Journal of Engineering Science, Vol. 10, 85–94, 2014

Artificial Neural Network Model Prediction of Glucose by Enzymatic Hydrolysis of Rice Straw

Erniza Mohd Johan Jaya, Nur Atiqah Norhalim and Zainal Ahmad*

School of Chemical Engineering, Universiti Sains Malaysia,
Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia

*Corresponding author: chzahmad@eng.usm.my

Abstract: *The aim of this paper is to predict the production of glucose using artificial neural network (ANN) and validation with the experimental values for hydrolysis process. The ANN consists of three layers which are input, hidden and output layer. The input layer is the manipulated variables in the case study, which are the activity of added cellulose, substrate initial concentration and hydrolysis time on the production of glucose while the output layer is the concentration of glucose. The performances of the model were evaluated using the coefficient of determination, mean square error and average relative deviation. The predictive model shows a good result as the coefficient of determination, 0.8361 was obtained with a small value of mean square error, 0.1947 and 5.644 as the average relative deviation. It clearly shows that ANN gives a good prediction on the enzymatic hydrolysis for the production of glucose.*

Keywords: Model prediction, artificial neural network, cellulose, enzymatic hydrolysis, rice straw

1. INTRODUCTION

Agriculture residue such as rice straw is one of the raw materials potentially able to produce high-value products. Rice straw is mainly used as a source of glucose which is in demand for various processes such fermentation. Sustainable development has become an issue that draws a considerable attention worldwide. This issue has driven researchers and engineers to work hard in converting lignocellulosic waste materials such as rice straws for better use. In this case, rice straw is hydrolysed to produce a more valuable product, which is glucose.

In order to get a better process, a model that represents the hydrolysis process should be developed. By using the MATLAB software to develop a model for hydrolysis process, the simulation can be done in a short time with efficient result. The program can be run numerous times with different operating conditions to obtain a better result. Also, the cost needed to run the program is significantly lower than running the process in a pilot plant or lab scale. After a

model is developed, a process with a high production rate should be achieved. From engineering perspective, this can be achieved with process optimisation. Operating process at its optimal state will help to reach maximum productivity with lowest possible cost while maintaining its quality.

In this paper, the model prediction for the enzymatic hydrolysis uses artificial neural network (ANN), an artificial intelligence tool commonly applied for a complex relationship. The ANN modelling can possibly be provided with more than two variables and can predict one, two or more outputs.[1] The ANN is effective for approximating nonlinear function, pattern recognition and classification problem.[7] Therefore, the ANN could recognise arbitrary nonlinear multiparametric discriminant functions directly from experimental value even without prior knowledge on the relationship of the process variables. Therefore, the ANN is used for stimulating and predicting the performance of the enzymatic hydrolysis of cellulose and then comparing the output performance with response surface model (RSM).

The paper aims to develop a predictive model for the production of glucose from the enzymatic hydrolysis of rice straw by using ANN. It also aims to do a comparative study between neural network model and RSM model for predicting glucose concentration during enzymatic hydrolysis of rice straw. The next section explains the procedure of the modelling using ANN and the verification of the prediction data obtained by ANN. Briefly, the ANN model's performance is evaluated based on correlation of determination (R^2), mean square error (MSE) and average relative deviation (ARD). The comparative study is done by comparing the prediction data with the experimental values and RSM values. The results obtained are discussed, and the final section concludes the findings.

2. MODEL PREDICTION ON ENZYMATIC HYDROLYSIS

The study is conducted by identifying the effects of different parameters on hydrolysis process. The selected parameters are used in the development of the model using ANN and the performance validated. The prediction data obtained are verified through comparative study with the experimental values and RSM.

2.1 Identification of the Effects of Different Parameters on the Enzymatic Hydrolysis Process

Based on the case study, there are three parameters which identify the contribution on the production of glucose of hydrolysis process. The parameters

involved are the activity of added cellulose (FPU), substrate initial concentration (g l^{-1}) and time (h). The predictive model was developed by using ANN based the experimental data provided from literature of previous study.

2.2 The Development and Validation of the Model

The prediction model was developed based on the actual values extracted from literature. The model is run using ANN. The type of ANN used in this study is a multilayer perceptron (MLP) such as multi-layer feed forward neural network based on back propagation learning rule to predict the production of glucose from the enzymatic hydrolysis of rice straw. There are a few considerations required to develop an ANN-based model, such model structure (architecture) and ANN model training.

The MLP with one hidden layer of sigmoidal neurons and a layer of linear output neurons was employed in this study. The MLP mainly consists of three layers; an input layer, an output layer and one or more hidden layers, whose numbers of neurons for each layer represented by N, M and K respectively. The different number of neurons allotted are interconnected by adjustable parameters (weights ad biases) associated with them. The relationship is expressed in the following equation:

$$g_k = F\left(\sum_{j=1}^{M} W_{kj}\, f\left(\sum_{i=1}^{N} w_{ji} + \theta_j\right) + b_k\right) \quad (j = 1, \ldots\ldots, M); (k = 1, \ldots\ldots, K) \tag{1}$$

where, w_{jt} is the weight connecting the *i*th neuron in the input layer and the *j*th neuron in the hidden layer; θj the bias of the *j*th neuron in the hidden layer; W_{kj} the weight connecting the *j*th neuron in the hidden layer and the *k*th neuron in the output layer; b_k the bias in the kth neuron in the output layer; and f(·) and F(·) the activation functions of the *j*th neuron in the hidden layer and of the *k*th neuron in the output layer, respectively.

The cross validation technique is used to find exact number of neurons in hidden layers and also to avoid the model from over-fitting while accomplish good generalisation from the training data set. These experimental data or data samples were required to split into training data set and a validation dataset. Then neural networks with different number of hidden nodes are trained with the training dataset. The performance of the model was evaluated on the ability to predict of the validation dataset by calculating mean square error (MSE) as express in Equation 2:

$$MSE = \sum_{k=1}^{K} (d_k - g_k)^2 \tag{2}$$

where g_k represents the prediction of the neural networks and d_k is the desired output, which in this study is glucose concentration Y (g l^{-1}). Figure 1 illustrates the general framework of the model-based approach proposed in this work. The inputs of the MLP neural network model are the activity of added cellulose, X_1 (FPU), substrate concentration, X_2 (g l^{-1}) and the hydrolysis time, X_3 (h) while the model output was glucose concentration, Y (g l^{-1}). The systematic approach developed can be used to build models with more than two inputs such in this case; enzyme concentration varies from $X_1 \ldots X_n$.

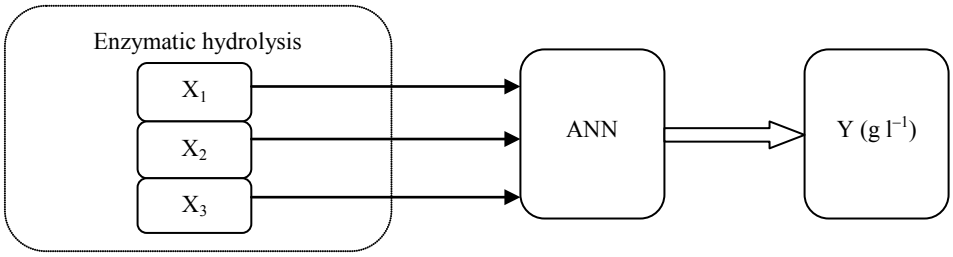

Figure 1: General framework of the model-based approach used in ANN and to optimise the enzymatic hydrolysis of rice straw.

2.3 The Verification of the Predicted Data

The predicted data obtained from the ANN model was compared with the experimental values and RSM values. The comparative step is done to test the capabilities of the model developed. Experimental data are obtained from literature[2] were used to train, validate and test artificial neural networks (MLP) for prediction of glucose concentration during enzymatic hydrolysis. The range of experiment data used is shown in Table 1.

Table 1: Range of experiment data used in this study.

Parameter	Unit	Range
Activity of added substrate	FPU	5–15
Substrate concentration, $C_{c,0}$	g l^{-1}	5–20
Time	h	3–9

3. RESULTS AND DISCUSSION

The model prediction is done using MATLAB R2009a software. It is used to generate a feed-forward back propagation neural network which predicts the production of glucose during enzymatic hydrolysis. The model prediction is accomplished when the number of neurons in the input layer corresponds to the

number of input variables, and the number of nodes in the output layer is equal to the number of output variables. The input variables in this study as mention in previous section were the activity of added cellulose (FPU), substrate concentration (g l^{-1}) and time (h) of hydrolysis also indicated as important factors to determine the extent of enzymatic hydrolysis. Meanwhile the output variable study was glucose concentration (g l^{-1}). The number of neurons in the hidden layer was determined by calibration through several run tests.

3.1 Neural Network Model Prediction

The range of input data of Table 1 (refer section 2.3) mentioned previously is used to develop an ANN model. The input data are randomly divided into three sets: training, validating and testing ones.[3] The first dataset was used to compute the gradient and update the network weights and biases, while the second dataset is used to prevent over fitting. The last dataset is not used during the training, but compared with different models.[4] Usually, 30% of data are used for testing and the remaining 70% for training and validation.[5]

In this study, multi-layer feed forward neural network based on back propagation learning rule is used to predict glucose production from enzymatic hydrolysis of rice straw. This ANN consists of one input and one output layer, with several—but usually only one—hidden layer[3] and information moves in only one direction, forward from the input layer through the hidden layer and then to the output.[6] Many studies show that three layers (only one hidden layer) could model the continuous functions of any accuracy. The schematic of the MLP network with one hidden layer is shown in Figure 2. In order to optimise the number of neurons in hidden layer, average relative deviation (ARD) of testing data versus the number neurons in hidden layer is plotted (Figure 3).

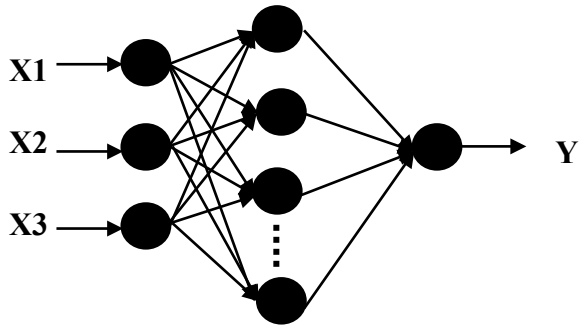

Figure 2: Schematic representation of ANN to stimulate the enzymatic hydrolysis of cellulose from rice straw.

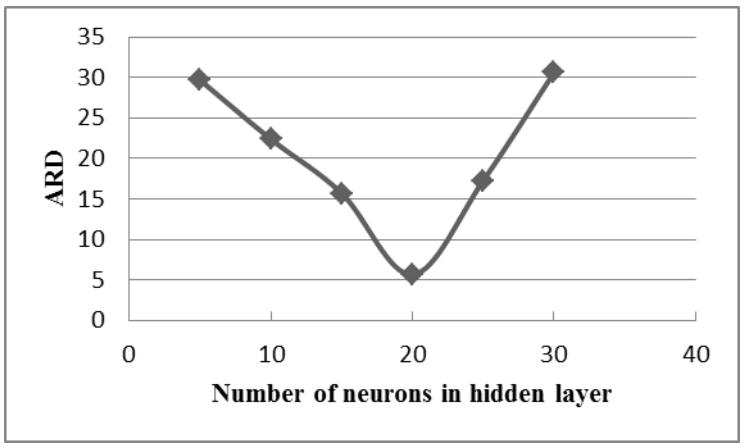

Figure 3: ARD of testing data versus the number of neurons in hidden layer.

In the present work, to test the prediction capabilities of the models, the predicted values obtained from ANN are compared with the experimental values. The coefficient of determination (R^2), mean square error (MSE), and average relative deviation (ARD) were determined. The R^2, MSE and ARD were calculated by following Equations (3), (4) and (5) respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_{pre,i} - X_{exp,i})}{\sum_{i=1}^{n}(X_{pre,i} - \bar{X})} \quad (3)$$

$$MSE = \frac{\sum_{i=1}^{n}(X_{pre,i} - X_{exp,i})^2}{n} \quad (4)$$

$$ARD = \frac{100}{n} \times \sum_{i=0}^{n} \left| \frac{X_{exp,i} - X_{pre,i}}{X_{exp,i}} \right| \quad (5)$$

where $X_{pre,i}$ is the predicted output from observation I; $X_{exp,i}$ the experimental (target) output from observation I; $\bar{X}$ the average value of experimental output; and n the total number of data. R^2 must be close to 1.0, the MSE and the ARD between the predicted and experimental data must be as small as possible. The result are summarised and presented in Table 2.

Table 2: Performance of neural network model.

Statistical parameter	Value
R^2	0.8361
MSE	0.1947
ARD	5.644

3.2 Comparison of ANN Model and RSM

A comparative study is performed between RSM and multi-layer feed forward neural network to estimate their abilities for predicting glucose concentration. The measured values of glucose concentration by ANN model are almost identical with experimental values as compared to the predicted values by RSM. Table 3 shows the measured values by ANN model and predicted values by RSM and comparison with experimental value obtained from literature.[2]

Table 3: Training samples for model building and comparison with experimental and RSM value obtained from literature.

X1 (FPU)	X2 (g l^{-1})	X3 (h)	Y (g l^{-1})				
			Experimental value	RSM		ANN	
				Model value	Relative error	Model value	Relative error
15	5	6	2.8600	2.5273	11.63	2.8631	0.11
10	10	6	3.3772	3.3772	0	3.3772	0
10	10	6	3.3772	3.3772	0	3.3772	0
5	20	6	3.6268	3.8273	5.53	3.6219	0.14
10	10	6	3.3772	3.3772	0	3.3772	0
5	5	6	1.6973	1.8934	11.55	1.6997	0.14
15	20	6	5.4556	5.3917	1.17	5.7286	5.00
10	20	3	4.1457	4.0844	1.48	3.4835	15.97
5	10	3	2.4550	2.2463	8.5	2.4528	0.09
10	10	6	3.3772	3.3772	0	3.3772	0
15	10	3	3.0972	3.2850	6.06	4.7174	52.31
10	5	9	2.6764	2.7308	2.03	2.6797	0.12
10	10	6	3.3772	3.3772	0	3.3772	0
10	5	3	1.8787	1.9619	4.37	1.4649	22.03
10	20	9	5.4820	5.4067	1.37	5.4823	0.01
5	10	9	3.4821	3.2943	5.39	3.4826	0.01
15	10	9	3.9351	4.1438	5.3	3.9353	0.01

Apparently, all the outputs of the neural network were much closer to experimental values than that responded by RSM. At sample of (X1 = 10, X2 = 10, X3 = 6), the relative errors for both RSM model and ANN model are zero, which is accurately predicted compared with experimental values of 3.3772. However, the ANN model generates a good prediction as most of the relative errors for the ANN model are lower than the RSM model in Table 3. Besides, only at samples of (X1 = 10, X2 = 20, X3 = 3), (X1=15, X2=10, X3=3) and (X1=10, X2=5, X3=3), the experimental error points produced by ANN exceeded that of RSM. Thus, the accuracy of neural network model is more desired. It can be said that ANN is capable of simulating the enzymatic hydrolysis of cellulose effectively.

Figure 4 shows the plot of predicted glucose concentration by ANN and RSM against the experimentally determined values. The results showed that ANN has predictions that are closer to the line of perfect prediction than those of RSM, where ANN has better prediction accuracy than RSM. ANN is a superior and more accurate modelling technique as compared to RSM as it represents the non-linearity in a much better way.[8–12] The ANN used is back-propagation network that consists of one input layer, one output layer and one and more hidden layers.

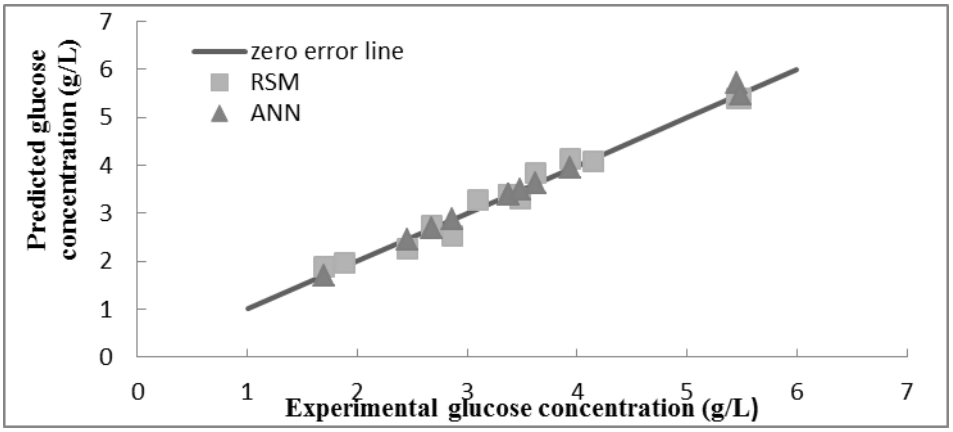

Figure 4: Values of glucose concentration predicted by ANN and RSM versus the experimental value.

3. CONCLUSION

This study finds that neural networks provide good fits to experimental data. The results show that the training of an artificial neural network with the experimental data from the enzymatic hydrolysis of cellulose has been quite

successful. ANN could exactly mimicked enzymatic reaction in the heterogeneous system. The ANN model is also clearly more accurate in comparison to RSM model. The neural network model is not complex because the estimation is realised by simple arithmetic operations. It is claimed that ANN models may act as a connection between isolated experimental data and led to a synergy between the independent studies. The applications of the artificial neural networks can be used for the on-line state estimation and control of enzymatic hydrolysis processes successfully.

4. ACKNOWLEDGEMENT

Sincere appreciation is extended to Universiti Sains Malaysia (USM) for providing the USM RUI 814076 grant as the financial support for this research.

5. REFERENCES

1. Rivera, E. C. et al. (2007). Development of adaptive modeling techniques to describe the temperature dependent kinetics of biotechnological processes. *Biochem. Eng. J.*, 36, 157–166.
2. Zhang, Y., Xu, J. L. & Yuan, H. Y. (2009). Modelling and prediction in the enzymatic hydrolysis of cellulose using artificial neural networks. Fifth International Conference on Natural Computation, Tianjin, China, 14–16 August.
3. Bucinski, A., Karamaç, M. & Amarowicz, R. (2005). Application of artificial neural networks for modelling pea protein hydrolysis by trypsin. *Pol. J. Food Nutr. Sci.*, 13, 163–168.
4. Lertworasirikul, S. (2008). Drying kinetics of semi-finished cassava crackers: A comparative study. *LWT-Food Sci. Technol.*, 41, 1360–1371.
5. Movagharnejad, K. et al. (2011). Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural network. *Energy*, 36, 3979–3984.
6. Mehdizadeh, B. & Movagharnejad, K. (2011). A comparison between neural network method and semi empirical equations to predict the solubility of different compounds in supercritical carbon dioxide. *Fluid Phase Equilib.*, 1, 40–44.
7. Ebrahimpour, A. et al. (2008). A modeling study by response surface methodology and artificial neural network on culture parameters optimization for thermostable lipase production from a newly isolated thermophilic *Geobacillus* sp. strain ARM. *BMC Biotechnol.*, DOI: 10.1186/1472-6750-8-96.

8. Dutta, J. R., Dutta, P. K. & Banerjee, R. (2004). Optimization of culture parameters for extracellular protease production from a newly isolated *Pseudomonas sp*. using response surface and artificial neural networks. *Process Biochem.*, 39, 2193–2198.
9. Singh, A., Majumder, A. & Goyal, A. (2008). Artificial intelligence based optimization of exocellular glucansucrase production from *Leuconostoc dextranicum* NRRL B-1146. *Biores. Technol.*, 99, 8201–8206.
10. Majumder, A. & Goy, A. (2008). Enhanced production of glucansucrase using response surface method. *Biores. Technol.*, 99, 3685–3691.
11. Movagarnejad, K. et al. (2000). A model for the rate of enzymatic hydrolysis of cellulose in heterogeneous solid–liquid systems. *Biochem. Eng. J.*, 4, 197–206.
12. Liu, G. C. & Wang, X. L. (2007). Optimization of critical medium components using response surface methodology for biomass and extracellular polysaccharide production by *Agaricus blazei*. *Appl. Microb. Biotechnol.*, 74, 78–83.