functions. The tansig and logsig have unique characteristics which showed that the activation function increased the network’s initial weight when approaching the optimal weights. Both activation functions increases correlatedly with the gradient function, while other activation functions especially purelin and sanlins does not produce such quality in characteristic and have shown opposite trend of gradient function (Hu, Xu, Wang, Xu, & Su, 2018). Therefore, in Table 3, the MSE values recorded by purelin and sanlins are significantly higher in comparison to the other two activation functions.
3.2 Selection of Training Runs

Figure 3: Effect of Training Runs on MSE in ANN Development

To select training runs, both models are simulated at increasing number of training runs. Both ANN with K-Fold cross validation and ANN are determined by using tansig activation function and Levenberg-Marquart training algorithm as a foundation. However, the ANN model required less than 1000 training runs to achieve its stopping criteria as the performance of the network has achieved its targeted value.

Based on Figure 3, the value of MSE is determined as a function in the number of training runs, in which the MSE value decreases as the number of iterations used increased until the iterations approached 4500 value. The MSE value increased significantly at 5000 iterations and therefore is not chosen. As the neural network undergoes different simulations of training runs, the optimum number of training runs can be obtained and would provide a better generalization on performance of the network (Torrecilla, Aragón, & Palancar, 2008). The modification of weight at each run is highly dependent on the learning rate and algorithm, as each weight changes can be converged either quickly or slowly thus greatly affecting the performance of the neural network (Carney & Cunningham, 1998).

Studies have shown that as the number of training runs increase, the slow rate of convergence during the weight modification of the neural required improvement as the number of runs increased (Hussain, Shafiqur Rahman, & Ng, 2002). A similar results trending have been shown by Chen, Ramaswamy and Ali (2001) on the modelling of osmo-convective blueberries drying through neural network development that focused on determining optimum number of iterations. The research has shown that as number of iterations increase, it leads to lower error value in neural network topology configuration. Torrecilla, Aragón and Palancar (2008) have stated that number of iterations required correlates with the size of the database and further emphasized that the 6000 iterations runs is needed to develop a neural network of 100 datasets, which is similar to the number of datasets in this study.

3.3 Selection of Number of Hidden Nodes and Hidden Layer
The selection of optimal number of hidden nodes and hidden layers is indicated in Table 3. The lowest MSE values and highest $R^2$ values of the validation results indicate that the optimal performance of the architecture design by number of hidden nodes and hidden layers. In general, the result showed that the ANN with K-Fold cross validation with configuration of 8 hidden nodes in the first layer and 2 hidden nodes in the second layer (3-8-2-3) produced the best results. The inconsistent trend of MSE values shown in Table 3 signified a few occurrences of fluctuation in MSE values in which increasing the number of hidden nodes and hidden layers allowed the learning process to escape from local minima or saturation points and proceed to a lower saturation point. In detail, the neural network backpropagation system is a gradient descent type, in which the network would be trapped at certain desired point called local minima and did not achieve global minimum point, which showed the lowest point of MSE value (Shukla, 2015). Different authors have proposed that increasing the number of hidden nodes would weaken the local minima problem, however stated that exceeding the number of hidden nodes in the neural would create more local minima (Choi, Lee, & Kim, 2008; Poston, Lee, Choie, & Kwon, 1991; Yu & Chen, 1995).

Further observations in Table 3 have shown that increased number of nodes produced inconsistent MSE values trend. Supporting statements from Assidjo, Yao, Kisselmina and Amané (2008) and Gorjian et al. (2011) have shown that overutilizing and underutilizing the number of neurons in the neural network produced higher error results. Underutilizing the number of neurons in the hidden layer will result in under-fitting, as the limited number of neurons is not capable of detecting signals in a big data set. The authors have also stated that overfitting occurred when the neural network with many neurons in the hidden layer and larger processing capacity are to be trained by a limited amount of dataset (Assidjo et al., 2008).

| Number of hidden layers | Number of neurons | ANN with K-fold Cross Validation | ANN |
|-------------------------|-------------------|---------------------------------|-----|
| 1st hidden layer | 2nd hidden layer | MSE | $R^2$ | MSE | $R^2$ |
| 1 | 5 | 0.201 | 0.547 | 0.309 | 0.514 |
| 1 | 6 | 0.231 | 0.342 | 0.604 | 0.543 |
| 1 | 7 | 0.230 | 0.337 | 0.598 | 0.871 |
| 1 | 8 | 0.268 | 0.346 | 0.759 | 0.324 |
| 1 | 9 | 0.287 | 0.344 | 0.488 | 0.529 |
| 1 | 10 | 0.342 | 0.492 | 0.527 | 0.614 |
| 1 | 11 | 0.372 | 0.779 | 0.625 | 0.628 |
| 1 | 12 | 0.295 | 0.783 | 0.339 | 0.708 |
| 1 | 13 | 0.242 | 0.420 | 0.154 | 0.895 |
| 1 | 14 | 0.222 | 0.280 | 0.362 | 0.692 |
| 1 | 15 | 0.561 | 0.216 | 0.271 | 0.550 |

Table 3: Results of ANN model with different number of hidden layers and hidden neurons
3.4 Comparison between both models

Overall, the optimum design for ANN with K-Fold cross validation is based on Levenberg-Marquart learning algorithm, transfer function of hyperbolic tangent sigmoid and 3-8-2-3 topology configuration. Based on the validation neural network results, the neural network design recorded a value of 0.064 for MSE and R² value of 0.855. As for the standalone ANN model applied with Levenberg-Marquart learning algorithm, the transfer function is using tangent-sigmoid and 3-6-2-3 topology configuration. The MSE and R² values of the model are 0.082 and 0.832, respectively. Therefore, the ANN with K-Fold cross validation slightly outperformed the other model, as it has resulted in lower MSE and higher R² values.
Studies have shown that R² and MSE are good indicators for neural network performance. High R² value and lower MSE value indicate the fitness level of the model. The mechanism behind K-Fold cross validation averaged the error estimation of all the trials done as the holdout method is repeated at K times (trials); such that at each time, one of the K subsets is used as the test set, while the remaining subsets (K-1) are formed as a training set (Alexander et al., 2015). Therefore, the average error residue generated is lower as the value for K time is higher. Lower residual error represents a better fit of the model based on particular data set and signified higher R² value and lower MSE value (Erenturk & Erenturk, 2007). However, studies found that cross-validation method provides an overly confident estimate for the predictive power of the model and its prediction capability external to the model remained unproven (Golbraikh & Tropsha, 2001).

Conclusion
The development of spray drying model using ANN technique for optimum topology configuration design using k-fold cross validation method was investigated. The artificial neural network using multilayer feedforward was constructed based on experimental data of spray drying coconut milk, using inlet temperature, maltodextrin and sodium caseinate concentrate as input parameters. The selected output parameters were outlet temperature, moisture content and surface free fat percentage. Based on the results, the ANN with K-Fold cross validation outperformed its counterpart, ANN based on comparison of MSE and R² values. The neural network with the best configuration was based on Levenberg-Marquart learning algorithm, transfer function of hyperbolic tangent sigmoid and 3-8-2-3 topology configuration using K-Fold cross validation. The establishment of optimum topology configuration was determined based on the lowest MSE value and highest R² value. However, the results did not prove that neural network design was the best approach, as they have high possibility to be trapped at multiple local minimum points. Other methods such as genetic algorithm and simulated annealing are proposed for the network to escape these points and converge to global minimum points (Choi et al., 2008).

References
Alexander, D. L. J., Tropsha, A., & Winkler, D. A. (2015). Beware of R ² : Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPPR Models. Journal of Chemical Information and Modeling, 55(7), 1316–1322. https://doi.org/10.1021/acs.jcim.5b00206

Aronoff, S., & Mackinney, G. (1996). Training Feedforward Networks with the Marquardt Algoritm. Brain and Cognition, 32(2), 273–344. https://doi.org/10.1006/brcg.1996.0066

Assidjo, E., Yao, B., Kisselmina, K., & Amané, D. (2008). Modeling of an industrial drying process by artificial neural networks. Brazilian Journal of Chemical Engineering, 25(3), 515–522. https://doi.org/10.1590/S0104-66322008000300009

Avila, E. L., Rodríguez, M. C., & Velásquez, H. J. C. (2015). Influence of Maltoextrin and Spray Drying Process Conditions on Sugarcane Juice Powder Quality. Rev.Fac.Nal.Agr.Medellín, 68(1), 7509–7520. https://doi.org/10.15446/rfnam.v68n1.47839

Behroozi-Khazaei, N., & Nasirahmadi, A. (2017). A neural network based model to analyze rice parboiling process with small dataset. Journal of Food Science and Technology, 54(8), 2562–
Beigi, M., & Ahmadi, I. (2018). Artificial neural networks modeling of kinetic curves of celeriac (Apium graveolens L.) in vacuum drying, 2061. https://doi.org/10.1590/fst.35717

Borra, S., & Di Ciaccio, A. (2010). Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Computational Statistics and Data Analysis, 54*(12), 2976–2989. https://doi.org/10.1016/j.csda.2010.03.004

Carney, J. G., & Cunningham, P. (1998). The Epoch Interpretation of Learning, 5. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.5940&rep=rep1&type=pdf

Chen, C. R., Ramaswamy, H. S., & Alli, I. (2001). Prediction of quality changes during osmo-convective drying of blueberries using neural network models for process optimization. *Drying Technology, 19*(3–4), 507–523. https://doi.org/10.1081/DRT-100103931

Choi, B., Lee, J.-H., & Kim, D.-H. (2008). Solving local minima problem with large number of hidden nodes on two-layered feed-forward artificial neural networks. *Neurocomputing, 71*(16–18), 3640–3643. https://doi.org/10.1016/j.neucom.2008.04.004

Erenturk, S., & Erenturk, K. (2007). Comparison of genetic algorithm and neural network approaches for the drying process of carrot. *Journal of Food Engineering, 78*(3), 905–912. https://doi.org/10.1016/j.jfoodeng.2005.11.031

Ferrari, C. C., Germer, S. P. M., Alvim, I. D., Vissotto, F. Z., & de Aguirre, J. M. (2012). Influence of carrier agents on the physicochemical properties of blackberry powder produced by spray drying. *International Journal of Food Science and Technology, 47*(6), 1237–1245. https://doi.org/10.1111/j.1365-2621.2012.02964.x

Gharsallaoui, A., Roudaut, G., Chambin, O., Voilley, A., & Saurel, R. (2007). Applications of spray-drying in microencapsulation of food ingredients: An overview. *Food Research International, 40*(9), 1107–1121. https://doi.org/10.1016/j.foodres.2007.07.004

Gnanasekharan, V., & Floros, J. D. (1995). Back Propagation Neural Networks: Theory and Applications for Food Science and Technology. *Food Flavors: Generation, Analysis and Process Influence, 1201–1222*. https://doi.org/10.1016/S0167-4501(06)80262-1

Golbraikh, A., & Tropsha, A. (2001). Beware of q2! *Journal of Molecular Graphics and Modelling, 20*(20), 269–276.

Gorjian, S., Hashjin, T. T., & Khoshtaghaza, M. H. (2011). Designing and optimizing a back propagation neural network to model a thin-layer drying process. *International Agrophysics,
Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural Network Design*. Boston, MA: PWS Publishing. https://doi.org/10.1007/BF00738424

Hu, J., Xu, L., Wang, X., Xu, X., & Su, G. (2018). Effects of BP Algorithm-based Activation Functions on Neural Network Convergence, 29(1), 76–85. https://doi.org/10.3966/199115992018012901007

Huang, L. X., Kumar, K., & Mujumdar, A. S. (2006). A comparative study of a spray dryer with rotary disc atomizer and pressure nozzle using computational fluid dynamic simulations. *Chemical Engineering and Processing: Process Intensification*, 45(6), 461–470. https://doi.org/10.1016/j.cep.2005.11.004

Hussain, M. A., Shafiur Rahman, M., & Ng, C. W. (2002). Prediction of pores formation (porosity) in foods during drying: Generic models by the use of hybrid neural network. *Journal of Food Engineering*, 51(3), 239–248. https://doi.org/10.1016/S0260-8774(01)00063-2

Jafari, S. M., Assadpoor, E., He, Y., & Bhandari, B. (2008). Encapsulation efficiency of food flavours and oils during spray drying. *Drying Technology*, 26(7), 816–835. https://doi.org/10.1080/07373930802135972

Jafari, S. M., Ganje, M., Dehnad, D., & Ghanbari, V. (2016). Mathematical, Fuzzy Logic and Artificial Neural Network Modeling Techniques to Predict Drying Kinetics of Onion. *Journal of Food Processing and Preservation*, 40(2), 329–339. https://doi.org/10.1111/jfpp.12610

Langrish, T. A. G., & Fletcher, D. F. (2001). Spray drying of food ingredients and applications of CFD in spray drying. *Chemical Engineering and Processing*, 40(4), 345–354. https://doi.org/10.1016/S0255-2701(01)00113-1

Patel, K., Chen, X. D., Jeantet, R., & Schuck, P. (2010). One-dimensional simulation of co-current, dairy spray drying systems – pros and cons. *Dairy Science & Technology*, 90(2–3), 181–210. https://doi.org/10.1051/dst/2009059

Poston, T., Lee, C.-N., Choie, Y., & Kwon, Y. (1991). Local minima and back propagation. *Neural Networks, 1991., IJCNN-91-Seattle International Joint Conference On*, ii, 173–176 vol.2. https://doi.org/10.1109/IJCNN.1991.155333

Sagar, V. R., & Suresh Kumar, P. (2010). Recent advances in drying and dehydration of fruits and vegetables: A review. *Journal of Food Science and Technology*, 47(1), 15–26. https://doi.org/10.1007/s13197-010-0010-8
Sheela, K. G., & Deepa, S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. *Mathematical Problems in Engineering, 2013*. https://doi.org/10.1155/2013/425740

Shukla, A. (2015). Identification of Global Minima of Back-Propagation Neural Network in the Prediction of Chaotic Motion, *112*(4), 1–5.

Stegmayer, G., Milone, D. H., Garran, S., & Burdyn, L. (2013). Automatic recognition of quarantine citrus diseases. *Expert Systems with Applications, 40*(9), 3512–3517. https://doi.org/10.1016/j.eswa.2012.12.059

Tallei, T. (2017). Characteristic Of Coconut Milk Powder Made By Variation Of Coconut-Water Ratio, Concentration Of Tween And Guar Gum. *Journal of Applied Science Research, 13*(6), 34–44.

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 and Engineering Chemistry Research, 47*(18), 7072–7080. https://doi.org/10.1021/ie8001205

Yu, X. H., & Chen, G. A. (1995). On the Local Minima Free Condition of Backpropagation Learning. *IEEE Transactions on Neural Networks, 6*(5), 1300–1303. https://doi.org/10.1109/72.410380

Zafisah, N. S., Yusof, Y. A., Ali, M. A., Roslan, N. S., Tahir, A., Aziz, M. G., & Chin, N. L. (2018). Processing of raw coconut milk for its value addition using spray and freeze drying techniques. *Journal of Food Process Engineering, 41*(1), 1–11. https://doi.org/10.1111/jfpe.12602

Zhang, Y., Wang, S., Ji, G., & Phillips, P. (2014). Fruit classification using computer vision and feedforward neural network. *Journal of Food Engineering, 143*, 167–177. https://doi.org/10.1016/j.jfoodeng.2014.07.001

**Acknowledgement**

The author is grateful to those who have contributed to this review paper. (No Vot. 9670000)