ISSN 1063-455X, Journal of Water Chemistry and Technology, 2019, Vol. 41, No. 4, pp. 219–227. © Allerton Press, Inc., 2019.

THEORETICAL PRINCIPLES OF WATER TREATMENT TECHNOLOGY

Artificial Neural Network Modeling of Cr(VI) Biosorption from Aqueous Solutions[1]

Farzaneh Mohammadi[a], Zeynab Yavari[a], Somaye Rahimi[a], and Majid Hashemi[b, *]

[a] *Environment Research Center and Department of Environmental Health Engineering, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran*

[b] *Environmental Health Engineering Research Center and Department of Environmental Health Engineering, School of Health, Kerman University of Medical Sciences, Kerman, Iran*

**e-mail: mhashemi120@gmail.com*

Received November 7, 2016
Revised April 27, 2017
Accepted January 1, 2019

Abstract—Artificial neural network (ANN) model was applied for predicting the biosorption capacity of excess municipal wastewater sludge for hexavalent chromium (Cr(VI)) ions from aqueous solution. The effects of initial concentration (5 to 90 mg/L), adsorbent dosage (2 to 10 g/L), initial pH (2 to 8), agitation speed (50 to 200 rpm) and agitation time (5 to 480 min) were investigated. The maximum amount of chromium removal was about 96% in optimum conditions. The experimental results were simulated using ANN model. Levenberg-Marquardt algorithm was used for the training of this network with tangent sigmoid as transfer function at hidden and output layer with 13 and 1 neurons, respectively. The applied model successfully predicted Cr(VI) biosorption capacity. The average mean square error is 0.00401 and correlation coefficient between predicted removal rate and experimental results is 0.9833.

DOI: 10.3103/S1063455X19040039

Keywords: biosorption, chromium (VI), neural network modeling, wastewater.

INTRODUCTION

Water environment pollution has become an important global environmental problem, partly due to the excessive emissions of heavy metal ion wastewater which have harmful effect on human health and other biological systems when they exceed the standard levels [1, 2]. Therefore, the generation and treatment of wastewater are considered as a serious ecological, economical and technical problem. Heavy metals such as Hg, Cd, Cr, As and Pb in small quantities have adverse effects on human and environmental health [3]. Chromium is a highly toxic pollutant discharged by the electroplating, metal finishing, leather tanning, photography, dye and textile industries [4]. Chromium exists in four different stable isotopes of various amounts of abundance (^{50}Cr, ^{52}Cr, ^{53}Cr and ^{54}Cr) and in several oxidation states, ranging from –2 to 6. However, only trivalent (Cr(III)) and hexavalent (Cr(VI)) are predominant in the environment [5].

Cr(VI) is believed to be much more toxic than Cr(III) even at trace levels as it is considered to be carcinogenic, mutagenic and teratogenic. It has been classified as a Group A human carcinogen by the USEPA. Chromium removal has become an emergency, high-priority environmental challenge [6]. Various methods have been used for heavy metal removal, including chemical precipitation, neutralization, membrane filtration and adsorption. Among these techniques, biosorption using low cast biosorbent has gained considerable attention because of high efficiency, low cost, more availability and ease of handling [7]. In recent years, sugarcane bagasse, Rice husk, sawdust, coconut husk, oil palm shell, neem bark, etc., have been used for the elimination of heavy metals from wastewater [8]. The use of non-living biomaterial containing metal-binding functional groups would have the advantage of not requiring utmost care and maintenance as well as being useful in remediating areas with high levels of heavy metal contamination [4–7].

The biosorption process is quite complex. This is due to the involvement of numerous parameters resulting in nonlinear relationships. It is difficult to simulate the batch biosorption results using the conventional mathematical modeling [8, 9]. Because of their reliable, strong and notable characteristics in capturing the nonlinear relationships existing between multi-input and output variables in complex systems, it has become evident that numerous applications of artificial neural network (ANN) have been successfully conducted in var-

[1] The text was submitted by the authors in English.

ious parts of environmental engineering field in the past decade [10], such as: prediction of sludge bulking in wastewater treatment plant [11]; river water quality modeling [12]; prediction of NOx emission [13]; modeling dioxin emission from solid waste incinerators [14]; modeling Acid Black172 and Congo Red biosorption from aqueous [15]; environmental pollution forecasts [16]; predictions of daily mean PM10 concentrations [17].

Although a number of investigations have been conducted on the removal of Cr(VI) ions from aqueous solutions using various adsorbents, every special material needs to be given a particular focus for investigation [1, 4, 6, 7, 18].

The present study describes the adsorption potential of excess municipal sludge in removal of Cr(VI) ions from aqueous solutions. The effects of various operational parameters, such as adsorbent dosage, initial concentration, initial pH, agitation speed and agitation time on Cr(VI) adsorption were also investigated. On the basis of batch adsorption experiments, we proposed a three-layer ANN model using a back propagation algorithm to predict the Cr(VI) removal efficiency of excess municipal sludge used as adsorbent materials in this work. Next, we conducted an optimization study to determine the optimal network structure. Finally, outputs obtained from the ANN modeling were compared with the experimental data.

EXPERIMENTAL

Preparation of Biosorbent

In this study, the excess sludge of south Isfahan wastewater treatment plant is used. The amount of produced sludge by this plant is 100 tons per day that is disposed without any special usage. The result showed that there was no detectable Cr(VI) levels present in excess sludge to have an effect on the batch experiment data. Prior to batch adsorption tests, excess sludge was dried in an oven (Nuve FN 500) at 103–104°C for 24 h. The dried biosorbent was sieved through a 150 μm sieve. This process was done due to higher efficiency of adsorption through shortening the diffusion path [18].

Preparation of Chemicals

Chromium stock was prepared according to method mentioned in Standard Methods for the Examination of Water and Wastewater Book [19]. Distilled water was used in order to prepare all solutions. 2.829 g potassium dichromate ($K_2Cr_2O_7$) with a purity of 99% and molecular weight of 19.294 g/mol was transferred to 1000 mL flask to prepare a Cr solution of 1000 mg/L and then its volume was arisen to 1000 mL using distilled water and was stored at 4°C. Other solutions were prepared with required concentration through dilution. NaOH and HCl solutions were used in order to adjust pH. Experimental instruments were washed by sulfuric acid 0.005 M and were rinsed several times with distilled water.

Batch Experiments of Biosorption

Biosorption experiments were conducted in batch system until reaching equilibrium. Initial concentrations of 5; 20; 50 and 90 mg/L of Cr(VI) were prepared through diluting stock solution. Batch experiments were carried out in a pH of 2; 4; 6 and 8 to determine the effect of initial pH on biosorption. 25 mL solutions were prepared at specific concentration and dry biosorbent of 2; 4; 6; 8 and 10 g/L were added to sample. Each experiment was performed in duplicate to observe the reproducibility and the mean value was used for each set of values. Experimental temperature was kept around 25°C. The optimal agitation time and optimal agitation speed were determined in a series of separate trials in order to reduce the number of experiments. Effects of various operating agitation time and speed ranging, respectively from 5 to 480 min and 50 to 200 rpm were also investigated in batch studies. Finally, the other samples were then placed in an orbital shaker and agitated up to a total optimal contact time of 120 min at a fixed optimal agitation speed of 200 rpm. Generally, 296 experiments were conducted in this study and 48 tests among them are used for optimizing agitation speed and time.

Then absorbent was separated by 0.45 μm filter from sample. The remaining chromium concentration in the sample was determined by colorimetric method and spectrophotometer at wavelength of 540 nm.

Percentage of Cr(VI) ions removal being the output parameter of the ANN model was considered as a measure of adsorption efficiency of excess municipal sludge. The efficiency of adsorption (R, %) was calculated as

$$R = \frac{(C_0 - C_e)\times 100}{C_0}, \quad (1)$$

where C_0 and C_e are the initial and the equilibrium Cr(VI) concentrations of the solution, respectively.

Neural Network Modeling of Cr(VI) Biosorption

ANN are computer models, based on a simplified modeling of the brain's biological functions, that's mean, having the ability to think, learn, remember reasons and solve problems. Neural network models are composed of neurons and weights, and are based on the principle that a highly interconnected system of simple processing elements can learn complex interrelationships between independent and dependent variables [19, 20]. Developing an artificial neural network is not simple, but involves many principle steps, which are; select a data producer, data production, data processing, neural network structure selection, training algorithm selection, neural network training, testing of the trained artificial neural networks, and using the trained ANN for simulation and validation [21].

ANN has a highly interconnected structure, and consists of a large number of simple processing elements called neurons, which are located in different layers in the networks; an input layer, output layer and intermediate units, called hidden layers. Each of these layers consists of neurons that are number of inter-connected processing units. These neurons are interrelated by passing the signals by adjusted weighted connections. The input layer received signals from external sources, and weighted individually, then transfers this information for processing to the hidden layers [22]. Hidden layers do preprocessing and transfer results to other hidden layers and output layer using transfer functions [20]. In feed forward neural networks data move forward and only in one direction and move from input layers to hidden layer and from hidden layer to output. There is no ring or cycle in these networks.

In this work, Neural Network Toolbox MATLAB R2013a software was used to predict the Cr (VI) ion removal rate from aqueous solution. The results of 80 experiments were used for development of neural network (ANN) and the results of 20 experiments were set aside for simulation at end. The results of experiments related to optimizing agitation time and speed were not entered into model and only the results of trials with optimal agitation time and speed, were entered as input and output. Input and output variables are given in Table 1.

Table 1. Input and output variables of ANN model

Variables	Experimental data	
Input Layer	Range	Unit
Initial Cr(VI) concentration	5, 20, 50, 90	mg/L
Initial pH	2, 4, 6, 8	–
Adsorbent dosage	2, 4, 6, 8 10	g/L
Contact time	120	min
Agitation speed	200	rpm
Cr(VI) removal rate	20.11–96.07	%

As there is no particular global strategy for selection of neural network structure and algorithms on different issues [9], researchers work on various parts of neural networks in order to achieve the best models in studied cases. In this study, back propagation neural networks (BPNN) with three layers as described in Table 2 were developed in order to determine the most appropriate transfer functions in hidden and output layers in predicting removal rate. The obtained Experimental data was introduced to network in the form of 580 input matrix and 180 output matrix.

The experimental data were randomly divided into three sets: 70; 15 and 15% of data sets were used as training, validation and testing, respectively. The Training data set is used to adjust the weights on the neural network. The Validation data set is used to minimize over fitting, with this data set the weights of the network are not adjusting, only verifying that any increase in accuracy over the training set actually causes an increase in accuracy over a data set that has not been introduced to the network before, or at least the network has not trained on it. If the accuracy over the training set increases, but the accuracy over then validation set remains the same or decreases, then there occurs overfitting of the neural network and the training should be stopped. The testing data set is used only for testing the final solution in order to confirm the actual predictive power of the network. The Testing set allows seeing if the training set was enough and whether the validation set did the Responsibility of preventing over fitting [23].

Table 2. Transfer functions used in neural networks of this research

Model	Transfer function	
	hidden layer	output layer
1	Linear (purelin)	Linear (purelin)
2	Tangent sigmoid (tansig)	Too
3	Linear (purelin)	Tangent sigmoid (tansig)
4	Tangent sigmoid (tansig)	Too

In this research, in the architecture of neural networks, the number of neurons in input layer is equal to 5 that equals to number of input model variables. Initially, the number of neurons in hidden layer is assumed to be 10; the number of hidden layer neurons was optimized after determining the best network among four models listed in Table 2. The number of neurons in output layer depends on the number of network output parameters. In this research, output model is the removal of Cr (VI). Thus, the number of output layer neurons is 1.

Neural network performance is evaluated through mean square error (MSE) and correlation coefficient (r); their equations are given below:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} \left(y_{\mathrm{obs},i} - y_{\mathrm{model},i}\right)}{n}; \tag{2}$$

$$r = \frac{\sum_{i=1}^{n} \left(y_{\mathrm{obs},i} - y_{\mathrm{obs,mean}}\right)\left(y_{\mathrm{obs},i} - y_{\mathrm{model},i}\right)}{\sqrt{\sum_{i=1}^{n} \left(y_{\mathrm{obs},i} - y_{\mathrm{obs,mean}}\right)^2 \left(y_{\mathrm{mode},i} - y_{\mathrm{model,mean}}\right)^2}}, \tag{3}$$

where n is the number of data points, $y_{\mathrm{obs},i}$, $y_{\mathrm{model},i}$, $y_{\mathrm{obs,mean}}$ and $y_{\mathrm{model,mean}}$ are the measured and predicted values and mean measured and mean predicted values of the processes, respectively [20].

Levenberg-Marquardt back propagation (LMBP) algorithm was used for network training. Figure 1 illustrates this algorithm stepwise. LMBP algorithm with memory reduction was used for the optimization. Based on previous experience, it was found that this algorithm achieves fast learning speed and high performance relative to other optimization algorithms [23].

Learning process in back propagation allows the neural network to produce a suitable response by adapting itself to motivation. The network adjusts its weights as it receives the input motivation. The learning process is done continuously until the actual response converges to the expected response.

During the training period, the weights of the neurons in the hidden and output layers are initialized with small pseudo-random numbers. Then the input pattern is propagated forward and the activation level of the hidden and output neurons is calculated using a transfer function. The weights are then updated starting from the output layer neurons backward to the hidden layer neurons. The error gradient between the actual and expected responses is then calculated using the delta rule. This causes the neuron to produce an error signal. This error signal changes its value depending on the weights of the neurons in each layer. The error triggers the network to adjust its weights and minimize the error rate until it converges to the suitable result.

The training process can be stopped using an error criterion. It could be either when the error is below or equal to predefined acceptable error rate; or by limiting the number of epochs. The second criterion confirms that the learning process will stop; however, there is no warranty that the network will be able to learn into its maximum potential [24]. Weight correction is performed any time by following equation:

$$\Delta W(t) = -\eta \mathrm{MSE} + \alpha \Delta W(t-1), \tag{4}$$

where η is training rate and α is momentum coefficient and both of them are in the range of 0 to 1 [25].

The training of an ANN is sensitive to training rate and momentum coefficient. The larger of training rate is selected, the quicker of the training, because large η value causes more changes to weights in the network. However, the training phase can cause fluctuation when η is selected too large. The momentum parameter is used to avoid the system from converging to a local minimum or saddle point. A high momentum parameter can also help to increase the speed of convergence of the system. However, setting the momentum parameter

too high can generate a risk of overshooting the minimum, which can cause the system to become unsteady. A momentum coefficient that is too low cannot reliably prevent local minima, and can also slow down the training of the system [24].

In this study, both of training and momentum parameters are set to 0.1, during the training phase. The mean square error is monitored during the training phase. The error usually decreases during the initial phase of training. However, when the network starts to over fit the data, the error on the training set will generally begin to rise. When the training error begins increasing for a selected number of iterations, the training is stopped and the weights at the minimum value of the training error are restored [25]. Figure 1 shows the neural network architecture and LMBP algorithm diagram in order to predict the percentage of chromium removal.

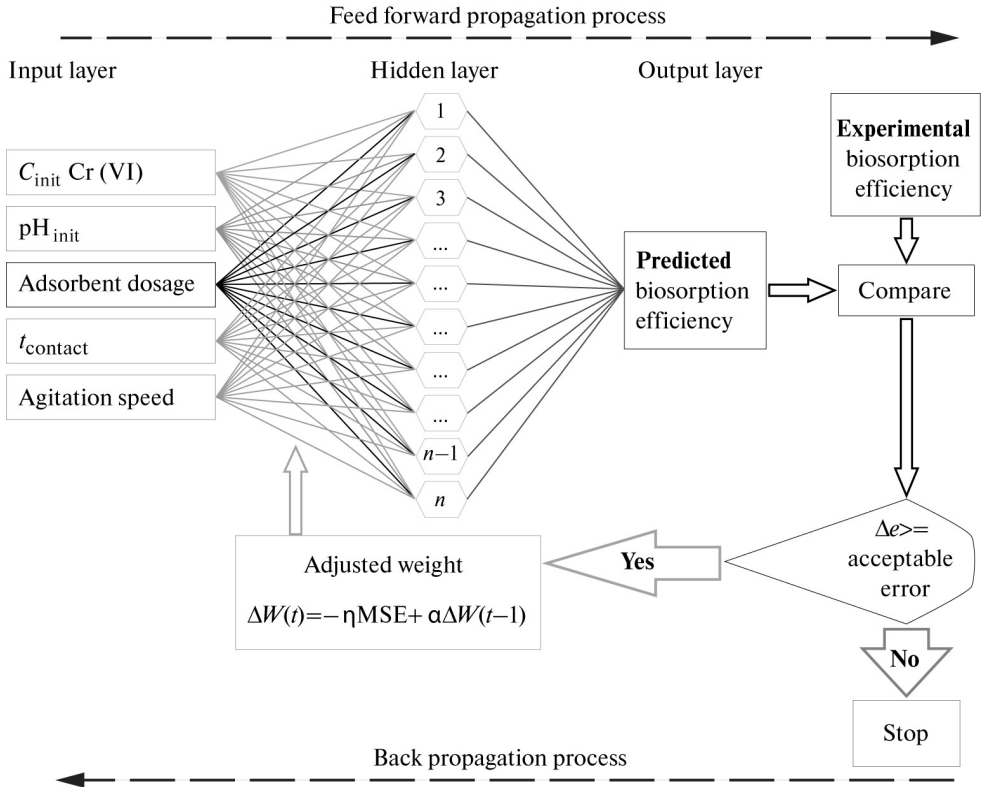

Fig. 1. Neural network architecture and LMBP algorithm diagram in order to predict the percentage of Cr removal.

RESULTS AND DISCUSSION

In the first step, networks with single hidden layer using different popular transfer functions (see Table 2) like sigmoid and purlin are developed for prediction of Cr(VI) removal. Figure 2 shows; the minimum value of MSE and *R*-value reached during training in four networks with different transfer functions.

It could be seen from Fig. 2 that network with Sigmoid Tangent in hidden and output layer, showed the least squares error and the highest correlation coefficient. Here the optimal topology of the developed ANN model is 5 : 10 : 1. The best state is occurred at Fourteenth epoch or in other words, 14th times that all of the training vectors are used to update the weights. MSE is equal to 0.013. This result seems reasonable because validation and test error have similar characteristics and it seems that over fitting problem does not occur. The next stage includes a series of analyzes on network response; in this regard, all training, validation and testing data is applied on network and a linear regression is obtained between network response and target vector that is experimental results. The regression coefficient and the correlation between expected response and actual response that is equal to 0.9830. the result is reasonable because *R*-value is close to 1 and there is a good correlation between network output and target vector.

After finding the most suitable transfer function in hidden and output layer, the aim of next step will be to determine the optimal number of neurons in network hidden layer, if based on the previous stage, transfer function of hidden and output layer would be Sigmoid Tangent. Neural networks were generated with the architecture explained above, and with number of neurons from 1 to 100 in order to determine the optimal

number of neurons in hidden layer. Minimum mean square error and correlation coefficient are shown in Table 3 for each of networks.

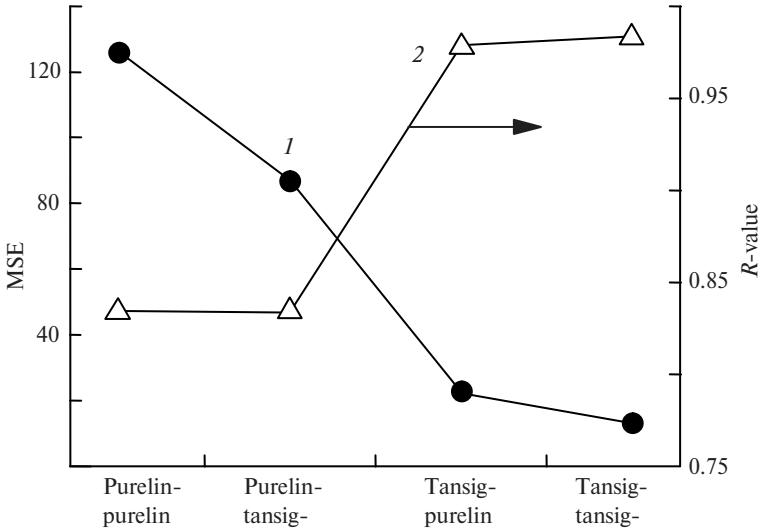

Fig. 2. The minimum error and correlation coefficient values of four models in order to determine the best combination of transfer function: *1*—MSE, *2*—*R*-value.

Table 3. Minimum error and correlation coefficient of neural networks that are designed for determining the optimal number of neurons

The number of neurons	1	2	3	4	5	6	7	8
MSE	0.008	0.010	0.086	0.048	0.066	0.019	0.058	0.022
R-value	0.841	0.916	0.923	0.928	0.912	0.961	0.933	0.971
The number of neurons	9	10	11	12	13	20	50	100
MSE	0.017	0.013	0.0036	0.0060	0.0040	0.028	0.0231	0.0348
R-value	0.976	0.9830	0.973	0.970	0.9833	0.972	0.936	0.703

Table 3 illustrates that, with increase in the number of neurons from 1 to 13, a gradual decrease was observed in the MSE. With 13 hidden neurons, the MSE reached its minimum value of 0.0040 and *R*-value reached its maximum value of 0.9833. Hence, the neural network containing 13 hidden neurons was chosen as the best case. When the number of neurons exceeded 13, the MSE showed a slight increase from 0.004 to 0.028 at 20 neurons. A further increase in the number of neurons from 20 to 100 resulted a sharp increase in the MSE.

The graphical outputs shown in Fig. 3 illustrate that the optimal network is a three-layer ANN, with tangent sigmoid transfer function (tansig) at hidden and output layer with 13 neurons in hidden layer.

As mentioned, the performance function used for training is based on the MSE between actual plant output and network predictions. Based on Fig. 3a, the training process was activated to achieve a performance target of 0.0040. The frequency of progress (in epochs) is set at 4. The learning rate was chosen to be 0.1. The value of this parameter was obtained after performing several trial and error runs. It was found that this value ensures stable fast learning.

The regression button in the training window performs linear regression between the network outputs and the corresponding targets. Figure 3b shows the results. It is observed that the output tracks the targets very well for training (R-value = 0.9802), validation (R-value = 0.9904), and testing (R-value = 0.9871). These values can be equivalent to a total response of R-value = 09833. In this case, the network response is satisfactory, and simulation can be used for entering new inputs.

After a network is trained, another 20 new input data is used for simulation. It must be noted that network has not been trained with this data and if it can predict the appropriate results it could confirm that network is reliable and usable. The simulation results of Cr(VI) removal efficiency are presented in Fig. 4 by plotting the

experimental data and ANN model predictions. It shows, that the predicted biosorption efficiencies of ANN results obtained are as close as with experimental data given to the network.

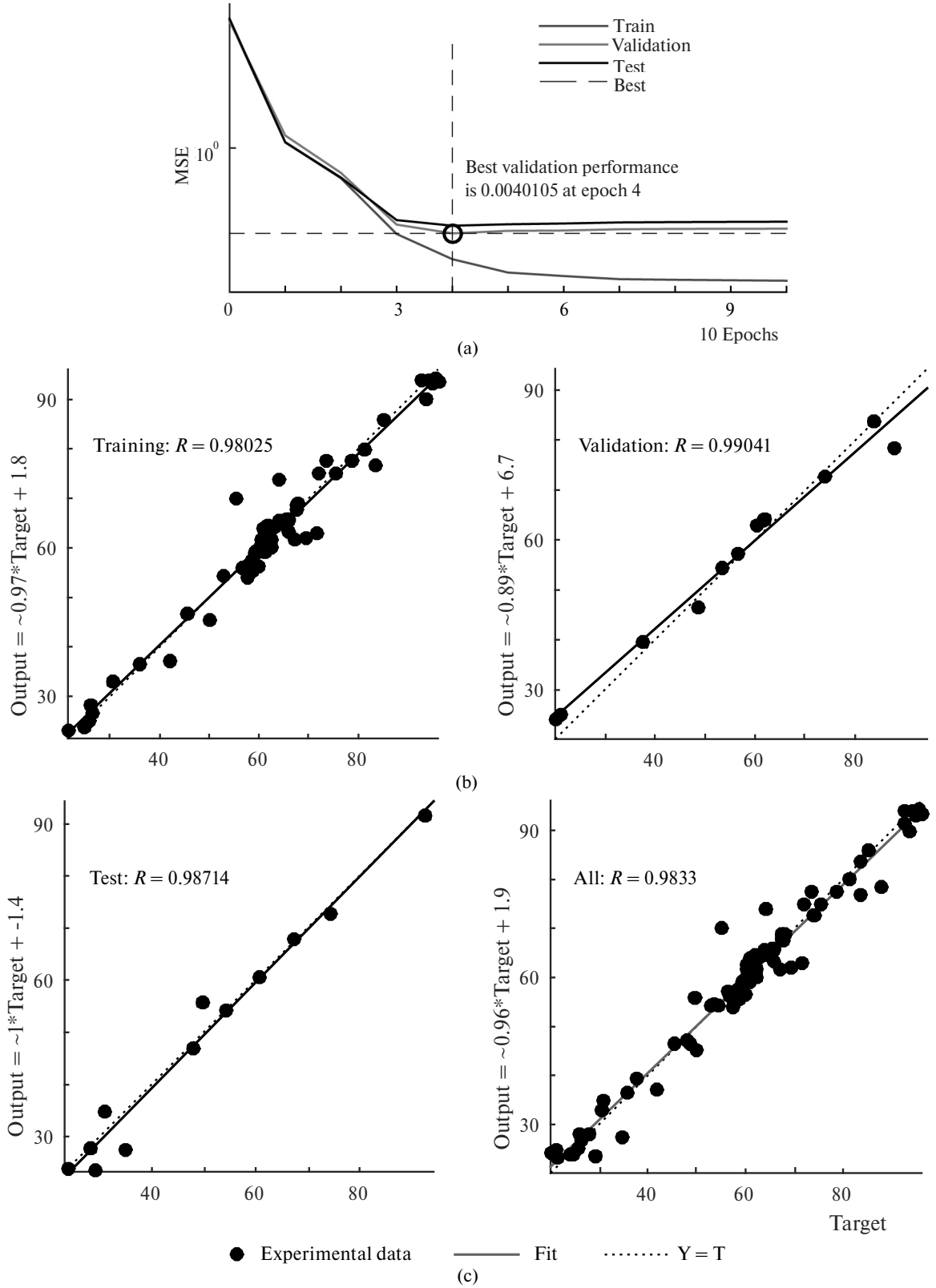

Fig. 3. Designed neural network that is designed for predicting the Cr (VI) removal rate (a); the performance of ANN model experiments (b) and predicted model regression (c).

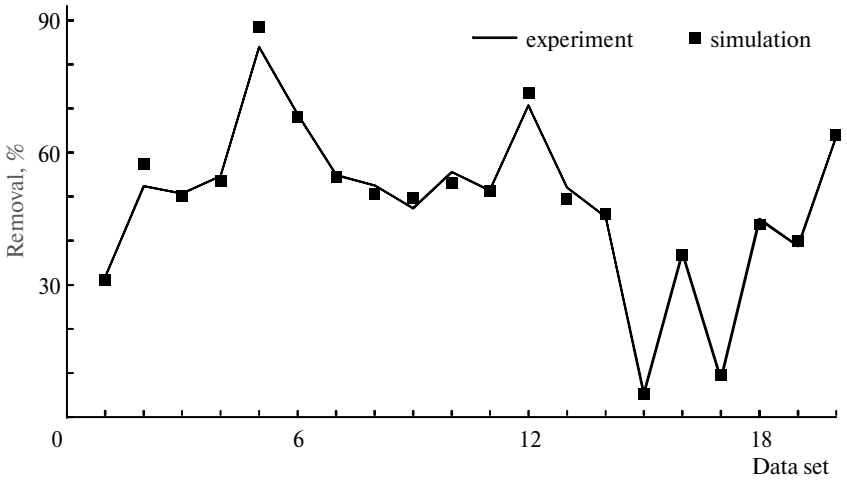

Fig. 4. Comparison of neural network simulation results that are designed for predicting Cr(VI) removal rate and experimental results.

CONCLUSIONS

In the present study, a neural network has been designed and demonstrated to predict the extent of Chromium ion adsorption with excess municipal wastewater sludge from aqueous solutions by taking into account the effect of initial Cr (VI) concentration, pH, adsorbent dosage, contact time and agitation speed. A simple back-propagation feed forward neural network using the Levenberg-Marquardt training algorithm is found to be very efficient to generalize and predict the degree of adsorption. The configuration of the neural network that gives the best prediction is the one with three layers consisting of 5; 13 and 1 neurons in each layer. Optimum transfer function in hidden and output layer is Sigmoid Tangent. ANN predicted results are very close to the experimental values. The average mean square error is 0.004 and correlation coefficient between predicted removal rate and experimental results is 0.9833. The present work suggests that neural network can be used as an efficient technique in modeling, estimation and prediction of biosorption process. Also, neural network can be considered as an efficient supplement for the conventional and complicated mathematical models in the prediction of bioprocess parameters.

ACKNOWLEDGEMENTS

The authors would like to thank Environment Research Center, Isfahan University of Medical Science, Isfahan, Iran, and Department of Environmental Health Engineering, School of Health, Research Center, Isfahan, Iran.

REFERENCES

1. Zhong, Q.Q., Yue, Q.Y., Li, Q., et al., *Carbohydr. Polym*., 2014, vol. 111, no. 788, pp. 788–796.
2. Malakootian, M., Dowlatshahi, Sh., and Hashemi, M., *J. Mazandaran Univ. Med. Sci*., 2013, vol. 23, pp. 69–78.
3. Meziane, F., Raimbault, V., Hallil, H., et al., *Sens. Actuators, B*., 2015, vol. 209, no. 1049, pp. 1–22.
4. Chakraborty, S., Dasgupta, J., Farooq, U., et al., *J. Membr. Sci*., 2014, vol. 456, no. 139, pp. 139–154.
5. Choppala, G., Bolan, N., and Park, G., *Adv. Agron*., 2013, vol. 120, no. 129, pp. 129–172.
6. Shi, M., Li, Z., Yuan, Y., et al., *Chem. Eng. J*., 2015, vol. 265, no. 84, pp. 84–92.
7. Ullah, I., Nadeem, R., Iqbal, M., and Manzoor, Q., *Ecol. Eng*., 2013, vol. 60, no. 99, pp. 99–107.
8. Hegazi, H., *HBRC J*., 2013, vol. 9, no. 3, pp. 276–282.
9. Ahmad, M., Haydar, S., Bhatti, A., and Bari, A., *Biochem. Eng. J*., 2014, vol. 84, no. 83, pp. 83–90.
10. Yetilmezsoy, K. and Demirel, S., *J. Hazard. Mater.*, 2008, 153, no. 1288, pp. 1288–1300.
11. Bagheri, M., Mirbagheri, S., Bagheri,Z., and Kamarkhani, A., *Process Saf. Environ. Prot*., 2015, vol. 95, no. 12, pp. 1–47.
12. Ding, Y.R., Cai, Y.J., Sun, P.D., and Chen, B., *J. Appl. Res. Technol*., 2014, vol. 12, no. 3, pp. 493–499.
13. Joo, S., Yoon, J., Kim, J., et al., *Appl. Therm. Eng*., 2015, vol. 80, no. 5, pp. 436–444.
14. Bunsana,S., Chenc, W., Chenc, H., et al., *Chemosphere*, 2013, vol. 92, no. 3, pp. 258–264.
15. Yang, Y., Wang, G., Wang, B., et al., *Biores. Technol*., 2011, vol. 102, pp. 828–834.
16. Popa, M., Ileana, I., Vosniako, F., et al., *J. Environ. Prot. Ecol*., 2011, vol. 12, no. 4, pp. 1948–1953.

17. Demir, G., Ozdemir, H., Ozcan, H.K., et al., *J. Environ. Prot. Ecol.*, 2010, vol. 11, no. 3, pp. 1163–1171.
18. Adeyinka, A., Llang, H., and Tina, G., *Scholl Eng. Technol.*, 2007, vol. 33, no. 2, pp. 1–8.
19. *APHA. Standard Methods for the Examination of Water and Wastewater*, 20th ed., Washington: Amer. Publ. Health Assoc., 2005.
20. Shanmugaprakash, M. and Sivakumar, V., *Biores. Technol.*, 2013, vol. 148, pp. 550–559.
21. Rafiq, M.Y., Bugmann, G., and Easterbrook, D.J., *Comput. Struct.*, 2001, vol. 79, no. 17, pp. 1541–1552.
22. Ozdemir, U., Azbay, B., Veli, S., and Zor, S., *Chem. Eng. J.*, 2011, vol. 178, no. 183, pp. 183–190.
23. Hegan, M. and Menhaj, H., *IEEE Transactions on neural network*, 1994, vol. 5, no. 6, pp. 989–993.
24. Giri, A., Patel, A., and Mahapatra, S., *Chem. Eng. J.*, 2011, vol. 178, no. 15, pp. 15–25.
25. Moreira, M. and Fiesler, E., *IDIAP Res. Institute*, Valais, Switzerland, 1995, pp. 1–29.