Comparative Study between Neural Network Model and Mathematical Models for Prediction of Glucose Concentration during Enzymatic Hydrolysis

M Nikzad
Lecturer,
Department of Chemical Engineering, Noshirvani University of Technology,
Babol, Iran.

K Movagharnejad
Associate Professor,
Department of Chemical Engineering, Noshirvani University of Technology,
Babol, Iran.

F Talebnia
Assistant Professor,
Department of Chemical Engineering, Noshirvani University of Technology,
Babol, Iran.

ABSTRACT
Enzymatic hydrolysis of cellulose is a complex process because of a number of inhibition and enzyme inactivation reactions which happen during hydrolysis. Artificial Neural Networks (ANNs) are very effective in developing predictive models for processes involving complex reaction kinetics that would otherwise be difficult to be modeled by more traditional deterministic approaches. The present investigation was carried out to study the application of Artificial Neural Network as a tool for predicting glucose production by enzymatic hydrolysis of pure cellulose and comparison with mathematical models and experimental results. A feed forward neural network with one hidden layer was trained and used to predict the glucose production. Comparing the R^2 (coefficient of determination), MSE (mean square error) and ARD (average relative deviation) values of the neural network model with the mathematical model, it was concluded that the neural network is more accurate than the mathematical models. The obtained results show that the ANN can be a useful method for the design of the enzymatic hydrolysis.

Keywords
Cellulose, Enzymatic hydrolysis, Artificial Neural Network, Modeling, Cellulase.

1. INTRODUCTION
Enzymatic hydrolysis of cellulose to produce reducing sugars has long been pursued because of its potential importance in bioethanol production. Bioethanol can be a sustainable substitute for petroleum-derived products [1]. The advantages of enzymatic hydrolysis of cellulose over other hydrolysis methods such as acid hydrolysis are lower utility (cooling water, gas, electricity) and disposal costs and no corrosion issues for the equipments [2]. The enzyme system for conversion of cellulose to glucose comprise endo-1,4-β-glucanase, cellobiohydrolase and β-glucosidase (also called cellobiase), which act sequentially and cooperatively to degrade crystalline cellulose to glucose [3]. In approaching the process design for the enzymatic hydrolysis of celluloses, it is important to be able to predict processing outputs in response to different input variables, such as substrate initial concentration, enzyme activity and hydrolysis time. The process of the enzymatic hydrolysis is highly complex, rendering it difficult to set a model for confident prediction. Enzymatic hydrolysis kinetics of cellulose have been studied for many years [4], but the development of accurate mathematical models for this reaction is still a critical challenge . An alternative approach to obtain a model of the reaction would be the use of Artificial Neural Networks (ANNs).

ANNs can handle incomplete data and deal with nonlinear problems. It can also perform prediction and generalization immediately after the training process [5]. Artificial NNs seem to be a feasible alternative in several instances, and their application for biotechnological processes is continuously growing [6]. With respect to biotechnological processes in particular, several studies can be found in literature, such as the description of the α-amilase inactivation, the prediction of the final concentration of ethanol in a batch fermentation process and as a soft-sensor [7-9]. ANN models are generally used for prediction, function approximation, classification, and clustering [10]. However, few papers were reported about ANN-basedmodel for enzymatic hydrolysis. The aim of the present study is to check the validity of ANN to predict the glucose production under various enzymatic hydrolysis conditions with available experimental data and compare ANN results with kinetic model results.

2. THEORETICAL
2.1 Mathematical modeling
Modeling of the enzymatic hydrolysis of cellulose has been studied in numerous published works. The mathematical models employed in this work for the enzymatic hydrolysis were adapted from Nidetzky et al. [11] and Movagharnejad et al. [12], where the conversion of microcrystalline cellulose (Merck , Germany) and microcrystalline cellulose (Sigmacell; Sigma, Deisenhofen, Germany) to glucose were investigated, respectively. The information about process conditions of the enzymatic hydrolysis can be found in Table 1. More details about the experiments are available in Nidetzky et al. [11] and Movagharnejad et al. [12].

Table 1: Process conditions in the enzymatic hydrolysis

Process parameter	Value	Ref
Temperature (°C)	50	
Initial solid concentration (g.L-1)	40-80-120-160	
Enzyme loading:		
Celluclast (CCN 3000/85-4) FPU/g	6.25-8.33-12.5-25	[11]
Novozym188 (DCN 003/87-11) IU/g	6.25-8.23-12.5-25	
Operating time (h)	200	
pH	4.8	
Temperature (°C)	50	
Initial solid concentration (g.L-1)	10-20-50	
Enzyme loading:		
Celluclast (CCN 3000/85-4) FPU/g	3.7-9.25-18.5	[12]
Novozym188 (DCN 003/87-11) IU/g	8-20-40	
Operating time (h)	24	
pH	5	

Note: FPU=filter paper unit; IU= International unit of enzyme activity.

Two mathematical model for enzymatic hydrolysis were used in this work [11, 12]. In the model conceived by Nidetzky et.al in order to describe the hydrolysis of microcrystalline cellulose, an effective substrate concentration was defined which deviates from bulk concentration of cellulose. They considered following assumptions in their model: two fractions of substrate, the enzyme-to-substrate ratio, and the effective substrate concentration. The results showed that the model predictions matched the experimental hydrolysis data very well over a wide range of substrate and enzyme concentrations. This model involves the hydrolysis kinetic for microcrystalline: definition of an effective substrate concentration [Eq. (1)], relation of substrate hydrolysis to product formation [Eq. (2)], and the effective substrate concentration in presence of cellobiose [Eq. (3)].

$$S_{eff} = aS_{eff,0}\exp\left(-k_{1,\max}\frac{E}{(K_1+E)}t\right)+ (1-a)S_{eff,0}\exp\left(-k_{2,\max}\frac{E}{(K_2+E)}t\right) \quad (1)$$

$$S_{eff,0} - S_{eff} = G \quad (2)$$

$$S_{eff,0} = S_0\frac{K_{eff}E_\beta^n}{\left(S_0+K_{eff}E_\beta^n\right)} \quad (3)$$

The mathematical model developed by Movagharnejad et al. was based on the shrinking particle concept. To consider the usually neglected particle size reduction during the hydrolysis, they assumed that these particles shrink. In another word, the particle size reduces while the particle shape remains constant. This assumption allowed them to simplify a very complicated three dimensional model to a much simpler one dimensional model. Their main assumptions may be listed as below:

1. A fraction of the cellulose enzymes is absorbed instantly on the particle surface
2. Some parts of the particles which are occupied by non-cellulosic materials are totally inert to the hydrolysis reaction.
3. The number of sites available for enzymes is proportional to the effective external surface of the particles.
4. Transfer of enzymes from the solution to the particle surface is very rapid so that it does not affect the hydrolysis rate.
5. Product inhibition negatively affects the rate of hydrolysis and enzyme activity.
6. Particles shrink during the hydrolysis reaction.

This mathematical model gives different equations for two stages of the process. In the first stage of the process the particles shrink without any considerable enzyme desorption (return of absorbed enzymes to the solution), while in the second stage a part of the absorbed enzymes return to the solution and the equations would be different.

The equation (4) is derived for the first stage:

$$(1-y)+\left(\alpha\theta_0+\frac{1}{K'_{eq}}\right)\ln\left(\frac{\alpha\theta_0+y-1}{\alpha\theta_0}\right)=\left(\frac{-K_d}{K'_{eq}}\right)t \quad (4)$$

Where t is time, y is the ratio of substrate concentration to the initial substrate concentration, and α, θ_0 , K'_{eq} and K_d are model parameters.

The following equation is developed for the second stage of the model:

$$\left(\frac{K_d}{3}\right)(t-t_1)=K'_{eq}\left[\left(\frac{-1}{3}\right)\left(z_1^3-z^3\right)+\left(\frac{\alpha}{2}\right)\left(z_1^2-z^2\right)-\alpha\left(z_1-z\right)\right] + \left(\alpha 3K'_{eq}+1\right)I_2+\alpha K'_{eq}I_1-\alpha^2K'_{eq}I_0 \quad (5)$$

Where $z=y^{1/3}$, andI_0, I_1 and I_2 are expressions which have different definitions in different conditions.

2.2 Neural Network Model

Artificial neural network (ANN) is a mathematical tool, which tries to represent low-level intelligence in natural organisms and it is a flexible structure, capable of making a non-linear mapping between input and output spaces [13]. Applications of ANN models include fermentation, extrusion processes, filtration, drying process, etc. [13-16]. Artificial neural networks have been successfully applied to modeling of various biological processes in recent years. They are one of the most popular artificial learning tools in biotechnology, with applications ranging from pattern recognition in chromatographic spectra and expression profiles, to functional

analyses of genomic and proteomic sequences. All neural networks consist of three main layers (input, hidden and 2 [17, 18] and many simple computational elements, \called nodes or neurons, arranged in layers and operating in parallel. The ANN weights, which define the strength of the connection between the nodes, are estimated from empirical data [6]. There are several types of Artificial Neural Network. The most common type of ANN in chemical engineering application is multilayer perceptron (MLP) [19]. In this study multi-layer feed forward neural network based on back propagation learning rule was used to predict glucose production of pure cellulose.

2.2.1 Multilayer perceptron (MLP) network

This NN consists of one input and one output layer, with several-but usually only one- hidden layer [20] and information moves in only one direction, forward from the input layer, through the hidden layer and then to the output [17]. Many studies showed 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 Fig. 1.

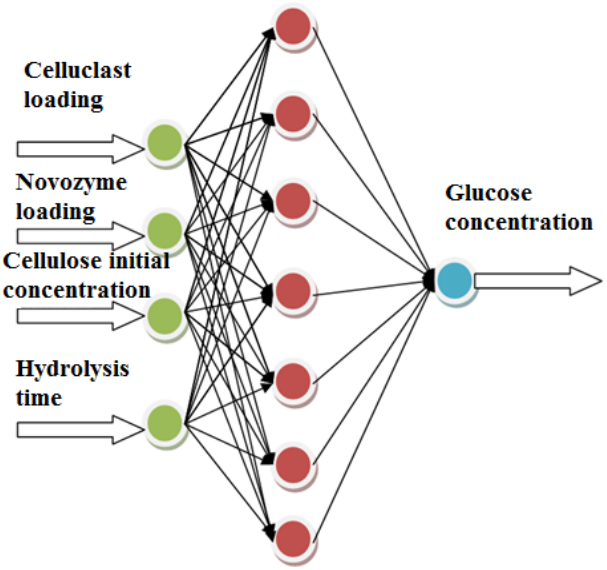

Fig. 1: Architecture of ANN used in developing model

It is necessary to train an artificial neural network before using it for a specific application [17]. The network needs to be trained using a training algorithm such as back propagation, cascade correlation and conjugate gradient [13] among which, the back propagation method is the most commonly used. MLP training begins by applying the input vector, to the input layer having network-processing element [19]. Training is a step by step method for the calculation of the weight factors and biases. During the training, network learns to generate new outputs through an iterative method [17]. Basically the objective of training patterns is to reduce the global error. The goal of every training algorithm is to decrease this global error by adjusting the weights and biases [13].

2.3 Verification of predicted data

In the present work, to test the prediction capabilities of the models, the predicted values obtained from mathematical models and ANN were compared with the experimental values. The coefficient of determination (R^2), mean square error (MSE), and average relative deviation (ARD) were determined and used to compare ANN and mathematical model. The R^2, MSE, and ARD were calculated by following Eq. (6), (7) and (9) respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(X_{pre,i} - X_{\exp,i}\right)^2}{\sum_{i=1}^{n}\left(X_{pre,i} - \overline{X}\right)^2} \quad (6)$$

$$MSE = \frac{\sum_{i=1}^{n}\left(X_{pre,i} - X_{\exp,i}\right)^2}{n} \quad (7)$$

$$ARD = \frac{100}{n} \times \sum_{i=1}^{n}\left|\frac{X_{\exp,i} - X_{pre,i}}{X_{\exp,i}}\right| \quad (8)$$

Where $X_{pre,i}$ is the predicted output from observation i, $X_{\exp,i}$ is the experimental (target) output from observation i, $\overline{X}$ is the average value of experimental output, and n is 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.

3. RESULTS AND DISCUSSION

3.1 Comparison of prediction ability of mathematical models

Two mathematical models were employed to model the enzymatic hydrolysis of microcrystalline cellulose. The Nidetzky's and Movagharnejad's models were compared for their prediction ability of glucose concentration.

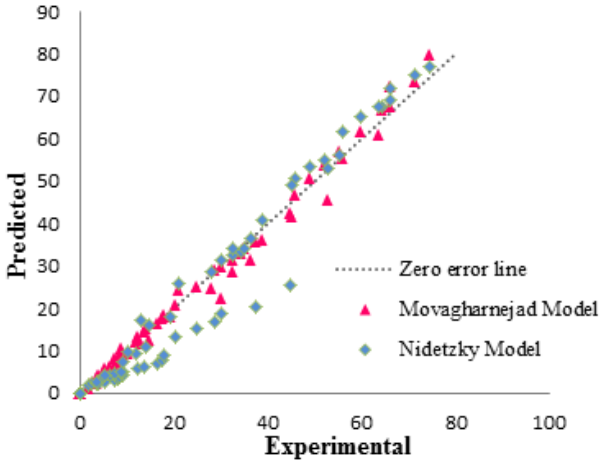

Fig. 2: The values of glucose concentration predicted by two mathematical models versus experimental values

Fig. 2. illustrates the comparative parity plot for Nidetzky's and Movagharnejad's predictions for the glucose concentration. Movagharnejad's model fitted the experimental data with an excellent accuracy, while Nidetzky's prediction showed greater deviations.

MSE, ARD, and R^2 values were calculated for the accuracy of each model. The mean square error (MSE) for Movagharnejad's model and Nidetzky's model is 4.9850 and 29.622, the coefficient of determination (R^2) is 0.9892 and 0.9363, and the average relative deviation (ARD) is 7.007 and 24.67, respectively. The statistical results of the two mathematical models showed that Movagharnejad's model is more accurate than Nidetsky's model.

3.2 Neural network model prediction

MATLAB 7.7.0. was used to generate a feed-forward back-propagation neural network which predicted the glucose production during enzymatic hydrolysis. The number of nodes in the input layer corresponds to the number of input variables; the number of nodes in the output layer is equal to the number of output variables. In this case, the input variables were the activity of added enzyme (Fpu/g substrate for celluclast and IU /g substrate for Novozyme 188), substrate initial concentration and time and the output variable was glucose concentration. The number of neurons in the hidden layer was determined by calibration through several run tests. The input data were randomly divided into three sets: training, validating and testing ones [21]. The first dataset was used for computing the gradient and updating the network weights and biases. The second dataset was used to prevent over fitting. The last dataset was not used during the training, but used to compare different models [10]. Usually 30% of data are used for testing and the remaining 70% for training and validation [18].

Experimental data obtained from literature [11, 12] were used to train, validation and test artificial neural networks (MLP) for prediction of glucose concentration during enzymatic hydrolysis. Totally, 65 data were collected for different experiments. Due to the different ranges of each input as seen in Table 2, the inputs were normalized into the interval [0 1] before feeding into the network by Eq. (9).

Table 2: Range of experimental data used in this study

Parameter	Time (hr)	Celluclast loading (FPU/g)	Novozyme loading (IU/g)	$C_{c,0}$
Range	0-172	3.7-25	6.25-40	10-160

$$(Scaled)_{value} = \frac{Actual_{value} - Min_{actualvalue}}{Max_{actualvalue} - Min_{actualvalue}} \times 0.8 + 0.1 \qquad (9)$$

In this network, Levenberg-Marquardt algorithm was used as a training algorithm–.The most common transfer function is sigmoid function. In this work, we used sigmoid function for transfer function in hidden layer and purelin function for transfer function of output layer. In order to optimize the number of neurons in hidden layer, average relative deviation (ARD) of testing data versus the number neurons in hidden layer is plotted (Fig. 3). ARD has been calculated by means of Eq.3. Results showed 4-7-1 is the best topology of the neural network.

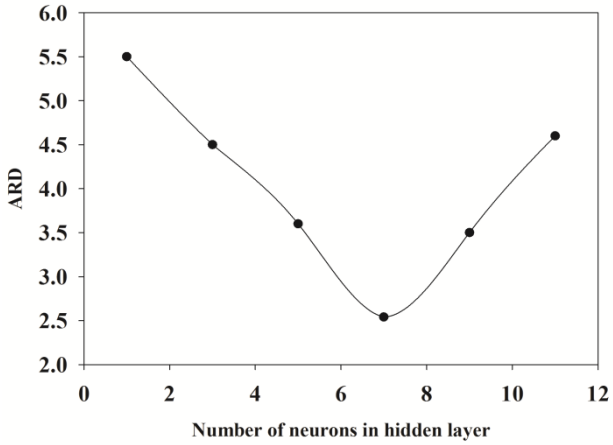

Fig. 3: ARD of testing data versus the number of neurons in hidden layer

The results for training and testing data are summarized and presented in Table 3. The scatter diagram of predicted values versus experimental values is also shown in Fig. 4. It shows that the neural network model prediction fits well with the experimental values.

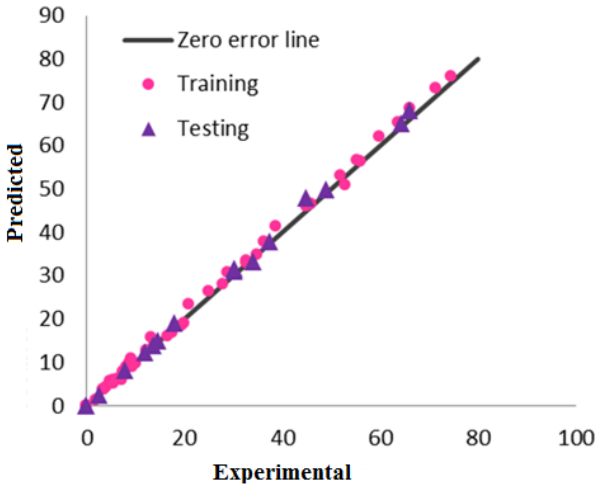

Fig. 4: The values of glucose concentration predicted versus experimental values for training and testing dataset using ANN

Table 3: Performance of Neural Network model

The best architecture	Statistical Parameter	Training data	Testing data	Overall
4-7-1	R^2	0.9969	0.9974	0.9970
	MSE	1.4642	1.1528	1.3934
	ARD	6.6240	2.5304	5.6504

3.3 Comparison of mathematical and ANN models

In this study, modeling of enzymatic hydrolysis of cellulose was investigated. A comparative study was performed between two mathematical models and multi-layer feed-forward neural network to estimate their abilities for prediction of glucose concentration. MSE, R^2 and ARD were used together to compare the mathematical models and ANN model. The Statistical values of mathematical models and ANN model are given in Table 4. Higher R^2, lower MSE and ARD values were obtained by ANN model compared to those of mathematical models. Thus the accuracy of neural network model was more desired. Fig. 5 shows the plot of predicted glucose concentration by mathematical models and ANN against the experimental values. The ANN represents a better agreement between observed and predicted values than mathematical models. On the other hand, we may also conclude that the ANN model is more general than these two mathematical methods. Each of these two models is compared by its own experimental data. It is clear that each researcher designs the experimental procedures and conditions in a way to obtain a better agreement with model predictions. So the model predictions in each of these conventional models are compared with their individual experimental data, but we did not perform any independent experiments for this study and so the data used for training, testing and validation of ANN model was extracted from the experimental results designed for each of these two previous conventional models. Therefore it can be claimed that the present model is not only more accurate but also more general. Numerous models have been reported in the scientific literature in recent years. Most of these models are accompanied by a set of experimental data, but it is difficult to compare or to make any correlation between the results from these independent and isolated datasets. We think that ANN models may act as something more basic than just a correlating or prediction device. ANN models may be considered as an interconnecting device between isolated experimental results and further work on this area may lead to a kind of synergy between isolated sets of experimental data. In this study, we have tried to make some connections between two different sets of data and corresponding applied models. These two individual studies have been selected among many others because the enzyme/substrate systems were very similar to each other and it would be logical to use them as the first step to make connection between the isolated sets of data about cellulose enzymatic hydrolysis. However, using more versatile sets of data might need to introduce more complicated ANN models with respect to inputs, outputs or architectures. The results of this work show the ability of an ANN-based model to deal with two datasets from separate works where the outcome of such model is superior to their own developed models.

Table 4: Statistical measures for mathematical models

Statistical Parameter	Nidetzky et al.,	Movagharnejad et al.,	ANN Model
R2	0.9363	0.9892	0.9970
MSE	29.622	4.985	1.3934
ARD	24.67	7.007	5.6504

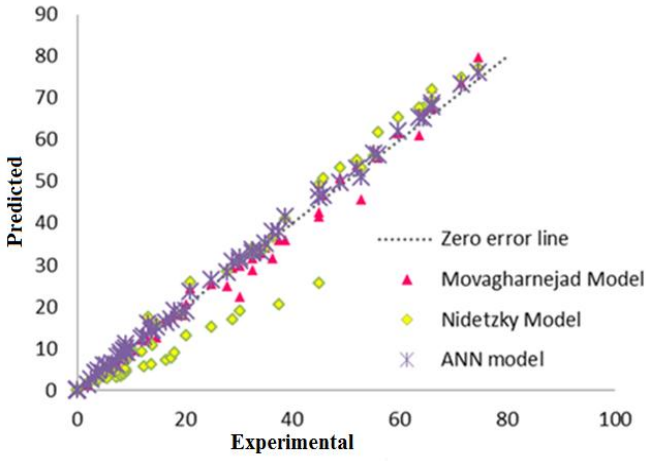

Fig. 5: The scatter plot of mathematical models and ANN model predicted values versus experimental values

4. CONCLUSION

Neural network and two mathematical models were compared for prediction of glucose concentration during the enzymatic hydrolysis of microcrystalline cellulose. Mathematical models are usually judged based on the agreement with experimental data. In this study, a unique experimental data set was used to compare different mathematical model and ANN model. Results of this work have clearly shown that the ANN model is more accurate in comparison to mathematical models. The neural network model is not complex because the estimation is realized 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.

5. REFERENCES

[1] L. R. Lynd, J. H. Cushman, R. J. Nichols, C. E. Wyman, "Fuel ethanol from cellulosic biomass", Science, 251 (1991) 1318-1323.

[2] Y. Sun J. Cheng, "Hydrolysis of lignocellulosic materials for ethanol production: a review", Bioresource technology, 83 (2002) 1-11.

[3] F. Corazza, L. Calsavara, F. Moraes, G. Zanin, I. Neitzel, "Determination of inhibition in the enzymatic hydrolysis of cellobiose using hybrid neural modeling", Brazilian Journal of Chemical Engineering, 22 (2005) 19-29.

[4] C. H. Stephens, P. M. Whitmore, H. R. Morris, M. E. Bier, "Hydrolysis of the amorphous cellulose in cotton-based paper", Biomacromolecules, 9 (2008) 1093-1099.

[5] S. A. Kalogirou, S. Panteliou, A. Dentsoras, "Artificial neural networks used for the performance prediction of a thermosiphon solar water heater", Renewable energy, 18 (1999) 87-99.

[6] J. A. Silva, E. H. Costa Neto, W. S. Adriano, A. L. O. Ferreira, L. R. B. Gonçalves, "Use of neural networks in the mathematical modelling of the enzymic synthesis of amoxicillin catalysed by penicillin G acylase immobilized in chitosan", World Journal of Microbiology and Biotechnology, 24 (2008) 1761-1767.

[7] A. H. Geeraerd, C. H. Herremans, L. R. Ludikhuyze, M. E. Hendrickx, J. F. Van Impe, "Modeling the kinetics of isobaric-isothermal inactivation of Bacillus subtilis α-amylase with artificial neural networks", Journal of food engineering, 36 (1998) 263-279.

[8] V. Saucedo, B. Eikens, M. Karim, "Identification techniques for a recombinant fed-batch fermentation for ethanol production", Advances in Bioprocess Engineering, 275 (1994).

[9] T. J. McAvoy, H. T. Su, N. S. Wang, M. He, J. Horvath, H. Semerjian, "A comparison of neural networks and partial least squares for deconvoluting fluorescence spectra", Biotechnology and bioengineering, 40 (1992) 53-62.

[10] S. Lertworasirikul, "Drying kinetics of semi-finished cassava crackers: A comparative study", LWT-Food Science and Technology, 41 (2008) 1360-1371.

[11] B. Nidetzky W. Steiner, "A new approach for modeling cellulase–cellulose adsorption and the kinetics of the cellulose in heterogeneous solid-liquid systems", Biochemical engineering journal, 4 (2000) 197-206.

[12] K. Movagarnejad, M. Sohrabi, T. Kaghazchi, F. Vahabzadeh, "A model for the rate of enzymatic hydrolysis of enzymatic hydrolysis of microcrystalline cellulose", Biotechnology and bioengineering, 42 (1993) 469-479.

[13] M. Kashaninejad, A. Dehghani, M. Kashiri, "Modeling of wheat soaking using two artificial neural networks (MLP and RBF)", Journal of food engineering, 91 (2009) 602-607.

[14] I. Bardot, N. Martin, G. Trystram, L. Bochereau, M. Rogeaux, J. Hossenlopp, "A New Approach for the Formulation of Beverages, II: Interactive Automatic Method", LWT-Food Science and Technology, 27 (1994) 513-521.

[15] M. Dornier, M. Decloux, G. Trystram, A. Lebert, "Dynamic modeling of crossflow microfiltration using neural networks", Journal of membrane science, 98 (1995) 263-273.

[16] P. Teissier, B. Perret, E. Latrille, J. Barillere, G. Corrieu, "A hybrid recurrent neural network model for yeast production monitoring and control in a wine base medium", Journal of biotechnology, 55 (1997) 157-169.

[17] B. Mehdizadeh K. Movagharnejad, "A comparison between neural network method and semi empirical equations to predict the solubility of different compounds in supercritical carbon dioxide", Fluid Phase Equilibria, 303 (2011) 40-44.

[18] K. Movagharnejad, B. Mehdizadeh, M. Banihashemi, M. S. Kordkheili, "Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural network", Energy(Oxford), 36 (2011) 3979-3984.

[19] A. Singh, D. Tatewar, P. Shastri, S. Pandharipande, "Application of ANN for prediction of cellulase and xylanase production by Trichoderma reesei under SSF condition", Indian Journal of Chemical Technology, 15 (2008) 53-58.

[20] D. E. Rumelhart, G. E. Hintont, R. J. Williams, "Learning representations by back-propagating errors", Nature, 323 (1986) 533-536.

[21] A. Bucinski, M. Karamaç, R. Amarowicz, "Application of artificial neural networks for modelling pea protein hydrolysis by trypsin", 13 (2004) 163-168.